Math 110 Packet
Version 1.0

Contents

UNIT 1

Affective Domain

Contents

Module 1 A Poster Session

Breakfast Cereals and Children's Health

Part I: Your instructor will show you a series of commercials for breakfast cereals. After viewing and discussing these commercials as a class, discuss the following four questions in your group. Make notes of key points raised in your group to prepare for the class discussion.

1) Should we be concerned about advertisers targeting children in breakfast cereal advertisements? Why or why not?

2) What impact, if any, do you think this advertising has on children's health? Be specific.

3) How do you define "unhealthy" cereals?

4) Do you think this is a world health issue or one confined to the United States? Explain.

Part II: For this activity your group will be working with a data set containing information about 24 breakfast cereals.

Question to be investigated:

Are cereals marketed to children less healthy than cereals marketed to adults?

Instructions: Answer this question and back up your answer with observations from the data.

Think It Through:

- As a group, decide which parts of the data set you plan to use. You do not need to explore all of the variables; choose one or a few.

- Explore the data by making graphs, diagrams, tables, and doing calculations that make sense to you. Make sure you are comparing cereals targeted at adults to those targeted at children. Try as many different ideas as you can think of. Different approaches will probably highlight different features in the data. Since this is just a preliminary investigation of this data, and you may not have time to do a thorough analysis, note ideas that surfaced in the group that could be important for future investigations.

Make a poster to summarize the group's initial findings:

- Write the initial conclusions. Make sure your group answers the question.

- The poster should illustrate or explain how your analysis of the data supports your answers and conclusions.

LMC Breakfast Cereal Data 1.1.2

	Name	Manufa...	Target	Shelf	Calories	Carbs	Fat	Fiber	Potassium	Protein	Sodium	Sugars	Vitamins	CRRating	Cups	Weight
1	Cap'n'Cr...	Quaker O...	child	middle	120	12	2	0	35	1	220	12	25	18	0.75	1
2	Cocoa P...	General ...	child	middle	110	12	1	0	55	1	180	13	25	23	1	1
3	Trix	General ...	child	middle	110	13	1	0	25	1	140	12	25	28	1	1
4	Apple Ja...	Kelloggs	child	middle	110	11	0	1	30	2	125	14	25	33	1	1
5	Corn Chex	Ralston P...	adult	bottom	110	22	0	0	25	2	280	3	25	41	1	1
6	Corn Flak...	Kelloggs	adult	bottom	100	21	0	1	35	2	290	2	25	46	1	1
7	Nut&Hon...	Kelloggs	adult	middle	120	15	1	0	40	2	190	9	25	30	0.67	1
8	Smacks	Kelloggs	child	middle	110	9	1	1	40	2	70	15	25	31	0.75	1
9	Multi-Gra...	General ...	adult	bottom	100	15	1	2	90	2	220	6	25	40	1	1
10	Cracklin' ...	Kelloggs	adult	top	110	10	3	4	160	3	140	7	25	40	0.5	1
11	Grape-Nuts	Post	adult	top	110	17	0	3	90	3	170	3	25	53	0.25	1
12	Honey N...	General ...	child	bottom	110	11.5	1	1.5	90	3	250	10	25	31	0.75	1
13	Nutri-Gra...	Kelloggs	adult	top	140	21	2	3	130	3	220	7	25	41	0.67	1.33
14	Product 19	Kelloggs	adult	top	100	20	0	1	45	3	320	3	100	42	1	1
15	Total Rai...	General ...	adult	top	140	15	1	4	230	3	190	14	100	29	1	1.5
16	Wheat C...	Ralston P...	adult	bottom	100	17	1	3	115	3	230	3	25	50	0.67	1
17	Oatmeal ...	General ...	adult	top	130	13.5	2	1.5	120	3	170	10	25	30	0.5	1.25
18	Life	Quaker O...	child	middle	100	12	2	2	95	4	150	6	25	45	0.67	1
19	Maypo	America...	adult	middle	100	16	1	0	95	4	0	3	25	55	1	1
20	Quaker O...	Quaker O...	adult	top	100	14	1	2	110	4	135	6	25	50	0.5	1
21	Muesli R...	Ralston P...	adult	top	150	16	3	3	170	4	150	11	25	34	1	1
22	Quaker O...	Quaker O...	adult	bottom	100		2	2.7	110	5	0		0	51	0.67	1
23	Cheerios	General ...	child	bottom	110	17	2	2	105	6	290	1	25	51	1.25	1
24	Special K	Kelloggs	adult	bottom	110	16	0	1	55	6	230	3	25	53	1	1

Name: Name of cereal
Manufacturer: Manufacturer of cereal
Target: Target audience for cereal (adult, child)
Shelf: Display shelf at the grocery store
Calories: Calories per serving
Carbs: Grams of complex carbohydrates in one serving
Fat: Grams of fat in one serving
Fiber: Grams of dietary fiber in one serving
Potassium: Milligrams of potassium in one serving
Protein: Grams of protein in one serving
Sodium: Milligrams of sodium in one serving
Sugars: Grams of sugars in one serving
Vitamins: Vitamins and minerals - 0, 25, or 100% of daily need in one serving
CRRating: Consumer Report rating
Cups: Number of cups in one serving
Weight: Weight in ounces of one serving

Module 2 Learning Strategies

Learning goal: Identify the learning opportunities available in this course.

Introduction: This activity is an opportunity to reflect on and discuss all of the ways in which this course is designed to support your learning and success.

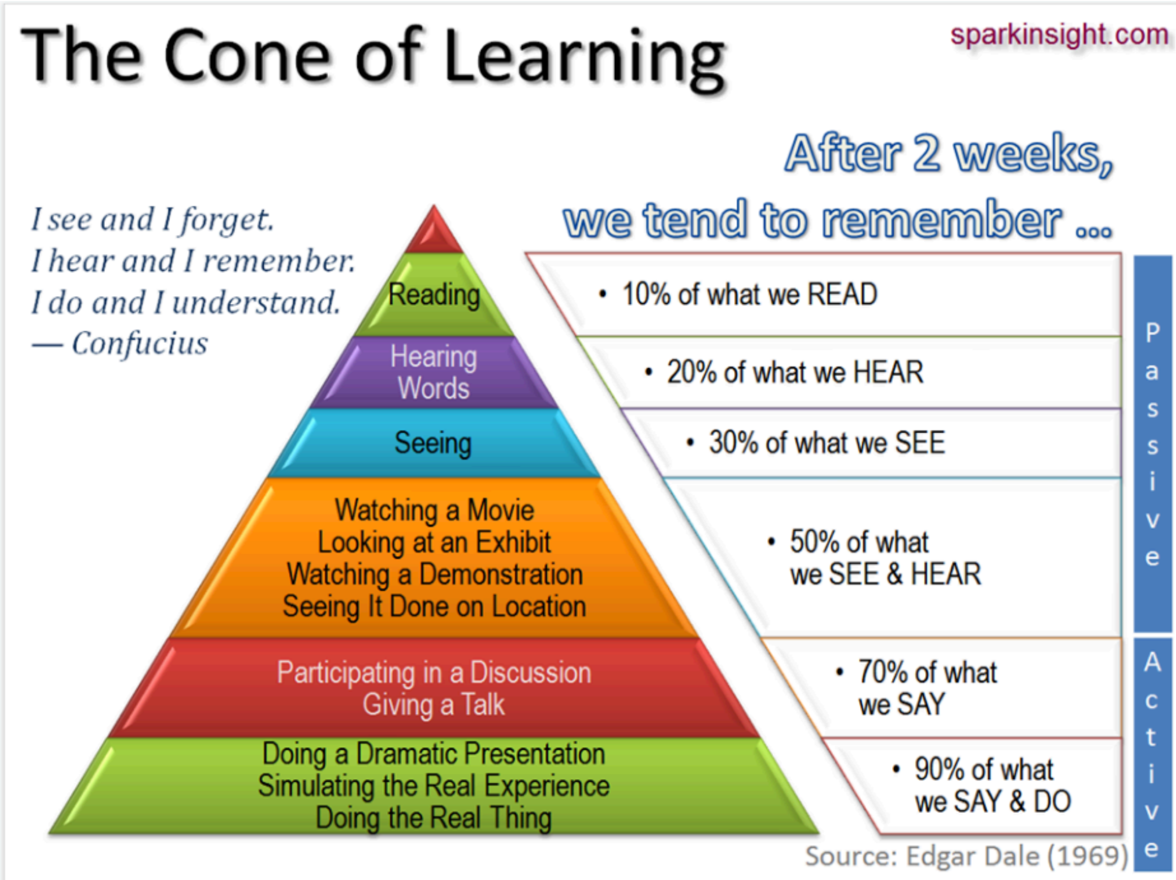
1) After college you may not remember all the facts and procedures that you have learned, but you will retain important skills that make you a powerful thinker and actor in the workplace and in life. What are these skills and how do you attain them through your college experience? This is our focus here.

   Fortune 500 companies are the companies with the largest earnings in the U.S. A survey of Fortune 500 leaders asked, what are the critical job skills that you value most in your employees? Here are the top ten skills identified by these industry leaders.

	1970	1999 [1]	2015 [2]
1	Writing	**Teamwork**	**Ability to Work in Team Structure**
2	Computtation Skills	**Problem Solving**	**Ability for Decision Making / Problem Solving**
3	Reading Skills	**Interpersonal Skills**	**Ability to Communication Verbally**
4	Oral Communications	**Oral Communications**	Ability to Plan / Organize / Prioritize
5	Listening Skills	Listening Skills	Ability to Obtain / Process Information
6	Personal Career Development	Personal Career Development	Abiltiy to Analyze Quantitative Data
7	Creative Thinking	Creative Thinking	Technical Knowledge Related to Job
8	Leadership	Leadership	Proficiency with Software Programs
9	Goal Setting / Motivation	Goal Setting / Motivation	Writing
10	**Teamwork**	Writing	Ability to Sell / Influence Others

   How will this class help you develop the top three most valued skills in the 2015 list? Think about how what you will be doing in class and in Canvas. Be specific.

2) What kind of activities contribute the most to long term memory and learning? Educational psychologists conduct research to answer this question. Edgar Dale, a famous educational psychologist, produced the following Cone of Learning based on his research into learning.



What learning opportunities in this class will contribute the most to your memory and learning? Think about how what you will be doing in class and in Canvas. Be specific.

3) Making mistakes, receiving feedback and using the feedback to improve your understanding and performance is crucial to learning.

   List all of the ways that you will receive feedback to improve your learning in this course. Think about in-class and out-of-class opportunities to receive feedback and use it to improve your understanding of course concepts. Be specific.

Module 3 Brainology and Growing Your Brain

Module 3.1 Brainology

Transforming Students' Motivation to Learn
Carol S. Dweck
Winter 2008

This is an exciting time for our brains. More and more research is showing that our brains change constantly with learning and experience and that this takes place throughout our lives.

Does this have implications for students' motivation and learning? It certainly does. In my research in collaboration with my graduate students, we have shown that what students believe about their brains — whether they see their intelligence as something that's fixed or something that can grow and change — has profound effects on their motivation, learning, and school achievement (Dweck, 2006). These different beliefs, or mindsets, create different psychological worlds: one in which students are afraid of challenges and devastated by setbacks, and one in which students relish challenges and are resilient in the face of setbacks.

How do these mindsets work? How are the mindsets communicated to students? And, most important, can they be changed? As we answer these questions, you will understand why so many students do not achieve to their potential, why so many bright students stop working when school becomes challenging, and why stereotypes have such profound effects on students' achievement. You will also learn how praise can have a negative effect on students' mindsets, harming their motivation to learn.

Mindsets and Achievement
Many students believe that intelligence is fixed, that each person has a certain amount and that's that. We call this a *fixed mindset*, and, as you will see, students with this mindset worry about how much of this fixed intelligence they possess. A fixed mindset makes challenges threatening for students (because they believe that their fixed ability may not be up to the task) and it makes mistakes and failures demoralizing (because they believe that such setbacks reflect badly on their level of fixed intelligence).

Other students believe that intelligence is something that can be cultivated through effort and education. They don't necessarily believe that everyone has the same abilities or that anyone can be as smart as Einstein, but they do believe that everyone can improve their abilities. And they understand that even Einstein wasn't Einstein until he put in years of focused hard work. In short, students with this *growth mindset* believe that intelligence is a potential that can be realized through learning. As a result, confronting challenges, profiting from mistakes, and persevering in the face of setbacks become ways of getting smarter.

To understand the different worlds these mindsets create, we followed several hundred students across a difficult school transition — the transition to seventh grade. This is when the academic work often gets much harder, the grading gets stricter, and the school environment gets less personalized with students moving from class to class. As the students entered seventh grade, we measured their mindsets (along with a number of other things) and then we monitored their grades over the next two years.

The first thing we found was that students with different mindsets cared about different things in school. Those with a growth mindset were much more interested in learning than in just looking smart in school. This was not the case for students with a fixed mindset. In fact, in many of our studies with students from preschool age to college age, we find that students with a fixed mindset care so much about how smart they will appear that they often reject learning opportunities — even ones that are critical to their success (Cimpian, *et al*., 2007; Hong, *et al*., 1999; Nussbaum and Dweck, 2008; Mangels, *et al*., 2006).

Next, we found that students with the two mindsets had radically different beliefs about effort. Those with a growth mindset had a very straightforward (and correct) idea of effort — the idea that the harder you work, the more your ability will grow and that even geniuses have had to work hard for their accomplishments. In contrast, the students with the fixed mindset believed that if you worked hard it meant that you didn't have ability, and that things would just come naturally to you if you did. This means that every time something is hard for them and requires effort, it's both a threat and a bind. If they work hard at it that means that they aren't good at it, but if they don't work hard they won't do well. Clearly, since just about every worthwhile pursuit involves effort over a long period of time, this is a potentially crippling belief, not only in school but also in life.

Students with different mindsets also had very different reactions to setbacks. Those with growth mindsets reported that, after a setback in school, they would simply study more or study differently the next time. But those with fixed mindsets were more likely to say that they would feel dumb, study *less* the next time, and seriously consider cheating. If you feel dumb — permanently dumb — in an academic area, there is no good way to bounce back and be successful in the future. In a growth mindset, however, you can make a plan of positive action that can remedy a deficiency. (Hong. *et al*., 1999; Nussbaum and Dweck, 2008; Heyman, *et al*., 1992)

Finally, when we looked at the math grades they went on to earn, we found that the students with a growth mindset had pulled ahead. Although both groups had started seventh grade with equivalent achievement test scores, a growth mindset quickly propelled students ahead of their fixed-mindset peers, and this gap only increased over the two years of the study.

In short, the belief that intelligence is fixed dampened students' motivation to learn, made them afraid of effort, and made them want to quit after a setback. This is why so many bright students stop working when school becomes hard. Many bright students find grade school easy and coast to success early on. But later on, when they are challenged, they struggle. They don't want to make mistakes and feel dumb — and, most of all, they don't want to work hard and feel dumb. So they simply retire.

It is the belief that intelligence can be developed that opens students to a love of learning, a belief in the power of effort and constructive, determined reactions to setbacks.

How Do Students Learn These Mindsets?

In the 1990s, parents and schools decided that the most important thing for kids to have was self-esteem. If children felt good about themselves, people believed, they would be set for life. In some quarters, self-esteem in math seemed to become more important than knowing math, and self-esteem in English seemed to become more important than reading and writing. But the biggest mistake was the belief that you could simply hand children self-esteem by telling them how smart and talented they are. Even though this is such an intuitively appealing idea, and even though it was exceedingly well-intentioned, I believe it has had disastrous effects.

In the 1990s, we took a poll among parents and found that almost 85 percent endorsed the notion that it was *necessary* to praise their children's abilities to give them confidence and help them achieve. Their children are now in the workforce and we are told that young workers cannot last through the day without being propped up by praise, rewards, and recognition. Coaches are asking me where all the coachable athletes have gone. Parents ask me why their children won't work hard in school.

Could all of this come from well-meant praise? Well, we were suspicious of the praise movement at the time. We had already seen in our research that it was the most vulnerable children who were already obsessed with their intelligence and chronically worried about how smart they were. What if praising intelligence made all children concerned about their intelligence? This kind of praise might tell them that having high intelligence and talent is the most important thing and is what makes you valuable. It might tell them that intelligence is just something you have and not something you develop. It might deny the role of effort and dedication in achievement. In short, it might promote a fixed mindset with all of its vulnerabilities.

The wonderful thing about research is that you can put questions like this to the test — and we did (Kamins and Dweck, 1999; Mueller and Dweck, 1998). We gave two groups of children problems from an IQ test, and we praised them. We praised the children in one group for their intelligence, telling them, "Wow, that's a really good score. You must be smart at this." We praised the children in another group for their effort: "Wow, that's a really good score. You must have worked really hard." That's all we did, but the results were dramatic. We did studies like this with children of different ages and ethnicities from around the country, and the results were the same.

Here is what happened with fifth graders. The children praised for their intelligence did not want to learn. When we offered them a challenging task that they could learn from, the majority opted for an easier one, one on which they could avoid making mistakes. The children praised for their effort wanted the task they could learn from.

The children praised for their intelligence lost their confidence as soon as the problems got more difficult. Now, as a group, they thought they *weren't* smart. They also lost their enjoyment, and, as a result, their performance plummeted. On the other hand, those praised for effort maintained their confidence, their motivation, and their performance. Actually, their performance improved over time such that, by the end, they were performing substantially better than the intelligence-praised children on this IQ test.

Finally, the children who were praised for their intelligence lied about their scores more often than the children who were praised for their effort. We asked children to write something (anonymously) about their experience to a child in another school and we left a little space for them to report their scores. Almost 40 percent of the intelligence-praised children elevated their scores, whereas only 12 or 13 percent of children in the other group did so. To me this suggests that, after students are praised for their intelligence, it's too humiliating for them to admit mistakes.

The results were so striking that we repeated the study five times just to be sure, and each time roughly the same things happened. Intelligence praise, compared to effort (or "process") praise, put children into a fixed mindset. Instead of giving them confidence, it made them fragile, so much so that a brush with difficulty erased their confidence, their enjoyment, and their good performance, and made them ashamed of their work. This can hardly be the self-esteem that parents and educators have been aiming for.

Often, when children stop working in school, parents deal with this by reassuring their children how smart they are. We can now see that this simply fans the flames. It confirms the fixed mindset and makes kids all the more certain that they don't want to try something difficult — something that could lose them their parents' high regard.

How *should* we praise our students? How *should* we reassure them? By focusing them on the process they engaged in — their effort, their strategies, their concentration, their perseverance, or their improvement.

"You really stuck to that until you got it. That's wonderful!"

"It was a hard project, but you did it one step at a time and it turned out great!"

"I like how you chose the tough problems to solve. You're really going to stretch yourself and learn new things."

"I know that school used to be a snap for you. What a waste that was. Now you really have an opportunity to develop your abilities."

Brainology

Can a growth mindset be taught directly to kids? If it can be taught, will it enhance their motivation and grades? We set out to answer this question by creating a growth mindset workshop (Blackwell, *et al*., 2007). We took seventh graders and divided them into two groups. Both groups got an eight-session workshop full of great study skills, but the "growth mindset group" also got lessons in the growth mindset — what it was and how to apply it to their schoolwork. Those lessons began with an article called "You Can Grow Your Intelligence: New Research Shows the Brain Can Be Developed Like a Muscle." Students were mesmerized by this article and its message. They loved the idea that the growth of their brains was in their hands.

This article and the lessons that followed changed the terms of engagement for students. Many students had seen school as a place where they performed and were judged, but now they understood that they had an active role to play in the development of their minds. They got to work, and by the end of the semester the growth-mindset group showed a significant increase in their math grades. The control group — the group that had gotten eight sessions of study skills — showed no improvement and continued to decline. Even though they had learned many useful study skills, they did not have the motivation to put them into practice.

The teachers, who didn't even know there *were* two different groups, singled out students in the growth-mindset group as showing clear changes in their motivation. They reported that these students were now far more engaged with their schoolwork and were putting considerably more effort into their classroom learning, homework, and studying.

Joshua Aronson, Catherine Good, and their colleagues had similar findings (Aronson, Fried, and Good, 2002; Good, Aronson, and Inzlicht, 2003). Their studies and ours also found that negatively stereotyped students (such as girls in math, or African-American and Hispanic students in math and verbal areas) showed substantial benefits from being in a growth-mindset workshop. Stereotypes are typically fixed-mindset labels. They imply that the trait or ability in question is fixed and that some groups have it and others don't. Much of the harm that stereotypes do comes from the fixed-mindset message they send. The growth mindset, while not denying that performance differences might exist, portrays abilities as acquirable and sends a particularly encouraging message to students who have been negatively stereotyped — one that they respond to with renewed motivation and engagement.

Inspired by these positive findings, we started to think about how we could make a growth mindset workshop more widely available. To do this, we have begun to develop a computer-based program called "Brainology." In six computer modules, students learn about the brain and how to make it work better. They follow two hip teens through their school day, learn how to confront and solve schoolwork problems, and create study plans. They visit a state-of-the-art virtual brain lab, do brain experiments, and find out such things as how the brain changes with learning — how it grows new connections every time students learn something new. They also learn how to use this idea in their schoolwork by putting their study skills to work to make themselves smarter.

We pilot-tested Brainology in 20 New York City schools. Virtually all of the students loved it and reported (anonymously) the ways in which they changed their ideas about learning and changed their learning and study habits. Here are some things they said in response to the question, "Did you change your mind about anything?"

I did change my mind about how the brain works…I will try harder because I know that the more you try, the more your brain works.

Yes... I imagine neurons making connections in my brain and I feel like I am learning something.

My favorite thing from Brainology is the neurons part where when u learn something, there are connections and they keep growing. I always picture them when I'm in school.

Teachers also reported changes in their students, saying that they had become more active and eager learners: "They offer to practice, study, take notes, or pay attention to ensure that connections will be made."

What Do We Value?

In our society, we seem to worship talent — and we often portray it as a gift. Now we can see that this is not motivating to our students. Those who think they have this gift expect to sit there with it and be successful. When they aren't successful, they get defensive and demoralized, and often opt out. Those who don't think they have the gift also become defensive and demoralized, and often opt out as well.

We need to correct the harmful idea that people simply have gifts that transport them to success, and to teach our students that no matter how smart or talented someone is — be it Einstein, Mozart, or Michael Jordan — *no one* succeeds in a big way without enormous amounts of dedication and effort. It is through effort that people build their abilities and realize their potential. More and more research is showing there is one thing that sets the great successes apart from their equally talented peers — how hard they've worked (Ericsson, *et al*., 2006).

Next time you're tempted to praise your students' intelligence or talent, restrain yourself. Instead, teach them how much fun a challenging task is, how interesting and informative errors are, and how great it is to struggle with something and make progress. Most of all, teach them that by taking on challenges, making mistakes, and putting forth effort, they are making themselves smarter.

Carol S. Dweck is the Lewis and Virginia Eaton Professor of Psychology at Stanford University and the author of Mindset: The New Psychology of Success (Random House, 2006).

For a list of references noted in this article, see http://www.nais.org/ismagazinearticlePrint.cfm?print=Y&ItemNumber=150509

To prepare for our discussion of this article, jot down notes in response to the following questions:

1. How can you tell if someone has a growth mindset?
2. How can you tell if someone has a fixed mindset?
3. According to Dweck, how do we acquire these mindsets?
4. Is it possible to change someone's mindset? If so, give some examples from the article. If not, explain why it is not possible.
5. How does this article relate to your personal experiences in school (or in learning math)?

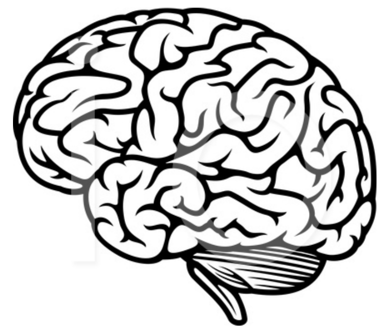
5

Module 3.2 Grow Your Brain

You Can Grow Your Brain

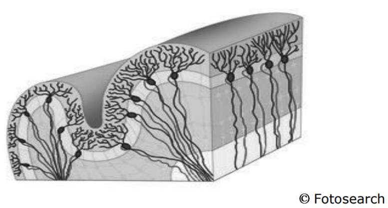
New research shows the brain can be developed like a muscle

Many people think of the brain as a mystery. We don't often think about what intelligence is or how it works. When you do think about what intelligence is, you might think that a person's intelligence is pre-determined at birth – either you are a "math person" or not – and stays that way for life. New research shows that the brain is more like a muscle – it changes and gets stronger when you use it. Scientists have been able to show how the brain grows and gets stronger when you learn.



Everyone knows that when you lift weights, your muscles get bigger and you get stronger. A person who can't lift 20 pounds when they start exercising can get strong enough to lift 100 pounds after working out for a long time. Muscles become larger and stronger with exercise. When you stop exercising, muscles shrink, and you get weaker. That's why people say, "use it or lose it."

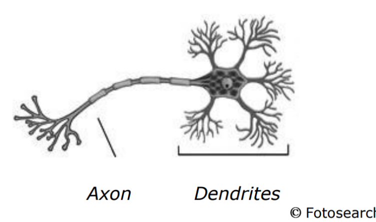
Most people don't know that when they practice and learn new things, parts of their brain change and get larger just like muscles. This is true even for adults. So, it's not true that some people "just can't learn". You can improve your abilities as long as you practice and use good strategies.



A section of the cerebral cortex

Inside the outer layer of the brain – the cortex – are billions of tiny nerve cells, called neurons. The nerve cells have branches connecting them to other cells in a complicated network. Communication between these brain cells is what allows us to think and solve problems.

When you learn new things, these tiny connections in the brain actually multiply and get stronger. The more you challenge your mind to learn, the more your brain cells grow. Things you once found very hard or even impossible to do – like speaking a foreign language or doing algebra – become easier. The result is a stronger, smarter brain.



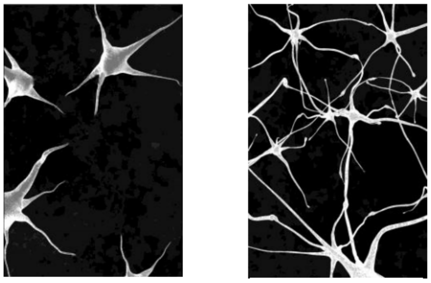
A typical nerve cell

How do we know that the brain can grow stronger?

Scientists started thinking the human brain could develop and change when they studied adult animal's brains. They found that animals who lived in a challenging environment with other animals and toys were different from animals who lived alone in bare cages. While the animals who lived alone just ate and slept all the time, the ones who lived with different toys and other animals were always active. They spent a lot of time figuring out how to use the toys and how to get along with the other animals.

Closer examination found these animals had more connections between nerve cells in their brains. The connections were bigger and stronger, too. In fact, their whole brains were about 10% heavier than the brain of animals living alone without toys. The adult animals who exercised their brains by playing with toys and each other were also "smarter" – they were better at solving problems and learning new things.

Effect of an Enriched Environment

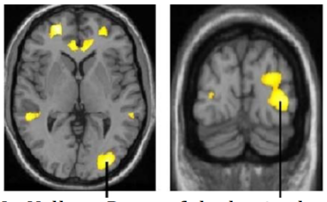


Nerves in brain of animal living in bare cage

Brain of animal living with other animals and toys

Can adults grow their brains?

Scientists have recently shown that adults can grow the parts of their brains that control abilities–like the ability to do math or juggle. In one study, scientists found a group of adults who were not jugglers. They taught half how to practice juggling in the right way. These people practiced for a long time and got much better at juggling. The other half didn't practice and didn't get better. Next, the scientists used a brain scanner to compare the brains of the two groups of people. They found that the people who learned how to juggle actually grew the parts of their brains that control juggling skills – the visual and motor areas. Their brains changed, so they actually had more ability.



In Yellow: Parts of the brain that grew when adults learned to juggle

This was surprising because these people said before the study that they couldn't juggle – just like some people say they're "not good at math." But when they learned good strategies for practicing and kept trying, they actually learned and grew their brains.

This can happen because learning causes permanent changes in the brain. The juggler's brain cells get larger and grow new connections between them. These new, stronger connections make the juggler's brain stronger and smarter, just like a weight-lifter's toned muscles.

The truth about “smart” and “dumb”

People aren’t “smart” or “dumb” at math. At first, no one can read or solve equations. But with practice, they can learn to do it. The more a person learns, the easier it gets to learn new things – because their brain “muscles” grow stronger.

This is true even for adults who have struggled for a long time to learn something. Dr. Wittenberg, a scientist from Wake Forest University, said “We used to think adults couldn’t form new brain connections, but now we know that isn’t true… The adult brain is like a muscle, and we need to exercise it.”

People who don’t know this can miss out on the chance to grow a stronger brain. They may think they can’t do it, or that it’s too hard. It does take work to learn, just like becoming stronger physically or becoming a better juggler does. Sometimes it even hurts. When you feel yourself get better and stronger, you realize that all the work is worth it.

A formula for growing brain: Effort + Good strategies + Help from others

Scientists found that learning to juggle is a lot like getting better at math. When people learn and practice new ways of doing algebra or statistics, they can grow their brains – even if they haven’t done well in math in the past.

Strengthening the “math” part of your brain usually happens when you try hard on challenging math problems. To grow your brain, you need to learn skills that let you use your brain in a smarter way.

If you use a bad strategy, you may not learn – even if you try hard. A few people study for math by doing the same set of easy problems and skipping the hard ones, or just re-reading the textbook, because it feels easier. When it comes time to do the test, they don’t do well because they didn’t work on the problems that stretched their brains and taught them new things. When this happens, they may even say “I’m just not smart at math.”

The truth is that everyone can become smarter at math if they practice in the right way. If a weight lifter watched other people exercise all day long, would he get stronger? If someone tried to learn how to juggle by just reading a book about juggling, they wouldn’t learn. You actually have to practice the right way – and usually that means the hard way – to get better at something.

In fact, scientists have found that the brain grows more when you learn something new, and less when you practice things you already know. This means that it’s not just how much time and effort you put in to studying math, but whether, when you study, you learn something new and hard. To do that, you usually need to use the right strategies. Luckily, strategies are easy to learn if you get help.

References:

- This article was retrieved from https://www.cmich.edu/ess/oss/Documents/Prepare%20for%20Success%20d4.pdf (2018, Dec 22).
- A similar version of this article was written by Lisa Blackwell (https://www.mindsetworks.com/websitemedia/youcangrowyourintelligence.pdf).
- Blackwell, L. A., Trzeniewski, K. H., & Dweck, C. S. (2007). Theories of intelligence and achievement across the junior high school transition: A longitudinal study and an intervention. Child Development, 78, 246-263
- Driemeyer, J., Boyke, J., Gaser, C., Buchel, C., May, A. (2008). Changes in Gray Matter Induced by Learning – Revisited. PLos One, 3, e2669. Doi:10.1371/journal.pone.0002669
- Nordqvist, C. (2004, Feb 1). "Juggling makes your brain bigger – New Study." Retrieved from http://www.medicalnewstoday.com/releases/5615.php

UNIT 2

Summarizing Data Graphically and Numerically

Contents

Module 4 Distributions of Quantitative Data

Module 4.1 Distributions of Quantitative Data: Introduction

Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objectives:

- Develop a way to describe and distinguish graphs of a quantitative variable.
- Identify reasonable explanations for what might explain the differences seen in different data sets.

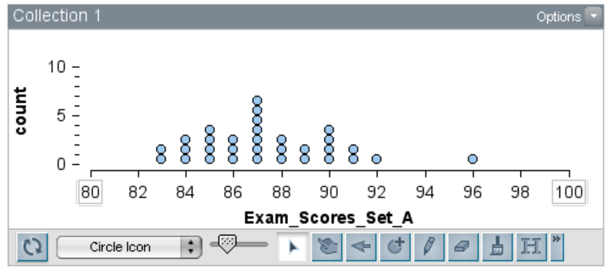
Introduction: Statisticians collect data in order to study and analyze groups. They focus on the group's data in aggregate, instead of focusing on individuals. For a statistician creating and comparing graphs is often the first step in analyzing data.

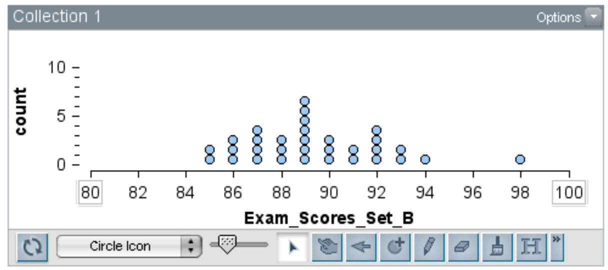
In this activity you will compare and contrast the graphs of hypothetical sets of exam scores. In other words, these graphs show made-up data. These hypothetical data sets have been constructed to help you begin to "see" like a statistician. By comparing these data sets, you will begin to develop an informal understanding of the key features of a graph that statisticians use to describe data. We will develop these ideas more formally in future lessons.

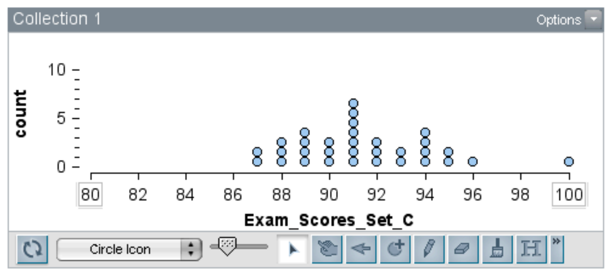
During this activity do your best to describe what you see. Jot down notes to capture your thinking as you go. You can use your notes during our class discussion of this activity. This activity does not require you to remember anything or to apply previous knowledge.

1) Compare and contrast the 3 graphs shown at the right.

    a) How are the graphs similar? How are they different? What is the most distinctive feature that distinguishes these three graphs from each other?







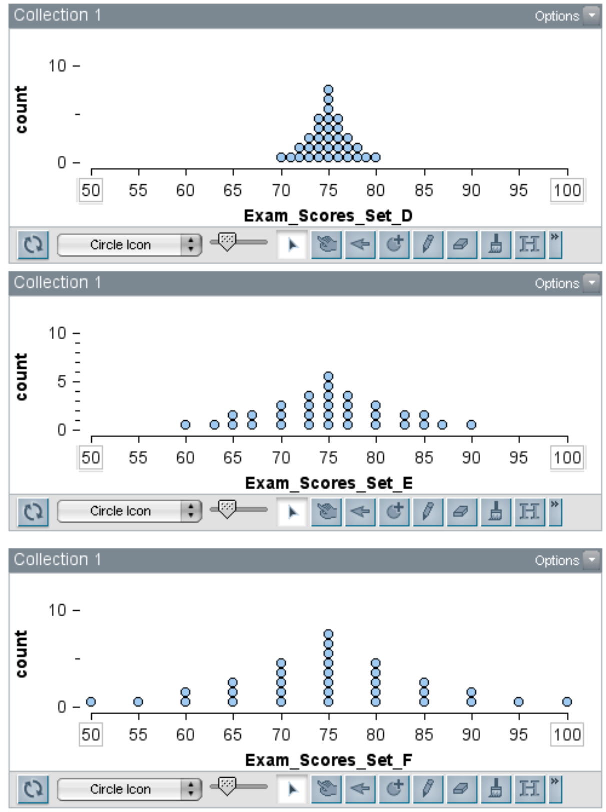
    b) For each graph, pick a single exam score to summarize the overall performance of the students. In other words, summarize each set of data with one number.
    Set A __________ Set B _________ Set C __________

    c) The average score for Set A is 87.4. What do you think the average score is for Set B and for Set C? (See if you can answer this without doing any calculations.)

    d) Which, if any, of the statements below is a reasonable explanation for the differences in the graphs? Why?

    - The graphs represent different classes. Different groups of students will obviously perform differently on an exam.
    - The graphs represent a single class after the teacher adjusted the grades. The teacher realized that some of the exam questions were not well written, so she adjusted the grades by adding points to the original exam scores.
    - There are no differences in the graphs. They look the same.

2) Compare and contrast the 3 graphs shown at the right.



a) How are the graphs similar? How are they different? What is the most distinctive feature that distinguishes these three graphs from each other?

b) *Thinking about center:* The average exam score is the same for each set of data (D, E and F). What do you think the average is?

c) *Thinking about spread:* Spread is a description of the variability we see in the data.
For exam set D the scores vary from a low of ________ to a high of _______.

For exam set E the scores vary from a low of ________ to a high of _______.

For exam set F the scores vary from a low of ________ to a high of _______.

d) Which of the statements below, if any, is a reasonable explanation for the differences in the graphs? Why?

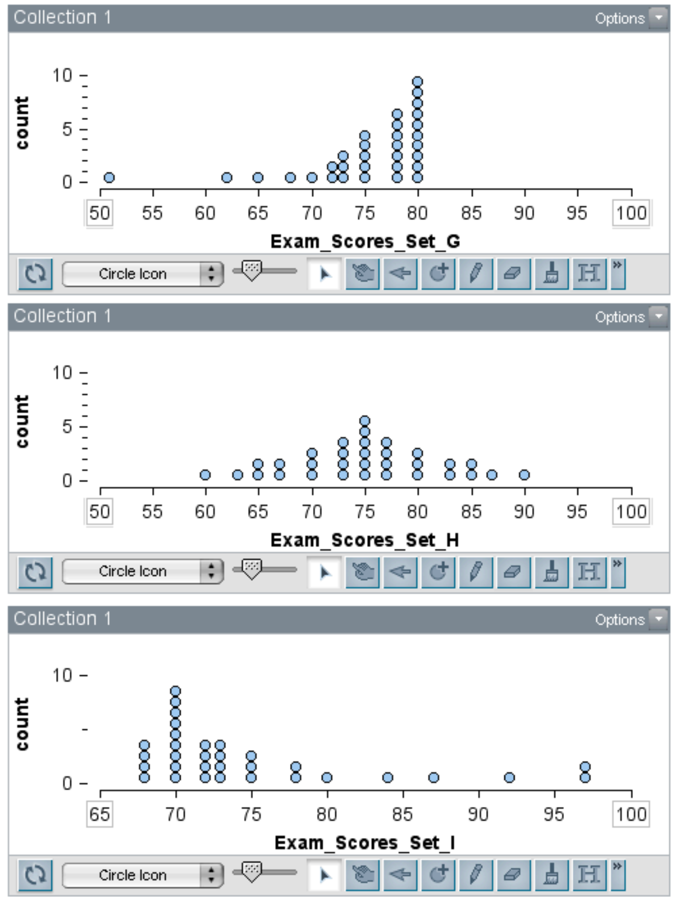
- The graphs represent a single class after the teacher adjusted the grades. The teacher realized that some of the exam questions were not well written, so she adjusted the grades by adding points to the original exam scores.

- The graphs represent 3 different classes with teachers that have different grading standards. One is an easy grader. One is a really hard grader.

- The differences could be explained by whether the teachers allowed the students to work together on the exam and how much time the students were given to finish the exam.

3) Compare and contrast the 3 graphs shown at the right.

a) In each of these 3 graphs, the average score is the same. The average score is 75. In what other ways are the graphs similar? How are they different? What is the most distinctive feature that distinguishes these three graphs from each other?



b) What might explain the differences in these graphs?

c) *Thinking about shape:* How would you describe the shapes (symmetric, skewed left, skewed right) of these three distributions of exam scores?

Module 4.2 Histograms

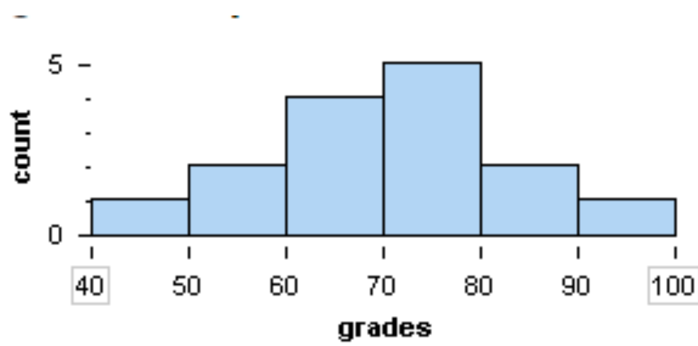
Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objective:

- Distinguish between categorical and quantitative variables;
- Identify graphs that represent the distribution of a quantitative variable;
- Analyze the distribution of a quantitative variable using a histogram. Describe shape, give a general estimate of center and the overall range, and calculate relevant percentages.

Overview:
In this activity you will again practice analyzing the distributions of quantitative variables using descriptions of shape, center and spread. This is the same type of thinking you did previously with dot plots, but this time the data will be summarized in a histogram.

This histogram shows the distribution of exam scores for 15 students in a Biology class.



a) How would you describe the shape of this distribution of exam scores? (Use course vocabulary.)

b) Give an interval that describes typical grades on this exam.

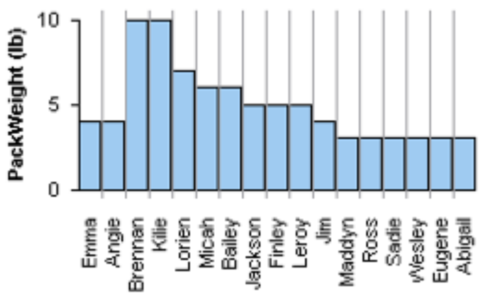
c) The *range* is the largest value minus the smallest value (Range = Max – Min). What is the largest the range could be? What is the smallest the range could be?

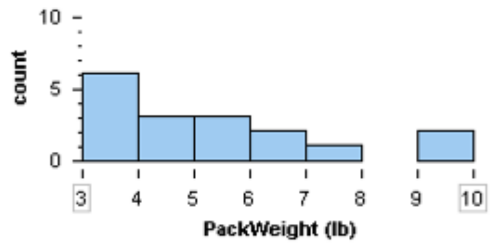
d) What percentage of the students made a D on the exam (a grade of 60-69%)?

e) What percentage of the students passed the exam with a 70 or better?

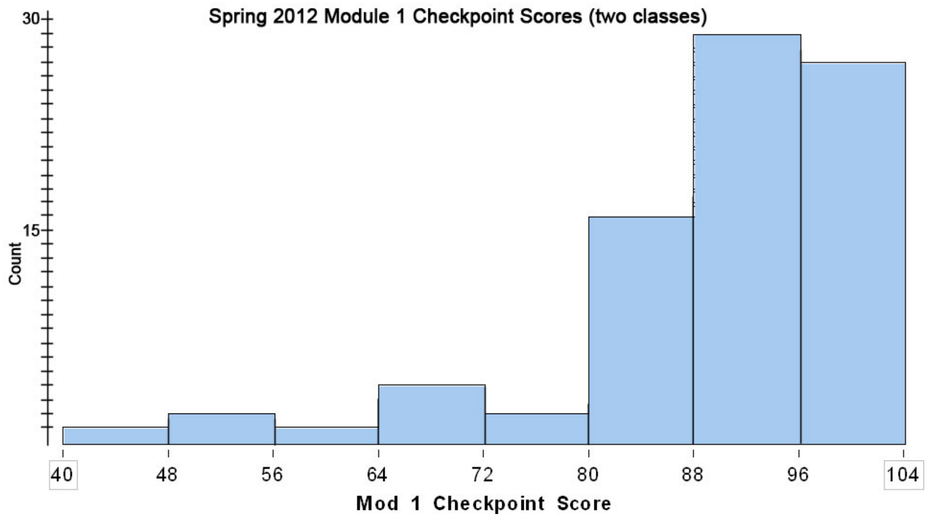
Group Work.

1) Which of the graphs below is a histogram? How do you know?





2) The following is a histogram indicating the distribution of scores on the Spring 2012 Module 1 Checkpoint for 82 students in Math 27 classes.



a) How would you describe the shape of this distribution of quiz scores? (Use course vocabulary.)

b) Give an interval that describes typical performance on this quiz.

For each of the following questions, answer the question if the histogram provides enough information to answer it. If not, write "not enough information".

c) What percentage of students scored below 80%?

d) How many students made an A (scored a 90% or higher)?

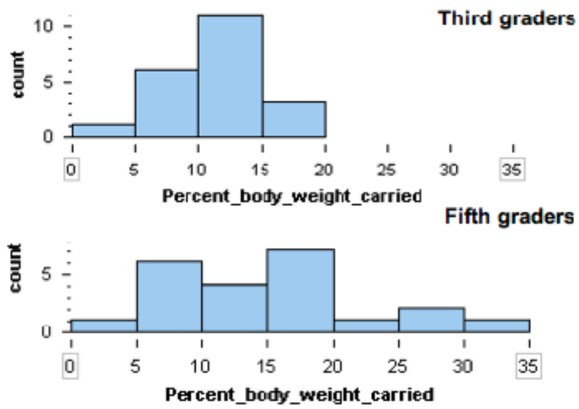
e) What is the lowest grade on the Module 1 Checkpoint?

f) What percentage of the students aced the quiz (a score of 100%)?

g) What is the average (mean) quiz score?

h) Did the majority of students pass the quiz (70% or better)?

3) The data graphed in these histograms describes 43 elementary school children. The variable is "percent of body weight carried in the school backpack." A child who weighs 60 pounds and carries 9 pounds has a variable value of 15% since 9÷60=0.15=15%. The American Chiropractic Association (ACA) recommends that children carry no more than 10% of their body weight.



% of body weight carried in backpack	Third graders	Fifth graders
0-5%	1	1
5-10%	6	6
10-15%	11	4
15-20%	3	7
20-25%	0	1
25-30%	0	2
30-35%	0	1
Total	21	22

a) Of the 3rd graders, how many are following the ACA recommendation?

b) Of the 3rd graders what percentage is following the ACA recommendation?

c) Of the 5th graders, what percentage is following the ACA recommendation?

d) Of all the children in this study, what percentage is NOT following the ACA recommendation?

e) Of the 5th graders who are NOT following the ACA recommendation, what percentage are carrying more than 25% of their body weight?

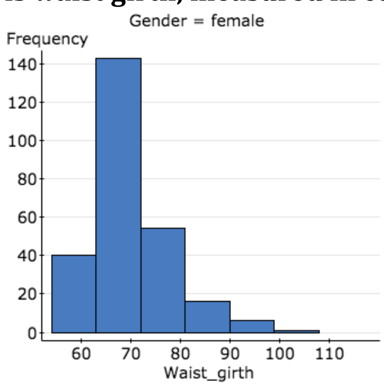
Module 4.3 Module 4 Lab

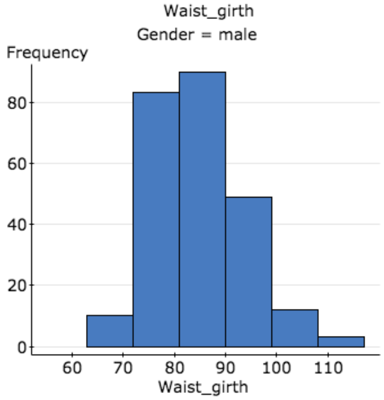
Name:____

Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objectives: Compare and contrast the distributions of a quantitative variable for two groups using histograms. Describe shape, give a general estimate of center and the overall range, and calculate relevant percentages.

1. **Here are data from adults (247 men and 260 women) who exercise regularly. The variable is waist girth, measured in centimeters.**





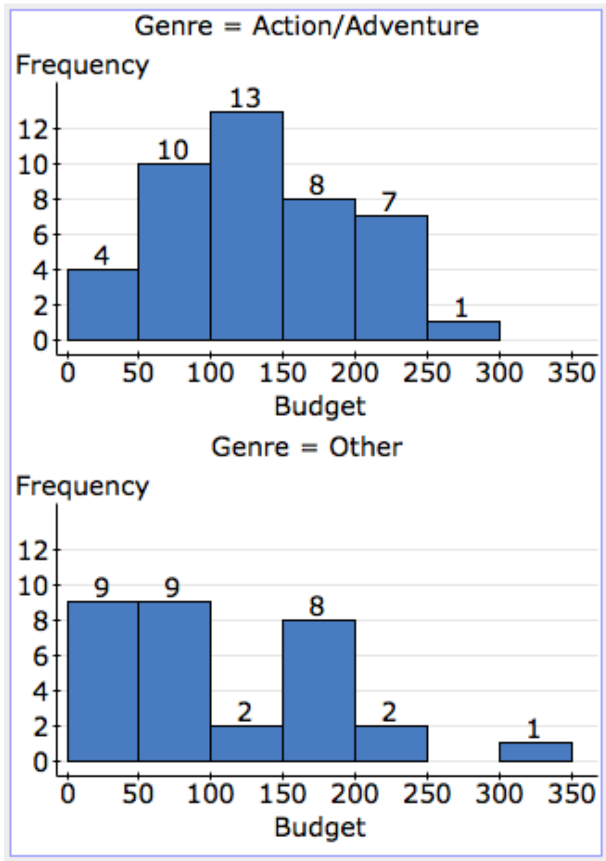
Are the following statements valid (true) or invalid (false)? Explain how the histograms support your answer.

(a) In this dataset, typical females have a smaller waist girth than typical males.

(b) There is less variability in waist girth for females.

(c) Here the distributions of waist girth measurement are skewed to the right for both males and females, with only a small percentage of each group having waist girths exceeding 99 cm.

2. **These histograms show the budget in millions of dollars for a sample of 74 movies listed in the top 100 USA box offices sales of all time. The movies are divided into two genres: Action/Adventure (with 43 movies) and Other (with 31 movies).**



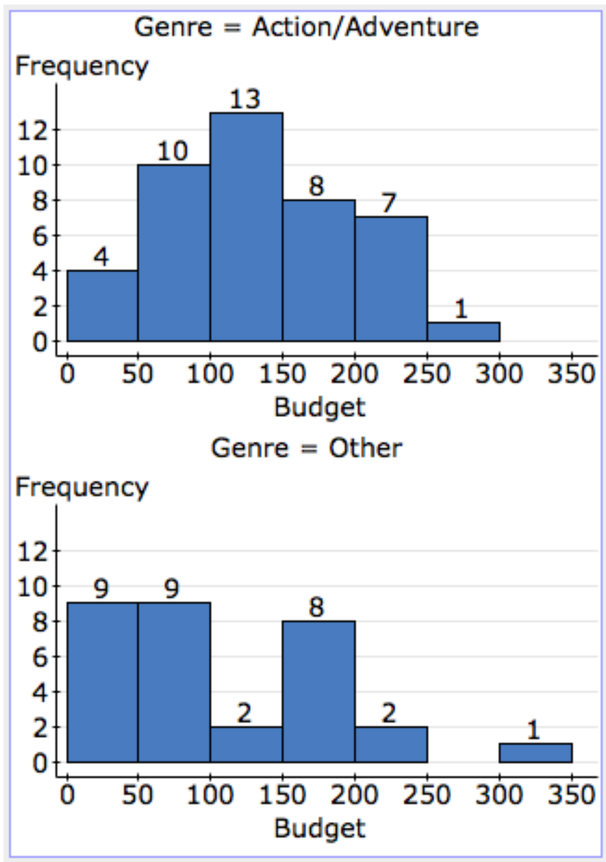
(a) Describe the shape of each distribution. What does the shape tell us about where most of the data fall?

(b) Which genre (Action/Adventure or Other) has the movie with largest budget?

(c) When we take all of the data into account, which genre tends to have larger budgets? (To answer this question, give an interval that represents typical budget amounts for each genre. Use these intervals to support your answer.)

(d) Which genre has more variability in budget amounts? (To answer this question, estimate the overall range of budget amounts for each genre. Use your estimates to support your answer.

(e) Pick the statement that you think is more strongly supported by the data:



- Action/Adventure movies tend to have larger budgets than other movies.
- Budget amounts are similar for Action/Adventure movies and Other movies.

For the statement you picked, support it with *at least three* precise observations from the histograms. Explain how your observations support the statement you chose.

Module 5 Measures of Center

Module 5.1 A Feel for Measures of Center

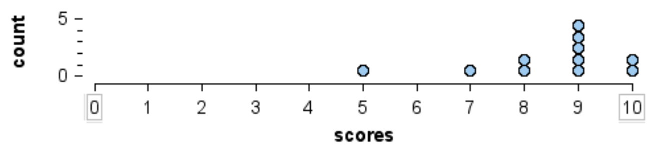
Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objectives:

- Find the mean and median from different representations of data.
- Develop number sense with mean and median by creating different data sets with a given mean or median.

Warmup Problem:

The dot plot gives quiz scores for a small class.



a) What is the mean? Show your work or explain how you got your answer.

b) What is the median? Show your work or explain how you got your answer.

c) Which measure (the mean or the median) is the better way to represent typical performance on this quiz? Why?

1) Here are two sets of exam scores, one for a class that has 4 students and one for a class that has 15 students.

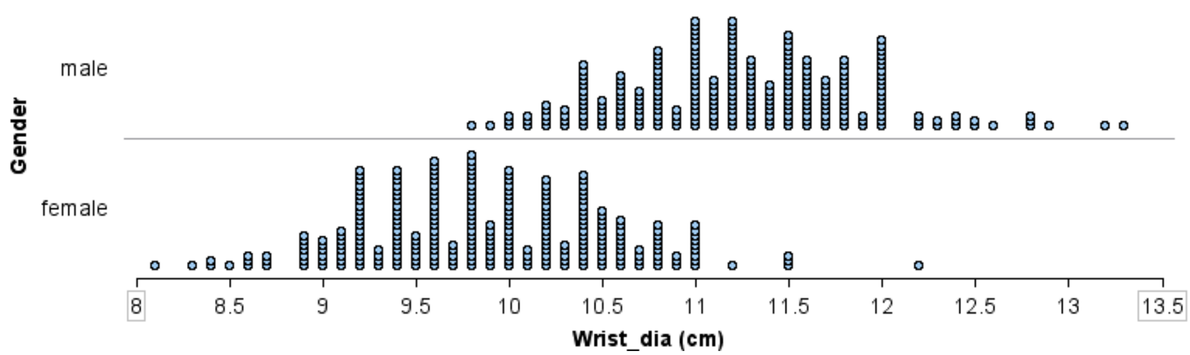
Class A: 80, 90, 90, 100

Class B: 60, 65, 65, 70, 70, 70, 75, 75, 80, 80, 80, 80, 80, 85, 100

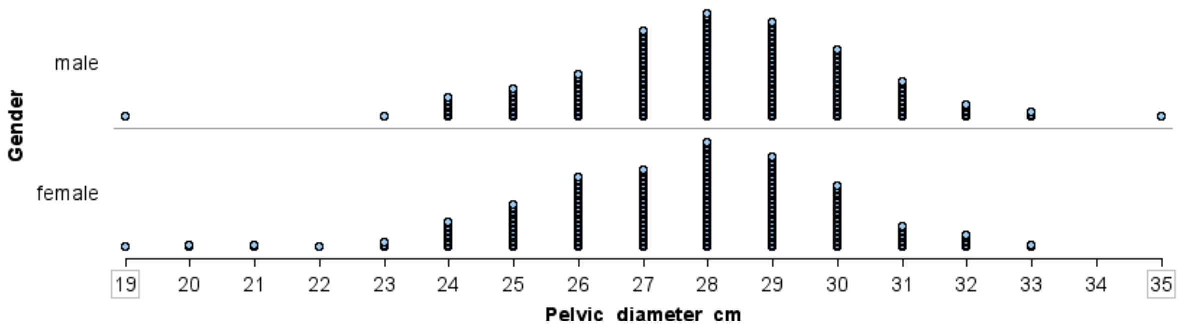
a) Without doing any calculations, which class do you think will have a larger mean? Why?

b) Now calculate the mean for each class. Which is larger? Why does this make sense?

2) For this data, is the mean wrist measurement for men (larger than, smaller than, or about the same as) the mean wrist measurement for women? (Obviously, you can't calculate the means, so jot down notes about how you thought about this.)



3) For this data, is the mean pelvic diameter for men (larger than, smaller than, or about the same as) the mean pelvic diameter for women? (Obviously, you can't calculate the means, so jot down notes about how you thought about this.)



4) This table gives quiz scores for a different class.

Scores	Number of Students
5	1
6	3
7	5
8	3
9	1

a) What is the mean? Show your work or explain how you got your answer.

b) What is the median? Show your work or explain how you got your answer.

c) Which measure (the mean or the median) is the better way to represent typical performance on this quiz? Why?

5) For this problem, use the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

   a) List five digits that have a median of 7 and a mean of 7 (repeats allowed). Find a different set of 5 digits that work.

   b) List five digits that have a median of 7 and a mean that is less than 7 (repeats allowed.) Give the mean of your 5 digits. Find a different set of 5 digits that work.

   c) List five digits that have a median of 7 and a mean that is more than 7 (repeats allowed.) Give the mean of your 5 digits. Find a different set of 5 digits that work.

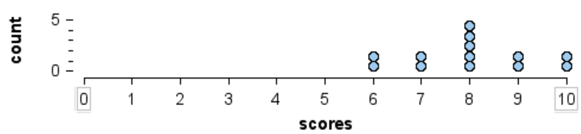
6) For this problem, use the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Construct a data set where neither the mean nor the median is a reasonable “typical” value.

Module 5.2 Shape and Measures of Center

Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

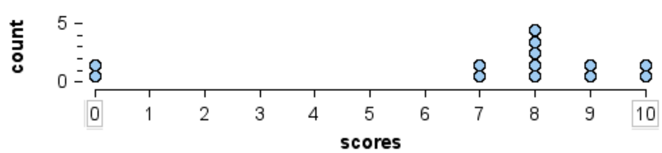
Specific Learning Objective: Relate measures of center to the shape of the distribution. Choose the appropriate measure for different contexts.

1) Here is a dot plot of Hilda's quiz scores in her math class.



   a) What is the shape of this distribution?
   b) Determine Hilda's median quiz score. Calculate Hilda's mean quiz score.

2) Oops, we made a mistake recording two of Hilda's scores. The two scores recorded as 6 should have been 0. Here's the revised dot plot of her quiz scores.



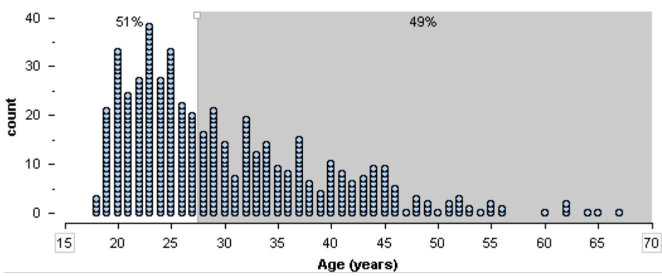
   a) What is the shape of this distribution?
   b) Find her revised median quiz score. Is it less than, more than, or about the same as her previous median quiz score? Why does this make sense?

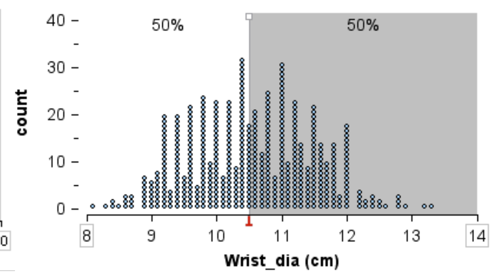
   c) Find her revised mean quiz score. Is it less than, more than, or about the same as her previous mean quiz score? Why does this make sense?

   d) Which measure of center (mean or median) would best represent her typical performance on the quizzes? Explain.

3) Write a few sentences that explain how the shape of the distribution might influence whether the mean is larger than, smaller than or about the same as the median.

4) For each distribution give an estimate for the median. Then say whether the mean is probably greater than, less than, or about equal to the median. Jot down some notes to explain how you figured this out.





5) Which of the following distributions is likely to have a mean that is smaller than the median? Jot down some notes to explain how you figured this out.

a) scores on the an OLI Checkpoint, where almost all of the grades were A's and B's (scores of 80 to 100), with one student making a zero.

b) repeated weighings of "one-pound" bags of peanut M&M's using a digital bathroom scale that reports weights to the nearest tenth of a pound.

c) salaries of NBA basketball players (This distribution is strongly right skewed by the few superstars who make much more money than the rest of the players.)

6) According to the U.S. Bureau of Census, the median family income in the United States in 2008-2012 was $53,046. Why do you think the median, rather than the mean, was reported by the U.S. Bureau of Census?

7) Summarize what you have learned about how to choose between the mean and median as the best measure of center based on the shape of the distribution.

Module 6 Measures of Spread about the Median

Module 6.1 Quantifying Variability Relative to the Median Part 1

Learning Goals:
- Create and interpret different graphs of a quantitative variable.
- Summarize and describe the distribution of a quantitative variable in context. Describe the overall pattern (shape, center and spread) and striking deviations from the pattern.

Specific Learning Objectives:
- Use quartiles to quantify variability relative to the median.
- Create and interpret boxplots, relate boxplots to histograms and dotplots.

1) Recall that the median is one way to summarize a distribution with a single number. Half of the data lies above the median and half lies below it. The median can be an actual data value or it can be a halfway mark between two values.

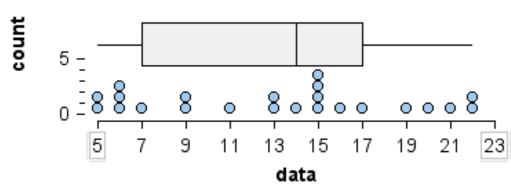
   a) Find the median for this set of 23 numbers:

   5 5 6 6 6 7 9 9 11 13 13 14 15 15 15 15 16 17 19 20 21 22 22

Statisticians measure spread relative to the median by marking the quartiles. Quartiles divide the data into four groups, with 25% of the data in each group (or as close to 25% as possible). The first quartile mark (Q1) has 25% of the data below it (or as close to that as possible). The second quartile mark (Q2, the median) has 50% of the data below it (or close to that as possible). The third quartile mark (Q3) has 75% of the data below it (or close to that as possible). You can think of Q1 as the median of the lower half of the data. Similarly, Q3 is the median of the upper half of the data.

   b) The 5-number summary is the minimum, Q1, Q2, Q3, maximum. Find the 5-number summary for the data set above.

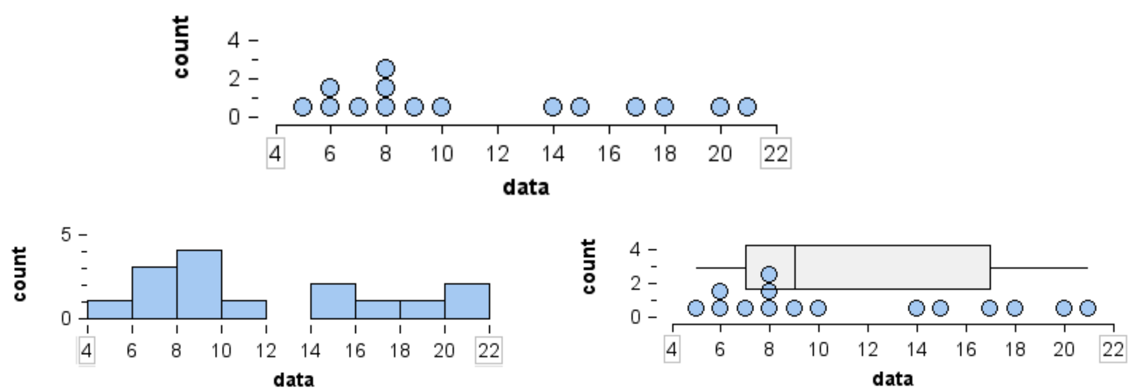
   c) The 5-number summary can be used to make a boxplot. Here is the boxplot for this set of numbers. Study it and see if you can explain how to make a boxplot based on your work in part b.



2) Draw a number line and make a boxplot to represent a data set with this 5-number summary: 3, 7, 9, 15, 20. (Make sure your number line has a constant scale. For example, you could mark off the number line in increments of 2.)

3) Here is a data set with its dot plot. We also made a histogram and a boxplot from the same data.

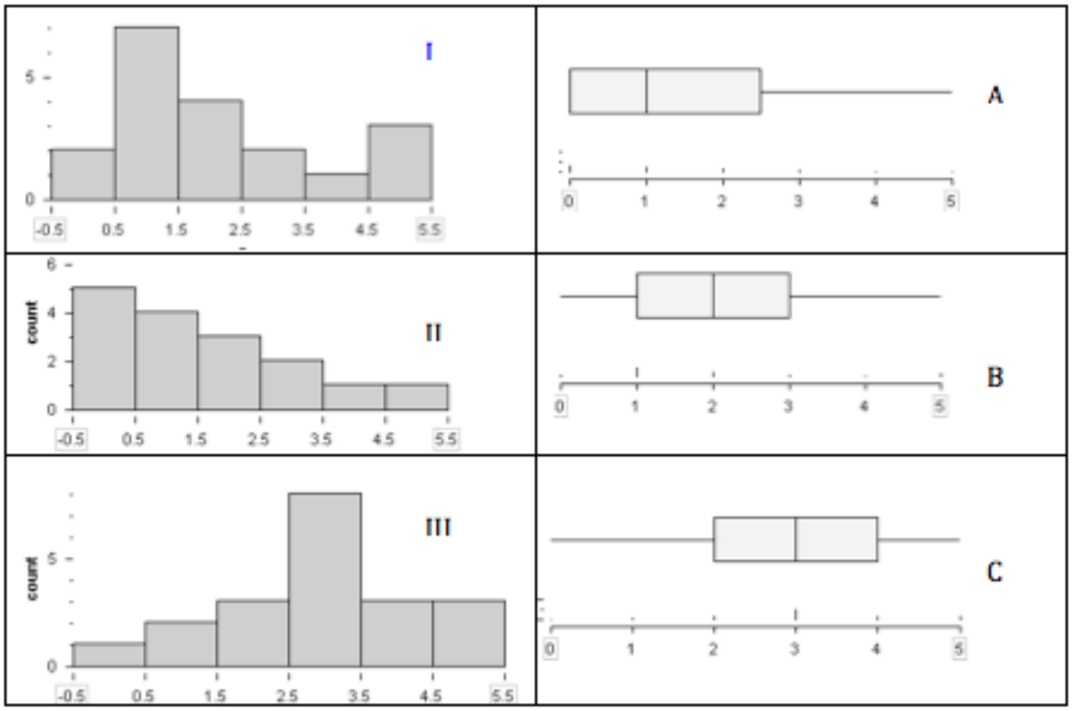
5 6 6 7 8 8 8 9 10 14 15 17 18 20 21



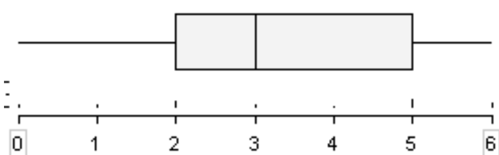
a) The graph that divides the data into four groups of equal counts is (circle one: histogram, boxplot). The graph that divides the data into bins of equal widths is (circle one: histogram, boxplot).

b) What is the 5-number summary for this data set? (Hint: you should be able to look at the boxplot to determine this.)

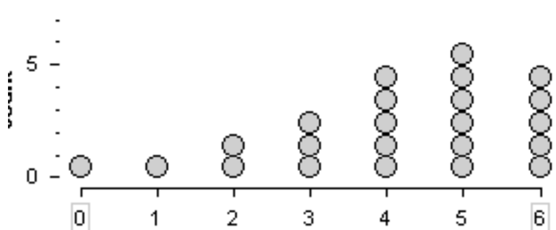
4) Match the histograms to the boxplots.



5) Make up a data set with 11 numbers that matches this boxplot. Make a histogram of your data.



6) Draw a boxplot for the data shown in the dot plot.



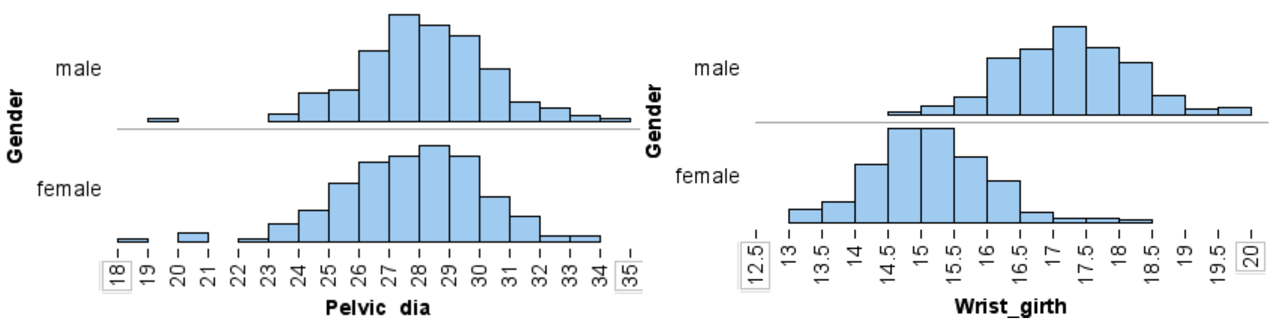
Module 6.2 Quantifying Variability Relative to the Median Part 2

Learning Goal: Compare distributions from two or more groups

Specific Learning Objective: Create and interpret boxplots to compare distributions.

Overview: In this activity we return to comparing and describing distributions. We will again use the ideas of shape, center and spread (as well as noting outliers.). But this time we will use the 5-number summary (Min, Q1, Q2, Q3, Max) to make these ideas more precise.

1) Here is data from 247 men and 260 women who exercise regularly. Pelvic diameter is a measurement from hipbone to hipbone in centimeters. Wrist girth is a measurement around the wrist, also in centimeters.



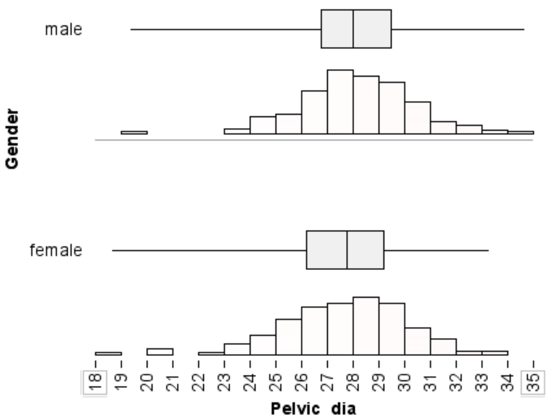
   a) Fill in the blanks with either pelvic diameter or wrist girth.

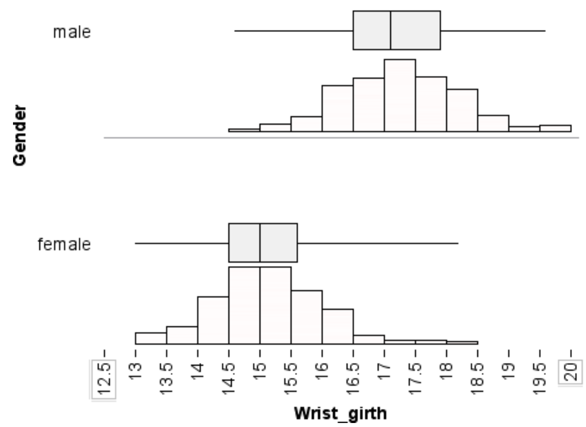
For the variable ________________________________, men and women have similar measurements.
For the variable ________________________________, men are substantially larger than women.

Previously, when we compared distributions of a quantitative variable, we described the shape and gave eyeball estimates of the center (a single typical measurement) and of the spread (both overall range and an interval of typical measurements.) Now we can make these estimates more precise using the 5-number summary as follows:

- The median can represent a typical measurement.
- The interval Q1 to Q3 can represent an interval of typical measurements.
- The interquartile range (IQR), which is Q3 minus Q1, describes the variability in the middle half of the distribution.

Here are the boxplots and the 5-number summaries for these distributions.





Pelvic Diameter (cm)						
Gender	N	Min	Q1	Q2	Q3	Max
Female	260	18.7	26.2	27.8	29.2	33.3
Male	247	19.4	26.8	28	29.5	34.7

Wrist Girth (cm)						
Gender	N	Min	Q1	Q2	Q3	Max
Female	260	13	14.5	15	15.6	18.2
Male	247	14.6	16.5	17.1	17.9	19.6

b) For the distributions that you identified as similar, compare the shapes, centers (medians and intervals of typical measurements) and spread (IQR and overall range). Make other observations using quartiles that support your conclusion. Then write a paragraph (or paragraphs) summarizing

Module 6.3 Module 6 Lab

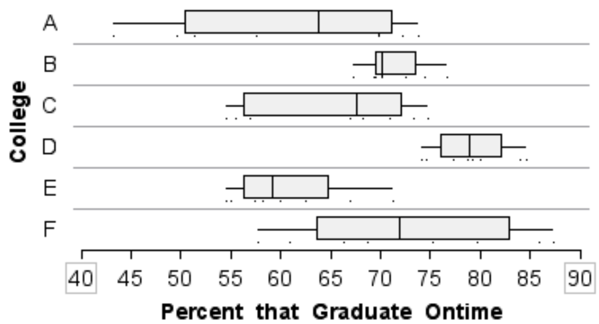
Unit 2 Module 6 Lab Assignment

Name: ______________________________

Learning Goal: Compare distributions from two or more groups.

Specific Learning Objective: Create and interpret boxplots to compare distributions.

1) The data graphed here is 8 years of data from six colleges. The variable is the percent of freshmen that graduate on-time.



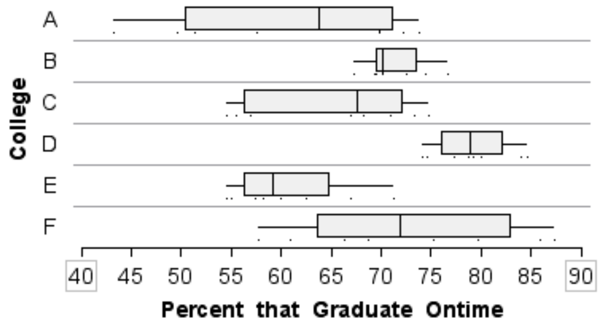
College	n	Min	Q1	Q2	Q3	Max
A	8	43.2	50.5	63.75	71.1	73.8
B	8	67.3	69.5	70.15	73.5	76.7
C	8	54.5	56.25	67.65	72.15	74.8
D	8	74.1	76	79	82.1	84.6
E	8	54.5	56.25	59.15	64.8	71.3
F	8	57.7	63.7	72	82.85	87.4

a) Which college had the highest on-time graduation rate during the years of this study? How do you know?

b) On average, which college had the highest on-time graduation rate? How do you know?

c) Now let's consider the variability in the data. Which college had the most consistent on-time graduation rates overall? Write a sentence precisely describing the variability for this college. (Use the 5-number summaries in the table.)

d) Which college had the smallest amount of variability relative to its median? In other words, which college has the least amount of variability in the middle half of its data? Write a sentence precisely describing the variability for this college. (Use the 5-number summaries in the table.)



College	n	Min	Q1	Q2	Q3	Max
A	8	43.2	50.5	63.75	71.1	73.8
B	8	67.3	69.5	70.15	73.5	76.7
C	8	54.5	56.25	67.65	72.15	74.8
D	8	74.1	76	79	82.1	84.6
E	8	54.5	56.25	59.15	64.8	71.3
F	8	57.7	63.7	72	82.85	87.4

For all of the colleges in this data set there is variability in on-time graduation rates from year to year. When we compare the colleges, there is also a lot of overlap in the distributions. Yet we can still distinguish substantial differences among some of them.

e) Pick two colleges that you think differ substantially in on-time graduation rates. Use the boxplots (and 5-number summaries) to provide at least 3 observations to support your choice.

f) Pick two colleges that you think have a similar distribution of on-time graduation rates. Use the boxplots (and 5-number summaries) to provide at least 3 observations to support your choice.

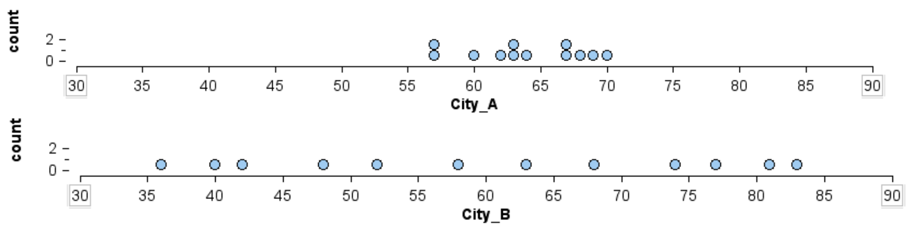
Module 7 Quantifying Variability Relative to the Mean

Module 7.1 Measuring Variability Relative to the Mean.

Learning Objective: Distinguish between graphs with large or small standard deviation using the concept of average deviation from the mean.

Warm-up:

1) The dot plots below show the average monthly high temperatures for New York City and San Francisco over a period of 10 years.



a) Is San Francisco City A or City B? How do you know?

b) One city has a median of 60.5°F; the other has a median of 63.5°F. Which is the median monthly high temperature in San Francisco? How do you know?

c) One city has an IQR of 6.5°F; the other has an IQR of 30.5°F. Which is the IQR for San Francisco? How do you know?

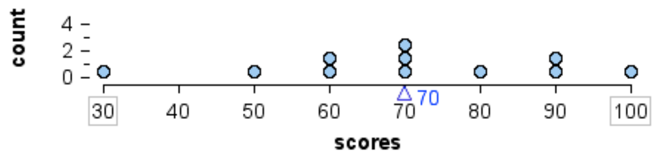
d) Here are the 5-number summaries for the two cities. Give intervals of typical average monthly high temperatures for the two cities.

City	Min	Q1	Q2	Q3	Max
A	57	61	63.5	67.5	70
B	36	45	60.5	75.5	83

e) Draw boxplots above the dot plots to summarize the monthly high temperatures.

The IQR is a way to measure variability relative to the median. How do we measure variability relative to the mean? That is the question we will investigate next. We will not start with a formula; instead we will work to build our intuition using graphs.

2) Here are exam scores for 11 students. The mean score is 70 points out of 100.



Which score varies the most from the mean? What is this score's distance from the mean?

Which score varies the least from the mean? What is this score's distance from the mean?

How far above the mean is the highest score?

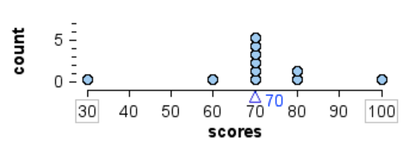
How many students had scores that vary 10 points from the mean?

Statisticians invented **standard deviation** to measure variability about the mean. ***The standard deviation is roughly the average distance that the data points vary from the mean.***

To estimate the standard deviation we will use the average distance from the mean. Fill in the table by finding the distance of each data point from the mean. What is the average distance that scores vary from the mean?

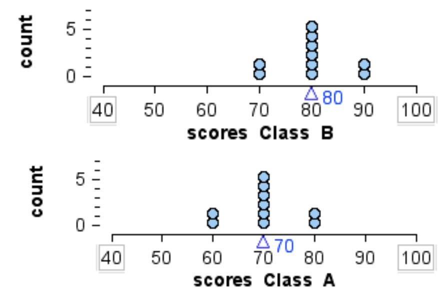
Data	30	50	60	60	70	70	70	80	90	90	100
Distance from mean		20						10			

3) Compare this graph of exam scores to the graph in #2. Here the mean is also 70. Do you think the average distance from the mean will be larger or smaller or the same? Why?

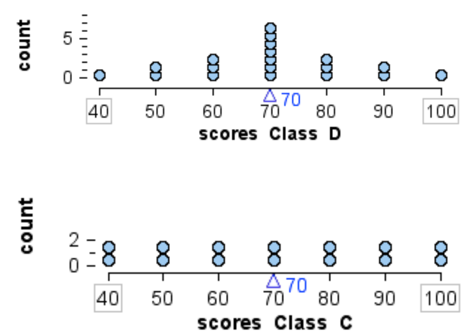


Group work: To develop your intuition about deviation from the mean, try to answer the following questions first without calculating anything, then check your intuition by calculating the average distance from the mean.

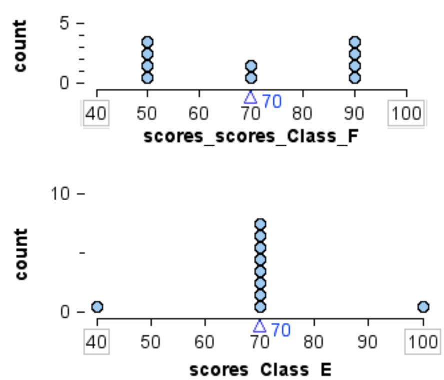
4) For each dot plot, the triangle marks the mean. Which class has the smaller average distance from the mean exam score? Or are the average distances from the mean equal? Why do you think so?



5) For each dot plot, the triangle marks the mean. Which class has the smaller average distance from the mean exam score? Or are the average distances from the mean equal? Why do you think so?



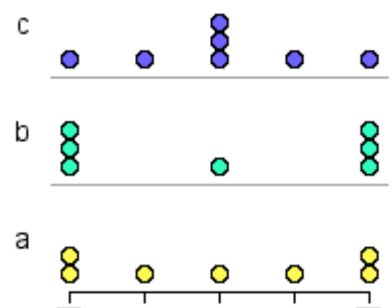
6) For each dot plot, the triangle marks the mean. Which class has the smaller average distance from the mean exam score? Or are the average distances from the mean equal? Why do you think so?



7) Draw a dot plot of 5 scores that has a mean of 70 and an average distance from the mean of zero. Is there more than one way to do this? Why or why not?

8) For the dot plots below we have removed the numerical values. We did this so that you can develop your intuition without doing any calculations. All three data sets have the same mean and are graphed on the same scale.

Which data set has the most variability about the mean? How do you know?



Which has the least? How do you know?

9) What have you learned so far about deviation from the mean? When comparing two graphs, what tips do you have for identifying the graph with the smaller deviation from the mean?

Module 7.2 The Standard Deviation

Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objective: Estimate and calculate the standard deviation from the mean.

The *average deviation from the mean* (ADM) is a rough estimate of the *standard deviation from the mean* (SD). Here are the formulas.

$$ADM = \frac{\sum |x-\bar{x}|}{n}, \qquad SD = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}.$$

Warm-up:

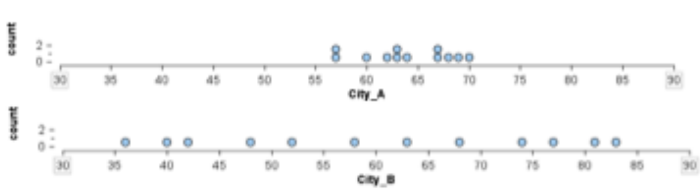
1) Use the standard deviation formula above to calculate the standard deviation for the following set of numbers. The mean $\bar{x}$ is 10.

$$\{5, 6, 10, 11, 18\}$$

Group work:

We will use technology to find the standard deviation of a data set, instead of calculating it by hand. So here we will practice problems that focus on the concept, instead of the mechanics.

2) Recall the City A is San Francisco and City B is New York City. The data shown here is the average of the highest temperatures for each month over a period of 10 years. If we calculate the SD for each distribution, we get 4.4 and 16.6 degrees. Which is the SD for San Francisco? How do you know? (See if you can answer this without doing any calculations!)

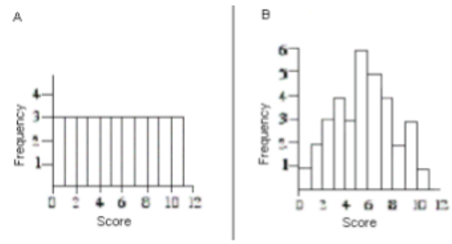


3) Which do you think will have a larger standard deviation? Why?

   a. The amount that a random sample of 30 LMC students spend per unit.

b. The amount that a random sample of 30 college students in the U.S. spend on per unit

4) Which distribution has the smaller standard deviation? Explain how you made your decision.



5) If the standard deviation of quiz scores on the Checkpoint 2.4 is zero, what do we know? Jot down a few notes to capture your thinking.

a. everyone made a 100% on the quiz
b. everyone failed the quiz
c. everyone made the same score on the quiz
d. it is impossible to tell

6) Describe in words what the standard deviation measures. (Think about how we have been estimating it in previous activities.)

Module 7.3 The Mean and Standard Deviation: Intervals of Typical Measurements

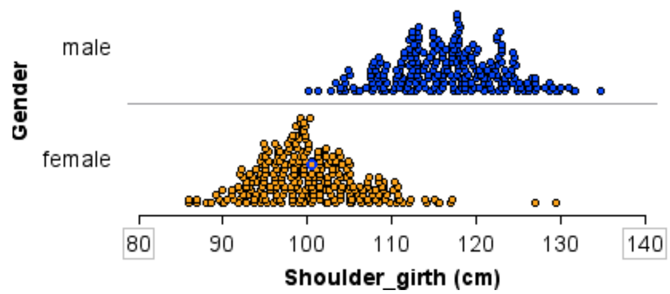
Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objective: Use the mean and standard deviation to create intervals of typical measurements.

When we compare two distributions, we want to describe an interval of typical data values. If we use the median as an average, we can use Q1 to Q3 to give an interval of typical data values. If we use the mean as an average, we can use the interval within one SD of the mean: (mean – SD, mean + SD).

Example: Here we have data on shoulder girth measurements for 247 men and 260 women who exercise regularly.

1) For this sample of men and women, would you argue that men tend to have shoulder girths that are larger than women? Why or why not?



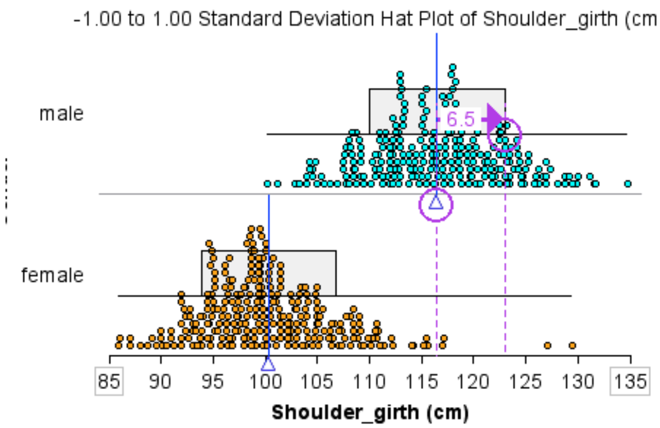
2) The mean shoulder girth for men is 116.5 cm vs. 100.3 cm for women. This tells us that on average men have larger shoulders. But there is variability in the data. To take variability into account, we can give an interval of typical measurements for each gender, instead of relying on a comparison of just the mean.

   Typical men have shoulder measurements within one standard deviation of the mean. The standard deviation for men is 6.5 cm.
   Mean – SD = 116.5 – 6.5 = 110
   Mean + SD = 116.5 + 6.5 = 123
   Based on this data, typical men have a shoulder girth between 110 and 123 cm.

   The standard deviation in women's shoulder girths is also 6.5cm. Find an interval of typical shoulder measurements for women using the mean and SD:

3) Do the intervals of typical measurements overlap? How does this observation support your answer to (1)?

4) Tinkerplots uses the mean and standard deviation (and the min and max) to draw a Standard Deviation Hat Plot.



   a) Label the mean and the standard deviation in the female distribution.

   b) Describe how to draw a Standard Deviation Hat Plot.

   c) What part of the hat plot represents typical measurements?

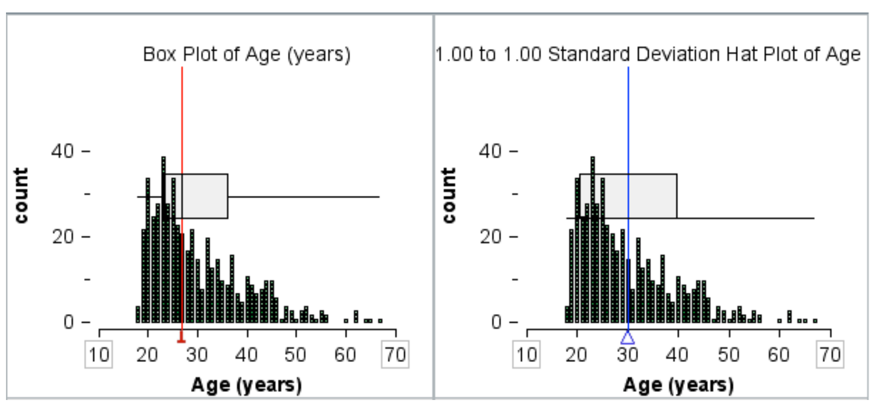
Group work:

5) The people in this sample are 18 to 67 years old. Here is a summary of the distribution of ages: Mean = 30, SD = 9.6; 5-number summary: 18, 23, 27, 36, 67

   a) Give an interval of typical ages for this sample using quartiles.

   b) Give an interval of typical ages for this sample using mean and SD.

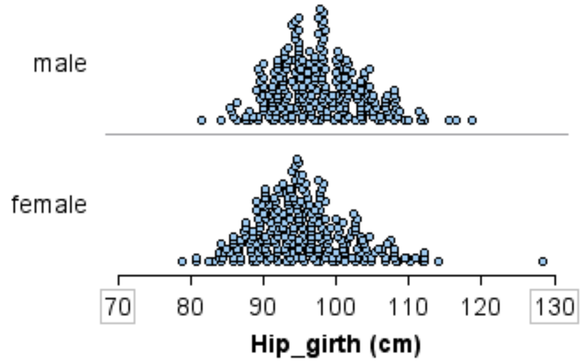
   c) Here is a boxplot and a standard deviation hat plot for this age distribution. In the boxplot label the median and other quartiles. In the standard deviation hat plot label the mean and SD.



   d) The median is not in the center of the box in the boxplot, but the mean is in the center of the standard deviation hat plot. Will the mean always be in the center of a standard deviation hat plot? Why or why not?

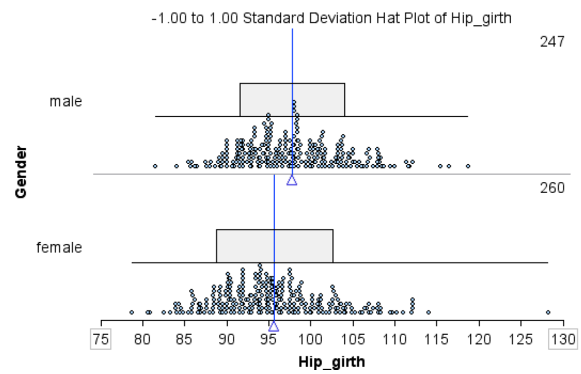
6) Hip girth is the distance around the widest part of the hips, given in centimeters.

a) For this sample, do you think men's hip measurements tend to be (larger, smaller, or about the same) as women's hip measurements? Why?



b) Calculate an interval of typical measurements using the mean and the standard deviation (mean ± SD) for males and females. These intervals are represented by the rectangle in the Standard Deviation Hat Plot.

Gender	Mean	SD
Male	97.8 cm	6.2 cm
Female	95.7 cm	6.9 cm



c) Do men tend to have larger hips than women? Use descriptions of shape, center (mean), and spread (SD) to support your choice. Incorporate your intervals of typical measurements from part (b), too.

Module 7.4 Module 7 Lab

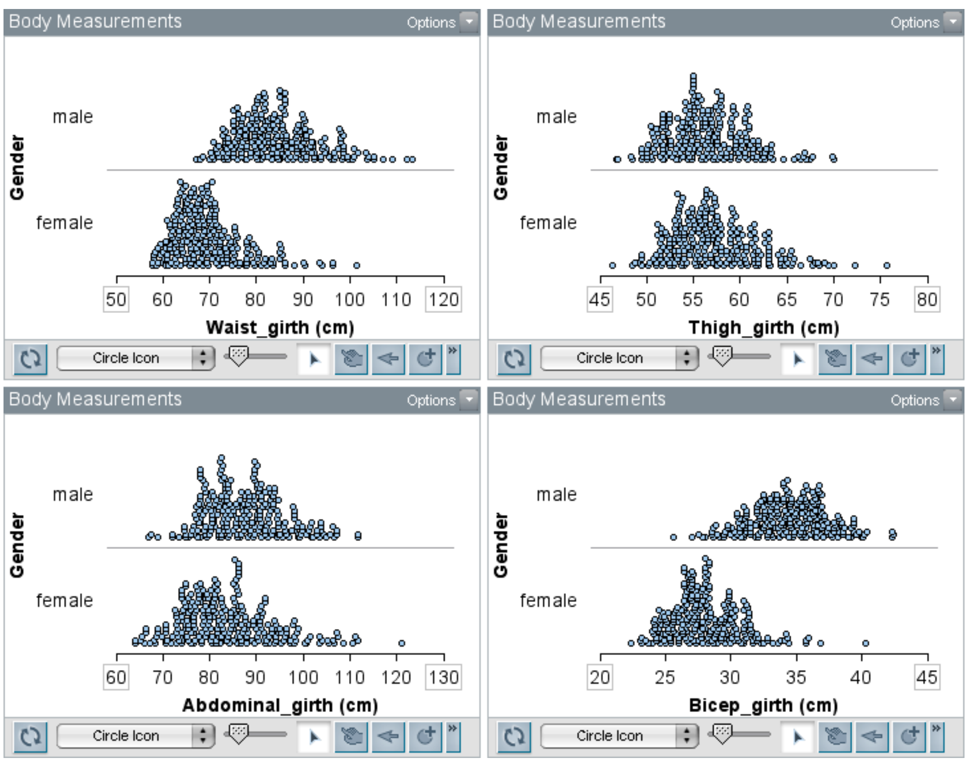
Name: ______________________________

Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objectives:

- Use the mean and standard deviation to create intervals of typical measurements.
- Use technology to generate graphs and numerical summaries to compare distributions of a quantitative variable.
- Compare and contrast different ways of identifying intervals of "typical" data values.

1) Here are the distributions for data we have on waist girth, thigh girth, abdominal girth and bicep girth measurements for 247 men and 260 women who exercise regularly.



a) For which of the four variables would you argue that that men tend to be larger than women? Why?

b) For which of the four variables would you argue that that men tend to be about the same size as women? Why?

c) Here are the means and SDs for each variable in centimeters. For three of the four variables males have a larger mean. **But using a single number (the mean) to represent the distribution does not take into account the variability in the data.** So we will calculate intervals of typical measurements for men and for women (mean ± SD) and look at the overlap (or lack of overlap) to compare the distributions.

	Waist girth		Thigh girth		Abdominal girth		Bicep girth	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Males	84.5	8.8	56.5	4.3	87.7	8.4	34.4	3.0
Females	69.8	7.6	57.2	4.6	83.7	9.9	28.1	2.7

Typical **waist** girth males: Typical **waist** girth females:

Substantial overlap in typical waist measurements? (Circle one: yes, no)

Typical **thigh** girth males: Typical **thigh** girth females:

Substantial overlap in typical thigh measurements? (Circle one: yes, no)

Typical **abdominal** girth males: Typical **abdominal** girth females:

Substantial overlap in typical abdominal measurements? (Circle: yes, no)

Typical **bicep** girth males: Typical **bicep** girth females:

Substantial overlap in typical bicep measurements? (Circle: yes, no)

d) Do the intervals you calculated in c) support your answers to a) and b)? Explain.

2) Suppose that you are designing a one-size-fits-most unisex belt for adults. A one-size- fits-most belt cannot fit everyone, but we want it to fit "typical" adults. We will use technology to identify "typical" intervals of waist measurements for the 507 adults in the data set *Body Measurement.txt* using ideas from Unit 2.

   a) For the waist girth, find an interval of typical measurements using quartiles.

   b) Find an interval of typical measurements using mean and standard deviation.

   c) Which interval do you think is the best to use in this situation to identify typical measurements? Why?

3) Open the *Body Temp and Heart Rate* data set.

   d) What heart rate do you think best represents the men in this sample? What about the women? Did you choose a mean or a median or some other number? Why?

   e) Which gender has more variability in their heart rates? Support your answer.

   f) Give an interval of "normal" heart rates for each gender using either quartiles or mean and SD. Briefly explain why you chose the measures you used.

   g) Based on this data, what do you think nurses should use as an interval of typical heart rates for adults? Should they use different intervals for men and women? Why or why not?

Module 7.5 Unit 2 Project

Instructions: In your group, choose one of the four options described below. Analyze the data using what you have learned in Unit 2. Each group will make a poster and present their analysis in a gallery walk. Based on feedback from the gallery walk, each group will revise their work and present an improved analysis to the class. Your instructor may also require each group member to write an analysis and submit it individually.

Poster Instructions: Your poster will include the following:

- A statement of the research question
- A description of the source of the data
- A description of the variables used in the analysis and an explanation of why your group chose these variables
- Graphs (with clear labels) and numerical summaries to support your analysis
- Explanations that reflect the use of Unit 2 concepts
- An answer to the question based on your analysis of the data

Option 1: Movies

Research question for Option 1: Which genres of movies are more successful: Action/Adventure movies or other types of movies?

Answer this question using the data set *Movies.txt*. Choose at least one quantitative variable that relates to your definition of success. Compare the success of action/adventure movies to other movies using the variable(s) you chose. Follow the instructions for the poster above.

Option 2: Body Temperatures

In current medical practice, 98.6°F is considered a normal body temperature, but this is an average. Normal body temperatures will vary throughout the day and are sensitive to hormone levels. However, an abnormally high or an abnormally low temperature may be a sign of illness.

Research question for Option 2:

- According to this data, what is a normal range for adult body temperature? What temperatures would you label as abnormally high or abnormally low?
- Should there be different normal temperature benchmarks for men and women?

To answer these questions analyze the adult body temperature data in *Body Temperature and Heart Rate.txt*. Follow the instructions for the poster above.

Option 3: Breakfast Cereals

Research question for Option 3: Are cereals targeted at children less healthy than cereals targeted at adults?

Answer this question using the data set *Cereals.txt*. Choose at least one quantitative variable that relates to your definition of healthy or unhealthy. Compare adult and child cereals using the variable(s) you chose. Follow the instructions for the poster above.

Option 4: Unisex Belt

Suppose that you are designing a one-size-fits-most unisex belt for adults. The belt should be long enough to fit typical adults without having an excessive amount of extra length.

Research questions for Option 4:

- What is the length of your belt? Where along the length of the belt are the holes and why?
- Based on this data, approximately what proportion of adults will be able to wear your belt?

Answer these questions using the data set *Body Measurement.txt*. Choose at least one quantitative variable that is relevant to the design of your belt. Follow the instructions for the poster above.

DEFINITION OF VARIABLES

Option 1: Movies in *Movies.txt*

This data set describes 75 movies listed in the top 100 USA box office sales of all time. Data was taken from IMDb.com in Spring 2014.

Variable **Variable Definition**
Year Year
Studio Name Studio Name
Studio Type Studio Type (Big 6, Other)
Genre Genre (Action/Adventure, Other)
Budget (millions $) Budgeted Cost to Produce (millions $)
US Box Office (millions $) US Box Office Revenues (millions$)
First Week End (millions $) First Week End Gross Box Office Revenues (millions $)
Movie_Length (minutes) Length (minutes)
Trailer_Length (seconds) Trailer Length (seconds)
Director Name of the Director
Director_Gender Gender of the Director
Director_Race Race of the Director
Star Name of the Star
Star_Gender Gender of the Star
Star_Race Race of the Star
Costar Name of the Costar
Costar_Gender Gender of the Costar
Costar_Race Race of the Costar
IMDb_Rating How it was rated by IMDb
Metascore How it was rated by Metascore
Metacriticcom_rating How it was rated by Metacriticcom
Rotten_Tomatoes How it was rated by Rotten Tomatoes
Number_of_Oscars Number of Oscars Won
Oscar_Nominations Number of Oscar Nominations
Oscar_Winner Did this movie win an Oscar?

Option 2: *Body Temperature and Heart Rate.txt*

Gender (male, female)
Temperature (degrees F)
Heart Rate (Beats per minute)

Option 3: *Cereals.txt*

Manufacturer: Manufacturer of cereal
Type: Cereal type (hot or cold)
Shelf: Display shelf at the grocery store
Target: Target audience for cereal (Child or Adult)
Calories: Calories per serving
Cups: Number of cups in one serving
Weight: Weight in ounces of one serving
Protein: Grams of protein in one serving
Fat: Grams of fat in one serving
Sodium: Milligrams of sodium in one serving
Fiber: Grams of dietary fiber in one serving
Carbs: Grams of complex carbohydrates in one serving
Sugars: Grams of sugars in one serving
Potassium: Milligrams of potassium in one serving
Vitamins: Vitamins and minerals - 0, 25, or 100% of daily need in one serving
Rating: Consumer Reports overall rating of nutritional value

Option 4: Body measurements in *Body Measurement.txt*

Variable **Variable Definition**
Gender Gender (male, female)
Age Age (Years)
Height Height (Centimeters)
Weight Weight (Kilograms)
Pelvic_dia Pelvic Diameter (Centimeters)
Chest_depth Chest Depth (Centimeters)
Chest_dia Chest Diameter (Centimeters)
Elbow_dia Elbow Diameter (Centimeters)
Wrist_dia Wrist Diameter (Centimeters)
Knee_dia Knee Diameter (Centimeters)
Ankle_dia Ankle Diameter (Centimeters)
Shoulder_girth Shoulder Girth (Centimeters)
Chest_girth Chest Girth (Centimeters)
Waist_girth Waist Girth (Centimeters)
Abdominal_girth Abdominal Girth (Centimeters)
Hip_girth Hip Girth (Centimeters)
Thigh_girth Thigh Girth (Centimeters)
Bicep_girth Bicep Girth (Centimeters)
Forearm_girth Forearm Girth (Centimeters)
Knee_girth Knee Girth (Centimeters)
Calf_girth Calf Girth (Centimeters)
Ankle_girth Ankle Girth (Centimeters)
Wrist_girth Wrist Girth (Centimeters)

UNIT 3

Examining Relationships: Quantitative Data

Contents

Module 8 Scatterplots, Linear Relationships, and Correlation

Module 8.1 Scatterplots and Correlation

Learning Goal: Use a scatterplot to display and describe the relationship between two quantitative variables.

Specific Learning Objectives:
- Identify explanatory and response variables;
- Read and interpret scatterplots;
- Identify direction, strength and form in scatterplots.

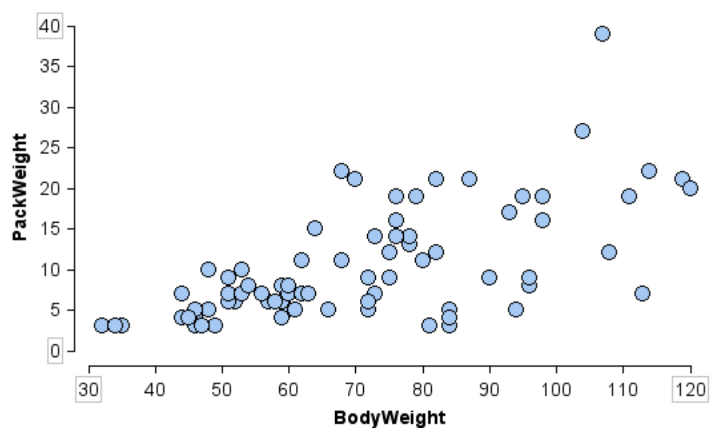
Introduction:

Previously we looked at data that comes from taking one quantitative measurement for each individual in a group. Now we will take two quantitative measurements for each individual and look at how the two measurements relate to each other. We will use a scatterplot to graph both measurements for each individual.

Here is a partial spreadsheet of body weights (in pounds) and backpack weights (in pounds) for a sample of elementary school students. The scatterplot contains data for all 79 students in the data set.

Backpack

	Name	BodyWeight	PackWeight
1	Angie	45	4
2	Emma	46	4
3	Sadie	32	3
4	Maddyn	47	3
5	Lorien	60	7
6	Bailey	52	6
7	Micah	57	6
8	Kilie	48	10
9	Abigail	46	3
10	Eugene	34	3



1) Do students who weigh more tend to carry more weight in their backpacks? Why do you think so?

Response variable: This is the variable that measures the outcome of a study. The response variable is also known as the dependent variable. It goes on the vertical axis in the scatterplot.

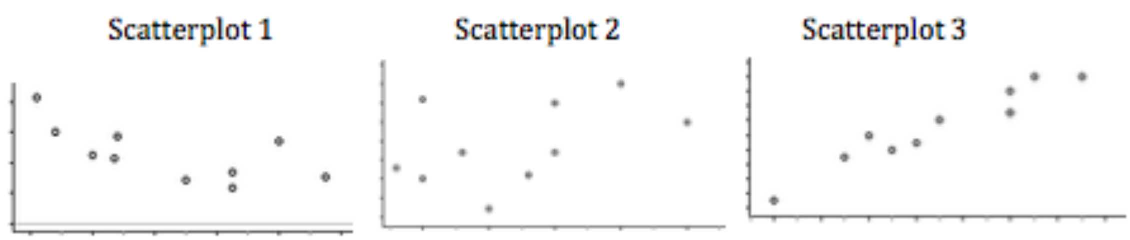
Explanatory variable: The explanatory variable may explain or contribute to the response variable; for this reason, we often use it to predict the values of the response variable. The explanatory variable is also known as the independent variable. It goes on the horizontal axis in the scatterplot.

2) Identify which variable goes on the horizontal axis in the scatterplot.

   a) For a group of students, a teacher compares homework grades to exam grades.

   b) For people at a bar, a researcher measures blood alcohol content (BAC) and the number of alcoholic drinks consumed.

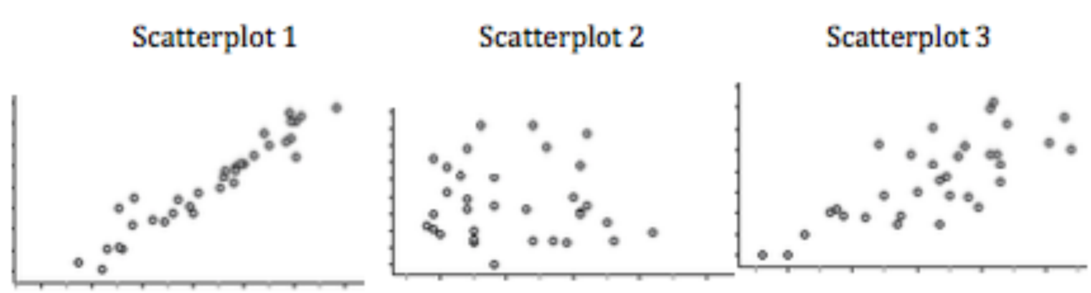
3) The scatterplots below differ in the DIRECTION of the association. The direction can be positive (upward) or negative (downward) or neither.



   a) Label each scatterplot as positive association, negative association, or neither.

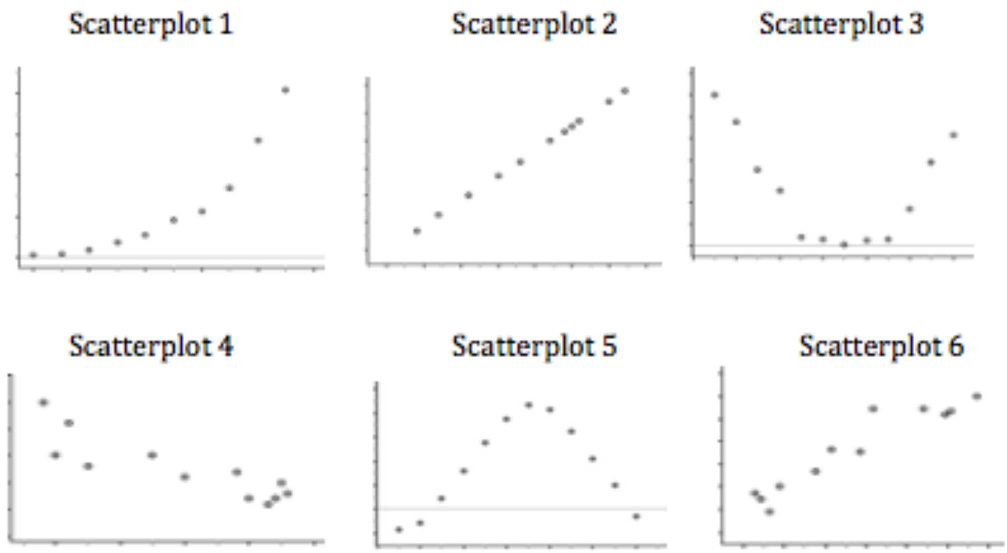
   b) For a group of people at a bar, do you think the relationship between BAC and number of alcoholic drinks will be positive, negative or neither? Why do you think so?

4) The scatterplots below differ in the STRENGTH of the association. The strength is how closely the data follow a pattern. Rank the scatterplots from strongest to weakest association.



5) These scatterplots differ in FORM. Some have a linear pattern; some have a curved pattern (also called curvilinear).

   a) Identify the scatterplots with a linear form.



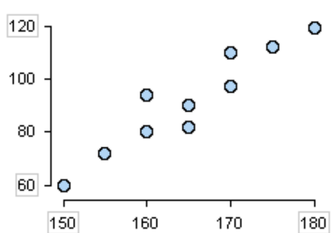
Module 8.2 Introduction to Linear Correlation: r

Learning Goal: Use a scatterplot to display the relationship between two quantitative variables. Describe the overall pattern and striking deviations from the pattern.

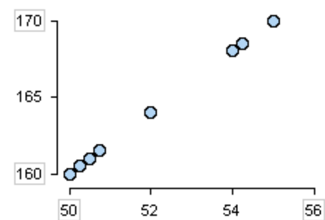
Learning Objective: Use r to describe strength and direction for a linear relationship.

Our goal in this activity is to look for patterns to try to figure out what the correlation r measures. *Below there are 9 scatterplots and 9 values of the correlation r. Try to assign each scatterplot an r-value from the list, using the following properties of r: 1) r is positive when the scatterplot is headed uphill and negative when the scatterplot is headed downhill; 2) the tighter the scatterplot is clustered about a line, the closer r is to 1 (headed uphill) or -1 (headed downhill).*

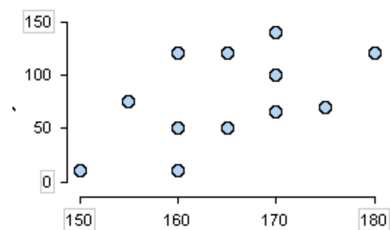
Scatterplot 1



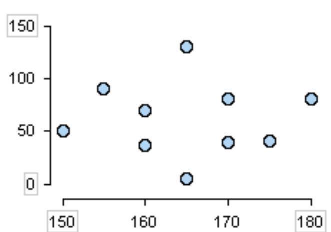
Scatterplot 2



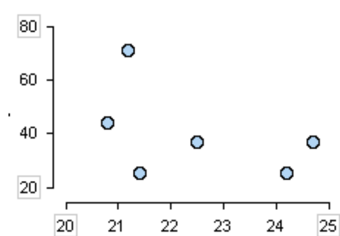
Scatterplot 3



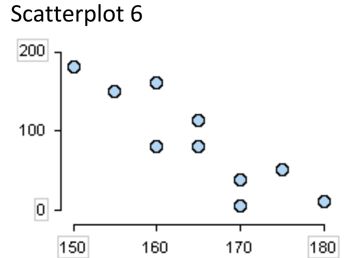
Scatterplot 4



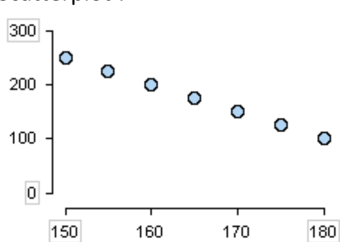
Scatterplot 5



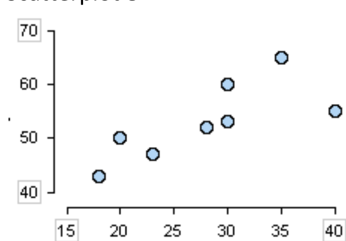
Scatterplot 6



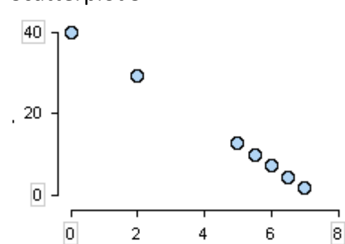
Scatterplot 7



Scatterplot 8



Scatterplot 9



Correlation coefficients (r-value)	-1	-1	-0.44	0.01	0.55	0.75	0.88	0.94	1

Module 8.3 Correlation is not Causation

Learning Goal: Distinguish between association and causation. Identify lurking variables that may explain an observed relationship.

A *lurking variable* is a variable that is not measured in the study. It is a third variable that is neither the explanatory nor the response variable, but it affects your interpretation of the relationship between the explanatory and response variable.

1) To understand the above ideas, read this excerpt from *A Mathematician Reads the Newspaper* by John Allen Paulos.

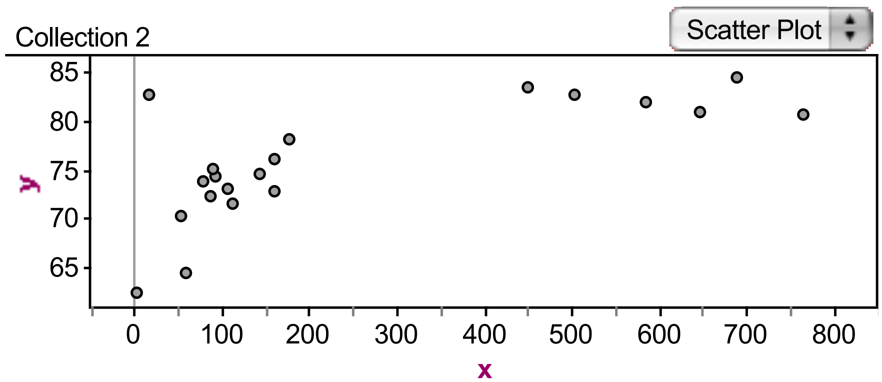
"A more elementary widespread confusion is that between correlation and causation. Studies have shown repeatedly, for example, that children with longer arms reason better than those with shorter arms, but there is no causal connection here. Children with longer arms reason better because they're older! Consider a headline that invites us to infer a causal connection: BOTTLED WATER LINKED TO HEALTHIER BABIES. Without further evidence, this invitation should be refused, since affluent parents are more likely both to drink bottled water and to have healthy children; they have the stability and wherewithal to offer good food, clothing, shelter, and amenities. Families that own cappuccino makers are more likely to have healthy babies for the same reason. Making a practice of questioning correlations when reading about "links" between this practice and that condition is good statistical hygiene." (p. 137)

a) In this example, "children with longer arms reason better than those with shorter arms," what is the explanatory variable? The response variable? The lurking variable?

b) Explain what it means to say "there is no causal connection" between these two variables.

c) What is "good statistical hygiene" to Paulos?

2) For the 20 countries with the largest population for 2009 the scatterplot shows
   x = internet users per 1000 people
   y = life expectancy (years)
   (World Almanac Book of Facts, 2009)



The correlation coefficient is 0.72, which is strong. Larger numbers of internet users per 1,000 correlate with longer life expectancy. Someone who confuses correlation with causation might suggest that an easy way to improve a country's life expectancy is to get more people onto the internet, which is a ridiculous cause-and-effect statement. Identify a lurking variable that might be explaining the strong association between life expectancy and the number of internet users per 1,000.

Module 9 Fitting a Line

Module 9.1 Introduction to Linear Regression

Learning Goal: For a linear relationship, use the least squares regression line to summarize the overall pattern and to make predictions.

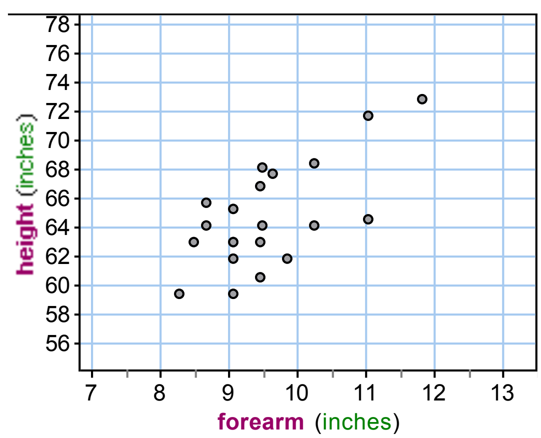
Introduction: Statistical methods are used in forensics to identify human remains based on the measurements of bones. In the 1950's Dr. Mildred Trotter and Dr. Goldine Gleser measured skeletons of people who had died in the early 1900s. From these measurements they developed statistical methods for predicting a person's height based on the lengths of various bones. These formulas were first used to identify the remains of U.S. soldiers who died in WWII and were buried in unmarked graves in the Pacific zone. Modern forensic scientists have made adjustments to the formulas developed by Trotter and Gleser to account the differences in bone length and body proportions of people living now. We will not use Trotter and Gleser's formulas in this problem, but we will use a similar process.

Amelia Earhart disappeared in 1937 while flying over the Pacific Ocean. In 1941 bones comprising about a third of a human body were discovered on an uninhabited Pacific island called Nikumaroro. Could these bones be the bones of the aviation pioneer Amelia Earhart?

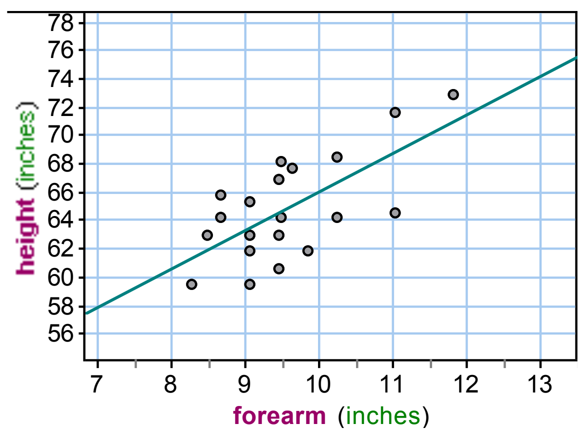
In 1941 an M.D. named Dr. Hoodless measured the bones. The length of the radius, which is the forearm bone, was 9.6 inches. Amelia Earhart said she was $5'8''$ tall, but other records suggest she might have been closer to $5'7''$.

1) Here is a scatterplot of the data collected from 21 female college students.

   a) Assuming the bone found on Nikumaroro belonged to a woman, how tall do you think she was? Why do you think so?



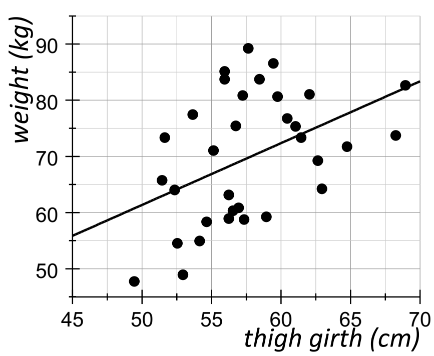
b) Use the line to predict the height of the person whose bones Dr. Hoodless measured. Plot this person in the scatterplot.
height = 39 + 2.7 (forearm)



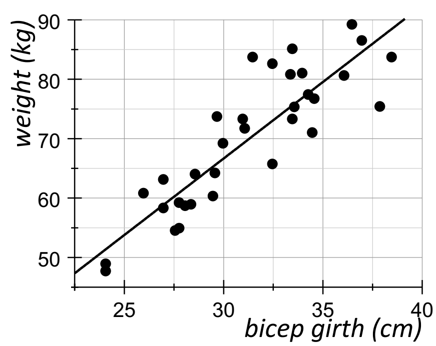
c) Based on this investigation, do you think the bones could belong to Amelia Earhart? Why or why not?

(Dr. Hoodless concluded that the bones belonged to a short, stocky European male. The bones are now missing, but modern analysis of the notes taken by Dr. Hoodless has revealed many discrepancies and his conclusions are hotly debated.)

2) The scatterplots below show body measurements in centimeters for 34 adults who are physically active. The line graphed in each scatterplot is the line of best fit. The equation for each line is given below the scatterplot.



Weight = 7.85 + 1.07 (thigh girth)



Weight = -13.45 + 2.67 (bicep girth)

a) Adriana has a thigh girth of 57 centimeters and a bicep girth of 25 centimeters. Predict Adriana's weight using both measurements. Show or explain your process.

b) Which prediction do you think is more accurate? Why?

Module 9.2 Interpreting the Constants

Learning Goal: For a linear relationship, use the least squares regression line to summarize the overall pattern.

Learning Objective: Interpret the rate of change (slope) and initial value (y-intercept) for regression lines.

Introduction:

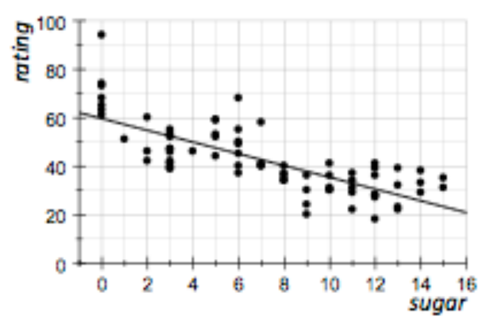
The y-intercept of the regression line is the predicted initial value of y when x is 0.

The slope is the predicted change in y divided by the change in x. It can be interpreted as the rate that our predictions for y change for each 1 unit increase in x.

Example:
Here are the Consumer Report ratings for 77 breakfast cereals. Sugar is measured in grams per serving. The Consumer Report rating is on a scale of 0 to 100, with 100 being the best score.

The equation of the regression line is *Predicted rating = 60 – 2.43(sugar).*

Let's interpret the meaning of the initial value in this context. The initial value (y-intercept) is 60. This is the prediction when x=0.



Predicted rating = 60 – 2.43·(*sugar*)

X is sugar (grams per serving) → (0, 60) ← Y is predicted CR rating

This tells us that if a cereal has 0 grams of sugar per serving, the predicted Consumer Report rating is 60.

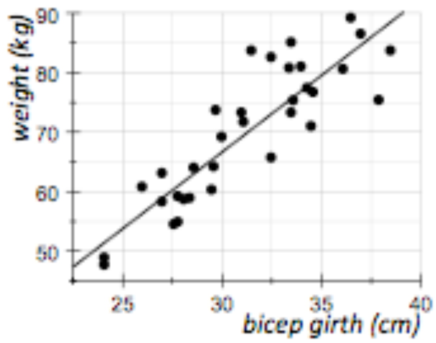
Now let's interpret the meaning of the rate of change (slope). The slope is –2.43. Think of this number as a rate (a ratio) $\dfrac{-2.43}{1}$

→ Change in Y: CR rating predictions
→ Change in X: grams of sugar in a serving

This rate tells us that when sugar amount increases from x to x+1, the predicted Consumer Report rating drops 2.43 points. In other words, our predictions for Consumer Reports ratings decrease 2.43 points for each additional gram of sugar in a serving of cereal.

Practice:

1) The regression line shown in the scatterplot has the equation: *Weight = -13.45 + 2.67 (bicep girth)*

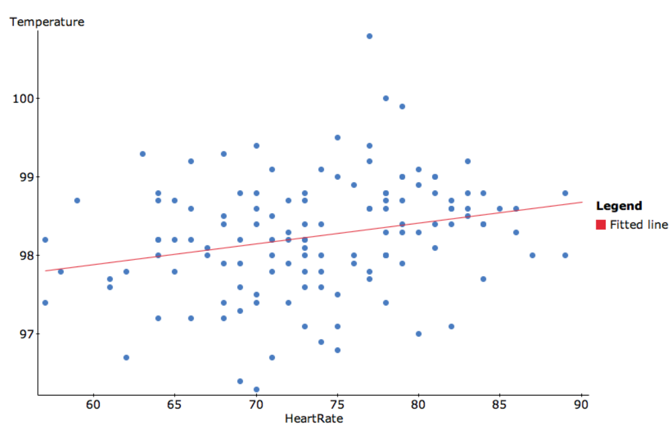


Interpret the y-intercept and the slope for the regression line using the context for the data.

2) For the 130 adults in this sample, heart rate is measured in beats per minute. Body temperature is measured in degrees Fahrenheit. StatCrunch gives the following linear regression results:

Dependent Variable: Temperature
Independent Variable: HeartRate
Temperature = 96.306754 + 0.026334549 HeartRate
Sample size: 130
R (correlation coefficient) = 0.2536564

Interpret the y-intercept and the slope for the regression line using the context for the data.



Module 9.3 Lab Assignment

Unit 3 Summary Lab Assignment

Name: ____

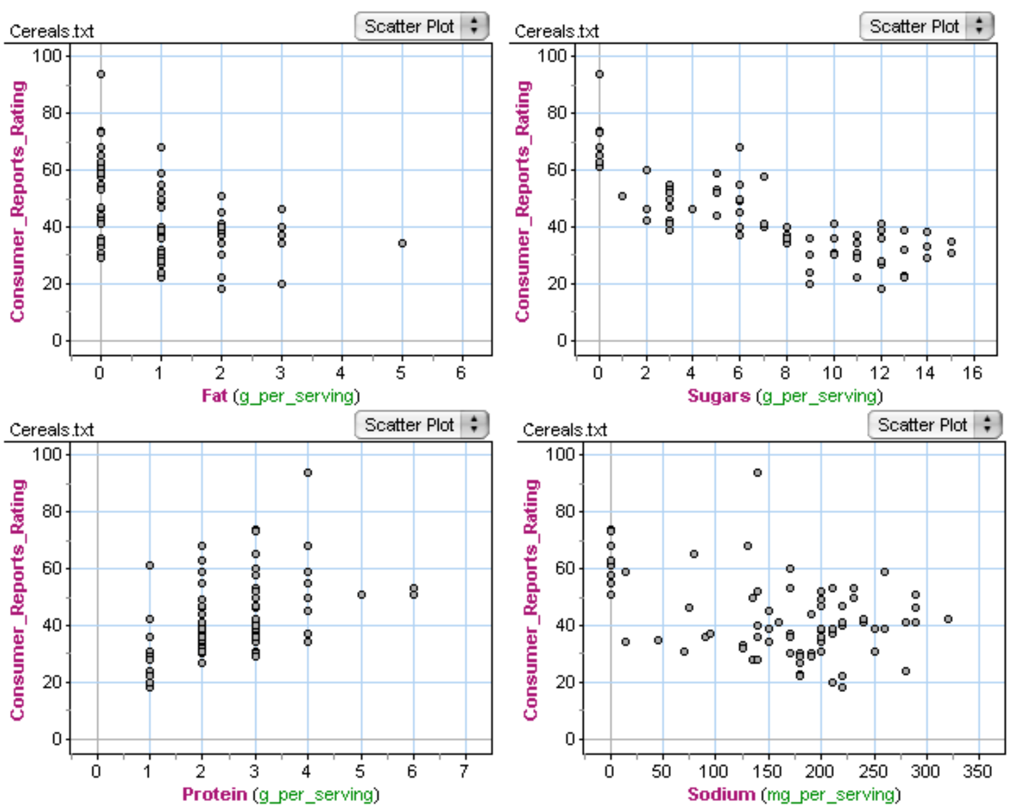
Learning Goal:

- Use a scatterplot to display the relationship between two quantitative variables. Describe the overall pattern and striking deviations from the pattern.
- For a linear pattern, use the least squares regression line to summarize the overall pattern and to make predictions.

Specific Learning Objective:

- Read and interpret scatterplots
- Describe the pattern in a scatterplot as positive or negative association, if appropriate.
- Identify the rate of change and initial value for predicted values in a least squares regression line.
- Interpret the rate of change (slope) and initial value (y-intercept) for regression lines.

1) These scatterplots show the relationship between various ingredients and Consumer Report Ratings for some breakfast cereals. Consumer Reports is a non-profit group that rates products to help consumers make smart purchases. Their rating system is based on an undisclosed formula.



Referring to the scatterplots on the previous page:

a) Captain Crunch has the lowest *Consumer Reports* rating of the 77 cereals in the data set. How much fat is in a serving of Captain Crunch?

b) In this set of 77 cereals, Product 19 has the most sodium in a serving. What is the *Consumer Reports* rating for Product 19?

c) All-Bran Extra Fiber is the cereal with the highest rating. How much sugar, fat, and sodium are in a serving of All-Bran Extra Fiber?

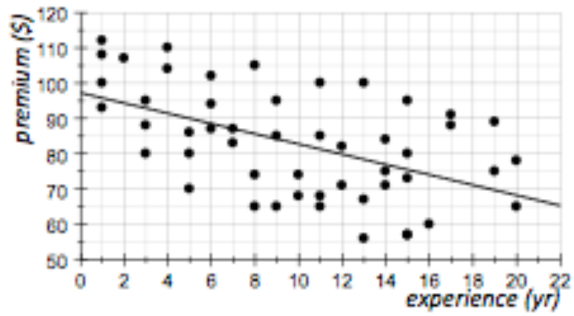
d) Which ingredients (fat, sugar, protein, sodium) are positively associated with Consumer Report Ratings? Which are negatively associated? Which have no association?

e) Do you think fiber would be positively or negatively associated with Consumer Report Ratings? Why?

2) In 2008, a statistics student gathered data on the monthly car insurance premiums paid by students and faculty at Los Medanos College. Her regression line is $\hat{y} = 97 - 1.45x$, where *y-hat* is the predicted monthly premium and *x* is years of driving experience.

Interpret the y-intercept **and** the slope for the regression line using the context for the data. In other words, your interpretations should refer to years of driving experience and monthly car insurance premiums.

Y-intercept:



Slope:

Module 9.4 Unit 3 Project

Instructions: In your group, chose one of the four options described below. Analyze the data using what you have learned in Unit 3. Each group will make a poster and present their analysis in a gallery walk. Based on feedback from the gallery walk, each group will revise their work and present an improved analysis to the class. Your instructor may also require each group member to write an analysis and submit it individually.

Poster Instructions: Your poster will include the following:

- A statement of the research question
- A description of the source of the data
- A description of the variables used in the analysis
- Scatterplots with clear labels that facilitate easy visual comparison
- Explanations that reflect the use of Unit 3 concepts (form, direction, strength, explanatory and response variables, regression equation, prediction, and r.
- An answer to the question based on your analysis of the data

Research Question 1:

Introduction: In the Movie data set, we have movie ratings from a variety of websites, such as *www.rottentomatoes.com*, *www.imdb.com*, and metascore ratings from *www.metacritic.com*.

Research questions (answer both):

- Which is a better predictor of Rotten Tomatoes movie ratings: IMDb ratings or Metascore ratings?
- For the variable that correlates with Rotten Tomatoes ratings more strongly, what is the predicted Rotten Tomatoes rating for a movie that has an IMDb rating of 7 and a Metascore of 80? Use Unit 3 concepts to comment on the accuracy of your prediction.

Investigate this question using the data set *Movies.txt*. Support your answer using concepts from Unit 3. Follow the instructions for creating a poster.

Research Question 2:

Introduction: Pediatricians use a child's measurements to predict his or her height and weight as an adult. Researchers studying child development measured a sample of 136 children at ages 2, 9, and 18. These children were born in 1928–29 in Berkeley CA. We will examine the relationships between body measurements for children and the corresponding measurement for an 18-year-old.

Use the data in the file titled *Child Development* to answer one of the following questions.

Research questions (choose one):

1. Can we more accurately estimate an 18-year-old's height or an 18-year-old's weight using the corresponding measurements for 9-year-olds?
2. Which is a better predictor of an 18-year-old's height: childhood height at age 2 or childhood height at age 9?
3. Is strength at age 9 a good way to predict strength at age 18? For which gender is strength at age 9 a better predictor of strength at age 18?

Support your answer using concepts from Unit 3. Follow the instructions for creating a poster.

After answering your research question, use the stronger relationship that you found to make a prediction using a regression line. Explain what you are doing in a way that a parent without statistical training could understand. Incorporate this into your poster and presentation.

Research Question 3:

Introduction: In a course called Introduction to Statistics at Carnegie Mellon University, a professor wants to gain insight into his students' performance on the final exam by analyzing exam grades from earlier in the semester. The data set describes 105 students. The data is in *gradebook.xls*. Support your answers using concepts from Unit 3. Follow the instructions for creating a poster.

Research questions (answer both):

- In general, which is the best predictor of a student's score on the final exam, the score on the first midterm or the second midterm? Use Unit 3 concepts to support your answer.
- What is the predicted score on the final exam for a student who scores a 77 on the first midterm and an 88 on the second midterm?

Support your answers using concepts from Unit 3. Follow the instructions for creating a poster.

Research Question 4:

Introduction: In forensic science, the identification of a dead body is often based on an incomplete skeleton. Forensic scientists use ideas from Unit 3 and other statistical techniques to predict the height and weight of a dead person based on measurements of bones or body parts. Use the file *Body Measurement.txt* to answer the following questions.

Research questions (answer both):

- In a war zone a foot and an ankle are recovered but no other parts of the body are found. The ankle girth is 25 centimeters. What is the predicted height and the predicted weight of this person? (Remember to always plot the data before making predictions!)
- Which is a more accurate prediction? Why do you think so?

Support your answer using concepts from Unit 3. Follow the instructions for creating a poster.

Definition Of Variables

Option 1: Movies in *Movies.txt*

This data set describes 75 movies listed in the top 100 USA box office sales of all time. Data was taken from IMDb.com in Spring 2014.

Variable **Variable Definition**

Year Year

Studio Name Studio Name

Studio Type Studio Type (Big 6, Other)

Genre Genre (Action/Adventure, Other)

Budget (millions $) Budgeted Cost to Produce (millions $)

US Box Office (millions $) US Box Office Revenues (millions $)

First Week End (millions $) First Week End Gross Box Office Revenues (millions $)

Movie_Length (minutes) Length (minutes)

Trailer_Length (seconds) Trailer Length (seconds)

Director Name of the Director

Director_Gender Gender of the Director

Director_Race Race of the Director

Star Name of the Star

Star_Gender Gender of the Star

Star_Race Race of the Star

Costar Name of the Costar

Costar_Gender Gender of the Costar

Costar_Race Race of the Costar

IMDb_Rating How it was rated by IMDb

Metascore How it was rated by Metascore

Metacriticcom_rating How it was rated by Metacriticcom

Rotten_Tomatoes How it was rated by Rotten Tomatoes

Number_of_Oscars Number of Oscars Won

Oscar_Nominations Number of Oscar Nominations

Oscar_Winner Did this movie win an Oscar?

Option 2: Body measurements in the file *Child_development*

Data from a study by Tuddenham, R. D., & Snyder, M. M. 1954. Physical growth of California boys and girls from birth to eighteen years. Child Development 1:183-364.

*Bodytype rates a person on a scale of slender (1) to fat (7).

Variable **Variable Definition**

CaseNo Number assigned to child

Gender Child's gender

Weight_2 Age 2 weight (kg)

Height_2 Age 2 height (cm)

Weight_9 Age 9 weight (kg)

Height_9 Age 9 height (cm)

*Leg_9*Age 9 leg circumference (cm)

Strength_9 Age 9 strength (kg)

Weight_18 Age 18 weight (kg)

Height_18 Age 18 height (cm)

Leg_18 Age 18 leg circumference (cm)

Strength_18 Age 18 strength (kg)

Bodytype_18 1-7 scale*

Option 3: Grades in *gradebook.xls*

Variable **Variable Definition**

Midterm1 Student's score on the first midterm (0-100 scale)

Midterm2 Student's score on the second midterm (0-100 scale)

Diff.Mid The difference between the two midterm exam scores (midterm1 - midterm2)

Extra credit Did the student turn in the extra credit assignment? (0=NO, 1=YES)

Final Student's score on the final (0-100 scale)

Class Student's class (1=Freshman, 2=Sophomore, 3=Junior, 4=Senior)

Option 4: Body measurements in *Body Measurement.txt*

These data were collected to investigate the correspondence between frame size, girths, and weight of physically active young men and women, most of whom were within normal weight range. A goal of this investigation was to develop predictive techniques to assess the lean/fat body composition of individuals. The individuals were not randomly sampled. Measurements were initially taken by Grete Heinz and Louis J. Peterson at San Jose State University and at the U.S. Naval Postgraduate School in Monterey, California. Later, measurements were taken at dozens of California health and fitness clubs by technicians under the supervision of one of these authors. See article in Journal of Statistics Education, Volume 11, Number 2 (2003), www.amstat.org/publications/jse/v11n2/datasets.heinz.html. Exploring Relationships in Body Dimensions, by Grete Heinz, Louis J. Peterson, and Carter J. Kerk.

Variable **Variable Definition**

Gender Gender (male, female)

Age Age (Years)

Height Height (Centimeters)

Weight Weight (Kilograms)

Pelvic_dia Pelvic Diameter (Centimeters)

Chest_depth Chest Depth (Centimeters)

Chest_dia Chest Diameter (Centimeters)

Elbow_dia Elbow Diameter (Centimeters)

Wrist_dia Wrist Diameter (Centimeters)

Knee_dia Knee Diameter (Centimeters)

Ankle_dia Ankle Diameter (Centimeters)

Shoulder_girth Shoulder Girth (Centimeters)

Chest_girth Chest Girth (Centimeters)

Waist_girth Waist Girth (Centimeters)

Abdominal_girth Abdominal Girth (Centimeters)

Hip_girth Hip Girth (Centimeters)

Thigh_girth Thigh Girth (Centimeters)

Bicep_girth Bicep Girth (Centimeters)

Forearm_girth Forearm Girth (Centimeters)

Knee_girth Knee Girth (Centimeters)

Calf_girth Calf Girth (Centimeters)

Ankle_girth Ankle Girth (Centimeters)

Wrist_girth Wrist Girth (Centimeters)

UNIT 4

Relationships in Categorical Data with an Introduction to Probability

Contents

Module 10 Two-Way Tables

Module 10.1 Relationships between Categorical Data

Learning goal: Use a two-way table to analyze the association between two categorical variables.

Introduction:

In the previous unit, we studied the association between two quantitative variables. Now we will examine the association between two categorical variables.

1) What is the difference between categorical and quantitative data? Give an example of each.

2) Which of the following survey questions yields categorical data?

   a) *Are you male or female?*
   b) *What is your height in inches?*
   c) *What is your college GPA?*
   d) *Where do you tend to sit in class: front, middle or back?*
   e) *Do you feel that you are overweight, underweight, about right in weight?*

3) Which of the following research questions involve an association between two categorical variables?

   a) *Are males or females more likely to feel that they are overweight?*
   b) *Is there a relationship between college GPA and where a student tends to sit in class?*
   c) *Is there a relationship between high school GPA and college GPA?*
   d) *Are students who sit at the front of class more likely to be shorter than students who sit at the back of class?*

4) Is there an association between gender and where college students sit in class? Let's see what this data suggests.

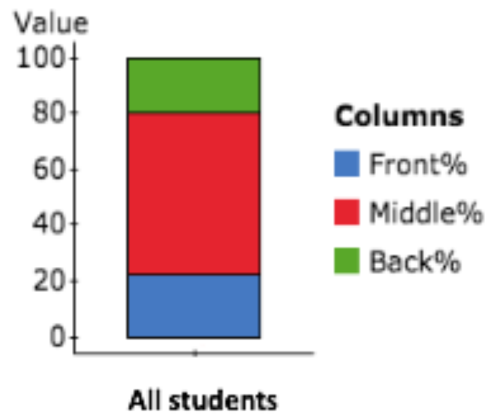
	Front	Middle	Back	Total
Female	37	91	22	150
Male	15	46	25	86
Total	52	137	47	236

a) Describe the distribution of class seating ignoring gender by calculating appropriate percentages and labeling the stacked barchart.

Of all the students in the study, what percentage sit in the front class?

In the middle of class?

In the back of class?



b) If gender is associated with class seating, we expect the seating distribution to be (circle one: the same, different) when we take gender into account.

c) Use percentages to describe the distribution of class seating for female students. These are called *conditional* percentages.

Of all the female students in the study, what percentage sit in the front class?

In the middle of class?

In the back of class?

d) Use *conditional* percentages to described the distribution of class seating for male students:

Of all the male students in the study, what percentage sit in the front class?

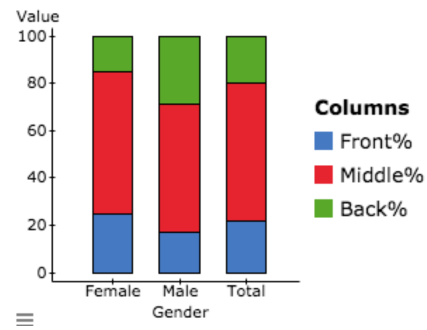
In the middle of class?

In the back of class?

e) Who is more likely to sit at the front of class, females or males?

Who is more likely to sit at the back of class, females or males?

f) Here are stacked bar graphs for all three class seating distributions: female, male, everyone.



This data suggests that gender is (<u>circle one</u>: not associated, weakly associated, strongly associated) with class seating for college students.

(Note: This is a judgement call. Later in the course will develop more concrete ways to determine if observed differences are significant.)

g) Why do we use percentages instead of counts to describe and compare the seating distributions?

5) Is there a relationship between a college student's major and whether they use the college's math tutoring services? Use this fictitious data to <u>calculate relevant percentages</u> to support your answer.

	Math tutoring NO	Math Tutoring YES	Total
STEM major	36	53	89
Non-STEM major	75	174	249
Total	230	108	338

a) What percentage of the students in this study use the college's math tutoring services?

b) Who is more likely to use math tutoring services, STEM or non-STEM majors?

c) This data suggests that major (STEM vs. non-STEM) is (<u>circle one</u>: not associated, weakly associated, strongly associated) with use of math tutoring services at this college.

Briefly support your choice.

Module 10.2 Introduction to Probabiity and Risk using Two-Way Tables

Learning Goal: Use a two-way table to estimate probability and risk.

Introduction to probability:

Previously, we used a two-way table to examine the association between two categorical variables. In this activity we will do the same calculations through a new lens; the lens is probability.

Later in the course we will study probability in more detail because it is the heart of statistical inference, but this is a natural place to begin to use probability language. We use the word probability to mean "likelihood" or "chance." For our purposes, probability describes the chance of a specific outcome if we are randomly selecting from an individual from a group.

When we ask questions such as "who is <u>more likely</u> to use math tutoring services, STEM or non-STEM majors," we are using probability language. Similarly, the question, "what percentage of the STEM majors use math tutoring" is the same as asking, "what is the probability that a STEM student uses the college's math tutoring services?" We do the same calculations to answer each of these questions.

We use the word *risk* when the probability describes a negative outcome. For example, we talk about the probability of winning the lottery and the risk of being struck by lightning.

1) This data comes from a Massachusetts study.

	Smoker Yes	Smoker No	Total
Low Birth Weight Yes	30	29	59
Low Birth Weight No	44	86	130
Total	74	115	189

a) For a smoker, what is the risk of having a low weight baby?

b) For a non-smoker, what is the risk of having a low weight baby?

c) How much does smoking increase the risk of having a low weight baby? What is the percent increase?

$$\textit{percent increase when smoking} = \frac{\textit{increase in risk}}{\textit{risk when not smoking}} =$$

Smoking increases the risk of having a low weight baby by _________%.

d) Other ways to compare risks:

- Calculate the percentage point difference by subtracting percentages:

  The risk associated with low birth weight is ______ percentage points higher for smokers.

- Calculate the relative risk by dividing the percentages:

  Smokers are _______ times more likely to have a low weight baby.

2) This table summarizes the results of a survey of firefighters in New York.

	No alcohol problems	Moderate to severe alcohol problems	Totals
Participated in 9/11 rescue	793	309	1,102
Did not participate in 9/11	441	110	551
Totals	1,234	419	1,653

a) Are the firefighters who participated in the 9/11 rescue at greater risk of alcohol problems? Support your answer.

b) How much does participating in the 9/11 rescue increase the risk of alcohol problems? What is the percent increase?

c) What is the relative risk of alcohol problems? Write a sentence to explain the relative risk.

3) StatCrunchU is a fictitious university of 46,000 students. In StatCrunch we can select a random sample of StatCrunchU students. Use this data to answer the questions for StatCrunchU students. Support your answers with appropriate calculations; include the ratios and explanations so that someone else can follow your reasoning.

	Student Loan no	Student Loan yes	Total
Job	17	65	82
No Job	77	41	118
Total	94	106	200

a) What is the probability that a StatCrunchU student has a student loan?

b) At StatCrunchU are students with jobs more likely to have student loans than students without jobs?

c) Are the following statements accurate based on this sample of StatCrunchU students? Support your answer.

   Students with a job are more than twice as likely to have a student loan than students without a job.

Module 10.3 Lab Assignment

Unit 4 Summary Lab Assignment

Name: ____

Learning Goal:

- Use a two-way table to analyze the association between two categorical variables.
- Create a hypothetical two-way table to answer complex probability questions.

Specific Learning Objective:

- Given a two-way table, calculate and interpret risk probabilities.
- Compare risk by calculating the percentage point difference in risk, the relative risk quotient, and percent increase or decrease in risk.

1) A study in Sweden looked at the impact of playing soccer on the incidence of arthritis of the hip or knee. They gathered information on former professional soccer players; people who played soccer but not professionally, and those who never played soccer.

	Professional soccer player	**Non-professional soccer player**	**Did not play soccer**	**Row Totals**
Arthritis	10	9	24	
No arthritis	61	206	548	
Column Totals				

a) Does this study suggest that the more soccer you play, the greater the risk of arthritis of the hip or knee? Be sure to identify the explanatory and response variables to inform how you construct your fractions for comparison.

b) What is the relative risk of developing arthritis if you play non-professional soccer? (Use non-players as the basis of your comparison.) Write a sentence that communicates the meaning of this number.

c) What is the relative risk of developing arthritis if you play professional soccer? (Use non-players as the basis of your comparison.) Write a sentence that communicates the meaning of this number.

2) The table summarizes the results from an observational study that followed a random sample of 8,474 people for about four years. All participants were free from heart disease at the beginning of the study. The variable Anger categorizes participants as Low Anger, Moderate Anger, High Anger based on the Speilberger Trait Anger Scale test, which measures how prone a person is to sudden feelings of anger. CHD stands for coronary heart disease, which is a count of the people who had heart attacks or needed medical attention for heart disease during the study.

	Low Anger	Moderate Anger	High Anger	Total
CHD	53	110	27	190
No CHD	3,057	4,621	606	8,284
Total	3,110	4,731	633	8,474

a) Does this data suggest that there is an association between anger and heart disease? Support your answer.

b) Use the method of your choice to compare the likelihood that a High Anger person has CHD to a Low Anger person. Write a sentence to communicate your results.

Module 10.4 Unit 4 Project

Module 10.5 Unit 4 Project

Instructions: In your group, choose one of the five options described below. Analyze the data using what you have learned in Unit 4. Each group will make a poster and present their analysis in a gallery walk. Based on feedback from the gallery walk, each group will revise their work and present an improved analysis to the class. Your instructor may also require each group member to write an analysis and submit it individually.

Poster Preparation Instructions: Your poster will include the following:

- A statement of the research question
- A description of the source of the data
- A description of the variables used in the analysis
- A contingency table (two-way table)
- Calculations of relevant percentages
- Pie charts or bar charts to support the analysis
- An answer to the question based on your analysis of the data

Option 1: Cheating
Research questions **(do both)**:

- Are college students willing to report cheating?
- Is the willingness to report cheating related to gender?

Investigate these questions for the students described in *body_image.xls*.

This data comes from a survey of university students at Carnegie Mellon University in Pittsburgh, PA.

Here are the survey questions and associated variables:
Are you a male or a female? Gender (male, female)
What is your height in inches? Height (in inches)
What is your GPA? GPA
What was your high school GPA? HS GPA
Where do you tend to sit in class? Seat (F=front, M=middle, B=back)
How do you feel about your weight? WtFeel (OverWt, AboutRt, UnderWt)
Would you report cheating if you witnessed it? (yes, no)

Option 2: Gender and Body Image

Research question: Do female college students tend to feel differently about their weight compared to male college students?

Investigate this question for the students described in *body_image.xls*.

This data comes from a survey of university students at Carnegie Mellon University in Pittsburgh, PA.

Here are the survey questions and associated variables:
Are you a male or a female? Gender (male, female)
What is your height in inches? Height (in inches)
What is your GPA? GPA
What was your high school GPA? HS GPA
Where do you tend to sit in class? Seat (F=front, M=middle, B=back)
How do you feel about your weight? WtFeel (OverWt, AboutRt, UnderWt)
Would you report cheating if you witnessed it? (yes, no)

Option 3: Breakfast cereals

Research question: Are cereals being intentionally marketed to children?

Investigate this question for the cereals described in *Cereals.txt*.

This data describes 75 breakfast cereals. A researcher collected this data from a popular grocery store chain.

Here is a description of the variables:
Name: Name of cereal
Manufacturer: Manufacturer of cereal
Type: Cereal type (hot or cold)
Shelf: Display shelf at the grocery store
Target: Target audience for cereal (Child or Adult)
Calories: Calories per serving
Cups: Number of cups in one serving
Weight: Weight in ounces of one serving
Protein: Grams of protein in one serving
Fat: Grams of fat in one serving
Sodium: Milligrams of sodium in one serving
Fiber: Grams of dietary fiber in one serving
Carbs: Grams of complex carbohydrates in one serving
Sugars: Grams of sugars in one serving
Potassium: Milligrams of potassium in one serving
Vitamins: Vitamins and minerals - 0, 25, or 100% of daily need in one serving
Rating: Consumer Reports overall rating of nutritional value

Option 4: Low Birth Weight

Research question: Is visiting a doctor during the early stages of pregnancy associated with a lower incidence of low birth weights?

Investigate this question for the women described in *low_birth_weight_study.txt*.

This data describes 189 pregnant women who participated in a medical study at a Massachusetts hospital.

Here are the variables:
AGE — Age of mother (in years)
LWT — Weight of mother at the last menstrual period (in pounds)
BWT — Birth weight of the baby (in grams)
Low_Wt — whether the baby was born weighing less than 2500 grams (No, Yes)
Smoker — Smoking status during pregnancy (No, Yes)
Labor — History of premature labor (No, Yes)
Visit — Did the mother visit a doctor during the first trimester of pregnancy (No, Yes)

Option 5: Movies

Research questions (choose one):
1. Are the Big 6 studios more likely to choose a person of color as a lead star?
2. Are women more likely to star in Action/Adventure movies or other types of movies?

Investigate these questions for the movies described in *Movies.txt*. This data set describes 75 movies listed in the top 100 USA box office sales of all time. Data was taken from IMDb.

Year: Year movie was released
Studio: Studios categorized as Big6 or Other
Studio Name: Name of studio producing the movie
Genre: Action/Adventure or Other
Budget: Movie budget in millions of dollars
US Box Office: total box office sales in millions of dollars
Opening Week: box office sales for opening week in millions of dollars
Movie Length: length of movie in minutes
Trailer Length: length of advertising trailer in seconds
Director: Name of the movie's director
Director Gender: male or female
Director Race: W (white) or POC (person of color)
Star: Name of the movie's lead star
Star Gender: male or female
Star Race: W (white) or POC (person of color)
Costar: Name of the movie's main costar
Costar Gender: male or female
Costar Race: W (white) or POC (person of color)
IMDb_Rating: Average IMDb user rating scale of 1-10
Metascore: Score out of 100, based on major critic reviews as provided by Metacritic.com
Metacriticcom_rating: Number of critic reviews used to calculate the Metascore
Rotten_Tomatoes: Score out of 100, based on authors from writing guilds or film critic associations
Number of Oscars: number of Oscars won by the movie
Oscar Nominations: number of Oscar nominations for the movie
Oscar Winner: whether the movie won an Oscar (yes or no)

UNIT 5

Probability and Probability Distributions

Contents

Module 11 Probability and Distributions

Module 11.1 Introduction to Probability

Learning Goal: Interpret probability as a long run relative frequency.

Introduction: Previously, in our discussion of categorical variables, we used the word *probability* to mean "likelihood" or "chance." The probability of an event is a number between 0 and 1. If the event never occurs (occurs 0% of the time), its probability is 0. If the event always occurs (occurs 100% of the time), its probability is 1.

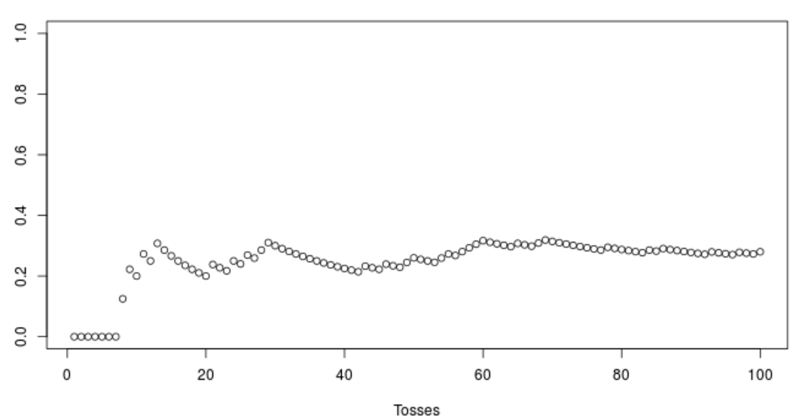
1) Consider the following three scenarios. Which of the three can we easily estimate the probability? Why is that?

- What is the probability that a toss of a fair coin lands on heads?
- What is the probability that the toss of a bottle cap lands right-side up?
- What is the probability that a randomly selected LMC student is a female?

2) A random event is unpredictable in the short run, but has a pattern in the long run that relates to probability. Your instructor will use an applet to simulate the flipping of an unfair coin. After the demonstration check your understanding by answering these questions:

a) Explain why flipping a biased coin is a random event.

b) Here a snapshot of the simulation for 100 tosses of an unfair coin. What is the probability of a getting a head with this coin?



3) If you pick a student at random from our class, what is the probability that she eats in the cafeteria today?

Module 11.2 Probability Distributions

Learning Goal: Use probability distributions to identify typical and unusual events and, when appropriate, to determine an expected value.

Introduction:

In our previous discussions of probability, we focused on determining the probability of one outcome at a time, such as estimating the probability of a head in the toss of a weighted coin. Now we shift our focus to describing the probabilities of all possible outcomes instead of the probability of just one outcome.

Statisticians describe outcomes as variable values. Each variable value has a probability. All possible variable values together with their probabilities are a *probability distribution*.

We can generate a probability distribution from theoretical probabilities (such as a fair coin or fair die) or from empirical probabilities (such as using relative frequency to find probabilities for a weighted coin or weighted die).

Examples of probability distributions:

<u>Probability distributions for a categorical variable</u>

The variable is *coin toss.* This is a categorical variable with variable values: Head, Tail. The probability distribution assigns each variable value (H or T) a probability.

Here is a probability distribution for a fair coin based on theoretical probabilities:

Coin Toss Outcome	Head	Tail
Probability	½= 0.5	½ = 0.5

Here is the probability distribution for a weighted coin based on empirical probabilities (relative frequencies):

Coin Toss Outcome	Head	Tail
Probability	201/1000=0.201	799/1000=0.799

Here is a probability distribution for blood type in the U.S. based on empirical probabilities (relative frequencies) from a large study by Stanford University's Blood Center (bloodcenter.stanford.edu):

Blood Type	O	A	B	AB
Probability	0.45	0.41	0.10	0.04

Check your understanding:

1) People with blood type O can donate blood to people with any other blood type. For this reason, people with blood type O are called universal donors.

    a) What is the probability that a randomly selected person from the United States is a universal donor?

    b) If there were 500,000 people in the Stanford study, how many had blood type O?

2) Based on these probability distributions, fill in the properties of a probability distribution below:

    - *All outcomes are assigned a probability. The probabilities are numbers between (and including) ______ and ________.* Why does this make sense?

    - *In a probability distribution, the sum of all of the probabilities is _________.* Why does this make sense?

Probability distributions for a quantitative variable

Here is the probability distribution for the roll of a fair die based on theoretical probabilities:

Die	1	2	3	4	5	6
Probability	1/6 ~ 0.167	1/6 ~ 0.167	1/6 ~ 0.167	1/6 ~ 0.167	1/6 ~ 0.167	1/6 ~ 0.167

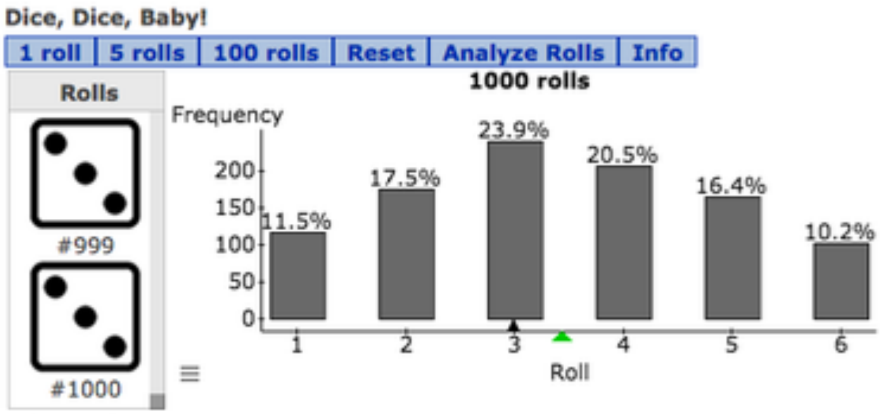
Here is the probability distribution for the number of Boreal owl eggs in a nest based on relative frequencies from large field studies in Canada.

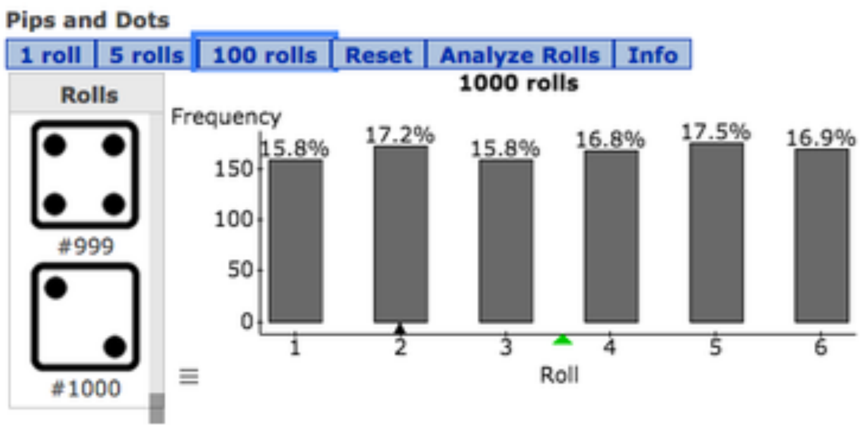
Number of eggs	0	1	2	3	4	5	6
Probability	0.20	0.10	0.10	0.25	0.25	0.05	0.05

Check your understanding:

3) Which is more likely: To find a boreal owl nest with 2 eggs or to find an empty boreal owl nest?

4) *Dice, Dice, Baby* and *Pips and Dots* are two companies that make dice. The graphs below show the probability distributions based on 1000 rolls of a die selected from each company. Is either company producing a fair die? Why do you think so?

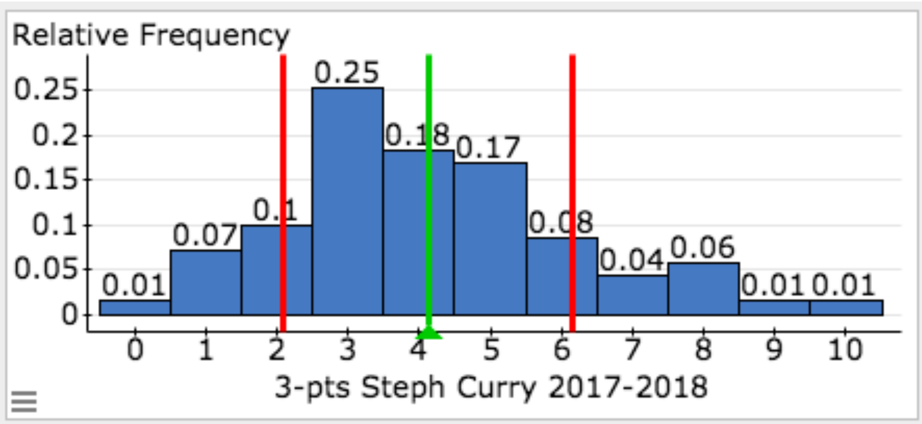




Describing the probability distribution for a quantitative variable

When the variable is quantitative, we can use shape, center and spread to describe the probability distribution. We can calculate a mean to describe the center and a standard deviation to describe the spread.

Here is an example. The variable is number of 3-point shots Steph Curry made during a game in 2017-2018 (including pre-season games.) The histogram of the probability distribution shows the mean and one standard deviation above and below the mean.



Summary statistics:

Column	n	Mean	Std. dev.
3-pts Steph Curry 2017-2018	71	4.1267606	2.0488992

Check your understanding:

5) Is it unusual for Steph Curry to not score a 3-point shot during a game? How do you know?

6) Typical values fall within one standard deviation of the mean.

   a) Give an interval that describes Steph Curry's typical number of 3-pointers during a game.

   b) What is the probability that Steph Curry hits within this interval during a game?

Expected value

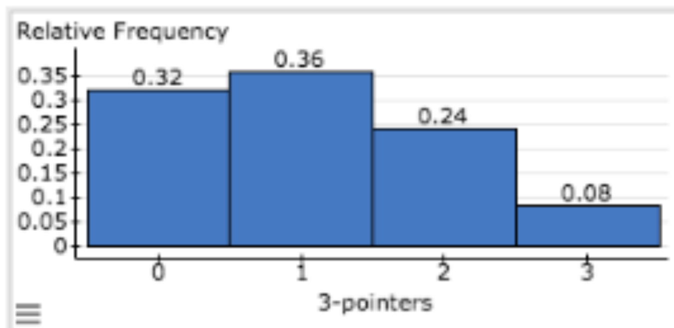
The mean of a probability distribution is called the *expected value.* It is a weighted mean, so values with a higher probability are more influential and values with a lower probability do not affect the expected value as much.

We will use technology to find the expected value, but, just this once, your instructor will demonstrate how to calculate the expected value by hand. The goal of this demonstration is to show how the expected value is a weighted mean.

7) Here is the number of 3-point shots made by a high school basketball player in 25 games, along with the associated frequency table and probability distribution.

0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 3 3

Number of 3-pointers scored per game	Frequency	Relative Frequency (Probability)
0	8	
1	9	
2	6	
3	2	



a) Estimate the expected value for 3-point shots for this player. Briefly explain what you are thinking.

b) Calculate the expected value (the weighted mean).

Module 12 Continuous Random Variables

Module 12.1 Probability Distributions for Continuous Random Variables

Learning Goals:

- Use a probability distribution for a continuous quantitative variable to estimate probabilities and identify unusual events.
- Use the mathematical model of a normal curve to estimate probabilities when appropriate.

Introduction

Previously, we examined probability distributions for categorical variables, such as blood type, and for quantitative variables, such as number of 3-pointers scored in a game.

For a quantitative variable, we can determine the mean (aka expected value) and standard deviation of the probability distribution.

In this activity we distinguish between two types of quantitative variables: discrete and continuous. For continuous variables we will learn to model the probability distribution with a mathematical curve. This curve will allow us to use technology to estimate probabilities.

Continuous vs. discrete quantitative variables

Discrete quantitative variables (also known as discrete random variables) have numeric values that can be listed and counted, for example number of 3-pointers scored in a game.

Continuous quantitative variables (aka c*ontinuous random variables*) have numeric values that cannot be listed and counted. Theoretically, we can measure a continuous variable to any desired accuracy, for example, height of a basketball player could be measured to the nearest foot, the nearest inch, the nearest ½-inch, etc.

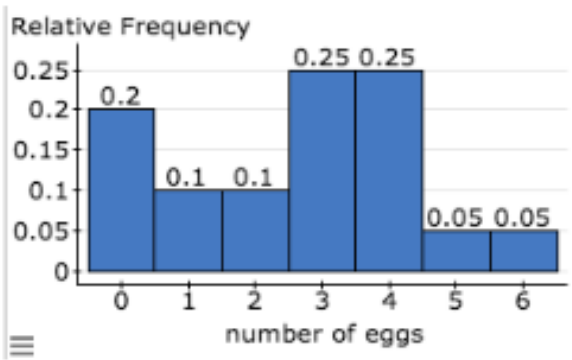
Check your understanding:

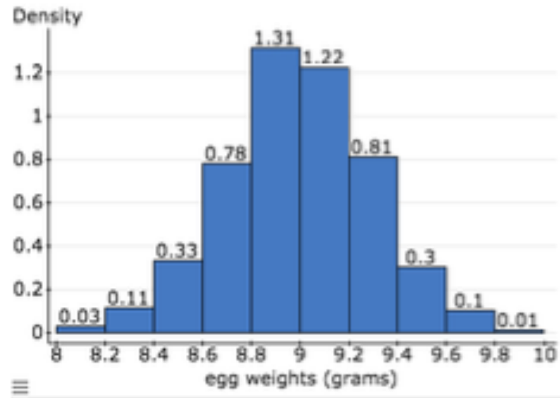
1) Here is a list of quantitative variables. Circle the variables that are continuous.

    - Foot length
    - Shoe size
    - Number of times a college student changes her major
    - Time it takes a college student to get to school
    - Number of Boreal owl eggs found in a nest
    - Weight of a Boreal owl egg

Probability Distributions

Below are two probability distributions: a discrete probability distribution that describes the number of eggs in a Boreal owl nest and a continuous probability distribution that describes the weight of Boreal owl eggs.





In a discrete probability distribution, the height of each bin is the relative frequency, which is the probability. For this reason, the sum of the heights of the bins is 1.

In a continuous probability distribution, we adjust the y-axis from a relative frequency to a density. This makes the area of each bin equal to the probability and the sum of the areas equal to 1. Later you will see why this is a helpful move.

A note about density: Density is a harder unit to understand than relative frequency, but they are related. For example, the urban cities have a higher density of people per square mile than rural communities. The density is relative frequency (%) per square mile.

In the egg weight probability density histogram, there is a larger density of eggs with weights near 9 grams. In other words, the % of eggs per gram is higher for eggs with weights near 9 grams than eggs with weights near 10 grams, just as in an urban city the % of people per square mile is higher compared to the rural areas.

Check your understanding:

2) Use the probability distributions above to answer these questions.

   a) What is the probability that a Boreal owl nest is empty?

   b) What is the probability that a Boreal owl egg weighs more than 9 grams but not more than 9.2 grams?

   c) Is it unlikely to find a Boreal owl egg that weighs less than 8.6 grams? Why do you think so?

A mathematical model of a bell-shaped probability distribution

Statisticians often represent probability distributions with smooth curves called probability density curves. These curves come from mathematical equations.

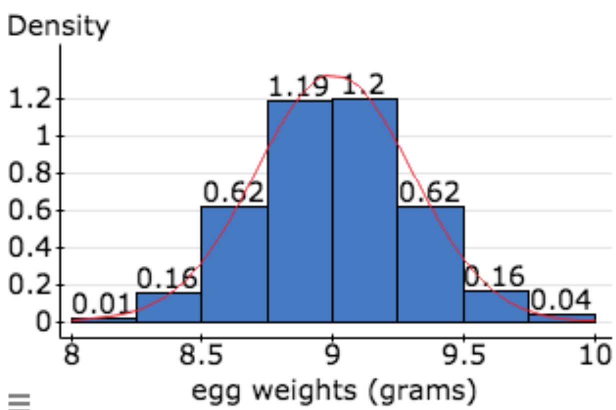
The area under the density curve represents the probability and the total area under the curve equals 1.

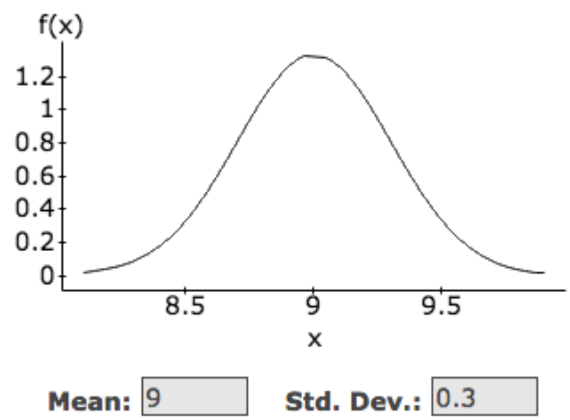
When the probability distribution is bell-shaped, we represent it with a bell-shaped density curve called a *normal* curve. The fact that this curve is called a *normal* curve indicates how prevalent it's use is.

Example:

Below is an example of a normal curve that models the probability distribution for Boreal owl egg weights. On the left is the probability density histogram with egg weights measured to the nearest 0.05 grams. You can see the normal curve sketched on the top of the histogram.

On the right is the normal curve without the histogram. We made this in StatCrunch. The normal curve has the same mean (expected value) and standard deviation as the probability histogram. It is a density curve so the total area under this curve equals 1.





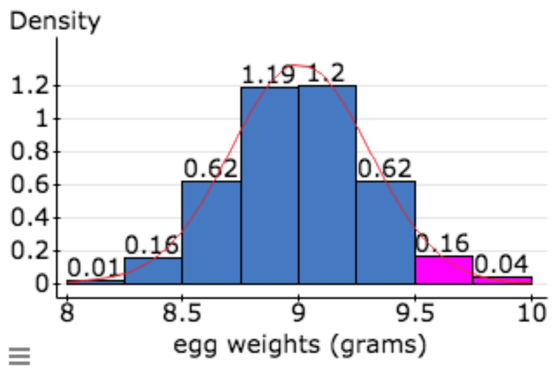
We can use StatCrunch to find the area under the normal density curve. This area represents probability.

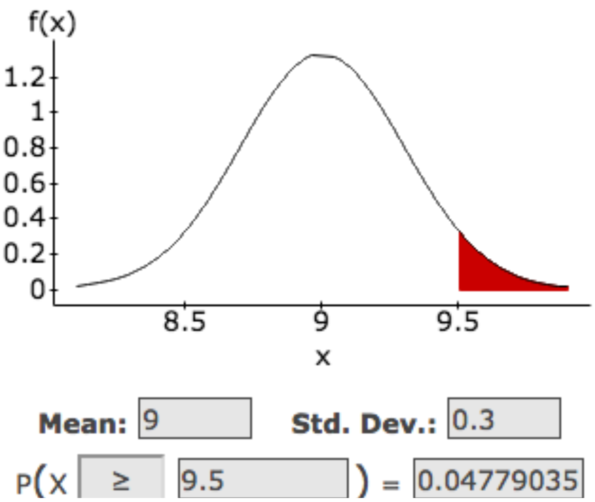
Check your understanding:

3) What is the probability that a Boreal owl egg weighs more than 9 grams? Use both the probability density histogram and the normal curve to find the probability.

4) What is the probability that a Boreal owl egg weighs more than 9.5 grams?

   Show that probability density histogram gives an answer close to the estimate given by StatCrunch.





Group work:

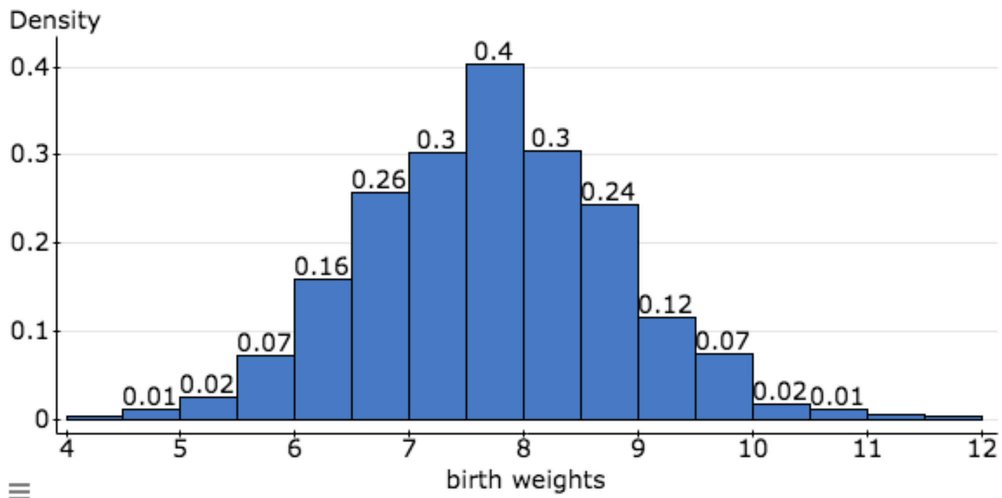
5) What is the probability that a Boreal owl egg weighs more than 8.5 grams but not more than 9 grams?

   a) Use the probability density histogram. Show your work.

   b) Verify that your answer is close to the StatCrunch estimate. Write down what you entered for Mean, Std. Dev., and copy the rest of the output.

   (StatCrunch instructions: Log into StatCrunch, choose Open StatCrunch. With the empty spreadsheet open, choose Stat, Calculators, Normal. You can figure out the rest.)

6) Here is a probability density histogram for birth weights based on data from a thousand babies of European heritage. The mean is 7.7 pounds with a standard deviation of 1.1 pounds.



a) Based on this data, what is the probability that a baby is born weighing less than 6 pounds? Use the probability density histogram. Show your work.

b) Verify that your answer is close to the StatCrunch estimate. Write down what you entered for Mean, Std. Dev., and copy the rest of the output using the probability notation used in StatCrunch; include the graph of the normal density curve.

c) What is the probability that a baby's birth weight is within one standard deviation of the mean? Show your work.

Module 12.2 Introduction to the Normal Distribution

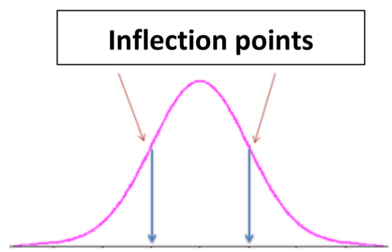
Learning Goal: Use the properties of the normal density curve to estimate probabilities using the Empirical Rule and standardized scores.

Introduction
In this activity we continue estimating probabilities using normal curves, but we will also investigate general properties of the normal curve.

General properties of normal curves:

Shape: Every normal curve has a bell-shape.
Center: Because of the symmetry in the curve, the mean is equal to the median; both are associated with the peak and divide the area under the curve in half.
Spread: As before, the standard deviation is roughly the average distance of values from the mean. It is associated with the inflection point on the curve where the curve transitions from concave down to concave up or vice versa.

For a density curves, we use Greek letters to represent the mean (μ) and the standard deviation (σ).

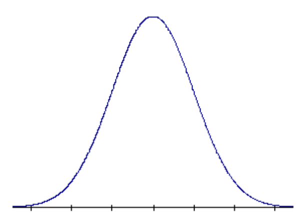


(Why the special notation? Because we are working with a mathematical model, not real data or relative frequencies from data.)

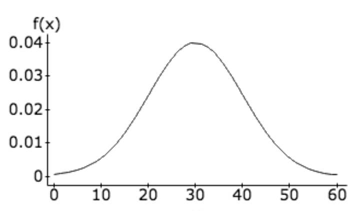
On the horizontal axis, label the mean as μ, the point that is one standard deviation below the mean as $\mu-\sigma$, and the point that is one standard deviation above the mean as $\mu+\sigma$.

Check your understanding:

1) For the normal curve at the right, mark a scale on the horizontal axis so that the mean is 6 and the standard deviation is 2.



2) Estimate the mean and standard deviation of the normal curve.



μ =

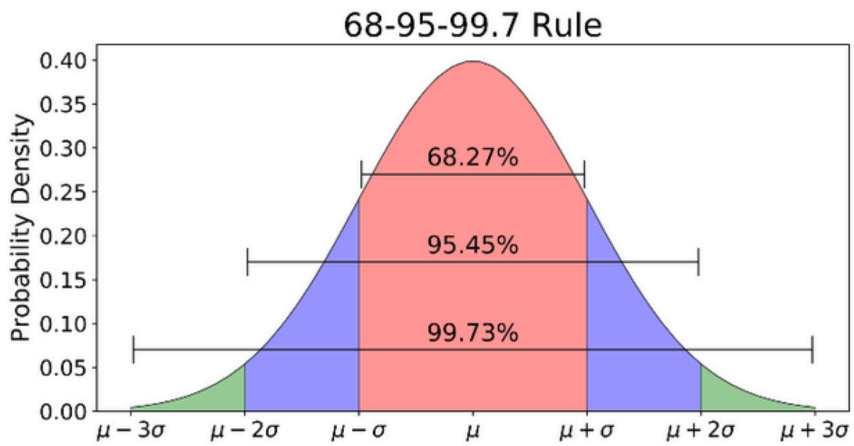
σ =

The Empirical Rule

For every normal curve, the probability that a value lies with one, two or three standard deviations from the mean is predictable and does not depend on the value of the mean or standard deviation. Specifically, the probability that a value …

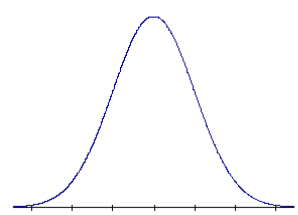
- Is within one standard deviation of the mean is 68% (this is the area between $\mu-\sigma$ and $\mu+\sigma$)
- Is within two standard deviations of the mean is 95% (area between $\mu-2\sigma$ and $\mu+2\sigma$)
- Is within three standard deviations of the mean is 99.7% (area between $\mu-3\sigma$ and $\mu+3\sigma$)

This special property is called the Empirical Rule or the 68-95-99.7 Rule.



Check your understanding:

3) Weights of 1-year old boys have a normal distribution; therefore, a normal curve is a good mathematical model for the probability distribution. The mean is 22.8 lbs and the standard deviation is 2.2 lbs.

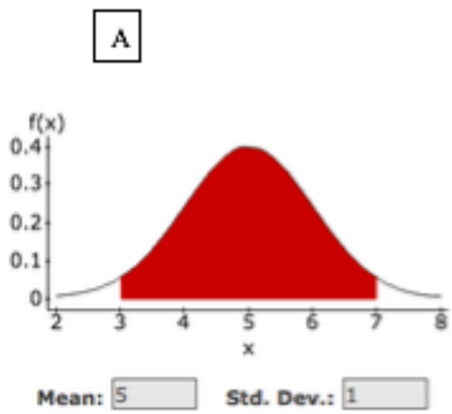


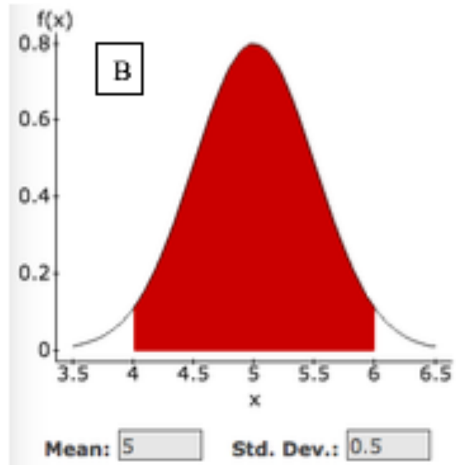
a) Label the horizontal axis of the normal curve to fit this scenario.

b) What is the probability that a randomly selected 1-year old boy weighs between 20.6 lbs and 25 lbs?

c) What is the probability that a randomly selected 1-year boy weighs more than 25 lbs?

4) Which probability represented by the shaded area is larger? Why do you think so?





Group work:

5) Birth weights have a normal distribution and so the probability distribution is normal in shape. The mean is 120 ounces and the standard deviation is 20 ounces.

   a) What is the probability that a randomly selected infant weighs less than 100 ounces? Show your work so that someone else can follow your thinking.

   b) Doctors define "normal" birth weight as weights that are within two standard deviations of the mean. (Here "normal" is the everyday use of that word, not the statistical use.)

      - What is the probability that a randomly selected baby will have a "normal" birthweight?

      - What is the range for "normal" birth weights by this definition?

   c) Doctors define "low birth weight" as a birthweight more than 2 standard deviations below the mean. What is the probability that a randomly selected baby has a low birth weight by this definition? Show your work.

Module 12.3 Standardizing Scores and the Standard Normal Probability Distribution

Learning Goal: Calculate and interpret a standardized score and use the Standard Normal probability density curve to estimate probabilities.

Introduction

In this activity we continue to estimate probabilities using normal curves but with a bit of a historical twist that will help us prepare for what will come later in the course.

We will learn to *standardize* a data value so that we can compare values from distributions with different means and standard deviations. Standardizing is a process that puts different variables on the same scale for easier comparison.

Standardized scores

In statistics a standardized score measures the distance between a value and the mean using standard deviation as a yard stick. In other words, a standardized score answers the question: how many standard deviations is this value from the mean?

Here is the formula for standardizing a value. The standardized score is called a Z-score.

$$Z = \frac{value - mean}{standard\ deviation}$$

Check your understanding:

1) Suppose that you took a test and scored an 85 out of 100 possible points. Suppose also that your class had a mean of 75 with a standard deviation of 10. Your friend took a test and scored a 16 out of 20 possible points. The mean for her class was 10 with a standard deviation of 5.

   Who performed better? To answer this question, we cannot compare your score of 85 to her score of 16 because the scales for the two tests differ: 100 possible points vs. 20 possible points. In addition, the mean and standard deviation are different for the two classes.

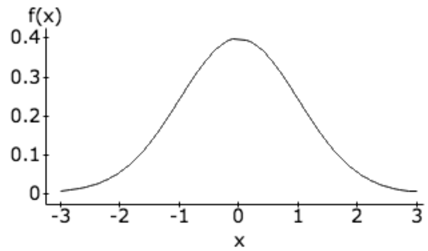
   a) Who scored a higher percentage? A percentage scales both scores to a score out of 100 so that we can more easily compare the scores. Find each percentage to answer the question.

   b) Who scored higher relative to their classmates? A Z-score scales both scores to a common measurement of "number of standard deviations from the class mean" so that we can compare the scores with standard deviation as the yardstick. Find each Z-score to answer the question.

The Standard Normal Curve

Before the advent of the sophisticated technology that is available today, statisticians used standardized values (Z-scores) to estimate probabilities.

The probability density curve for Z-scores is a normal curve with a mean of 0 and a standard deviation of 1. This curve is called the Standard Normal Curve.



Without the benefit of applets and statistical programs like StatCrunch, statisticians used tables to find probabilities.

Obviously, they could not produce a table for a normal curve with every possible mean and standard deviation, so they standardized data-values and used a single table associated with the Standard Normal Curve. For a given Z-score, they could look up the associated probability for values greater that Z-score.

Check your understanding:

2) Use the Empirical Rule and the Standard Normal Curve to answer the following:

a) P(-1 ≤ Z ≤ 1) =

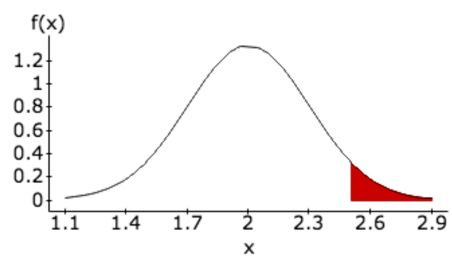
b) P(Z ≥ 1) =

c) P(-2 ≤ Z ≤ 2) =

d) P(Z ≤ -2) =

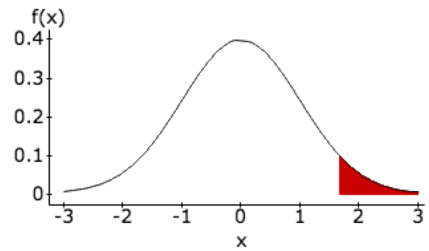
3) Suppose that the distribution of chicken egg weights is normal with a mean of 2 oz. and standard deviation of 0.3 oz. The minimum weight of a jumbo egg is 2.5 oz.

a) What is the Z-score for the smallest jumbo egg? What does the Z-score tell us? Show the Z-score on the probability distribution of egg weights.



b) What is the probability that a randomly selected egg is classified as a jumbo egg?

Use StatCrunch and the Standard Normal Curve to estimate the probability. Fill in the missing values and label the probability in the image.

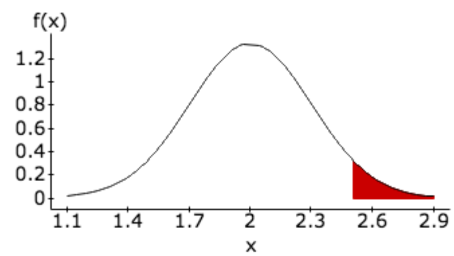


(In StatCrunch, choose Open StatCrunch, Stat, Calculator, Normal. Note that the default settings give the Standard Normal Curve.)

Mean: ______ Std. Dev.: ______ Here X represents ______

P (X ______) = ______

c) Now find the probability again using the same StatCrunch Normal Calculator, but this time enter the mean and standard deviation for the probability distribution of egg weights as we have done before. Fill in the missing values and label the probability in the image.



Mean: ______ Std. Dev.: ______

Here X represents ______

P (X ______) = ______

d) Theoretically, both methods should give the same probability, but the answers you found may differ a bit. What adjustment could you make to rectify this?

When do you use a Z-score and the Standard Normal Curve instead of a normal probability distribution based on a given mean and standard deviation?

The answer is that you can use either method.

Why do we teach the Z-score and the Standard Normal Curve when we really don't need it to find a probability?

The answer is that we will use Z-scores later in the course. In addition, when we get into statistical inference methods, we will always be calculating a test statistic (a Z-score or some other metric) and using a density curve based on those scores to find probabilities. Standardized scores are our first introduction to a test statistic and the Standard Normal Curve is the first example of a density curve based on a test statistic. Later, we will use probability distributions for different test statistics that are not normal curves.

Group work

4) Weights of one-year-old boys are normally distributed with a mean of 22.8 pounds and a standard deviation of 2.15 pounds. (Source: About.com). Children whose weights are more than two standards deviations from the mean are unusually large or small.

   a) What is the probability that a randomly selected one-year-old boy will be unusually small based on the definition above? Show your work.

   b) Ann's one-year-old son weighs 18 pounds. What is his Z-score? Is he unusually small relative to other one-year-old boys?

   c) Why is Ann's son's Z-score negative?

   d) What is the probability that a randomly selected one-year-old boy weighs less than Ann's son? Document your use of StatCrunch to find the answer.

Module 12.4 Unit 5 Lab

Background: Between the years of 1997 and 2003, the World Health Organization collected data on head circumference, arm circumference, and other body measurements for 8440 children from Brazil, Ghana, India, Norway, Oman and the USA. All of the children were healthy and breastfed by mothers who did not smoke. Children in the sample lived in socioeconomic conditions favorable to normal physical growth. Variables such as parental education level and income level were used to exclude children whose growth was "environmentally constrained"; however, low-weight full term babies were included if their mothers met socioeconomic and nutritional criteria. In short, the study design attempted to include healthy children from a wide range of ethnic and cultural groups whose home environments reasonably assured the conditions for healthy physical growth.

This data was used to produce growth charts that are part of every pediatrician's toolkit for monitoring a child's overall health. Governments and the United Nations also rely on growth charts to describe a country's overall health.

(http://www.who.int/childgrowth/standards/second_set/technical_report_2.pdf)

In this project we will focus on the head circumference data. Body measurements are usually bell-shaped, so we will use a normal model to represent the distribution of head circumferences in each of the following problems.

For full credit, show your work. Include calculations and/or a sketch of the normal curve with area shaded. State whether you used the StatCrunch Normal Calculator or the OLI Normal Calculator or the Empirical Rule.

1) According to the WHO report, girls who are one month old have a mean head circumference of 36.55 centimeters with a standard deviation of 1.17 cm.

   a) Ann is concerned that her daughter's head is small. Her daughter has a head circumference of 34.25 centimeters when she is one-month old. If we consider measurements more than 2 standard deviations from the mean as unusual, is Ann's daughter's head measurement unusually small? Support your answer.

   b) According to Medscape.com, microcephaly is a head circumference more than two standard deviations below the mean. What percentage of 1-month old girls will be categorized as having microcephaly? How do you know?

2) According to the WHO report, girls who are 2-years old have a mean head circumference of 47.18 cm with a standard deviation of 1.40 cm. A two-year old girl with Down's Syndrome has a head circumference of 44.5 cm. Children with Down's Syndrome typically have smaller heads, so this is not surprising.

   a) Relative to the WHO data, what is this girl's z-score? What does the z-score tell us?

   b) Using the WHO data in a normal model, what percentage of the girls has a head circumference that is smaller than the girl with Down's Syndrome?

UNIT 6

Types of Statistical Studies and Producing Data

Contents

Module 13 Producing Data for a Statistical Study

Learning Goals:

- Given a statistical research question, identify the population(s) and the variable(s).
- Categorize variables as quantitative or categorical.
- Distinguish between observational studies and experiments.

Introduction:

The goal of Statistics is to draw a conclusion about a large population based on a small sample. This is called *statistical inference.*

Up to this point in the course, we have learned many of the building blocks for statistical inference. For example, we know the difference between categorical and quantitative variables. This will be very important in identifying which statistical inference method to use.

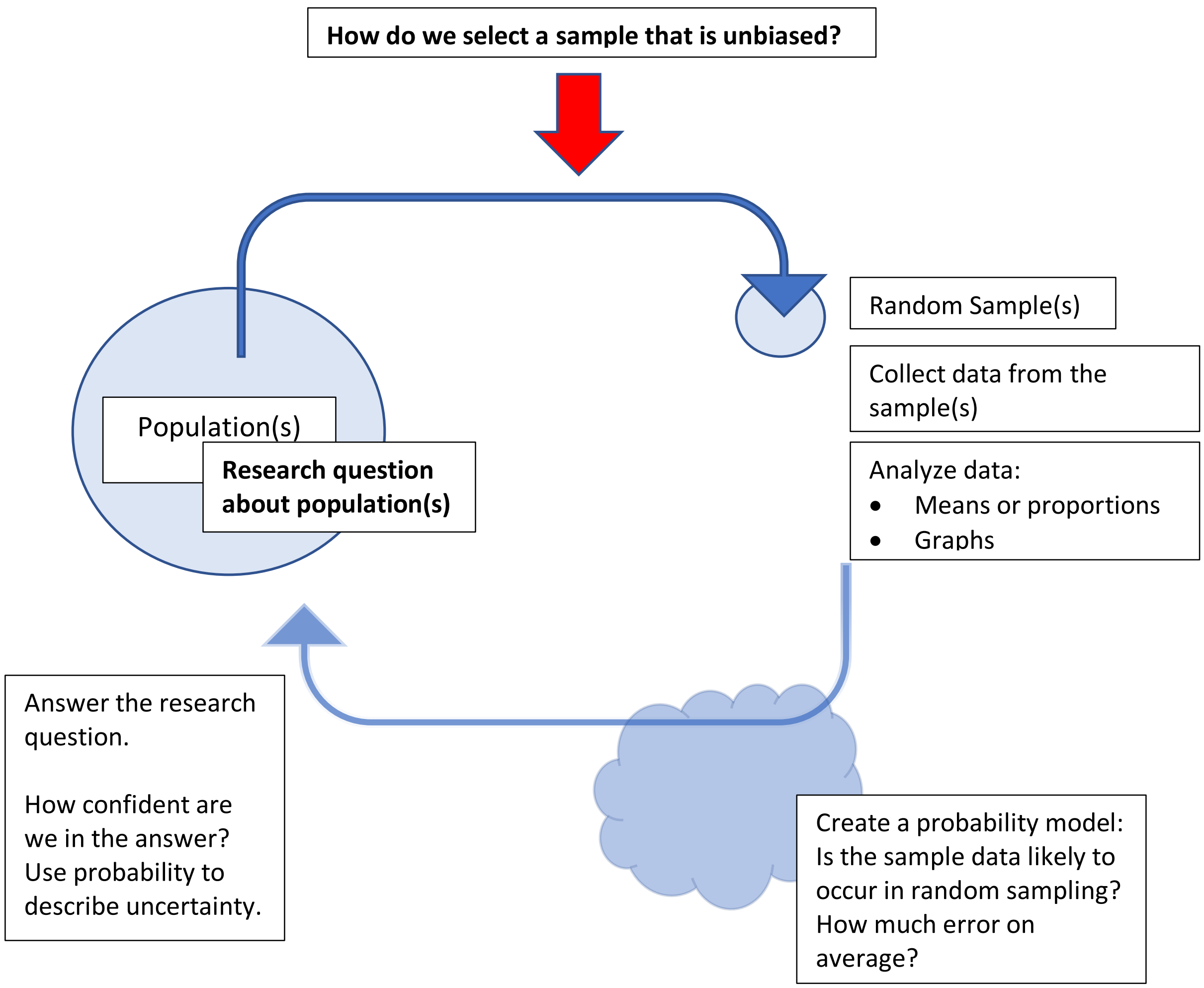
For each type of data (quantitative vs. categorical), we know how to summarize it using numerical measures (e.g. means and standard deviations for quantitative data or percentages for categorical data) and graphs (e.g. histograms or bar charts.) We know how to explore relationships between two variables using scatterplots, correlation, and linear regression if the variables are quantitative or using conditional percentages if the variables are categorical. We will continue to explore data and summarize it as part of statistical inference.

We have also learned about probability and probability distributions. Probability is based on relative frequency, which is the long-term frequency of an outcome in repeated random sampling. We recently finished learning about how to model probability distributions with a density curve to estimate probability with area under the curve. We will be using probability density curves with every statistical inference method.

In this Unit we tackle the last building block for statistical inference: data collection. In order for our conclusions about a population to be valid, we must have data that is not biased in any way. This Unit focuses on responsible data production.

The diagram on the next page ties all of these ideas together into the Big Picture of a Statistical Investigation.

Big Picture of a Statistical Investigation



Example: Suppose that our research question is "What is the average amount of money that full-time LMC students spend on textbooks in a semester?"

In this example, the population is ______________________________. To answer this question, we select a random sample from the population and ask each student in the sample ____.

The data we collect is (circle one: categorical or quantitative) and we summarize it using (circle one: a mean or a percentage).

Eventually, we will learn to create a probability distribution and probability density curve that will help use determine if the sample we collected is likely to occur given our assumptions about the population. From this we will learn to draw a conclusion about the population.

Check your understanding:

1) Dissect the research questions to identify the population, variable and variable type.

Research question	**Population**	**Variable**	**Variable Type and Summary**
What is the average number of hours that community college students work each week?	Community college students	Number of hours a student works each week	Quantitative Use a mean
What proportion of all U.S. college students are enrolled in a community college?			
Is the average course load for a California community college student at least 12 units?			
Do the majority of LMC students qualify for federal student loans?			

2) The next set of research questions involve two variables. We use a categorical explanatory variable to create the comparison groups; we can think of these as separate populations. Individuals in the samples give information relevant to the response variable and this data is analyzed and compared. The response variable can be categorical or quantitative.

Research question	**Explanatory variable and comparison groups (two populations)**	**Response variable (data analyzed and compared); type and summary**
Are college athletes more likely to receive academic advising than students who are not athletes?	College Athlete? (Yes, No) Two populations: college athletes compared to non-athletes	Receive academic advising? (Yes, No) Categorical Compare proportions
In community colleges, do female students have a higher GPA than male students?		
Is chemotherapy or radiation a more effective treatment for shrinking the size of cancerous liver lesions?		
Are elementary school children who drink soda with their lunch more likely to be categorized as overweight?		

Observational Studies vs. Experiments

When the research question involves two variables, the goal of the study is to determine if the explanatory variable correlates with, or even causes, a change in the response variable.

In an *observational study*, the explanatory variable is some pre-existing characteristic that is used to divide the individuals into groups, e.g. gender or drinking soda at lunch. The researcher cannot control who is in each group. By contrast, in the *experiment*, the researcher assigns individuals to the groups and each group receives a different treatment, e.g. chemotherapy or radiation.

In both types of studies, the researcher compares the responses of the groups. In an observational study, researchers may take steps to reduce the influence of these other factors on the response, but it is difficult in an observational study to get rid of all the factors that may have an influence. For example, when examining medical records, researchers may remove people from a cancer study who have a family history of cancer, but there may be other factors affecting cancer rates, such as diet and exercise, that are not measured and cannot be accounted for or removed. For this reason, an observational study can provide evidence of an association between two variables, but it provides, at best, weak evidence of a cause-and-effect relationship. The observed association may be confounded by unmeasured variables.

Unlike an observational study, an experiment can provide evidence of a cause-and-effect relationship between the variables because the researchers can manipulate and control more of the *confounding variables* that might influence the response variable. In a well-designed experiment, they can conclude the differences in the response are due solely to the treatments they imposed.

Check your understanding:

3) In the 1980's doctors routinely prescribed hormone replacement therapy for women in menopause. A series of studies in the 1990's based on women's medical records showed that women taking hormone replacements also had a reduction in heart disease. But women who take hormones are different from other women. They tend to be richer and more educated, to have better nutrition, and to visit the doctor more frequently. These women have many habits and advantages that contribute to good health, so it is not surprising that they have fewer heart attacks.

    a) Are the 1990 studies observational studies or experiments? Why do you think so?

    b) In these studies what is the explanatory variable? What is the response variable? What are some of the confounding variables?

    c) Can we conclude from these studies that the hormones caused the reduction in heart attacks? Why or why not?

4) In 2002, the Women's Health Initiative sponsored a large-scale study to examine the health implications of hormone replacement therapy. In this study, researchers randomly assigned over 16,000 women to one of two treatments. One group took hormones. The other group took a *placebo*. A placebo is a pill with no active ingredients that looks like the hormone pill.

   The 2002 study was *double-blind*. *Blind* means that women did not know if they were receiving hormones or the placebo. *Double-blind* means that the information was coded, so researchers administering the pills did not know which treatment the women received.

   After 5 years, the group taking hormones had a *higher* incidence of heart disease and breast cancer. In fact, the differences were so significant that the researchers ended the study early. As a result, hormone replacement therapy is now rarely used.

   a) Is the 2002 study an observational study or an experiment? Why do you think so?

   b) What is the explanatory variable? What is the response variable?

   c) Explain how random assignment might control the effects of some of the confounding variables you identified in #3.

   d) What else did researchers do to control the effects of other variables on the response?

   e) Why did this study lead to the elimination of use of hormone replacement therapy despite the fact that the studies in the 1990's supported hormone replacement therapy?

Module 14 Observational Studies and Sampling

Learning Goals:

- Define statistical bias.
- Recognize biased sampling plans.
- Explain the purpose of random sampling in an observational study.

Introduction:

We now focus on how to collect reliable and accurate data for an observational study.

In an observational study, we analyze a sample data and draw a conclusion about a population. For example, we want to use an exit poll to predict who will win an election. The sample needs to be a subset of the population. In addition, the responses of the sample should be representative of the responses of the population.

A *sampling plan* describes exactly how we will choose the sample. A sampling plan is *biased* if it systematically favors certain outcomes.

Bias often occurs in situations where individuals self-select into the sample. For example, a poll conducted by a conservative call-in radio program may overestimate opposition to proposed gun legislation because only conservatives with strong opinions against gun control will take the time to participate.

Statisticians use random sampling in an attempt to eliminate bias. In random sampling all individuals in the population have an equal chance of being selected. In general, larger random samples produce more reliable results than smaller ones.

Random sampling also guarantees that the sample results do not change haphazardly from sample to sample. When we use random selection, the variability we see in sample results is due to chance and the results obey the mathematical laws of probability. This is important in statistical inference.

Check your understanding:

1) Suppose that we want to estimate the mean number of text messages sent by LMC students each day. Which sampling plan will produce the most reliable estimate? Why do you think so?

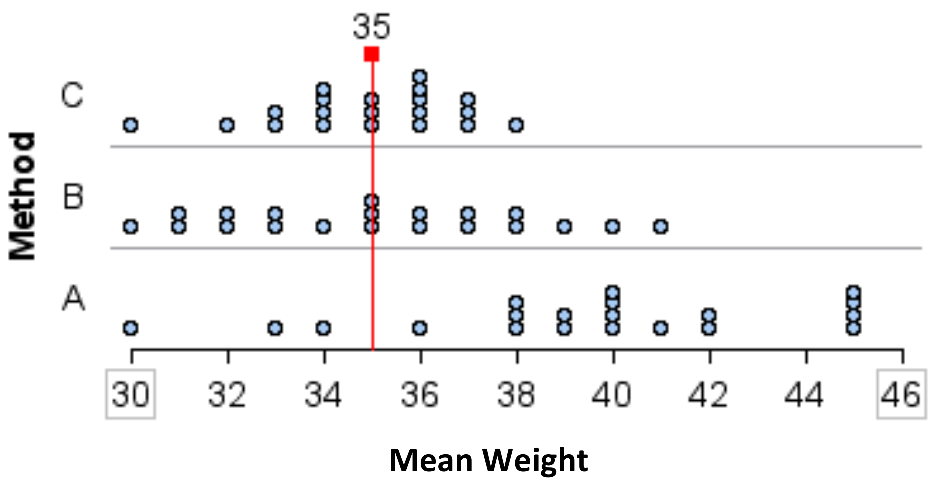
   a) Select 50 students at random from the list of students' Insite email addresses. (All LMC students have Insite email addresses.)
   b) Select 100 students at random from the list of students' Insite email addresses.
   c) Select the first 200 students who you see texting in the LMC Quad.
   d) Select 300 students at random who follow LMC on twitter.

Group work:

2) A 5th grade class conducts a study to estimate the mean weight of a snail. They have a box containing 100 snails; each snail has a number on its shell. The 20 children do the study using three different sampling methods:

Method A: Each child picks 5 average looking snails, weighs each one, then averages the 5 weights.
Method B: Each child draws 5 numbers out of a hat, locates the 5 snails labeled with those numbers, weighs each one, then averages the 5 weights.
Method C: Each child draws 10 numbers out of a hat, locates the 10 snails labeled with those numbers, weighs each one, then averages the 10 weights.

Here are the results:



a) The mean weight of the 100 snails is 35 grams. Which of the sampling method is producing the most biased estimates of the overall mean weight? How can you tell by looking at the graphs?

b) How does the graph produced by Method B and C illustrate that random samples produce unbiased weight estimates?

c) How does the graph produced by Method C illustrate that larger random samples produce more accurate estimates than smaller random samples?

Module 15 Collecting Data—Conducting an Experiment

Module 15.1 Experiments and Random Assignment

Learning Goals:

- Identify confounding variables and identify features in experiment design that control the effects of confounding variables.
- Explain the purpose of random assignment in an experiment.

Introduction:

In our previous discussion of the hormone replacement studies, we examined how an experiment can control the effects of confounding variables better than an observational study. For this reason, an experiment can provide evidence of a cause-and-effect relationship between the explanatory variable and the response variable.

In an experiment, researchers attempt to control the effects of confounding variables using *direct control* and *random assignment.*

Direct controls address potential confounding variables up front. For example, if we think a family history of cancer will increase cancer rates, we can remove people with a family history of cancer from the experiment. But what if we do not know about the confounding variable ahead of time? For example, if a gene influences cancer rates, but we do not know that. This is why a well-designed experiment uses random assignment.

How does random assignment control the effects of confounding variables?

Here is how it works. Random assignment creates treatment groups that are similar to each other. It does not eliminate the effect of the confounding variable; instead it equalizes the effect across the treatment groups.

When we have similar groups relative to the confounding variable at the start of the experiment, then we can assume that the confounding variable is not explaining the differences we see in treatment groups at the end of the experiment. Therefore, we can conclude that the observed differences are the result of the treatments which are associated with the explanatory variable. This allows us to say that the explanatory variable is responsible for differences in the responses. This is what we mean by a cause-and-effect relationship.

Like random sampling in observational studies, random assignment in an experiment is also necessary for statistical inference. Why? Because statistical inference is based on probability and probability is based on random events. If an experiment does not use random assignment, the data from that experiment cannot be used to draw conclusions about the relationship between the variables for a larger group of subjects.

Check your understanding:

1) This diagram is adapted from Ramsey and Schafer's *The Statistical Sleuth*.

Conclusions Permitted by Different Study Designs				
		How are individuals assigned to groups?		
		Random assignment	**Not random**	
How are individuals selected?	**Random selection**	Select a random sample from one population. Randomly assign individuals to treatment groups.	Select random samples from existing populations or groups.	**Make an inference about population(s). Draw a conclusion about the population(s).**
	Not random	Find a group of individuals. Randomly assign them to treatment groups.	Examine available individuals from different populations or groups.	
		Make an inference about a treatment effect. Draw cause-and-effect conclusions.		

a) In the diagram, which descriptions correspond to experiments? Which to observational studies? Label these.

b) How does this diagram highlight key ideas from Unit 6 (Modules 13-15)?

c) In the diagram, add the phrase "potential for sampling bias" and the phrase "potential for confounding by other variables" in the empty boxes where it makes sense.

d) In the diagram, circle the description of a study that produces data that cannot be used in statistical inference. Is there more than one?

Group work:

2) A high school student named David Merrell did an experiment to examine if music affects the time it takes rats to run a maze. (If you are curious, you can watch a conference presentation by Merrell https://www.youtube.com/watch?v=78QtW_AxWFU (21:16)).

   The <u>explanatory variable</u> was exposure to music. He had three treatment groups: *one group* listened to heavy metal music by the group Anthrax. A *second group* listened to Mozart. The *third group* never heard music.

   a) Which group is the control group? What is the purpose of a control group in this experiment?

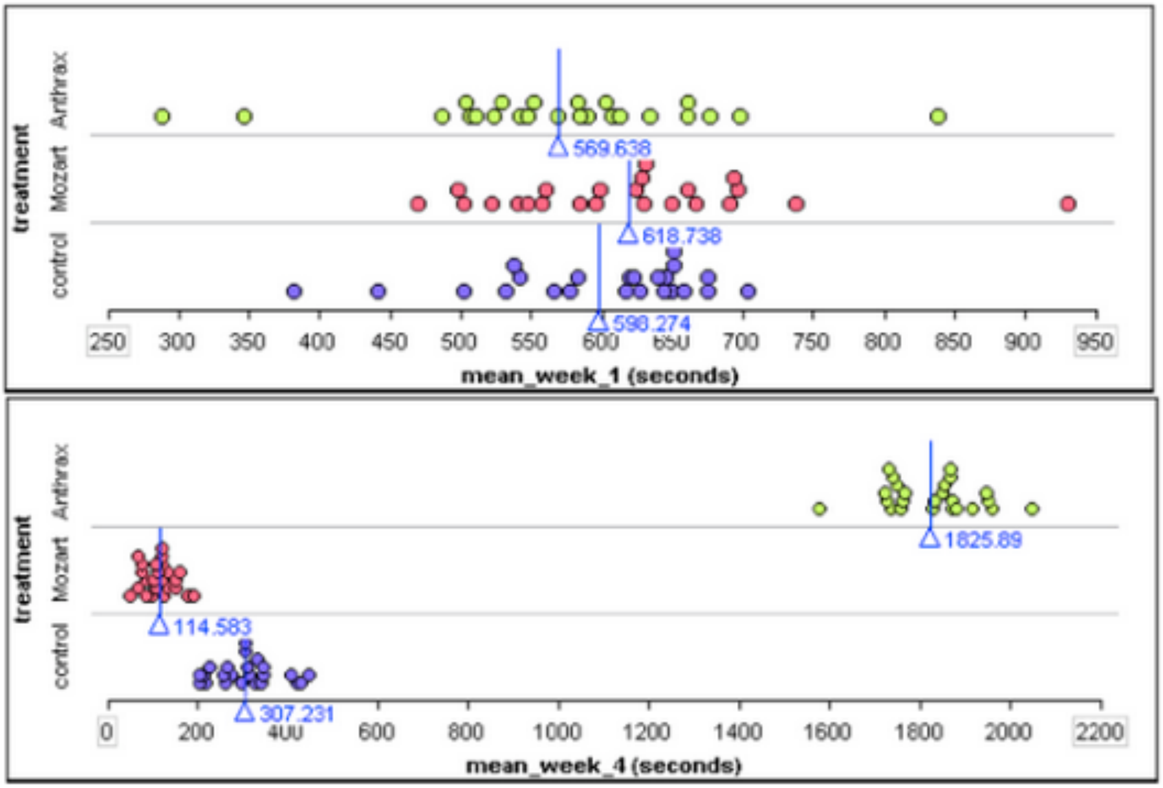
   The <u>response variable</u> was the average time (in seconds) to complete three runs. Every week each rat ran the maze three times. Merrell recorded each rat's average time for the week.

   Merrell included many direct controls in an attempt to minimize the impact of confounding variables on the rats' performance.

   b) Match each confounding variable with the direct control that addresses it. Circle confounding variables that cannot be addressed by one of the direct controls listed.

Confounding variable	**Direct control**
Some rats may be faster learners.	Use rats with born on the same day. At the start of the experiment remove rats with weights outside the normal range.
Overweight rats or old rats may be slower.	House all rats in identical individual cages and give them the same amount of food and light.
Living conditions may affect a rat's performance.	Train all of the rats to run the same maze before the experiment begins.
Some rats may be faster due to hereditary factors.	In the treatment groups, rats listen to music at 70 decibels for 10 hours a day for a month.
The volume of the music rather than the type of music may affect a rat's performance.	Blinding: the person measuring the run times does not know which treatment the rat received.

Here are the average run times (3 runs) for the 1st week (start of the experiment) and the 4th week (end of the experiment.) Each dot represents one rat. Blue lines mark the overall mean for all rats.



c) Merrell claims that he randomly assigned rats to treatment groups. Does the data support his claim? Why or why not?

d) Does exposure to music affect the rats' performance in the maze? How does the data support your answer?

Module 15.2 Lab Assignment: Practice with Confounding Variables and Design

Name: ____________________________________

Learning Goal:

- Distinguish observational studies from experiments.
- Identify explanatory, response, and potential confounding variables in a study.
- Identify features of experiment design that control the effects of confounding variables.

(This activity is adapted from an activity by Beth Chance and Allan Rossman in *Workshop Statistics*.)

1) Near-sightedness typically develops during the childhood years. Recent studies have investigated a possible association between near-sightedness and the use of night- lights with infants. In a study by Quinn, Shin, Maguire, and Stone (1999), researchers surveyed parents of 479 children visiting a pediatric ophthalmology clinic. One of the questions was "Under which lighting condition did/does your child sleep at night?" before the age of 2 years. The parents chose between "room lighting," "a night light," and "darkness." Based on the child's most recent eye examination, researchers separated the children into three groups: near-sighted, normal refraction, or far-sighted.

a) Who are the individuals of interest in this study?

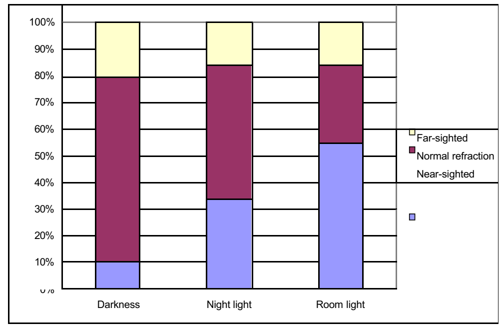
b) What is the explanatory variable in this study? Is it quantitative or categorical?

c) What is the response variable in this study? Is it quantitative or categorical?

d) Is this an observational study or an experiment? How can you tell?

The following table and graph display the data from this study.

	Darkness	Night light	Room light	TOTALS
Far-sighted	40	39	12	91
Normal refraction	114	115	22	251
Near-sighted	18	78	41	137
TOTALS	172	232	75	479



e) What does the data suggest about the association between near-sightedness and higher levels of light exposure? Support your answer.

f) Is it valid to conclude that sleeping in a lit room or with a night light **causes** an increase in a child's risk of near-sightedness? Why or why not?

g) Identify at least one confounding variable. Explain how this variable might be responsible for the association we see between the explanatory variable and the response variable in this study.

Module 15.3 Unit 6 Project

In this project you will design an observational study and an experiment to answer a research question.

Instructions: Make a poster to represent your observational study and your experiment.

Things to ponder and to represent in your poster or presentation:

1) Observational study:
   - What is the research question?
   - What is the population? What is the variable?
   - Will you be working with categorical or quantitative data?
   - How will you produce your data: survey? Observation? Other?
   - If you are surveying, what question(s) will you ask in the survey? If you are observing, what are you measuring or recording?
   - How will you randomly select your sample?

2) Experiment:
   - What is the research question?
   - What are the explanatory and response variables? Is the response variable categorical or quantitative?
   - How will the response variable be measured or observed?
   - What a potential confounding variable? How will you control the effect of this variable on the responses?
   - How will you use random assignment?

Choose one of the research questions below or write your own.

1) Which paintball is the best one: Valken Infinity or Wrek Elite?

   A paintball has an outer shell made of gelatin and it is filled with a water-based paint. A high-quality paintball is tough and durable and will not break inside the paintball gun when struck by the firing pin. But it will break when it hits its target. High-quality paintballs are also spherical without dimples or swellings. Paintballs deteriorate with age. Fresh paintballs should bounce and remain intact on the 1^{st} or 2^{nd} bounce when dropped from a 6-foot height.

2) Which diaper is better: Pampers Swaddlers or Huggies Snug & Dry?

   What makes a good baby diaper? Absorbency, leakage control, fit, strength of the adhesive tab, environmental impact (decomposition rate), other?

3) Does embedded tutoring improve students' grades in a college math class?

   Embedded tutoring is an intervention in which a peer tutor attends the class and provides support for students. It works best when the class is interactive as opposed to a lecture format.

4) Is there a halo effect in college professors' treatment of students?

   The halo effect is a type of cognitive bias in which our overall impression of a person influences how we feel and think about his or her character or abilities. Essentially, your overall impression of a person ("He is nice!") impacts your evaluations of that person's specific traits ("He is also smart!"). The opposite effect is also true. Negative feelings about one characteristic lead to negative impressions of an individual's other features.

UNIT 7

Linking Probability to Statistical Inference

Contents

Module 16 Introduction to Inference

Learning Goal: Recognize situations where statistical inference is, and is not, appropriate.

Introduction: We are now transitioning into the last part of the course: statistical inference. Let's briefly discuss the difference between descriptive statistics (the early part of this course) and inferential statistics.

Differences between Descriptive and Inferential Statistics

For descriptive statistics, we choose a group that we want to describe and then measure all subjects in that group. The statistical summary describes this group with complete certainty. For example, we can calculate the mean height of first graders at the local elementary school.

For inferential statistics, we define the population and then devise a sampling plan that produces a representative sample from the population. For example, we might want to estimate the mean height of first graders statewide by selecting a random sample of first graders from across the state.

A study using descriptive statistics is simpler to perform. However, if you need evidence that an effect or relationship between variables exists in an entire population rather than only your sample, you need to use inferential statistics.

With inferential statistics the statistical results incorporate the uncertainty that is inherent in using a sample to understand an entire population. Describing this uncertainty is where probability comes in. Because probability is involved, a statistical study involves data that comes from some random process, such as random sampling or random assignment.

In statistical inference, we will use probability density curves. These curves are mathematical models that represent the long-run behavior of random samples. With these models we will be able to describe and quantify the potential error in a random sample and the uncertainty we may have in the inference about a population that is based on a single random sample.

Check your understanding:

Which of the following studies involve statistical inference?

a) Who is the better homerun hitter: Barry Bonds or Babe Ruth?
b) Is the Contra Costa school bond measure expected to pass in the next election?
c) Is there a relationship between political party affiliation and views on gun control in the United States?
d) Is there a relationship between gender and views on vaping in our class?
e) What percentage of customers expect name brands to provide customer service on Facebook?

Module 17 Distribution of Sample Proportions

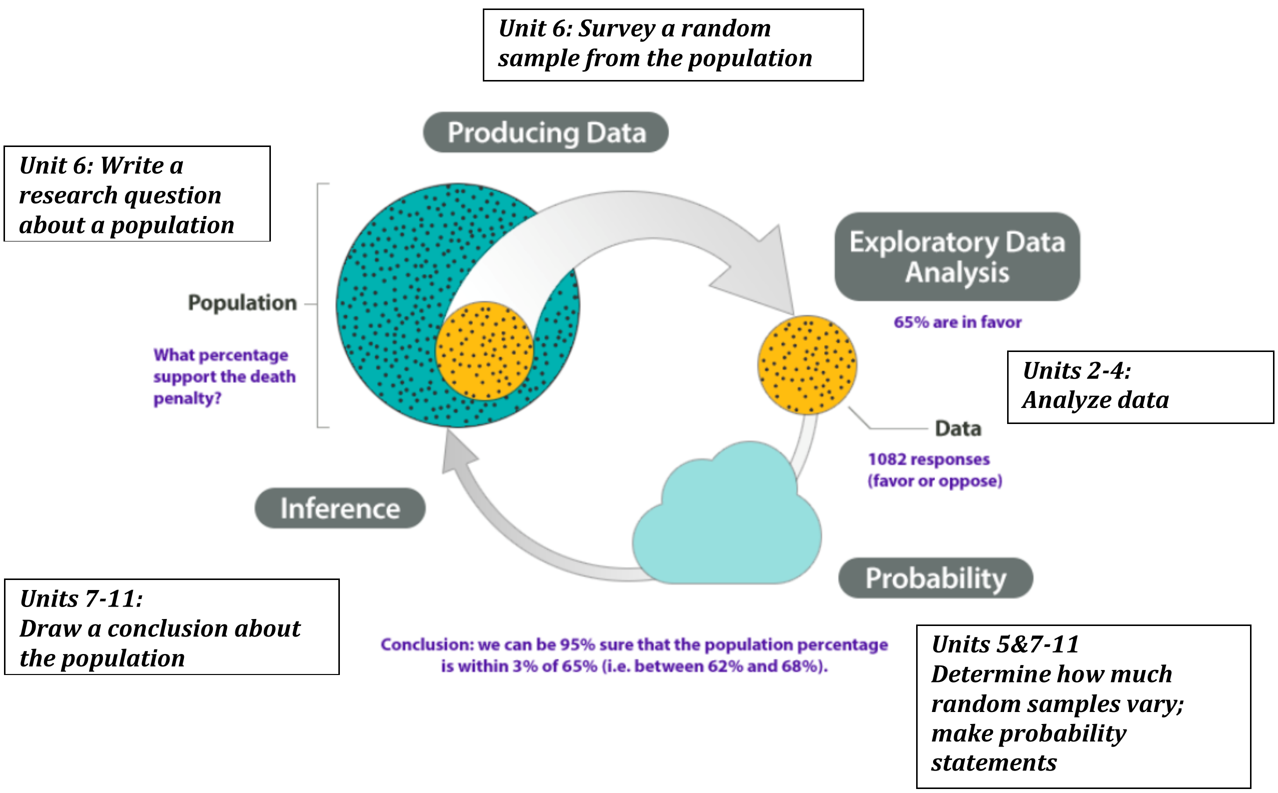
Module 17.1 Understanding Sampling Variability: Intro. to Sampling Distributions

Learning Goals:

- Use simulation and probability to draw conclusions based on data.
- Describe the sampling distribution for sample proportions and use it to identify unusual (and more common) sample results.

Overview:

1) Use the diagram to describe the process of statistical investigation.



Introduction to this activity[1]

Before we can draw inferences about a large population, we need understand how samples relate to the population and how much samples will vary. To do this, we will view our class as a population and try to determine the percentage of our class that has a cat.

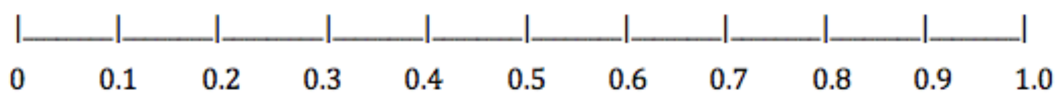
This is of course artificial because our population is very small. If we really wanted to know the percentage of our class with cats, we would just take a census of the entire population and find out! But our goal is to examine how random samples from this population behave so that we can begin to understand how much

[1] This activity is adapted from a workshop presentation by Beth Flynn.

variability occurs in random samples. This will help us estimate the amount of error we expect in random samples when estimating a population proportion.

Questions to answer during the demonstration

2) Make a dotplot of the distribution of sample proportions based on the random samples from your class. Make a title for your graph and label the axis with a description of what the numbers represent.



3) What does each dot represent in the dotplot?

4) Describe the distribution of sample proportions (shape, center, spread).

5) What is the mean and SD of the distribution? What do these numbers tell us?

6) What are typical sample proportions in this distribution? Why do you think so?

7) What sample proportions are unlikely? Why do you think so?

8) Given the samples we have collected so far, what do you think is the proportion of all students in our population that have a cat? In other words, what is your estimate for the population proportion? Why do you think so?

9) How accurate do you think your estimate of the population proportion is based on the variability we see in our random samples? How much error is there on average in these samples?

10) Since we have such a small population, let's take a census. What is the actual population proportion that has a cat?

11) How much error was there in our estimate of the population proportion? Why did this error occur?

Module 17.2 Sampling Distribution for a Population Proportion

Learning Goal: Describe the sampling distribution for sample proportions and use it to identify unusual (and more common) sample results.

Introduction: In this activity we will continue to investigate ways to use random sampling to draw an inference about a population.

Last time we gathered many random samples from a population (our class) in order to examine how random samples behave and to estimate the proportion of the population that was pierced.

This time we will begin with a hypothesis about the population. We will examine the behavior of random samples from the population when the hypothesis is true.

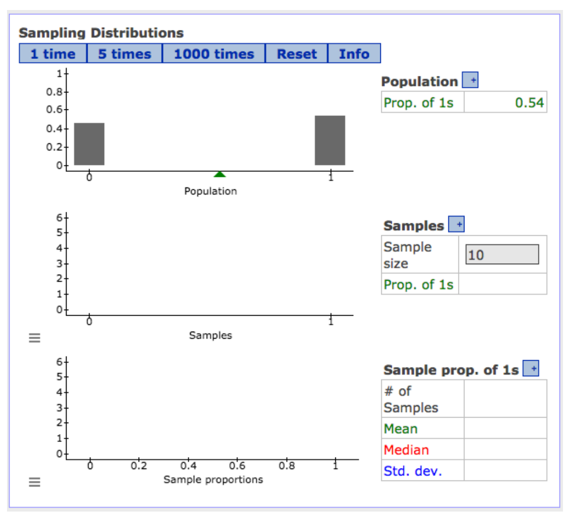
Here is a fact: According to the 2010 U.S. Census, about 54% of California residents were born in California.

Let's see if what is true for all California residents is also true for LMC students. In other words, our hypothesis is that 54% of LMC students were born in California.

In this activity we will use a simulation that is similar to what we have already done. But this time we will be investigating the variability in samples assuming our hypothesis about the LMC population is true.

Answer the following questions as your instructor does a simulation in StatCrunch.

Questions to answer during the demonstration

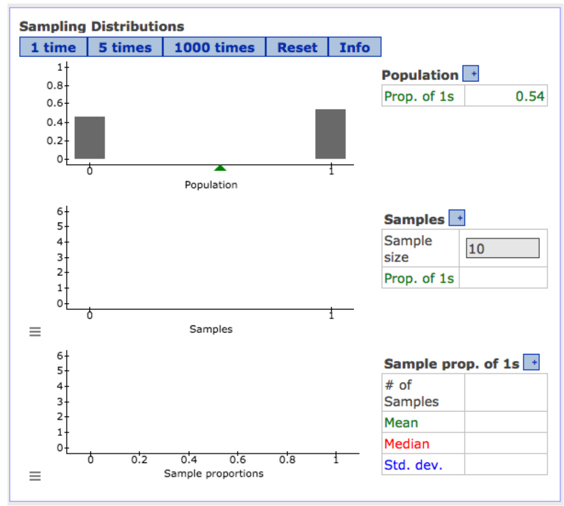


1) Your instructor set up the simulation by choosing *binary* because this categorical variable has two options: Born in CA (yes or no). Your instructor also entered p=0.54 to represent the hypothesis about the proportion of the LMC student population born in CA.

   a) How is this hypothesis represented in the graph of the population?

   b) What does a 0 represent? What does a 1 represent?

2) Your instructor will now use the simulation to select a random sample of 10 LMC students from the population that has 54% born in CA (p=0.54).



a) What proportion of the students in this random sample was born in CA? (Fill in the appropriate blank in the StatCrunch output.)

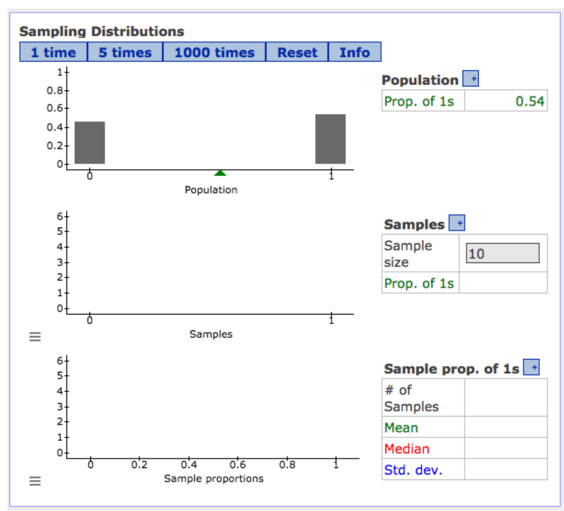
b) Make a graph in the StatCrunch printout to represent this sample.

c) How many students were surveyed in this sample?

d) Put a dot in the bottom graph to represent this sample.

e) How much error is there in this random sample? (How far does it deviate from our hypothesized population proportion of 0.54?)

3) Your instructor will now use the simulation to select a 2^{nd} random sample of 10 students from the LMC population.



a) What proportion of the students in this second random sample was born in CA? (Fill in the appropriate blank in the StatCrunch output.)

b) Make a graph in the StatCrunch printout to represent this sample.

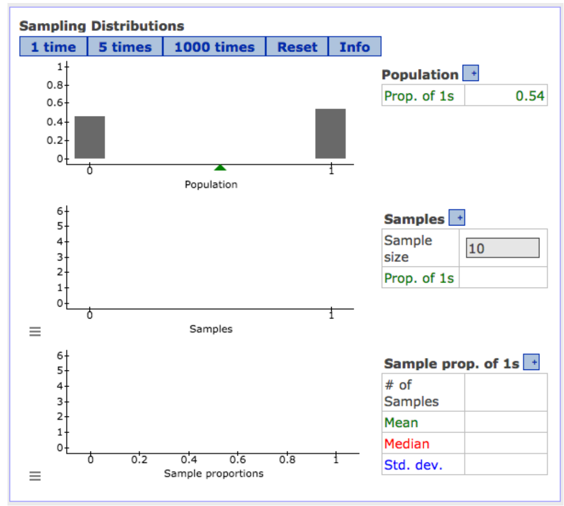
c) How many students were surveyed in this sample?

d) Put a dot in the bottom graph to represent this sample. Also, put a dot to represent the 1^{st} sample so that the bottom graph now has two dots.

e) How much error is there in this second random sample? (How far does it deviate from our hypothesized population proportion of 0.54?)

4) And again ... your instructor will use the simulation again to select a third sample.

   a) As we continue to select random samples, the sample proportions will vary. What stays the same in the simulation and does not change?



   b) As before, fill in the sample proportion for the 3rd random sample and make a graph of the sample results in the StatCrunch printout.

   c) Represent this sample proportion, along with the two previous one in the bottom graph.

   d) Which of three sample proportions has the greatest amount of error? Which has the least?

   e) In the StatCrunch printout, fill in the mean, median and standard deviation based on your instructor's display. What do each of these numbers tell us?

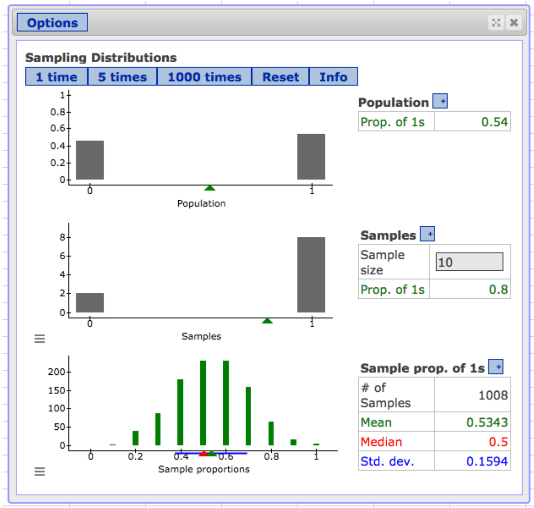
5) Your instructor will now select five additional samples. Answer these questions based on what you see in the simulation.

   a) What are we hypothesizing is true about the population?

   b) How many random samples have we collected? How many students are in each sample?

   c) What do you think the blue line is telling us in the graph of the distribution of sample proportions?

6) Your instructor will now select 1,000 additional samples. The distribution of sample proportions shown here will be very similar to your instructor's results.



a) How many random samples have we collected?
   How many students are in each sample?

b) Describe the shape of the distribution of sample proportions at this point.

c) What is the mean of the sampling distribution now? Why does this make sense?

d) On average approximately how much error is there in these samples?

e) Give an interval to represent typical sample proportions.

f) Identify a few sample proportions that are unusual.

7) Your instructor will select a random sample of 10 students from the class.

a) Is the class' sample proportion unusual when compared to the samples in the simulation? How do you know?

b) How much error is there in our class sample proportion relative to our hypothesized population proportion of 0.54? Is this much error unusual when sampling from a population of students in which 54% are born in CA?

c) Do you think our class sample could have come from a population in which 54% are born in California? Or do you think our class sample comes from a population with a different proportion of native Californians?

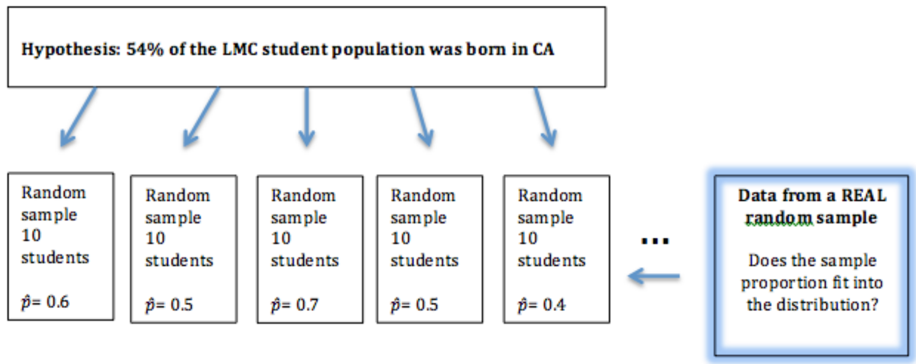
Module 17.3 Effect of Sample Size on the Sampling Distribution

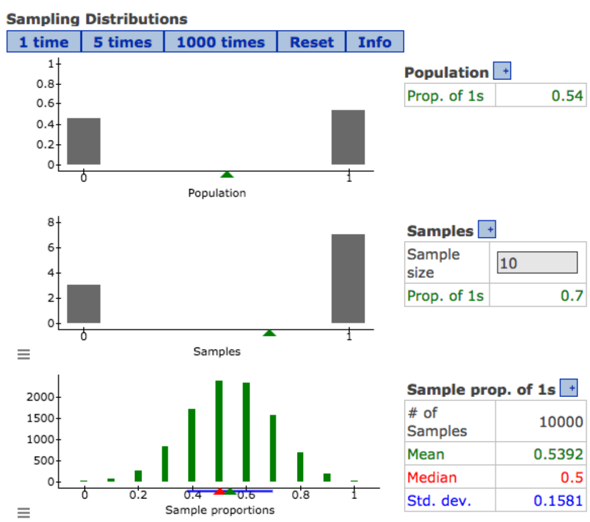
Learning Goal: Describe the sampling distribution for sample proportions and use it to identify unusual (and more common) sample results.

Learning Objective: Describe the effect of sample size on the distribution of sample proportions.

Introduction:

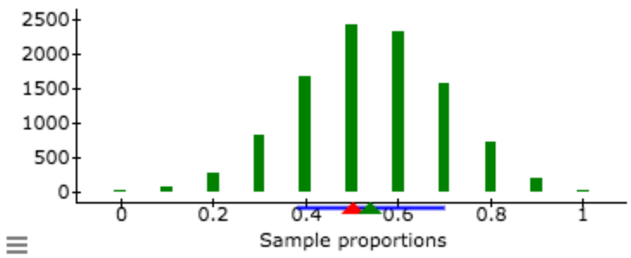
Your class will discuss these diagrams to review what we have learned from recent activities. Take notes … contribute to the discussion!





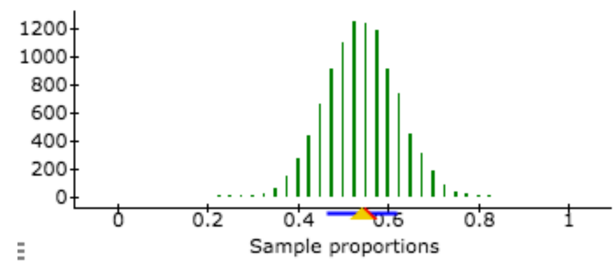
1) In the simulations we have conducted so far, we have always used a sample size of 10 students. Do you think larger samples will be better? Why or why not?

2) Let's repeat the simulation with a larger sample size. Your instructor will run a simulation selecting random samples of 40 students from a hypothetical student population in which 54% are born in California. The resulting distribution of sample proportions will be similar to the second picture below. In your group, use the images below to compare and contrast the distribution of sample proportions for n=10 (on the top) to the distribution of sample proportions for n=40 (on the bottom.)



Sample prop. of 1s

# of Samples	10000
Mean	0.541
Median	0.5
Std. dev.	0.1584



Sample prop. of 1s

# of Samples	10000
Mean	0.5402
Median	0.55
Std. dev.	0.0788

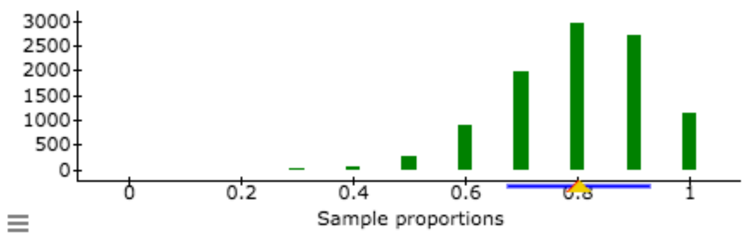
a) Compare the means for the two distributions. What do you notice? Why does this make sense?

b) Compare the standard deviations of the two distributions. What do you notice? Why does this make sense?

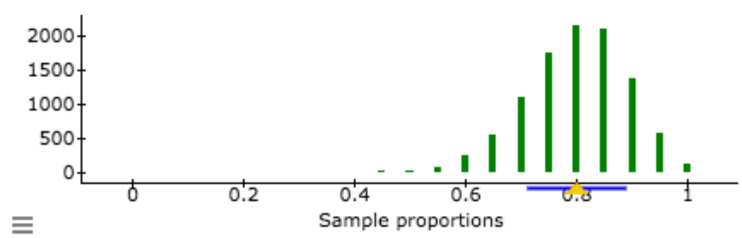
3) Now we are changing the context of our simulations.

Suppose that we think that 80% of LMC students will vote in the next election. So our hypothesis is p=0.80.

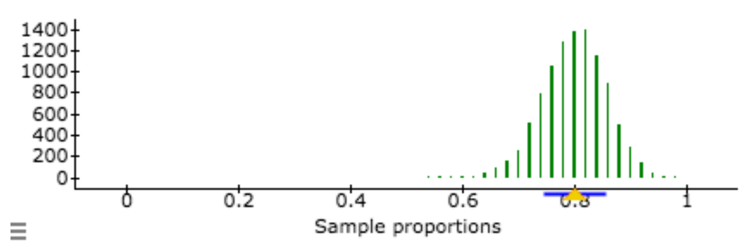
We created four distributions of sample proportions by selecting random samples from a population with p=0.80. One distribution came from samples of 10 students. We used samples of 20 students, samples of 50 students, and samples of 100 students to create the other three distributions. These are pictured below.



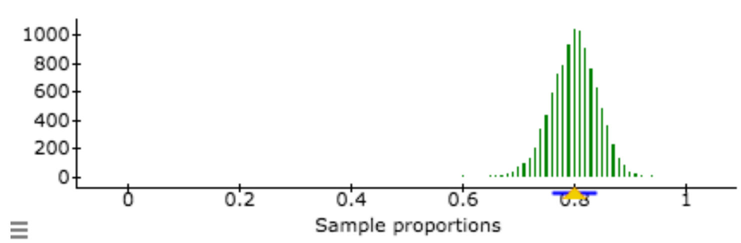
Sample prop. of 1s	
# of Samples	10000
Mean	0.8022
Median	0.8
Std. dev.	0.1273



Sample prop. of 1s	
# of Samples	10000
Mean	0.8002
Median	0.8
Std. dev.	0.0891



Sample prop. of 1s	
# of Samples	10000
Mean	0.7995
Median	0.8
Std. dev.	0.0566



Sample prop. of 1s	
# of Samples	10000
Mean	0.8001
Median	0.8
Std. dev.	0.0402

a) Start by labeling each distribution with the size of the samples represented (n=10, n=20, n=50, n=100). Briefly explain how you made your decision.

b) Compare the means of the distributions. What do you notice? Why does this make sense?

c) Compare the standard deviations of the four distributions. What do you notice? Why does this make sense?

d) Our intuition tells us that larger samples have less error on average. How do the sampling distributions support this statement?

e) Which sample sizes keep the average error within 10% of our hypothesized 0.80? Which sample sizes keep the average error within 5%? Explain how you know.

f) Compare the shapes of the four distributions. What do you notice about the shape as n increases? Which sample sizes have distributions that look normal in shape?

4) Let's think about the effect of sample size on our conclusions.

Suppose that we poll a random sample of 10 real students and 70% (7 out of 10) say that they plan to vote in the next election. Do you think that this sample came from a population with 80% planning to vote? Why or why not?

Suppose that we poll a random sample of 100 real students and 70% (70 out of 100) say that they plan to vote in the next election. Do you think that this sample came from a population with 80% planning to vote? Why or why not?

Module 17.4 Mathematical Model for the Distribution of Sample Proportions

Learning Goals:

- Use a mathematical model of the normal curve to represent the distribution of sample proportions for a given scenario.
- Use a z-score and the standard normal model to estimate probabilities of specified events.

Introduction:

In the last activity we examined how the size of the random samples affects the average amount that the sample proportions vary from the population proportion. We discovered that large samples do a better job estimating the population proportion in the long run. In other words, large samples tend to have sample proportions that are closer to the population proportion, which means smaller amounts of error on average.

Why do we care about this?

- Because we use samples to estimate population proportions. So we need to understand how much error to expect on average.

- Also when we want to test a hypothesis about a population, we need some way of judging whether the real sample data could have come from the population described in our hypothesis. Therefore, we need to know how much sampling error will arise naturally from the sampling process in order to determine if the sample proportion from the real data is unusual or not.

Up to this point we have used simulations to estimate how much random samples behave. Let's summarize what we have learned about the distribution of sample proportions from these simulations. Fill in the blanks.

- *Center:* Proportions from random samples will vary but in the long run, sample proportions average out to the ____; therefore, the mean of the sample proportions equals __________________________.
- *Spread:* Larger random samples will vary (*circle one*: less, more, the same amount) compared to smaller samples. Therefore, increasing the sample size will (*circle one*: increase, decrease, not affect) the standard deviation of sample proportions.
- *Shape:* For larger samples, the shape of the distribution of sample proportions is approximately _______________________.

At a time when technology was less advanced (or non-existent!), statisticians were not able to run simulations. Instead, they developed mathematical models to describe the distribution of sample proportions. A mathematical model is an equation and its associated curve.

Here are the features of the mathematical model we will use to describe the distribution of sample proportions:

- *Center:* Mean of the sample proportions is p, the population proportion.

- *Spread:* Standard deviation of the sample proportions depends on the population proportion, p, and the size of the sample, n. The standard deviation equals $\sqrt{\frac{p(1-p)}{n}}$.

- *Shape:* For large samples, the mathematical model is a normal curve. The normal curve is a good model for the distribution of sample proportions only when $np \geq 10$ and $n(1-p) \geq 10$. If these conditions are not met, then we have to rely on simulations instead of the normal curve.

Examples:

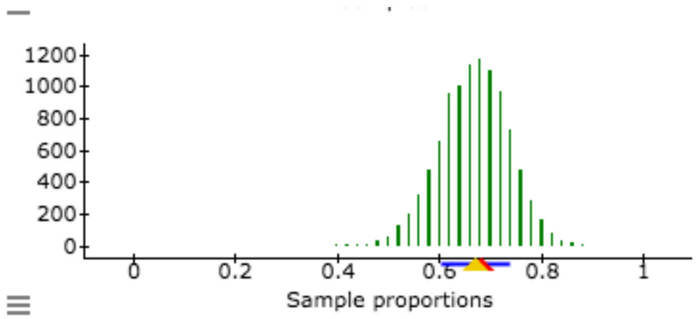
1) Use the mathematical model for the distribution of sample proportions to answer the following questions:

   Suppose that on the next ballot there is proposition on legalizing marijuana. In order for the proposition to pass, 67% of voters must vote yes. Prior to the election, a local newspaper conducts a poll of 50 randomly selected voters.

   a) Is a normal model a good fit for the distribution of sample proportions?

   b) In the long run, how much error do we expect on average in this situation based on the mathematical model?

c) We ran a simulation in StatCrunch with p=0.67 and n=50. Explain how the distribution of sample proportions from the simulation relates to the normal model.



Sample prop. of 1s

# of Samples	10000
Mean	0.6686
Median	0.68
Std. dev.	0.0671

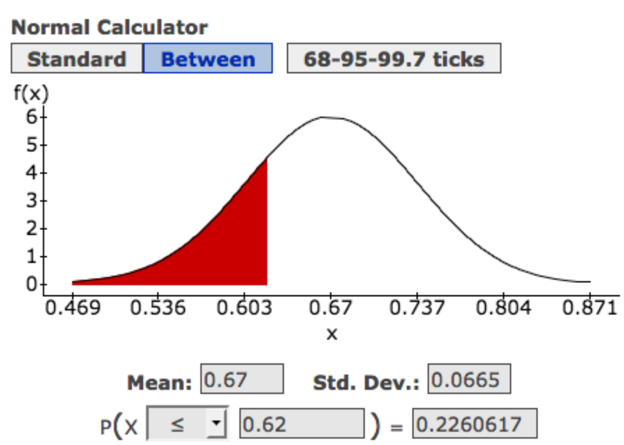
2) Now suppose prior to the election, the newspaper conducts a poll of 50 randomly selected voters and finds that 62% plan to vote yes on the proposition to legalize marijuana.

Since a 67% majority is needed to pass the proposition, this poll suggests that the proposition will not pass. However, if the 5% error is within expected amount of error for this situation, then this poll does not necessarily mean that the proposition will not pass. In other words, a sample with 62% voting YES may be fairly likely when sampling from a population with 67% voting YES. Let's see if this is the case.

a) Is 62% an unusual result for a random sample of 50 people when sampling from a population that has 67% supporting the proposition? How do you know?

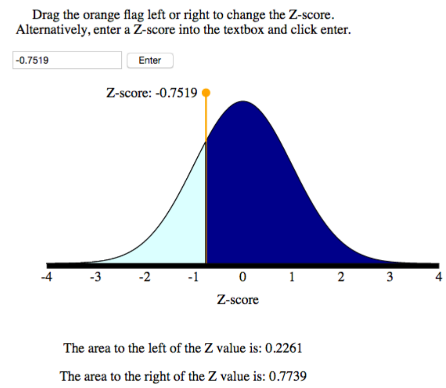
b) What is the probability that a random sample of 50 voters will have 62%, or even fewer, planning to vote YES? (Use the StatCrunch Normal Calculator to find the probability.)

How does this probability support your answer in (a)?



c) In OLI probability is determined using the standard normal curve for Z-scores.

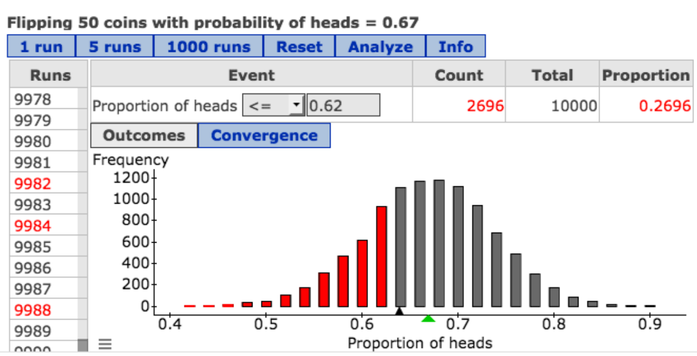
- Verify that the Z-score for a sample with 62% voting YES is approximately z = -0.7519.



- Why do we use the area <u>to the left</u> of z=-0.7519 to estimate the probability that 62% or fewer plan to vote YES?

- How does the probability using the standard normal curve compare to the probability using the normal curve in (b)?

d) We conducted a simulation to estimate the same probability. In this simulation a coin flip represents a vote. A head represents a YES vote. The coin is weighted so that 67% of the time it lands on a head. This represents 67% of the population voting YES. A sample consists of 50 coin tosses to represent a random sample of 50 people. The proportion of heads in 50 tosses represents the sample proportion voting YES. In the simulation, we selected 10,000 random samples (which means the coin was flipped 50 x 10,000 = 500,000 times!)



- Circle the samples in the distribution with proportions less than or equal to 0.62.
- According to the simulation, what is the probability that a random sample will have 62% or fewer planning to vote YES?

Notice that the probability estimates differ slightly when we use a mathematical model of the normal curve and when we use a simulation. This is because the normal model is a good fit for the distribution of sample proportions, but not a perfect fit! Anyway, both are estimates that come from different approaches, so don't get hung up on nailing down the probability exactly because this is not possible. What is important is to be able to use the probability estimate to draw a conclusion. In this case, both answers suggest that it is not that unusual to get a sample proportion of 62% or fewer from a population that would pass the proposition at 67%.

Group work:

1) In describing the LMC student population, suppose that a reporter writes, "The majority of LMC students have not used the CORE for academic assistance." The LMC Experience conducts a poll of 30 randomly selected LMC students to investigate whether this is true. In the poll, 30% (9 of the 30) report never using services in the CORE.
   a) Is 30% an unusual result if the reporter's claim is true? How do you know?

   b) Verify that a normal model fits the distribution of sample proportions in this case.

   c) Use the StatCrunch Normal Calculator (or the OLI z-score applet) to estimate the probability that a random sample of 30 students will have 30%, or even fewer, who have used services in the CORE, assuming that the reporter's statement about all LMC students is true. (Also sketch the normal curve and show the mean and standard deviation, or your z-score calculation.)

      (StatCrunch instructions: From the StatCrunch log in page, choose Open StatCrunch to open an blank spreadsheet. Under **Stat**, **Calculators**, choose **Normal**. Set the mean and standard deviation according to the formulas. Set the inequality as indicated in the probability question and enter the sample proportion. Hit Compute.)

2) In an article titled "Tatoos Becoming More Accepted at Work", CBS News reported in 2007 that 23% of college students were tattooed. Let's use this as a hypothesis for the proportion the population of LMC students who are tattooed.

   a) Suppose we randomly select 30 LMC students and find that about 33% are tattooed (10 out of 30). Is this an unusual result if our hypothesis is true? How do you know?

   b) Verify that a normal model is NOT a good fit for the distribution of sample proportions in this situation.

   c) Because the normal model is not a good fit, we must use a simulation to estimate probabilities. Conduct a coin-flipping simulation in StatCrunch to answer the following question: What is the probability that 33% or more are tattooed in a random sample of 30 if our hypothesis is true?

      Explain how you used the results of the simulation to estimate the probability. Include enough information about the set-up of the simulation that someone else could replicate your work.

      (StatCrunch instructions: From the StatCrunch log in page, choose Open StatCrunch to open an blank spreadsheet. Under **Applets**, **Simulation**, choose **Coin flipping**. Under **Simulate coin tosses**, set the probability of heads to the hypothesized population proportion. Under **Tally heads in tosses**, click **Proportion**. Set the inequality as indicated in the probability question and enter the sample proportion.)

Module 18 Introduction to Statistical Inference

Module 18.1 Introduction to Confidence Intervals

Learning Goal: Find a confidence interval to estimate a population proportion when conditions are met. Interpret the confidence interval in context.

Learning Objective: Use information from a media report of a survey to construct and interpret a confidence interval.

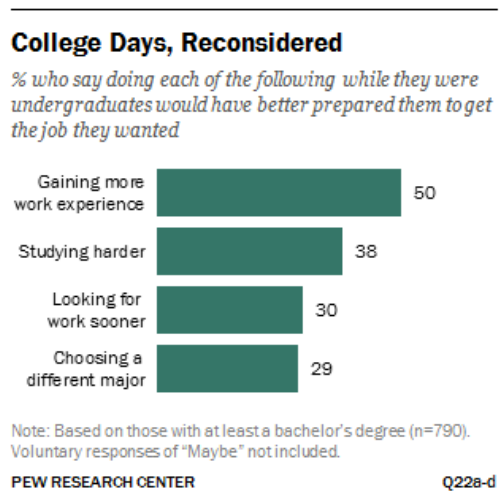
Introduction: Our goal in statistical inference is to *infer* something about a population from a sample. For example, suppose that researchers want to estimate the proportion of the population of U.S. adults that support the death penalty. It is too expensive and difficult to survey every U.S. adult. So researchers will poll a random sample of U.S. adults and use the sample proportion to draw a conclusion about the population proportion.

In this activity we will learn about one type of statistical inference called a *confidence interval.* We construct a confidence interval when our goal is to estimate a population proportion.

Example: Constructing a confidence interval from a poll.

The Pew Research Center asked college graduates what they could have done, while still in school, to be better prepared for their dream job.

We will use the results of this poll to estimate the proportion of the population of all college graduates who wished they had studied harder.
In the report, the researchers state that the poll had a margin of error of 4.3 percentage points at the 95% confidence level.



Source: http://www.pewresearch.org/methodology/u-s-survey-research/sampling/

1) What percentage of the sample of 790 college graduates said they wished they had studied harder?
2) Our estimate for the population proportion will be an interval of values, instead of a single number. Statisticians call this interval a *confidence interval.* The confidence interval is based on a sample proportion and a margin of error. Here is the formula: *sample proportion ± margin of error*

   Write the confidence interval to estimate the proportion of the population of all college graduates who wished they had studied harder.

3) Interpreting the confidence interval: For the population of all college graduates, we are 95% confident that the proportion that wish they had studied harder in college is between ________ and _________.

   Note: We will discuss the phrase '95% confident' in depth later.

4) In order to be confident that the sample represents the population without bias, we must have a random sample. Read the following excerpt from the Pew Research methodology for this *College Days* study.

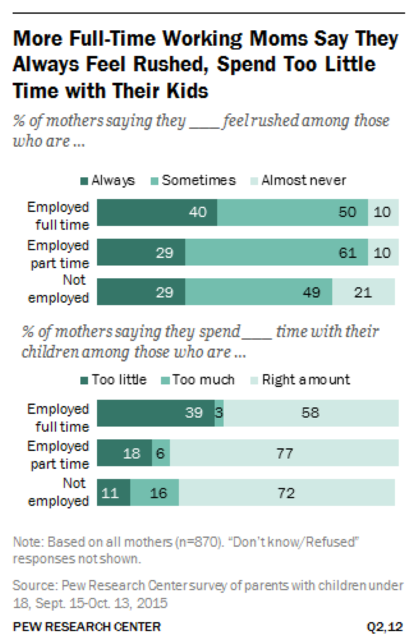
   "A majority of Pew Research Center surveys are conducted among the U.S. general public by telephone using a sampling method known as random digit dialing or "RDD." This method ensures that all telephone numbers in the U.S – whether landline or cellphone – have a known chance of being included. As a result, samples based on RDD should be unbiased, and a margin of sampling error and a confidence level can be computed for them.

   Is this a random sample? How do you know?

Group work:

1) The graph pictured here summarizes some of the results of a poll about parenting conducted by the Pew Research Center.

   Source:http://www.pewsocialtrends.org/2015/11/04/raising-kids-and-running-a-household-how-working-parents-share-the-load/



   a) The report states, "the error attributable to sampling is plus or minus 3.9% at the 95% level of confidence."

      For the population of all mothers employed full time, calculate a confidence interval to estimate the proportion of the who feel that they spend too little time with their children.

   b) Interpret the confidence interval:
      For the population of all mothers employed full time, we are 95% confident that between ________ and _________ feel they spend too little time with their children.

c) In order to be confident that the sample represents the population without bias, we must have a random sample. Read the following excerpt from the Pew Research methodology.

   "The analysis in this report is based on telephone interviews conducted from Sept. 15 to Oct. 13, 2015, among a nationally representative sample of 1,807 parents, 18 years of age or older, with children under 18, living in all 50 U.S. states and the District of Columbia. A combination of landline and cell phone random digit dial (RDD) samples was used. Interviews were conducted in English and Spanish. The survey results pictured here came from the responses of 870 mothers."

   Is this a random sample? How do you know?

2) Based on the McClatchy-Marist Poll summarized below, can we conclude that the majority of registered voters nationwide support a ban on the sale of assault weapons and semi-automatic weapons? Why or why not?

McClatchy-Marist Poll. July 5-9, 2016. N=1,053 registered voters nationwide. Margin of error ± 3.

"Do you think Americans are safer with more guns or fewer guns?"

	More guns %	Fewer guns %	Number is about right (vol.) %	Unsure %
ALL	45	46	3	5
Democrats	16	77	2	4
Republicans	79	16	3	3
Independents	44	44	5	7

"Do you favor or oppose a law to ban the sale of assault weapons and semi-automatic rifles?"

	Favor %	Oppose %	Unsure %
ALL	51	46	3
Democrats	74	23	3
Republicans	30	66	4
Independents	48	49	3

3) According to a Gallup poll conducted in September 2016, "64.2% of U.S. adults report that they ate healthy all day yesterday."

   Results are based on telephone interviews conducted as part of the Gallup-Healthways Well-Being Index survey, with a random sample of 2,415,499 adults, aged 18 and older, living in all 50 U.S. states and the District of Columbia. For results based on the total sample of national adults, the margin of sampling error is ± .08 percentage points at the 95 percent confidence level.

   a) Construct a confidence interval to estimate the proportion of U.S. adults who ate healthy yesterday.

   b) Interpret your confidence interval.

   c) This is a very precise confidence interval with a very, very small margin of error (0.08% is less than 1/10 of one percent). Why do you think the margin of error is so small for this poll?

4) What is the purpose of a confidence interval?

Module 18.2 Finding a 95% Confidence Interval

Learning Goal: Find a confidence interval to estimate a population proportion when conditions are met. Interpret the confidence interval in context.

Specific Learning Objectives:

- Construct a 95% confidence interval from scratch.
- Check conditions that allow the use of a 95% confidence interval.

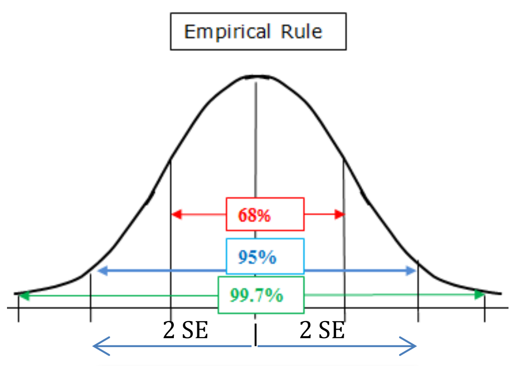
Introduction

In this activity we will learn how to construct a confidence interval from scratch.

Recall that the confidence interval is *sample proportion ± margin of error.*
For a 95% confidence interval the margin of error is two standard errors.

sample proportion ± margin of error
sample proportion ± 2 SE

This makes sense because of the Empirical Rule. If we randomly select a sample, there is a 95% chance that the sample proportion lies within two standard errors of the population proportion. Of course, this is only true if the distribution of sample proportions is normal in shape.



We learned previously that the standard error (standard deviation) for the distribution of sample proportions is $\sqrt{\frac{p(1-p)}{n}}$, where p is the population proportion and n is the sample size.

Putting the pieces together gives the formula for the 95% confidence interval:

Sample proportion ± margin of error
Sample proportion ± 2 standard errors

$$\text{Sample proportion} \pm 2\sqrt{\frac{p(1-p)}{n}}$$

IMPORTANT NOTE: This formula can only be used if a normal model is a good fit for the distribution of sample proportions.

You may realize that the formula for the confidence interval is a bit odd, since our goal in calculating the confidence interval is to estimate the population proportion, p. Yet the formula requires that we know p in order to calculate the standard error. **For now, we use an estimate for *p* from a previous study when calculating the confidence interval, as we have done before.** This is not the usual way statisticians estimate the standard error, but it captures the main idea and allows us to practice finding and interpreting confidence intervals. Later, we explore a different way to estimate standard error that is commonly used in statistical practice.

Example: The National Health Interview Survey for 2011-2014 estimates that 37% of 18-24 year olds are not meeting the Physical Activity Guidelines for Americans. These guidelines include recommendations for both aerobic and muscle-building exercise.

Suppose that we survey 30 randomly selected LMC students in this age category and find that 45% are not meeting the Physical Activity Guidelines for Americans. Find the 95% confidence interval and interpret it.

- Verify that the conditions for the use of the normal model are met based on the NHIS study.

- Calculate the standard error based on the NHIS study.

- Find the 95% confidence interval and represent it as a line segment.

- Interpret the interval.

Here is a link if you want to read about the Physical Activity Guidelines later: https://health.gov/paguidelines/

Group Work:

1) Using the data from the American Community Survey (IPUMS), the Pew Research Center reported that 32.1% of 18- to 34-year-olds was living with their parents in their parents' home in 2014. For the first time in recorded history, this was the most common living arrangement for this age group.

   What percentage of 18- to 34-year-olds is living at home with their parents this year?

   To answer this question, suppose that this year researchers conduct another survey with a random sample of 1,100 adults in this age group. Suppose that 35% are living at home with their parents. Assuming that the variability in random samples will be the same as in the 2014 study, find the 95% confidence interval for this year and interpret it.

   a) Verify that the conditions for the use of the normal model are met based on the Pew Research Center study.

   b) Calculate the standard error based on the 2014 study.

   c) Find the 95% confidence interval for this year and represent it as a line segment.

   d) Interpret your 95% confidence interval.

2) What percentage of Contra Costa County adults supports a ban on assault-style weapons?

   Suppose that we survey a random sample of 100 adults living in Contra Costa County and find that 62% support a ban on assault-style weapons.

   According to a 2015 study by the Pew Research Center, 57% of U.S. adults favor a ban on assault-style weapons. Assuming that the variability in random samples will be the same in Contra Costa County as in the U.S., find the 95% confidence interval to estimate the proportion of CCC adults that favor a ban on assault-style weapons.

a) Verify that the conditions for the use of the normal model are met using the 2015 study.

b) Why do we check conditions for use of the normal model?

c) Calculate the standard error based on the 2015 study. What does this number tell us?

d) Find the 95% confidence interval for Contra Costa County.

e) Write a sentence to interpret the interval.

f) Are we confident that the majority of Contra Costa County residents support a ban on assault-style weapons? Why or why not?

g) Are we confident that the percentage of Contra Costa County residents that supports a ban is greater than the percentage nationwide as reported by the Pew Research Center? Why or why not?

Module 18.3 What Does "95% Confident" Really Mean?

Learning Goal: Interpret the confidence level associated with a confidence interval.

Learning Objective: Explain the meaning of "95% confident."

Introduction:
When we say "we are 95% confident that between 44% and 52% of Americans drink soda every day, what do we really mean by the phrase "we are 95% confident"? Let's find out.

To investigate the meaning of "95% confident", we will assume that we know the population proportion for some context and then we will examine how often a confidence interval contains this population proportion.

Simulation:
According to the Community College Research Center, 80% (or more) of community college students intend to earn at least a bachelor's degree.
Source: http://ccrc.tc.columbia.edu/media/k2/attachments/what-we-know-about-transfer.pdf

For our simulation, let's assume that 80% is a population parameter. In other words, we are going to assume that the finding in this study actually holds true for the entire population of community college students.

Based on this assumption, let's simulate collecting random samples of 100 students and calculate some confidence intervals. Our goal is to see how often a confidence interval will accurately estimate the population proportion. By this we mean that the confidence interval contains the population proportion of 0.80.

1) Verify that the distribution of sample proportions for this situation can be modeled by a normal curve.

2) Use the StatCrunch Sampling Distribution applet to select a random sample with p=0.80 and n=100. Calculate the sample proportion and the associated confidence interval.

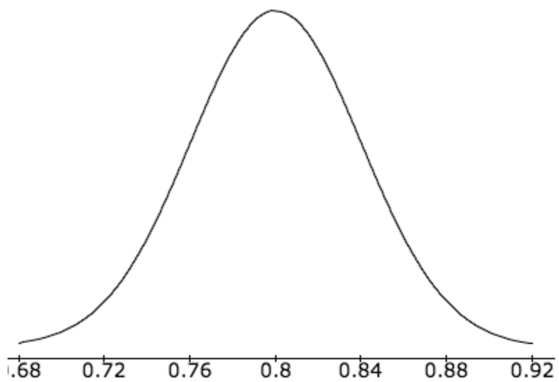
Your instructor's sample proportion and confidence interval:

Interpretation:

Your sample proportion and confidence interval:

Interpretation:

3) Here is the normal curve with mean of p=0.80 and standard error of 0.04, which models the distribution of sample proportions. We will plot a few confidence intervals below it.

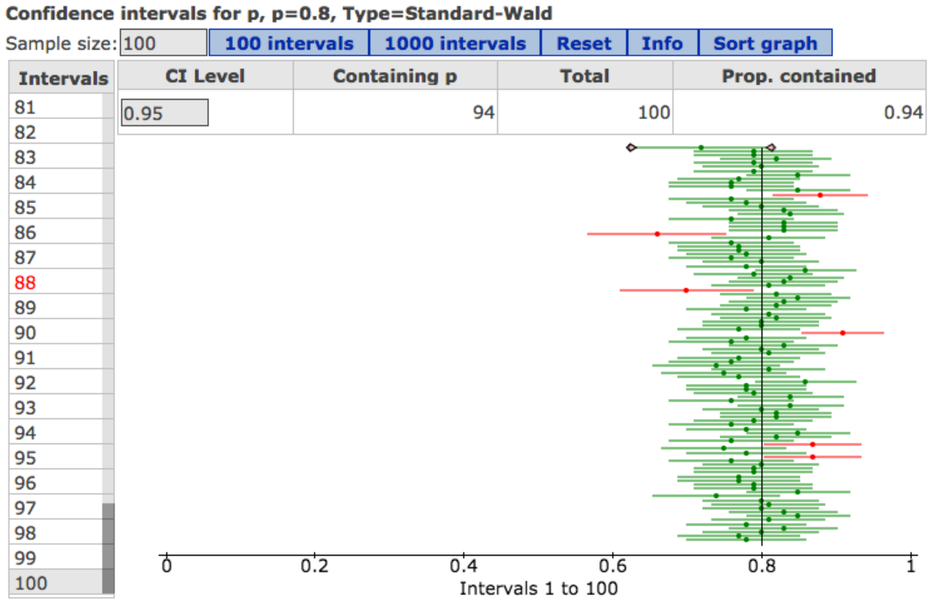


All of the confidence intervals are the same length. Why does this make sense?

4) Which of the confidence intervals that we plotted accurately estimate the population proportion? (In other words, which of the confidence intervals that we plotted contain p=0.80?)

5) What proportion of the class got a confidence interval that did NOT contain p=0.80?

6) The image below represents 100 confidence intervals with p=0.80 and n=100.



a) What does each line segment represent?

b) Why is there a dot in the middle of each line segment? What does it represent?

c) Circle a confidence interval that does not accurately estimate the population proportion. Why doesn't this interval contain p?

d) What percentage of the confidence intervals plotted here accurately estimate the population proportion? (In other words, what percentage of the confidence intervals plotted here contain p=0.80?)

e) In the long run, if we collected thousands and thousands of random samples, what percentage of the associated confidence intervals will contain p? Why do you think so?

f) What do you think "95% confident" means?

7) A student selected a random sample with a sample proportion of 0.85 and correctly calculated the associated confidence interval as 0.77 to 0.93. He also correctly interpreted the interval as "Of all community college students, we are 95% confident that the proportion who intend to get at least a bachelor's degree is between 0.77 and 0.93."

   When asked to interpret the meaning of the phrase "95% confident," the student wrote "95% confident means that 95% of the time the population proportion lies between 0.77 and 0.93." But this is incorrect.

   Use the image of the StatCrunch simulation on the previous page to explain why his interpretation of "95% confident" is incorrect.

Module 18.4 Introduction to a Hypothesis Test

Learning Goal: Test a hypothesis about a population proportion using a simulated sampling distribution or a normal model of the sampling distribution. State a conclusion in context.

Introduction:

Now we focus on the second type of inference: hypothesis testing. We are going to work on the logic of a hypothesis test now and address the more formal aspects of it later.

In hypothesis testing, we make a claim about a population proportion and use a sample proportion to test it. This is very similar to the thinking we did with simulations in the previous module.

Example: According to a 2013 survey of college students conducted by Citi and Seventeen Magazine, 20% of college students have a credit card when they start college. Is the percentage higher for community college students this year?

Here is our claim: More than 20% of community college freshmen have a credit card this year.

We start with the hypothesis that 20% of community college freshmen have a credit card. Our claim is that the percentage is higher. If our claim is true, then our hypothesis about the population is wrong.

To test the hypothesis, we select a random sample of community college freshmen. Suppose that the data shows that 25% of the sample has a credit card.

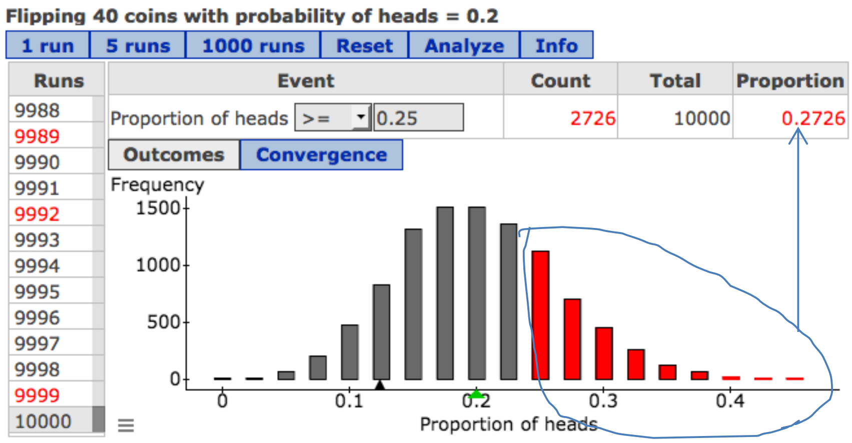
1) Does this sample support our claim? Why or why not?

But wait! Maybe when we collect random samples from a population with 20% owning a credit card, we will find that a sample proportion of 25% is not that unusual. Maybe the 5% error is not that big relative to the typical error we will see.

Surprise, surprise … we need to determine that amount of the variability expected in random samples drawn from a population where 20% have a credit card before we can draw any conclusions.

We now have two ways to examine variability: simulation or the normal model (if conditions are met.) We know that the amount of variability in random samples depends on the sample size. Let's examine what happens when we select samples of 40 freshmen (n=40). Then let's see what happens when we collect larger samples of 300 (n=300)

2) *Sample size of 40:* Here are the results from a coin-flipping simulation in StatCrunch with probability of a head set to 0.20. Each "run" is 40 coin flips representing a random sample of 40 community college freshmen.

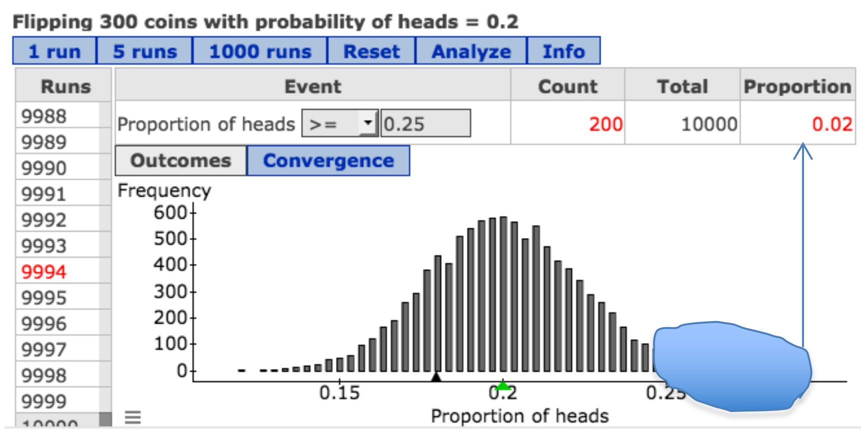


a) With a sample of 40 freshmen, is a proportion of 0.25 unusual when the population proportion is 0.20? How do you know?

b) Based on the simulation, what is the probability that a random sample of 40 has 25% or more with a credit card? How do you know?

c) What does this suggest about our claim? Does our sample data provide strong evidence that more than 20% of the entire population own a credit card? Why or why not?

3) *Sample size of 300:* Here are the results from a coin-flipping simulation in StatCrunch with probability of a head set to 0.20. Each "run" is 300 coin flips representing a random sample of 300 community college freshmen.



a) With a sample of 300 freshmen, is a proportion of 0.25 unusual when the population proportion is 0.20? How do you know?

b) What is the probability that a random sample of 300 has 25% or more with a credit card? How do you know?

c) What does this suggest about our claim? Does our sample data provide strong evidence that more than 20% of the entire population own a credit card? Why or why not?

Summary:

Are we right that more than 20% of the all community college freshmen own a credit card? In other words, is our claim true?

- If the sample proportion of 0.25 is fairly typical, then the associated probability is large. This means that it is not surprising to see random samples with 25% or more owning a credit card when 20% of the population owns one. In other words, even though more than 20% of the sample had a credit card, **the deviation is not large enough** to support our claim that the population proportion is also larger than 20%.

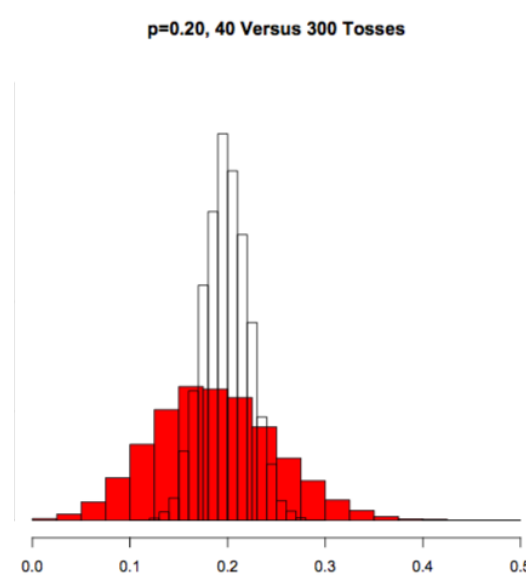
  Therefore, we conclude that our **claim is not true.**

- If the sample proportion of 0.25 is unusual, then the associated probability is small. This means that it is surprising to see random samples with 25% or more owning a credit card when 20% of the population owns one. In other words, this **deviation is large enough** to support our claim that the population proportion is also larger than 20%.

  Therefore, we conclude that our **claim is true.**

4) How can the same sample result (a sample proportion of 25% in this case) lead to different conclusions about the population proportion?

   We put both sampling distributions on the same axis for easier comparison. Use this image to explain why a sample proportion of 25% gives different conclusions when n=40 and n=300.



Group work:

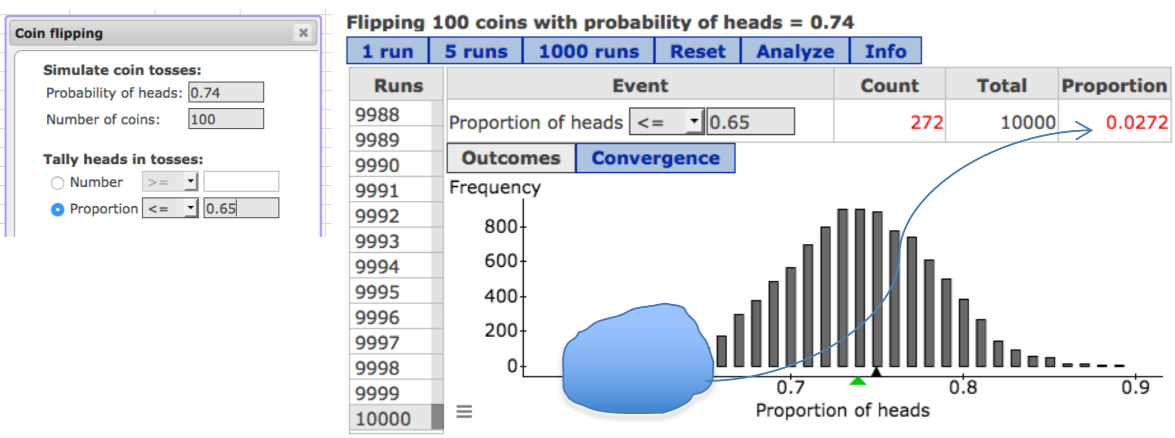
5) "Authoritative numbers are hard to come by, but according to a 2002 confidential survey of 12,000 high school students, 74 percent admitted cheating on an examination at least once in the past year."
Source: http://abcnews.go.com/Primetime/story?id=132376&page=1

A Principal of a local high school is appalled by this statistic. She wants to determine if the situation is any better at her school.

She hires a firm to conduct a confidential survey of a random sample of 100 students at her school. Sixty-five percent of the 100 students admit to cheating on exams.

a) Identify the following for this scenario:
   - The hypothesized population proportion
   - The sample proportion from the data
   - The sample size

b) The consultant runs a simulation and gets the following results. Explain how your answers in (a) relate to this simulation.



c) Is a sample proportion of 65% unusual when drawing random samples of 100 students from a population in which 74% cheat on exams? How do you know?

d) Based on this simulation, what is the probability that a random sample of students will have 65% or fewer cheating on exams if 74% of the population of students is cheating? How do you know?

e) What does this simulation suggest about the principal's hope that cheating is less prevalent at her school?

f) Could the principal have used a normal model instead of a simulation? Why or why not?

g) The principal wants an estimate of the proportion of students who cheat on exams at her school. Calculate a 95% confidence interval and interpret it. Does the confidence interval support your answer to (e)?

Module 18.5 Unit 7 Lab

Background: Researchers in the Psychology Department at Yale University investigate the development of moral, social and cognitive behaviors in infants as a way of gaining insight into the underpinnings of human behavior.

In an introduction to a famous study published in Nature in 2007, Yale researchers wrote, "The capacity to evaluate other people is essential for navigating the social world. Humans must be able to assess the actions and intentions of the people around them, and make accurate decisions about who is friend and who is foe, who is an appropriate social partner and who is not."[1]

The following summary of this study is from Science Daily[2]

"The 2007 study by Yale University researchers provided the first evidence that 6- and 10-month-old infants could assess individuals based on their behavior towards others, showing a preference for those who helped rather than hindered another individual. … In the original experiment, infants watched a wooden toy (i.e., the "climber") attempt to climb a hill. They viewed two social interactions; one in which a "helper" toy nudged the climber up the hill, and another in which a "hinderer" toy nudged the climber down the hill.

After viewing these two scenarios, the infants were presented with a tray; on one side of the tray was the helper and on the other side was the hinderer. Amazingly, the majority of infants picked the helper over the hinderer. To further elucidate infants' moral reasoning abilities, a "neutral" toy (i.e., a toy that neither helped nor hindered) was pitted against the helper or hinderer. When the neutral character was paired with the helper, the infants preferred the helper; when paired with the hinderer, they preferred the neutral character. … The paper concluded that the experiments show that infants can evaluate individuals based on how they interact with another individual, and that their ability to do this is 'universal and unlearned'."

If you are interested in seeing short videos of this experiment, here is a link: http://www.yale.edu/infantlab/socialevaluation/Helper-Hinderer.html

In the Yale study, the infants watched the videos several times. Afterwards, when the infants were shown a tray with two pieces of wood, one shaped like the helper and the other shaped like the hinderer in the video, 14 of the 16 infants chose the helper over the hinderer. The Yale researchers concluded that infants prefer the helper toy and that this implies that infants have an innate ability to evaluate individuals based on how they interact with another individual.

---

[1] http://www.nature.com/nature/journal/v450/n7169/full/nature06288.html

[2] http://www.sciencedaily.com/releases/2012/08/120815093230.htm

But wait … how do we know that infants really prefer the helper toy when we only have results from a single sample? Could these results occur if infants are just arbitrarily choosing a toy? What's the probability that this happens? We will conduct a simulation to answer these questions and see if simulation supports the conclusions reached by the Yale researchers.

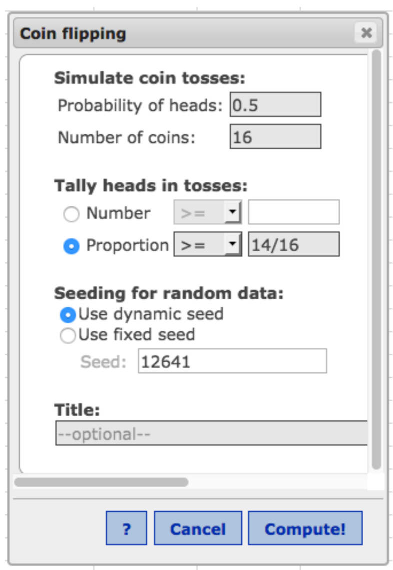
Simulation: The StatCrunch applet simulates flipping a coin. It makes sense to use a coin flip in the simulation because in the experiment the infants had a choice between two toys (helper and hinderer).

Let's assume that the population of all infants does <u>not</u> have a preference. If this is true, then their choices can be simulated with a coin flip.

In the simulation each coin toss will represent one infant's choice. A head tells us that the infant chose the helper toy. A tail tells us that the infant chose the hinderer toy. Set up the applet as shown here.

<u>Setting up the simulation</u>: Log into StatCrunch. Choose Open StatCrunch. Choose Applets, Simulation, Coin Flipping.

Set up the simulation as shown below:



Probability of heads: 0.5 (Because we are assuming that the infants are arbitrarily choosing, in the long run we expect 50% to choose the helper toy.)

Number of coins: 16 (There are 16 infants in each sample)

Tally heads in tosses: Proportion >= 14/16 = 0.875 (Because we want to see if the results from the Yale study are unusual, we want to find the probability that a sample will have a proportion of 14/16 or greater.)

Hit Compute!

Click on "1 run" and watch what happens. Repeat this three or more times until you understand how the simulation works.

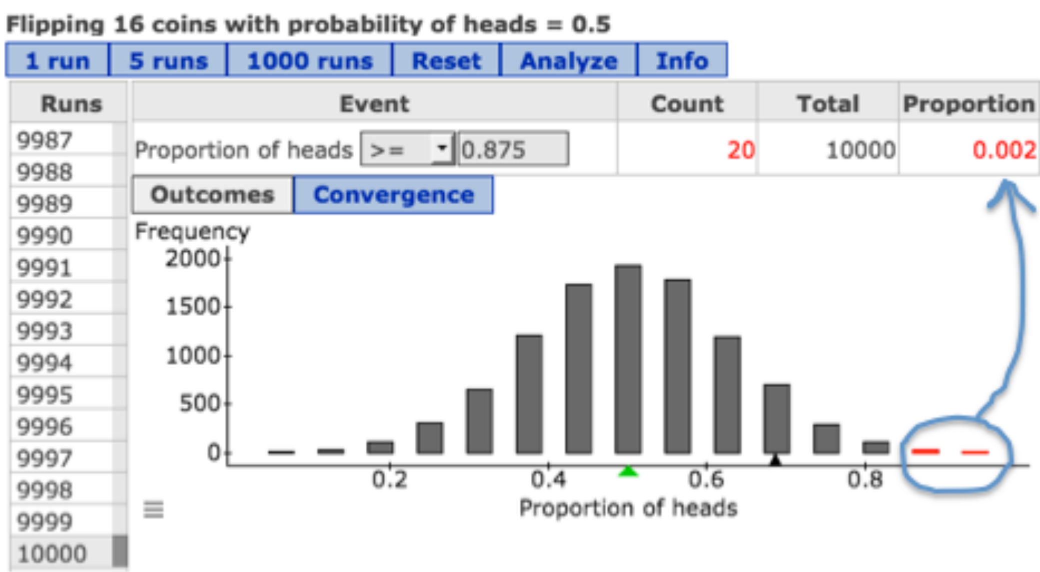
Questions about the distribution of sample proportions:

1) Here are the results from one sample of 16 coin tosses.

   a) What is the sample proportion for the 16 coin tosses pictured?

   b) Write a sentence to explain the meaning of the sample proportion in the context of infants and toys.

2) Here is the graph from a simulation of 10,000 samples (called "runs" in the StatCrunch applet.)



   a) If we had a dot plot instead of a histogram, there would be 10,000 dots. What would a dot represent? (One infant, a sample of 16 infants, a sample of 10,000 infants, or other?)

   b) In the graph the horizontal axis says "proportion of heads". Write a more accurate label for these values in terms of babies and toy preference.

c) For these 10,000 samples, apply the theory you learned to estimate the mean and the standard deviation of the distribution of sample proportions. Explain or show how you got your estimates.

d) Is a normal model a good fit for this distribution? Why or why not?

Questions about making an inference based on a distribution of sample proportions:

3) In the Yale experiment, 14 out of 16 infants chose the helper toy. This a sample proportion of 0.875.

   a) Is 14 out of 16 unusual in a random sample if we assume p=0.5? Explain how you know.

   b) What is the probability that in a random sample of 16 infants, 14 or more will choose the helper toy? Explain how you determined the probability.

c) The Yale researchers claim that infants recognize, and make judgments about, other individuals based on how they interact with each other. To show this, their experiment must produce strong evidence that the infants are not arbitrarily picking one toy over the other. Does their experiment provide strong evidence in support of their claim that infants have a preference for the helper? Why or why not?

4) Researchers in the Department of Psychology at the University of Otago in New Zealand have challenged the findings from the Yale study. Read the short article from Science Daily and describe the confounding variables that are the focus of the New Zealand critique. Summarize what the New Zealand researchers did to investigate the effects of these confounding factors.

http://www.sciencedaily.com/releases/2012/08/120815093230.htm

UNIT 8

Inference for One Proportion

Contents

Module 19 Estimating a Population Proportion

Learning Goal: Construct a confidence interval to estimate a population proportion when conditions are met. Interpret the confidence interval in context.

Specific Learning Objectives:
- Calculate confidence intervals using a sample proportion (instead of p from a previous study) to estimate the margin of error
- Calculate confidence intervals for different levels of confidence
- Describe the effect of increasing sample size on the margin of error

Introduction
We will continue our discussion of confidence intervals in this activity.

1) What is the purpose of a confidence interval?

2) Which of the following questions would be answered by calculating a confidence interval?

   - What is the average amount of money that community college students receive in financial aid?
   - Do the majority of community college students qualify for federal student loans?
   - What proportion of CA community college students qualify for the BOG fee waiver?

3) When we want to estimate a population proportion with a 95% confidence interval, we used the formula: *Sample proportion ± margin of error*

$$\hat{p} \pm 2\sqrt{\frac{p(1-p)}{n}}$$

   a) This interval is estimating a population proportion, p. Yet p appears in the formula. How did we handle this previously?

   b) The margin of error is based on the standard error. What does the standard error tell us?

   c) Why is the margin of error 2 times the standard error, instead of 3 times or something else?

   d) What conditions have to be met before we can use this formula?

So what's new in this Module? We will not have a previous study to estimate a value for the population proportion in order to calculate the standard error. Instead we will estimate the standard error using the sample proportion.

Examples:

4) In an evaluation of academic counseling services on campus, the Chair of the Counseling Department randomly selects 100 students and emails them a survey. Of the 65 who responded, 12 rated their counseling experience as "unsatisfactory".

   Based on these results, estimate the proportion of the entire student body that has had "unsatisfactory" academic counseling.

   a) Notice that we do not have a previous study that gives us an estimate for p. Therefore, we will estimate p using the sample proportion $\hat{p}$. This changes the way that we verify that the normal model is a good fit for the distribution of sample proportions.

      Verify that conditions are met for use of the normal model using $\hat{p}$ as an estimate for p.

   b) Construct the confidence interval: $\hat{p} \pm 2\sqrt{\frac{p(1-p)}{n}}$ becomes $\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

   c) Interpret the confidence interval in the context of counseling.

5) While 95% confidence intervals are the most common in the media, it is possible to adjust the formula to calculate confidence intervals with other levels of confidence.

   a) What would change in our previous example if we calculated the 99.7% confidence interval?

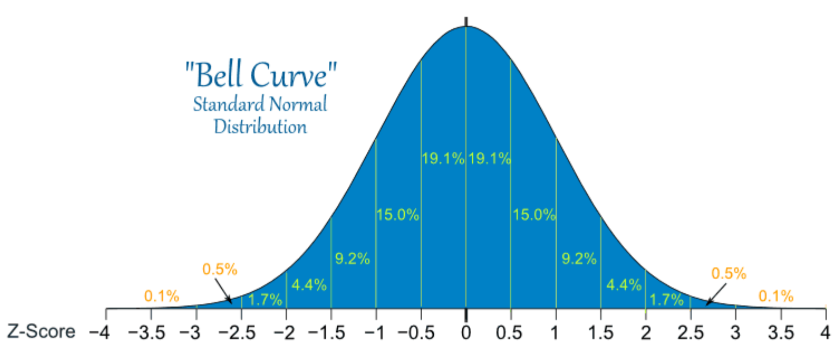
   b) Calculate the 99.7% confidence interval.

   c) Does the error get larger or smaller? Why does this make sense?

   d) Why might someone want to calculate a 99.7% confidence interval instead of a 95% confidence interval?

6) In real life statistical practice, occasionally, you might see a 90% confidence interval or 99% confidence interval.

   a) In general, what is formula for the margin of error for a 90% confidence interval? Use the standard normal curve to estimate it.



Summary:
The general formula for a confidence interval for a population proportion is $\hat{p} \pm Z_c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. The notation Z_c stands for "critical z-score". The critical z-score is determined by the confidence level.

Confidence level	68%	90%	95%	99%	99.7%
Critical z-score	1	1.645	2 or 1.96	2.576	3

Group work: Information in these problems was obtained from Gallup's *Confidence in Institutions* poll at http://www.gallup.com/poll/1597/confidence-institutions.aspx. Gallup's polling methodology uses random-digit-dial methods, so we can treat these samples as random samples. From information given in the methodology section, assume that 450 U.S. adults were surveyed in June 2003 and in June 2012.

7) "Americans' confidence in the U.S. military, the institution with the highest confidence rating in June 2012 (75% of adults surveyed reported a great deal or quite a lot of confidence in the U.S. military), is down only seven percentage points from an all-time high of 85% in June 2003."

    a) Use a 90% confidence interval to estimate the proportion of the population of U.S. adults in 2003 that had a high level of confidence in the U.S. military (answered "great deal" or "quite a lot" of confidence).

    b) Use a 90% confidence interval to estimate the proportion of the population of U.S. adults in 2012 that had a high level of confidence in the U.S. military.

    c) Based on your confidence intervals, could the proportion of U.S. adults with a high level of confidence in the U.S. military be the same in 2003 and as in 2012? Why or why not?

    d) What does a statistician mean by "90% confident?"

8) According to the same survey, "The church or organized religion has lost twenty-four percentage points with 44% of those surveyed reporting a great deal or

quite a lot of confidence in the church in June 2012 compared to 68% in May 1975."

The following StatCrunch print-outs show the 90% and 99% confidence intervals for estimating the proportion of U.S. adults with a high level of confidence in churches or organized religion in 2012.

One sample proportion confidence interval:
p : Proportion of successes
Method: Standard-Wald

90% confidence interval results:

Proportion	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
p	198	450	0.44	0.023399905	0.40151058	0.47848942

One sample proportion confidence interval:
p : Proportion of successes
Method: Standard-Wald

99% confidence interval results:

Proportion	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
p	198	450	0.44	0.023399905	0.37972584	0.50027416

a) Write the 90% and 99% confidence intervals from the StatCrunch print-outs.

b) Which interval has the largest margin of error? How do you know?

c) Which interval is the most likely to actually contain the proportion of U.S. adults with a high level of confidence in churches or organized religion in 2012? How do you know?

d) Which interval do you think is best for estimating the proportion of U.S. adults with a high level of confidence in churches or organized religion in 2012? Why?

e) What could Gallup do to decrease the margin of error in the 99% confidence interval?

Module 20 Hypothesis Testing

Module 20.1 Hypothesis Testing

Learning Goals:

- Given a claim about a population proportion, determine null and alternative hypotheses.
- Recognize the logic behind a hypothesis test and how it relates to the P-value.
    - Compare P-values to a level of significance to draw a conclusion
    - State conclusions to hypothesis tests using the language of “reject the null” or “fail to reject the null” and statistical significance

Introduction:

In the last activity we focused on constructing confidence intervals. In this activity we will focus on hypothesis testing. Confidence intervals and hypothesis testing are the two types of statistical inference we will study this semester.

We use a hypothesis test when we want to find a “yes” or “no” answer to a claim about a population parameter.

1) Which of the following questions can be answered by a hypothesis test?

- What is the average amount of money that community college students receive in financial aid?
- Is the proportion of CA community college students that qualify for the BOG fee waiver greater than 40%?
- Do the majority of community college students qualify for federal student loans?

Now we will walk through an example of a hypothesis test. The hypothesis test is the same thinking we did in Activity 18.4, so we will review a problem from that activity and add to it the vocabulary and notation of a hypothesis test.

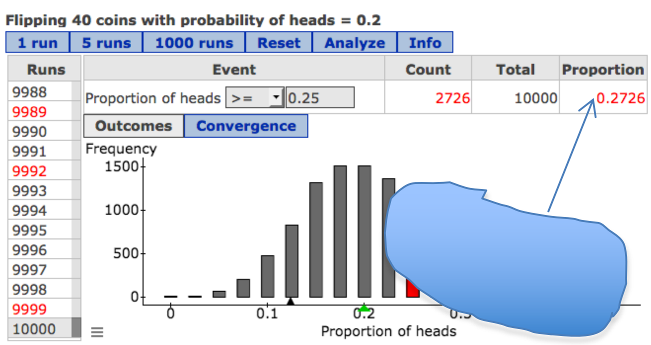
Steps in a hypothesis test:
Step 1: Determine the hypotheses.
Step 2: Collect the data and report the sample results.
Step 3: Assess the data.
Step 4: State a conclusion

Example: According to a 2013 survey of college students conducted by Citi and Seventeen Magazine, 20% of college students have a credit card when they start college. Is the percentage higher for community college students this year?

In a survey of a random sample of 40 community college freshmen, 25% have a credit card.

<table>
<tr><th>What we did before</th><th>Notation and vocabulary of a hypothesis test</th></tr>
<tr><td>We start with the hypothesis that 20% of community college freshmen have a credit card this year. Our claim is that the percentage is higher.</td><td>Step 1: Determine the hypotheses.
Ho: p=0.20
Ha: p>0.20
p=proportion of community college freshmen with a credit card this year</td></tr>
<tr><td>To test the hypothesis, we select a random sample of community college freshmen. Suppose that the data shows that 25% of the sample has a credit card.</td><td>Step 2: Collect the data and report the sample results.
$\hat{p}$ = 0.25</td></tr>
<tr><td>Obviously, 25% is greater than 20%. But we need to determine if this 5 percentage point difference is typical or unusual when we look at samples coming from the population.

Here is the distribution of 10,000 sample proportions from samples of 40 community college freshmen.



Visually, we can see that 0.25 is not unusual in this distribution so the associated probability is large.

About 27% of the time, we expect a random sample to have 25% or more with a credit card when 20% of the population owns one.</td><td>Step 3: Assess the data.

Find the P-value.

The P-value is 0.27, which is large.

This indicates that the survey result of 25% is not unusual when sampling from a population where Ho is true.</td></tr>
<tr><td>Are we right that more than 20% of the all community college freshmen own a credit card? In other words, is our claim true?

Even though more than 20% of the sample had a credit card, the difference is not large enough to support our claim that the <u>population proportion</u> is also larger than 20%.

Therefore, we conclude that our claim is not supported by the data.</td><td>Step 4: State a conclusion
The sample evidence is not statistically significant. The observed difference between 0.25 and 0.20 can be attributed to sampling variability.

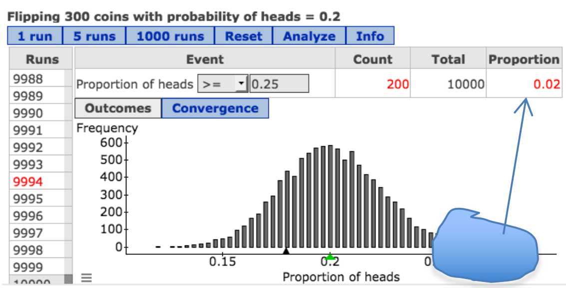
We do not have enough evidence to conclude that more than 20% of the population of community college freshmen owns a credit card.

Fail to reject Ho.</td></tr>
</table>

You may be wondering … how does the conclusion change when the P-value is small?

In 18.4 we also ran a simulation of 10,000 samples of 300 students for this scenario. Let's *assess the data* and *state a conclusion* with these larger samples.

Visually, we can see that 0.25 is unusual in this distribution so the associated probability is small.



About 2% of the time, we expect a random sample to have 25% or more with a credit card when 20% of the population owns one.

Step 3: Assess the data.

Find the P-value.

The P-value is 0.02, which is small.

This indicates that the survey result of 25% is unusual when sampling from a population where Ho is true.

Are we right that more than 20% of the all community college freshmen own a credit card? In other words, is our claim true?

It is surprising to see random samples with 25% or more owning a credit card when 20% of the population owns one. In other words, this *difference is large enough* to support our claim that the <u>population proportion</u> is also larger than 20%.

Therefore, we conclude that our *claim is true.*

Step 4: State a conclusion

The sample evidence is statistically significant. The observed difference between 0.25 and 0.20 cannot be attributed to sampling variability.

We have enough evidence to conclude that more than 20% of the population of community college freshmen owns a credit card.

Reject Ho in favor of Ha.

You may be wondering … how small does a P-value have to be in order to accept the claim as true and reject the null hypothesis?

Since the hypothesis test asks us to answer "Is the claim true?" (yes or no), it is common practice to agree ahead of time on the definition of "unusual" sample results so that we can answer this question without any haggling. This definition of "unusual" is called a **significance level.**

If the P-value is small enough to be less than the significance level, then the sample data is considered unusual. We conclude the sample data is statistically significant. This is strong evidence that our claim is true. (Reject Ho in favor of Ha.)

If the P-value is larger than the significance level, then the sample data is not considered unusual. We conclude the sample data is not statistically significant. We do not have sufficient evidence to support our claim. (Fail to reject Ho.)

Group work

2) In 2001 polls indicated that 74% of Americans favored mandatory testing in public schools as a way to rate the school. Your local school board conducts a survey to determine if a smaller proportion of residents in the district now support the use of mandatory test results in rating schools. They find that 68% of the residents are in support this year.

a) Which one of the following pairs of hypotheses fit this scenario?

- Ho: $p=0.74$, Ha: $p<0.68$
- Ho: $p=0.68$, Ha: $p<0.68$
- Ho: $p=0.74$, Ha: $p<0.74$
- Ho: $p=0.68$, Ha: $p<0.74$

b) What does p represent in the hypotheses? Write a sentence that clearly defines p.

3) Forty-six percent of high school students ages 12 to 17 in the United States have had sexual intercourse, according to a 2014 study by the Official Journal of the American Academy of Pediatrics (OJAAP).

Is this percentage higher in inner city high schools?

Suppose that a study finds that 52% of inner city high school students have had sexual intercourse. The sample data is used to test the hypotheses: Ho: $p=0.46$, Ha: $p>0.46$. The P-value is 0.02.

a) Which conclusion is supported by this P-value?

- The sample evidence is unusual when sampling from a population with p=0.46, so the data is statistically significant.
- The sample evidence is not unusual when sampling from a population with p=0.46, so the data is not statistically significant.

b) Which conclusion is supported by this P-value?

- Reject Ho in favor of Ha
- Fail to reject Ho

c) What can we conclude? (Choose all that apply)

- The difference between 0.46 and 0.52 is statistically significant.
- This survey provides enough evidence to conclude that the percentage of all high school students engaging in sexual intercourse is higher in the inner cities.
- This study does not suggest that inner city high school students are engaging in sexual intercourse at higher rates.
- The difference between 0.46 and 0.52 can be explained by expected sampling variability.

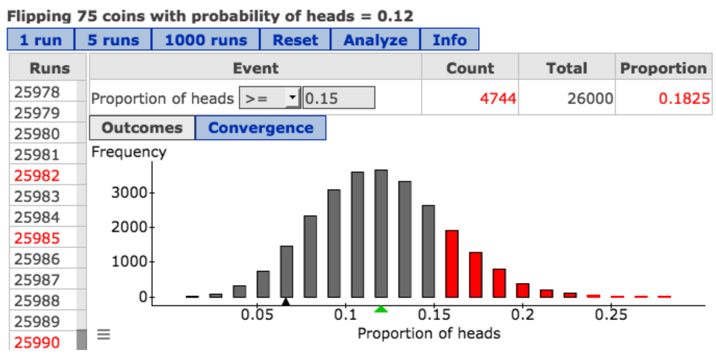
4) A 2011 California Student Survey found that in any 30-day period, almost 12 percent of students in the 9th grade admit to using drugs at least once at school.
Source: http://www.huffingtonpost.com/marsha-rosenbaum/drug-education_b_3906983.html

To investigate this issue in your community, suppose that the school board hires a firm to conduct a survey of a random sample of 9th graders to determine if drug usage is greater than 12% at schools in the district. Seventy-five students respond to the anonymous survey and 11 admit to drug usage at school. (11 out of 75 is about 15%)

The sample data is used to test the hypotheses:

Ho: p=0.12 (12% of the population of 9th graders have used drugs at school)
Ha: p>0.12 (More than 12% of the population of 9th graders have used drugs at school)

Here is the StatCrunch image from the simulation of selecting 26,000 random samples of 75 students.



a) What is the P-value?

b) Which conclusion is supported by this P-value?

- The sample evidence is unusual when sampling from a population with p=0.12, so the data is statistically significant.

- The sample evidence is not unusual when sampling from a population with p=0.12, so the data is not statistically significant.

c) Which conclusion is supported by this P-value?

- Reject Ho in favor of Ha

- Fail to reject Ho

d) What can we conclude? (Choose all that apply)

- The difference between 0.12 and 0.15 is statistically significant.

- This survey provides enough evidence to conclude that the percentage of 9th graders using drugs at school is higher than 12%.

- This study does not suggest that the drug usage rate for 9th graders in this district is higher than 12%.

- The difference between 0.12 and 0.15 can be explained by expected sampling variability.

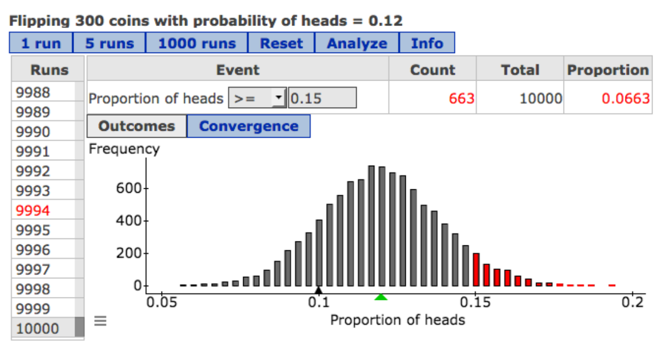
5) The school board is happy to hear that drug usage by 9th graders in the district is not higher than the statewide rate. But a concerned citizen raises a concern about the relatively small sample size. The school board agrees to redo the study with a larger sample.

Because of the controversy over the first study, the school board sets a 5% significance level to alleviate disagreements over the conclusions from the second study.

This time the research firm is able to get 300 9th graders to respond; 45 admit to drug use at school, which is again 15%. The sample data is again used to test the hypotheses:

Ho: p=0.12 (12% of the population of 9th graders have used drugs at school)
Ha: p>0.12 (More than 12% of the population of 9th graders have used drugs at school)

Here is the StatCrunch image from the simulation of selecting 10,000 random samples of 300 students.



a) What is the P-value?

b) Considering the significance level, which conclusion is supported by this P-value?

- The sample evidence is unusual when sampling from a population with p=0.12, so the data is statistically significant.

- The sample evidence is not unusual when sampling from a population with p=0.12, so the data is not statistically significant.

c) Considering the significance level, which conclusion is supported by this P-value?

- Reject Ho in favor of Ha

- Fail to reject Ho

d) What can we conclude? (Choose all that apply)

- The difference between 0.12 and 0.15 is statistically significant.
- This survey provides enough evidence to conclude that the percentage of 9th graders using drugs at school is higher than 12%.
- This study does not suggest that the drug usage rate for 9th graders in this district is higher than 12%.
- The difference between 0.12 and 0.15 can be explained by expected sampling variability.

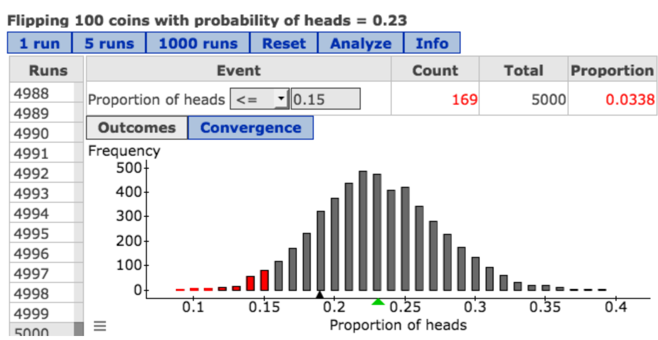
6) In an article titled "Tattoos Becoming More Accepted at Work", CBS News reported in 2007 that 23% of college students were tattooed. Let's use this as a hypothesis for the proportion the population of LMC students who are tattooed. Test this hypothesis against a level of significance of 5%.

Suppose that in a random sample of 100 LMC students, 15% are tattooed. What can we conclude?

a) State the hypotheses (use Ho and Ha notation)

b) What is the sample proportion that we will use to test the hypotheses?

c) Determine the P-value based on the simulation.



d) State a conclusion

7) What questions do you have at this point about hypothesis testing?

Module 20.2 Type I and Type II Errors

Learning Goal: Recognize Type I and Type II errors.

Our conclusion in a hypothesis test is based on probability. Because of this, there is a chance that our conclusion is wrong.

1) What can go wrong?

   a) Indicate which cells in the table are correct decisions.

Our actions

		We reject Ho (in favor of Ha)	We fail to reject Ho (not enough evidence in favor of Ha)
Unknown reality	Ho is true.		
	Ho is false. (Ha is true)		

   b) If Ho is true, but we reject it, this is Type I error.
   If Ho if false, but we fail to reject it, this is a Type II error.
   Label each type of error in the table.

2) In testing for bacterial counts on meat, we have the following hypotheses
   Ho: Meat is not spoiled.   Ha: Meat is spoiled.

   Describe Type I and Type II errors in this context.

3) In the early 1900s researchers used Zener cards to try to identify people with ESP. A person would try to identify the symbol on a hidden card.

   Someone who is just guessing has a 1 in 5 chance of a correct guess, which is 0.20. Imagine a person correctly identified 5 symbols out of 10 tries.

Image courtesy of http://en.wikipedia.org/wiki/Extrasensory_perception.

   Ho: person is guessing (p=0.20) (does not have ESP)
   Ha: person is not guessing (p>0.20) (may have ESP)

   With 10 tries, the distribution associated with guessing gives a P-value of 0.009 for this person.

a) What do we conclude based on the P-value?

b) Which type of error is possible here?

4) Thousands of women every year use home pregnancy tests, which are sold in drug stores and supermarkets. Home pregnancy tests give a result of positive (pregnant) or negative (not pregnant). A systematic review published in 1998 showed that home pregnancy test kits, when used by experienced technicians, are almost as accurate as professional laboratory testing (97.4%). When used by consumers, however, the accuracy fell to 75%: the review authors noted that many users misunderstood or failed to follow the instructions included in the kits. Improper usage may cause both false positives (Type I error) and false negatives (Type II error).

Describe both types of error relative to the hypotheses below. Use the terms "false positive" and "false negative."
Ho: woman is not pregnant
Ha: woman is pregnant

Module 20.3 P-Values and What They Mean

Learning Goal: Recognize the logic behind a hypothesis test and how it relates to the P-value.

Learning Objective: Interpret a P-value as a probability in the context of a statistical study.

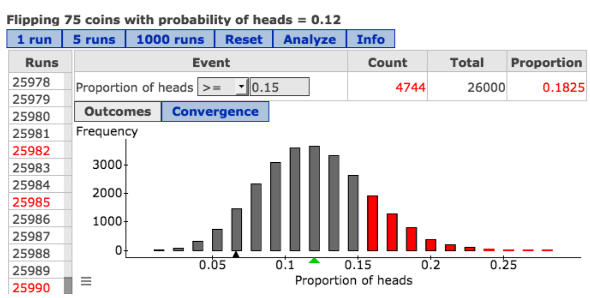
Introduction:
We now know how to use the P-value to draw a conclusion from a hypothesis test. In this short activity we will focus on the meaning of the P-value.

Let's return to a familiar example: A 2011 California Student Survey found that in any 30-day period, almost 12 percent of students in the 9th grade admit to using drugs at least once at school.
Source: http://www.huffingtonpost.com/marsha-rosenbaum/drug-education_b_3906983.html

To investigate this issue in your community, the school board hires a research firm to conduct a survey of a random sample of 9th graders in the district to determine if drug usage is greater than the statewide rate of 12%. Seventy-five students respond to the anonymous survey and 11 admit to drug usage at school during the last month. (11 out of 75 is about 15%)

The sample data is used to test the hypotheses:
Ho: p=0.12 (12% of the population of 9th graders have used drugs at school)
Ha: p>0.12 (More than 12% of the population of 9th graders have used drugs at school)

Here is the StatCrunch image from the simulation of selecting 26,000 random samples of 75 students.



The P-value is about 0.18.

What can we conclude?

In this activity we will focus on the **meaning of the P-value**, instead of focusing on how to use it to draw a conclusion.

What does P-value = 0.18 mean?

Here is one way to express it: *If the population proportion is 0.12, we expect to see sample proportions vary from this. But will sample proportions as large or larger than 0.15 occur very often? How often? What's the probability? The probability is 0.18.*

Here is another way to describe the P-value with a bit more context: *In random sampling from a population of 9th graders with 12% drug usage rate at school, we expect to see variability. With random samples of 75 students, there is an 18% chance that a sample will have 15% or more reported using drugs at school.*

Here is another way to describe the P-value. This one is very concise: *There is an 18% chance that a random sample of 75 ninth-graders will have 15% or more reported using drugs at school when we sample from a population of ninth-graders in which 12% are using drugs.*

Notice that each of these descriptions includes the following information:

- the null hypothesis
- the description of the samples included in the P-value
- the probability

The last two descriptions are better because they add context about

- the population
- the variable

1) Go back to the three interpretations of the P-value 0.18 and identify the null hypothesis, the description of the samples included in the P-value and the probability in each interpretation.

In general, a P-value is the probability that a sample statistic (such as $\hat{p}$) is as extreme or more extreme than the statistic from the observed sample, if the null hypothesis is true.

When you interpret a P-value, don't use this general interpretation. Instead describe it in a way that references the context of the problem, as shown in the examples above.

Group work:

2) "Authoritative numbers are hard to come by, but according to a 2002 confidential survey of 12,000 high school students, 74 percent admitted cheating on an examination at least once in the past year."
Source: http://abcnews.go.com/Primetime/story?id=132376&page=1

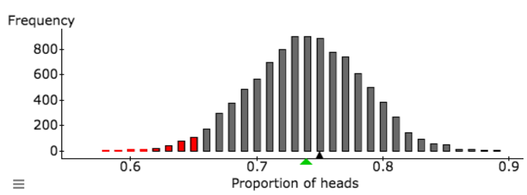
A Principal of a local high school hopes that cheating is less prevalent at her school. She hires a research firm to conduct a confidential survey of a random sample of 100 students at her school. Sixty-five percent of the 100 students admit to cheating on exams.

In a test of the hypotheses: Ho: p = 0.74, Ha: p < 0.74, the P-value is 0.027.

a) In the following interpretation of the P-value, identify the null hypothesis, the description of the samples included in the P-value, and the probability.

There is a 2.7% chance that a random sample of 100 students will have less than 65% admitting to cheating on exams if 74% of the student population is cheating on exams.

b) In the simulated distribution of sample proportions, circle the samples included in the P-value. Label the hypothesized population proportion and the P-value.



3) *Do you feel that the death penalty acts as a deterrent to the commitment of murder, that it lowers the murder rate, or not?* According to Gallup polls, 64% of adults in 2011 answered "no, does not".
Source: http://www.gallup.com/poll/1606/death-penalty.aspx

Suppose that in a random sample of adults this year, 68% answer "no, does not."

Has the percentage of the public with opinion increased since 2011?
We test the following hypotheses: Ho: p = 0.64, Ha: p > 0.64. The P-value is 0.20.

Which of the following interpretations of the P-value are accurate and complete? For those that are not accurate, explain why.

- The probability that more than 64% of adults will answer "no" this year is 0.20.

- The probability that more than 68% of adults will answer "no" this year is 0.20.

- If random samples are greater than 68%, then there is a 20% chance that the null hypothesis is true this year.

- There is a 20% chance that random samples will have more than 68% answering "no" if 64% of the population has this opinion.

- If the null hypothesis is true, then the probability that random samples will have $\hat{p}$ greater than 0.68 is 0.20.

4) In a Gallup poll about the cause of record highs in 2015 temperatures, 49% of adults answered "human-caused climate change," 46% chose "natural changes in the Earth's temperatures," and 5% had no opinion.

Suppose that in a random sample of adults this year, 55% attribute warmer weather patterns to "human-caused climate change."

Is the percentage of the public with this opinion higher this year than in 2015? We test the following hypotheses: Ho: p = 0.49, Ha: p > 0.49. The P-value is 0.03. Interpret the P-value as a probability statement.

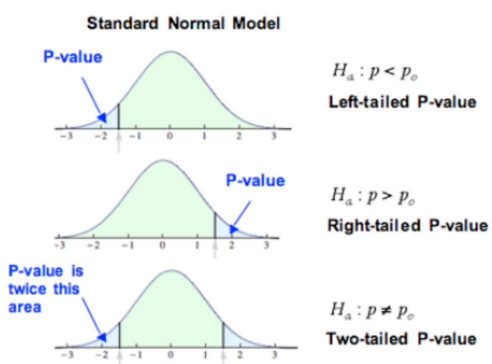
Module 21 Hypothesis Test for a Population Proportion

Module 21.1 Hypothesis Testing for a Population Proportion

Learning Goals:

- Recognize when a situation calls for testing a hypothesis about a population proportion.
- Conduct a hypothesis test for a population proportion. State a conclusion in context.

Introduction:
In this activity we continue our work with hypothesis testing. This first example is what we call a "two-sided" test. We use a two-sided test when our claim is that there is a change, as opposed to an increase or a decrease, in a population proportion. In this situation, the alternative hypothesis has a "≠" symbol, instead of a "greater than" or "less than" symbol.



Example: CCSSE is the Community College Survey of Student Engagement. CCSSE was launched in 2001, with the intention of producing new information about community college quality and performance. In the CCSSE student survey data is used to help community colleges set benchmarks to improve student learning and retention.

According to the 2016 CCSSE data from about 430,000 students nationwide, about half of the students reported that they "often" or "very often" prepared two or more drafts of a paper or assignment before turning it in.

Are the results similar at LMC? Specifically, let's test the claim that LMC is different. Use a 5% level of significance.

Step 1: Determine the hypotheses.

Ho: p = 0.50
Ha: p ≠ 0.50

Step 2: Collect the data and report the sample results.

Suppose that in a random sample of 200 LMC students, 56% report that they "often" or "very often" prepared two or more drafts of a paper or assignment before turning it in.

Step 3: Assess the data.

Are these results statistically significant? Or is the 6% difference between 56% and the hypothesized 50% just due to the variability we expect to see when we collect random samples? These questions can be answered with a P-value.

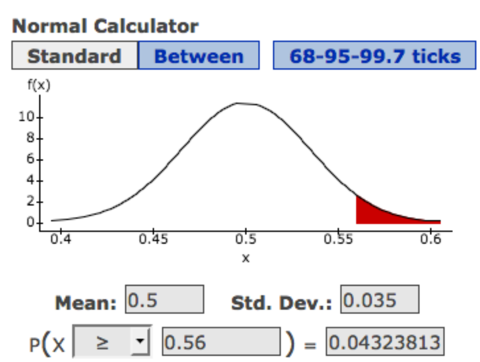
- Verify that the conditions are met for the use of a normal model for the distribution of sample proportions.

- What are the mean and standard deviation of the distribution of sample proportions in this situation?

- Are the results statistically significant?

  Use the StatCrunch Normal Calculator to find the P-value by doubling the P-value associated with a one-sided test.

  What is the P-value?



Step 4: State a conclusion

Group work

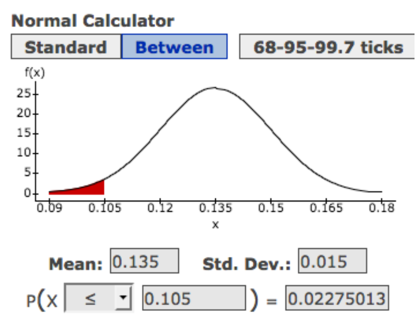
1) According to the 2016 CCSSE data from about 430,000 community college students nationwide, about 13.5% of students reported that they "often" or "very often" come to class without completing readings or assignments.

   Are the results similar at community colleges in California? Specifically, let's test the claim that California students are different. Use a 5% level of significance.

   Suppose that the CCSSE is given in California to a random sample of 500 students and 10.5% report that they "often" or "very often" come to class without completing readings or assignments.

   a) State the hypotheses and write a sentence to explain what p represents.

   b) Verify that the normal model is a good fit for the distribution of sample proportions.



   c) Verify that the mean and standard deviation of the distribution of sample proportions is as shown in the Normal Calculator. Explain or show the formula(s) with numbers plugged in.

   d) Use the image of the Normal Calculator to find the P-value.

   e) State your conclusion.

2) Do you think your conclusion in the previous problem would change if the 10.5% came from a random sample of 100 students instead of a random sample of 500 students? Why or why not?

3) A 2014 Fox News Poll reported that 20% of voters is tattooed. Has the percentage changed this year?

Suppose that in a survey of 1000 voters this year, 22.5% are tattooed.

We use StatCrunch to run the hypothesis test (Stat, Proportion Stats, One sample, With summary)

One sample proportion hypothesis test:
p : Proportion of successes
H_0 : p = 0.2
H_A : p ≠ 0.2

Hypothesis test results:

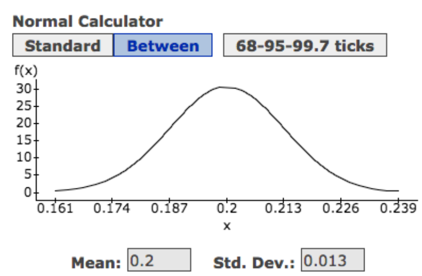
Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	225	1000	0.225	0.012649111	1.9764235	0.0481

a) Write a sentence to explain what p represents in the hypotheses.

b) What wording in the problem statement suggests that the alternative hypothesis should be “≠”?

c) The yardstick in the normal curve is standard error. Shade the total area under the normal curve that represents the P-value.

d) Does the StatCrunch printout give the P-value or do we have to double this number? How do you know?



e) What do we need to know about this sample in order to conduct a hypothesis test?

4) According to the 2016 CCSSE data from about 430,000 community college students nationwide, about 26% of students reported that they never made a class presentation.

   The Communications Department at a local community college is trying to encourage faculty in other departments to require class presentations. They think that class presentations are less prevalent at their college than at other colleges.

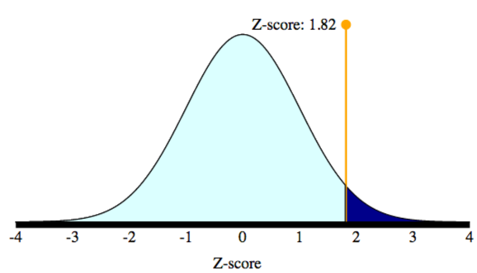
   To study this issue they survey a random sample of 100 students and find that 34% have never made a class presentation. This sample supports what the Communications Department thinks is true, but is this 8 percentage point difference significant? We have to conduct a hypothesis test to know.

   a) State the hypothesis. Write a sentence to describe what p represents.

   b) Verify that the normal model is a good way to estimate the P-value.

   c) Verify that the standard error is approximately 0.044. Show the formula with correct numbers plugged in. What does this number represent?

   d) This time we used the OLI Normal Distribution Calculator, which is based on z-scores. Verify that the z-score for the sample is about 1.82. Show the formula with correct numbers plugged in.



The area to the left of the Z value is: 0.9656

The area to the right of the Z value is: 0.0344

   e) What is the P-value?

   f) What conclusion can we draw?

Module 21.2 Cautionary Notes about Drawing Conclusions from a Hypothesis Test

Learning Goals:

- Distinguish statistical significance from practical importance.
- From a description of a study, evaluate whether the conclusion of a hypothesis test is reasonable.

Previously, we discussed Type I and Type II errors, which arise naturally when we are randomly sampling and setting significance levels. There is always the possibility that a Type I or Type II error has occurred and statisticians can estimate the probability of each type of error.

But there are other more serious problems with hypothesis testing that are also more prevalent. In this activity we will examine a variety of real studies that highlight potential problems with statistical studies.

1) The following is an excerpt from *Statistics* by Freedman, Pisani, and Purves.

   In some cases of cirrhosis of the liver, the patient may start to hemorrhage and bleed to death. One treatment involves surgery to redirect the flow of blood through a portacaval shunt. The operation to create the shunt is long and hazardous. Do the benefits outweigh the risks?

   In 51 studies of the portacaval shunt, 34 (66%) reported results that were characterized as "strongly enthusiastic" about the shunt. Only 6 studies (12%) were characterized as "no enthusiasm" for the shunt.

   Here is an analysis of these studies as published in *Gastroenterology*

	Degree of enthusiasm for the shunt by researchers		
	Strong	Moderate	None
No controls	24	7	1
Controls, but not randomized	10	3	2
Randomized controlled	0	1	3

   a) What is the purpose of a control group in an experiment? Give an example of a control group that could be used in a portacaval shunt experiment.

   b) What is the purpose of randomization in an experiment?

   c) Based on these studies, do you think that doctors should use portacaval shunts? Why or why not?

2) Researchers performed a randomized controlled double blind experiment to investigate the use of cholestyramine in reducing blood cholesterol levels and thus reducing heart attacks. There were 3,806 subjects, who were all middle-aged men at high risk of heart attack; 1906 were chosen at random for the treatment group (cholestyramine) and the remaining 1900 were assigned to a control group (placebo).
Source: "The Lipid Research Clinics Primary Prevention Trial Results,"Journal of the American Medical Associations vol.251 (1984) pp 351-64.

The subjects were followed for 7 years. The drug reduced cholesterol levels by about 8% . Futhermore, there was a lower incidence of heart attacks in the treatment group (155 or 8.1%) vs. the control group (187 or 9.8%).

This 1.7% difference was statistically significant with a P-value of about 0.035 for a one-tailed test. The article was published in the *Journal of American Medical Association,* whose editors are quite strict about using significance levels of 0.05.

a) Would the experiment's results have been significant if a two-tailed test was conducted? How do you know?

b) The researchers said this was "strong evidence" that cholestryamine reduces heart attacks by reducing cholesterol levels. Why might a skeptical reader think that the researchers are overstating their results?

c) Cholestryamine can have the following side effects: constipation, stomach/abdominal pain, gas, nausea, and vomiting. The cost per year is approximately $1,350. If you were at high risk of a heart attack and also had high cholesterol, would you take cholestryamine based on what you know at this point? Why or why not?

3) This is a true story … In a meeting at a community college on the impact of tutoring, a researcher said that tutoring had a statistically significant impact on improving GPA. His sample size was over 2,000 students. A math professor in the audience asked how much GPAs improved. The response: 0.05 GPA points.

   a) How can such a small increase in GPA be statistically significant?

   b) Should the college increase tutoring funds, at the expense of funding other programs, based on these results? Why or why not?

4) Read through these summaries (footnotes on the next page)

- A 1996 study on the effects of nicotine on cognitive performance revealed that findings of nicotine or smoking improving performance were more likely to be published by scientists who acknowledged tobacco industry support.[1]
- A 2003 study of published research on antidepressants found that studies sponsored by manufacturers of selective serotonin reuptake inhibitors (SSRI) and newer antidepressants tended to favor their products over alternatives when compared to non-industry-funded studies. Also, modelling studies funded by industry were more favorable to industry than studies funded by non-industry sponsors.[2] In general, studies funded by drug companies are four times more likely to favor the drug under trial than studies funded by other sponsors. [3]
- A 2006 review of experimental studies examining the health effects of cell phone use found that studies funded exclusively by industry were least likely to report a statistically significant result.[4]
- The US Food and Drug Administration (FDA) determined in 2008 that the bisphenol A (BPA) in plastic containers is safe when leeched into food, citing chemical industry studies. Independent research studies reached different conclusions[5] with over 90 percent of them finding health effects from low doses of BPA.[6]
- Two opposing commercial sponsors can be at odds with the published findings of research they sponsor. A 2008 Duke University study on rats, funded by the Sugar Association, found adverse effects of consuming the artificial sweetener Splenda. The manufacturer, Johnson & Johnson subsidiary McNeil Nutritionals LLC, responded by sponsoring its own team of experts to refute the study.[7]
- A 2012 analysis of outcomes of studies pertaining to drugs and medical devices revealed that manufacturing company sponsorship "leads to more favorable results and conclusions than sponsorship by other sources."[8]

   a) What is the point being made by these reviews of statistical studies?

b) Your instructor will assign your group to read one or more sections of Hilda Bastin's article, *'They would say that, wouldn't they?' A reader's guide to author and sponsor biases in clinical research.*

*Source: Hilda Bastian (December 2006). "'They would say that, wouldn't they?' A reader's guide to author and sponsor biases in clinical research". J R Soc Med. **99** (12): 611–614. doi:10.1258/jrsm.99.12.611. PMC 1676333. PMID 17139062.*

Prepare a summary of your section for class discussion.

1. *Christina Turner; George J Spilich (1997). "Research into smoking or nicotine and human cognitive performance: does the source of funding make a difference?". Addiction. **92** (11): 1423–1426. doi:10.1111/j.1360-0443.1997.tb02863.x. PMID 9519485.*
2. *C. Bruce Baker; Michael T. Johnsrud; M. Lynn Crismon; Robert A. Rosenheck; Scott W. Woods (2003). "Quantitative analysis of sponsorship bias in economic studies of antidepressants". The British Journal of Psychiatry. **183** (6): 498–506. doi:10.1192/bjp.183.6.498.PMID 14645020.*
3. *Becker-Brüser W (2010). "Research in the pharmaceutical industry cannot be objective". Z Evid Fortbild Qual Gesundhwes. **104** (3): 183–9. PMID 20608245.*
4. *Anke Huss; Matthias Egger; Kerstin Hug; Karin Huwiler-Müntener; Martin Röösli (2006-09-15). "Source of Funding and Results of Studies of Health Effects of Mobile Phone Use: Systematic Review of Experimental Studies". Environmental Health Perspectives. **115** (1): 1–4.doi:10.1289/ehp.9149. PMC 1797826. PMID 17366811.*
5. *vom Saal FS, Myers JP (2008). "Bisphenol A and Risk of Metabolic Disorders". JAMA. **300** (11): 1353–5. doi:10.1001/jama.300.11.1353. PMID 18799451.*
6. *David Michaels (2008-07-15). "It's Not the Answers That Are Biased, It's the Questions". The Washington Post.*
7. *Stephen Daniells (2009-09-25). "Splenda study: Industry and academia respond". Foodnavigator.com.*
8. *Lundh, A; Sismondo, S; Lexchin, J; Busuioc, OA; Bero, L (Dec 12, 2012). "Industry sponsorship and research outcome.". The Cochrane database of systematic reviews. **12**: MR000033. doi:10.1002/14651858.mr000033.pub2. PMID 23235689.*

JOURNAL OF THE ROYAL SOCIETY OF MEDICINE Volume 99 December 2006

'They would say that, wouldn't they?' A reader's guide to author and sponsor biases in clinical research

Hilda Bastian

J R Soc Med 2006;**99**:611–614

'This study was funded by (Company A). Professor XYZ has received honoraria and travel support for lectures and advisory boards, as well as research grants, from (Company A) and (Company B).'

Commercial sponsorship of clinical research, especially for drugs, is ubiquitous. One of the solutions to some of the dilemmas arising from this is full disclosure of authors' financial interests and relationships with sponsors. But is disclosure enough, and what should we as readers make of the fine print at the end of journal articles? Is sponsorship bias the only bias we should watch out for?

The arguments for and against are familiar. In one corner, people accuse researchers of having sold out; no one should believe anything that Professor XYZ now says. From the other corner, people energetically defend academics' integrity and argue that we can pretty much trust them no matter where their money comes from. And anyway, the journal can be trusted not to let anything too dubious get through, can't it?

For readers, wherever our views might stand along the spectrum between these polarized positions, questions remain. Should we just gloss over the disclosure of interest fine print and somehow be reassured just because it is there, or should we be scrutinizing it? Is there a point at which this information should signal the warning *lector emptor* (let the reader beware), and if so, can we tell?

Clinical research does not just serve clinical interests. It also plays a major economic role. And research provides the building blocks of academic careers. That means that, inevitably, authors have interests of their own, although the weight of these no doubt varies greatly among individuals. Those interests do not always conflict with the interests of patients and clinicians, but sometimes they might. Journals have, to a greater or lesser degree, shifted the responsibility to the reader: only a minority of journals might actually decide against publication purely on conflict of interest grounds.[1,2] Most journals and all readers will not know if the amounts of declared financial interest involved are trivial or major. It is an honour system, and it is in the disclosure to readers that journals see their ethical obligations largely filled. There is no routine or random investigation, and generally no policy on sanctions—picture major international sport before random doping tests and the barring of offending players from future competition.

IQWiG – The German Institute for Quality and Efficiency in Health Care, Cologne, Germany

E-mail: hilda.bastian@iqwig.de

FINANCIAL BIAS

Fortunately research bias itself is the subject of research. Articles on this subject, too, come from polarized positions, but over time, some fairly clear messages have emerged. The question most studied here is whether or not drug companies have an influence on the research they sponsor on their products. One-third of trials are sponsored by drug companies.[3] Somewhere between one-third to three-quarters of the trials in the five biggest general medical journals are industry-financed (*Annals of Internal Medicine, BMJ, JAMA, Lancet* and *New England Journal of Medicine*).[4] But does it influence the results?

The most recent review of studies that compared trials sponsored by drug companies with those that were not concluded that sponsored trials were more likely to say that the sponsor's product was effective.[5] Recent studies of systematic reviews came to the same conclusion.[6,7] For trials, this was not because of problems within the conduct of the research. If anything, the quality of these trials was better than that of non-sponsored trials. Simply not publishing studies that fail to show a benefit might be a factor here, so-called publication bias.[8] In the end, though, it has been suggested that sponsors are by and large getting the results they want, not by suppressing unpalatable studies or 'fiddling' the results, 'but rather by asking the 'right' questions',[9] for example, by comparing the drug to a smaller dose of a competitor's drug. Biased reporting of outcomes has also been shown to affect the outcomes of trials.[10] In many fully legitimate ways, trials can be done in the most ideal circumstances, giving a treatment a chance to perform at its personal best.

All of which suggests the best strategy for readers is to maintain some healthy scepticism, especially with early positive results. A promising treatment is often in fact merely the larval stage of a disappointing one. At least a third of influential trials suggesting benefit may either

JOURNAL OF THE ROYAL SOCIETY OF MEDICINE Volume 99 December 2006

ultimately be contradicted or turn out to have exaggerated effectiveness.[11]

Critical readers need to be conscious of two other aspects that are characteristic of industry-conducted or sponsored trials, and which could be influencing them. The first is the extent to which the overall clinical research agenda is dominated by research into drugs because of the availability of this sponsorship. And the second is the full impact of marketing, which begins with the article itself. The proportion of trials where industry employees are named co-authors appears to be increasing.[3]

Less obvious may be the influence of professional medical writers paid for by industry funds. This might at times be exerting an influence on framing and even content. It might often lead to an improvement in the quality of writing. A professionally written article might quite simply be a more compelling one, and that effect may be boosted by post-publication marketing. Articles that are discussed in the general media may end up being more influential in the scientific literature, too.[12] At the time of writing, out of the five big general medical journals, only the *BMJ* and *Lancet* explicitly asked authors questions about the involvement of professional writers who may or may not be listed as co-authors. *JAMA* recently updated its policy,[13] but did not add this question.

FINANCIAL ENTANGLEMENTS

Research is not as helpful in relation to the other parts of the financial declaration made by our fictional Professor XYZ at the head of this article. Nor is there a consensus of opinion of what these financial connections mean. There is a truism that says most people avoid biting the hand that feeds them—and in this case, it is a hand that pours champagne, too. That suggests that it is self-censorship and a relationship of dependence and privileges that might be important. Social interaction and the collegiality that emerges from working together might blunt the willingness to be too critical. But it is also argued that as long as people are taking money from multiple commercial sources, then they have hedged their bets and are consequently more independent. In the end, on this matter, as readers we are left to our own judgment. There is no evidence to suggest this strategy protects researchers (or anyone else) from bias.

While the requirements for adequate author independence in relation to data are fairly similar between major journals, the same cannot be said about other aspects of authors' financial interests. Table 1 shows the types of interests the five major medical journals specifically ask of their authors (although all ask general questions). There are few issues in common. That might help explain why these declarations vary so much between journals, with some fairly short and others often quite extensive. Journals vary in other matters too: for most, the information about conflict of interest is buried in fine print at the end of the article, but some have it prominently at the front.

University standards do not always provide much of a safeguard either. A study of academic medical centres in the US found some major differences in what was deemed acceptable in contracts between industry and clinical trialists.[14] For example, 50% of the universities allowed sponsors to draft the manuscript, and almost 15% found it acceptable to allow sponsors the authority to revise manuscripts or decide whether they could be published.

Table 1 **Specific questions asked by the major five medical journals on authors' financial interests in July 2006**

	Ann Int Med	*BMJ*	*JAMA*	*Lancet*	*NEJM*
Fee from other than the journal for writing or being named on the paper				•	
Paid professional writer		•		•	
Employment	•	•	•	•	
Consultancy	•	•	•	•	•
Holding stocks, shares or equity	•	•	•	•	•
Patents/patent applications			•	•	•
Company board position				•	
Expert testimony	•	•	•	•	
Funds for grants and/or research and/or staff and/or equipment	•	•	•	•	•
Travel grants		•		•	
Speaking fee		•			•
Honoraria	•			•	
Royalties	•		•		

Ann Int Med: past five years and pending (including grants); *BMJ*: last five years; *JAMA*: last five years and foreseeable future; *Lancet*: within three years of submitted work, presently or applying for; *NEJM*: time not specified

JOURNAL OF THE ROYAL SOCIETY OF MEDICINE Volume 99 December 2006

Interesting for readers, too, would be the medical journals' policies in relation to their own editorial staff, but their internal practices are not as readily available as their requirements of authors. The *Lancet* has published its quite strict standards. Editors at the *Lancet* may not accept offers for travel, accommodation, hospitality or gifts.[15] Such policies exist at other journals, too, but the policies are not generally publicly available. Conflict of interest statements for editorial staff of the *BMJ* are published online.[16]

What about research participants? A study of potential clinical research participants in the US found that people with chronic illness believed they should be told about financial entanglements of researchers before they are asked to participate in a trial.[17] Most would still participate, but depending on the specific scenario, from 2% to 32% would not participate if they knew of financial benefits to the researchers. The scenarios that attracted the highest frequency of concern were those involving personal income for the researchers.

What do researchers and readers think? If the debates on this are anything to go by, many researchers believe their entanglements with industry have no influence, and others believe those researchers are deluding themselves.[18] Recent editorials in *The Wall Street Journal*[19] and *New York Times*[20] suggest the medical journals have a real credibility problem in the wider public sphere on this issue.

Readers are likely to also have divergent views. We definitely have unequal knowledge. We cannot really expect to understand the extent of an academic's dependence on industry from the way the data are generally presented to us. Insiders can often tell you that an academic is now wholly dependent on large sums of industry money, but the average reader cannot know if financial entanglements are in the trivial or mind-boggling range. It would be helpful for many of us as readers if there was a way of classifying this. Establishing a random audit system is worth considering, along with negative consequences for the future if a researcher transgresses. But perhaps most urgently of all, readers might be helped by a better *lector emptor*: serious conflicts of interest should be up at the level of the abstract—ideally within the abstract—so that few readers will miss it.

SELF-PROMOTION BIAS

Inevitably, researchers and many other knowledge workers potentially have an interest in promoting themselves, or at least in peddling their views, theories and hobby horses. Research grants and publications are both major sources of life-blood and currency for academic careers. Publications, particularly in the more important journals like the one you are reading now, carry influence because of their impact on you, the reader. But they are also necessary to amass credit for the progress of academic careers. The article being cited in other articles in turn accrues the academic version of frequent flyer points in the Science Citation Index.

Authors will often be building on their own body of work, which means they will inevitably be at least a bit self-referential. But one study has found some evidence that a high proportion of self-citation within an article might indicate a lower quality of article.[20] Another has found no association between self-citation in the diabetes literature with methodological rigour.[21] Still, 'Even if I say so myself' is not a very high level of evidence. Of course a certain amount of self-citation is legitimate and essential. On the other hand, self-citation also 'perpetuates one's interpretations or opinions' and it might 'falsely validate the conclusions of an author or group'.[21] Self-citation has been calculated as an average of 20% of all citations in one study,[21] and 36% in another.[20]

'EUREKA!' BIAS

Archimedes' famous cry of discovery and intellectual excitement brings us to a key bias that is possibly implicated

Key questions

From journal to journal, the answers to questions about financial entanglement may be sprinkled through different parts of a research article, including: authorship affiliations and contributorship, the methods section, declaration of interests section, acknowledgements and funders' section. Check all the fine print if you want to assess potential for author and sponsor bias. You need to see the full article—the abstract will not include such information.

- How much independence from the funders did the researchers have, and did they control the data, its interpretation and the publication?
- Is there a systematic review on the topic the research is addressing, and if so, how do the results of this study fit into the other evidence? If there is no systematic review, are there trials?
- What do accompanying editorials, letters and analyses in secondary evaluation publications (such as *ACP Journal Club* and *Evidence Based Medicine*) say?
- Is the range of opinion on the issue shown? Do they tell you what the other schools of thought on this issue are?
- Who says so? Is there a high proportion of self-citation? Are there many unsupported claims?
- If you disagree with the author, what evidence do you have to support your position?

in all author and sponsor biases—and it is one that readers can share, too. We can all believe too quickly that which we most earnestly hope to be true. One of the key reasons blinding is such a critical part of trial methodology is because we know that bias is human, and blinding helps protect us from investigator bias. While this kind of bias does not always mean that the benefits of a treatment end up being exaggerated, sometimes it does.[22,23] While some people are nihilistic, most of us want to find something that *helps*. The research suggests, though, that when we see a claim of benefit, the strength of the claim is weakened when the research has been underwritten by the product's manufacturer.

It will not help, perhaps, if we lose all optimism and hope for progress, but healthy scepticism is essential too. In the end, it is not surprising when a manufacturer publishes a study that says its product is better than another, or someone says their own pet theory is the best idea since sliced bread: well, they would say that, wouldn't they?

Competing interests HB accepts no income other than her salary from an independent, non-profit, evidence-based medicine research institute established under national statute. She accepts no travel or hospitality support from industry, either directly or indirectly. She is a member of advisory committees to two medical journals, the *BMJ* and *Controlled Trials*.

REFERENCES

1 Ancker J, Flanagin A. A comparison of conflict of interest policies at peer-reviewed journals in multiple scientific disciplines. Paper presented at *The Fifth International Congress on Peer Review and Biomedical Publication*; Chicago, 17 September 2005. Available at http://www.ama-assn.org/public/peer/abstracts.html#LinkTarget_6571 (accessed 23 July 2006)

2 Cooper RJ, Gupta M, Wilkes MS, Hoffman JR. Conflict of Interest Disclosure Policies and Practices of Peer-Reviewed Biomedical Journals. Paper presented at *The Fifth International Congress on Peer Review and Biomedical Publication*; Chicago, 17 September 2005. Available at http://www.ama-assn.org/public/peer/abstracts. html#LinkTarget_6571 (accessed 23 July 2006)

3 Buchkowsky SS, Jewesson PJ. Industry sponsorship and authorship of clinical trials over 20 years. *Ann Pharmacother* 2004;**38**:579–85

4 Egger M, Bartlett C, Juni P. Are randomised controlled trials in the BMJ different? *BMJ* 2001;**326**:1167–70

5 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;**326**:1167–77

6 Jørgensen AW, Hilden J, Gøtzsche PC. Cochrane reviews compared with industry-sponsored meta-analyses and other meta-analyses of the same drugs. *BMJ* 2006;**333**:782

7 Yank V, Rennie D, Bero LA. Are authors' financial ties with pharmaceutical companies associated with positive results or conclusions in meta-analyses on antihypertensive medications? Paper presented at The *Fifth International Congress on Peer Review and Biomedical Publication*; Chicago, 17 September 2005. Available at http://www.ama-assn.org/public/peer/abstracts.html#LinkTarget_6571 (accessed 27 February 2006)

8 Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997;**9**(Suppl):15–21

9 Smith R. Medical journals are an extension of the marketing arm of pharmaceutical companies. *PLoS Medicine* 2005;**2**:364–6

10 Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;**291**:2457–65

11 Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005;**294**:218–28

12 Phillips DP, Kanter EJ, Bednarczyk B, Tastad PL. Importance of the lay press in the transmission of medical knowledge to the scientific community. *N Engl J Med* 1991;**325**:1180–3

13 Flanagin A, Fontanarosa PF, DeAngelis CD. Update on *JAMA*'s conflict of interest policy. *JAMA* 2006;**296**:220–1

14 Mello MM, Clarridge BR, Studdert DM. Academic medical centers' standards for clinical-trial agreements with industry. *N Engl J Med* 2005;**352**:2202–10

15 James A, Horton R. *The Lancet*'s policy on conflicts of interest. *Lancet* 2003;**361**:8–9

16 BMJ Journals. Declaration of competing interests (editorial). Available at http://bmj.bmjjournals.com/aboutsite/comp_editorial.shtml (accessed 27 February 2006)

17 Kim SYH, Millard RW, Nisbet P, Cox C, Caine ED. Potential research participants' views regarding researcher and institutional financial conflicts of interest. *J Med Ethics* 2004;**30**:73–79

18 BMJ. Time to untangle doctors from drug companies. *BMJ Theme Issue* 2003;**326**:Issue 7400

19 Armstrong D. Medical reviews face criticism over lapses. *Wall Street Journal* 19 July 2006

20 (Anonymous editorial). Our conflicted medical journals. *New York Times* 23 July 2006

21 Aksnes DW. A macro study of self-citation. *Scientometrics* 2003;**56**:235–46

22 Gami AS, Montori VM, Wilczynski NL, Haynes RB. Author self-citation in the diabetes literature. *CMAJ* 2004;**170**:1925–27

23 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995:**273**:2546–7

24 Balk EM, Bonis PA, Moskowitz H, *et al*. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;**287**:2973–82

Module 21.3 Unit 8 Lab

For the Unit 8 lab you will use a random sample of 100 students from a StatCrunchU. StatCrunchU is a fictitious virtual population of 46,000 students.

Instructions for accessing your data: To access StatCrunchU, log into StatCrunch and click on **Resources**. Scroll down to the heading **Take a sample from StatCrunchU** and click on **StatCrunchU**.

You will see a student survey with 6 questions. This is a fictitious survey that was answered by each of the 46,000 fictitious students at StatCrunch U.

Below the survey you can set the sample size. Set this to 100 and click Survey. A spreadsheet will appear with the survey results for your random sample of 100 StatCrunchU students.

Note: Your sample is a random sample; therefore, your results will differ somewhat from other students' results.

Instructions for the lab assignment:

1) How many females are in your sample? How many males? What proportion of your sample is female? What proportion of your sample is male? (StatCrunch steps: Stat, Tables, Frequency)

2) Is there an equal proportion of men and women at StatCrunchU?

   a) Show that the conditions are met for the use of a normal model for a hypothesis test ($np \geq 10$ and $n(1-p) \geq 10$, where p is from the null hypothesis).

   b) Use StatCrunch to test the claim that the proportion of females at StatCrunchU is equal to the proportion of men. Paste the StatCrunch printout below. (StatCrunch steps: Stat, Proportion Stats, One Sample, With Data. To copy click on Options.)

   c) Write a conclusion to your hypothesis test referring to females at StatCrunchU.

   d) Explain what the P-value means as a probability that refers to random samples of 100 StatCrunchU students.

3) What are the proportions of females and males at StatCrunchU?

   a) Show that the conditions are met for the use of a normal model for a confidence interval (count of successes and failures are greater than 10).

   b) Determine a range of plausible values for this proportion by using StatCrunch to find a 95% confidence interval. Paste the StatCrunch printout below. (StatCrunch steps: Stat, Proportion Stats, One Sample, With Data.)

   c) Interpret your interval referring to females at StatCrunchU.

   d) Explain what is meant by "95% confident."

4) Does your confidence interval support your hypothesis test? Explain.

Module 21.4 Unit 8 Project

Instructions:

Your group will be assigned one of the research questions below. Each group will prepare a poster. Your instructor may also require each group member to write an analysis and submit it individually.

This project requires you to learn some new StatCrunch moves. See instructions at the end of this assignment.

Your poster will include the following:

- Names of group members
- Your sample proportion with a pie chart or bar chart
- StatCrunch output (include ALL of the information from the printout)
- Explanation of all of the numbers in the StatCrunch output
- Sketch a normal curve using p as stated in the cited study with SE from the StatCrunch output
- For confidence intervals:
    - A sketch of the confidence interval appropriately located below the normal curve.
    - An answer to the research question that interprets the confidence interval in terms of the population of StatCrunchU students
    - An explanation of the meaning of "95% confident"
- For hypothesis tests:
    - P-value shaded in the normal curve
    - An answer to the research question that draws a conclusion from the hypothesis test in terms of the population of StatCrunchU students
    - An interpretation of the P-value as a probability statement about random samples of 1000 StatCrunchU students

StatCrunch instructions are at the end of this assignment.

Option 1: Hypothesis test
According to the CSU Mentor, about 87% of CSU East Bay undergraduates are full-time students. "Full-time" is defined as taking 12 or more units in a semester. Is the proportion of StatCrunchU students who are full-time less than 87%?
Source:
http://www.csumentor.edu/campustour/undergraduate/9/csu_east_bay/csu_east_bay5.html

Option 2: Confidence interval
What proportion of StatCrunchU students are full-time students? "Full-time" is defined as taking 12 or more units in a semester.

Option 3: Hypothesis test
According to a 2015 report from Georgetown University's Center on Education and the Workforce (CEW), more than 70% of college students work.
Source: https://cew.georgetown.edu/wp-content/uploads/Press-release-WorkingLearners__FINAL.pdf

Is the proportion of StatCrunchU students who work different from the 2015 CEW national estimate?

Option 4: Confidence interval
What proportion of StatCrunchU students work?

Option 5: Hypothesis test
According to the Institute for College Access and Success, in 2015 68% of college graduates have student loan debt.
Source: http://ticas.org/posd/map-state-data

Is the proportion of StatCrunchU students with loans different from the 2015 College Access and Success estimate?

Option 6: Confidence interval
What proportion of StatCrunchU students has student loans?

Option 7: Hypothesis test
According to the U.S. Department of Education's National Center for Education Statistics, 41% of graduating seniors had credit card debt in the year 2000.
Source: http://www.pirg.org/highered/BurdenofBorrowing.pdf

Is the proportion of StatCrunchU students with credit card debt greater than the U.S. Department of Education's estimate for the year 2000?

Option 8: Confidence interval
What proportion of StatCrunchU students has credit card debt?

Instructions for accessing your data:
To access StatCrunchU, log into StatCrunch and click on **Resources**. Scroll down to the heading **Take a sample from StatCrunchU** and click on **StatCrunchU**.

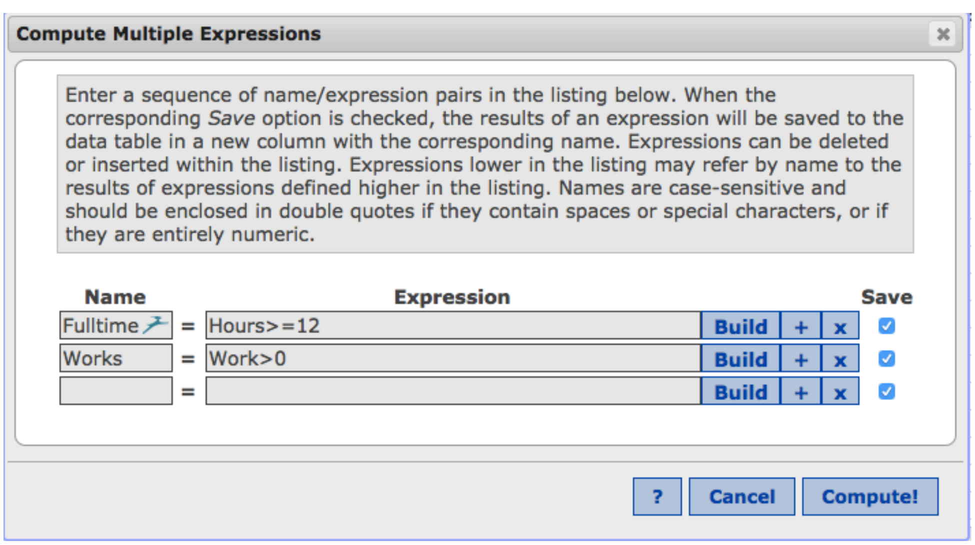
You will see a student survey with 6 questions. This is a fictitious survey that was answered by each of the 46,000 fictitious students at StatCrunch U.

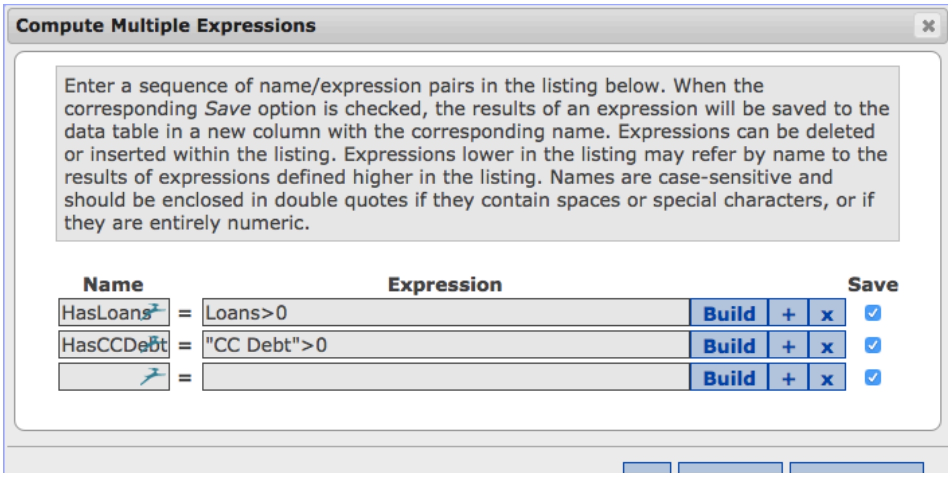
Below the survey you can set the sample size. Set this to 1000 and click Survey. A spreadsheet will appear with the survey results for your random sample of 1000 StatCrunchU students.

Instructions for creating a new variable:
All of the research questions in this lab involve proportions. Proportions summarize categorical data. However, most of the data in the spreadsheet is quantitative. For example, credit hours, work, loan debt, and credit card debt are all quantitative variables. Therefore, for all of the research questions in this project, you will need to convert the quantitative data into categorical data.

We will create additional columns in the spreadsheet to accommodate the necessary categorical variables for our analyses. An easy way to create one (or several) categorical variables, each based on values in a quantitative variable, is to use the Compute Multiple Expressions feature. The next several screenshots will walk you through the procedure.

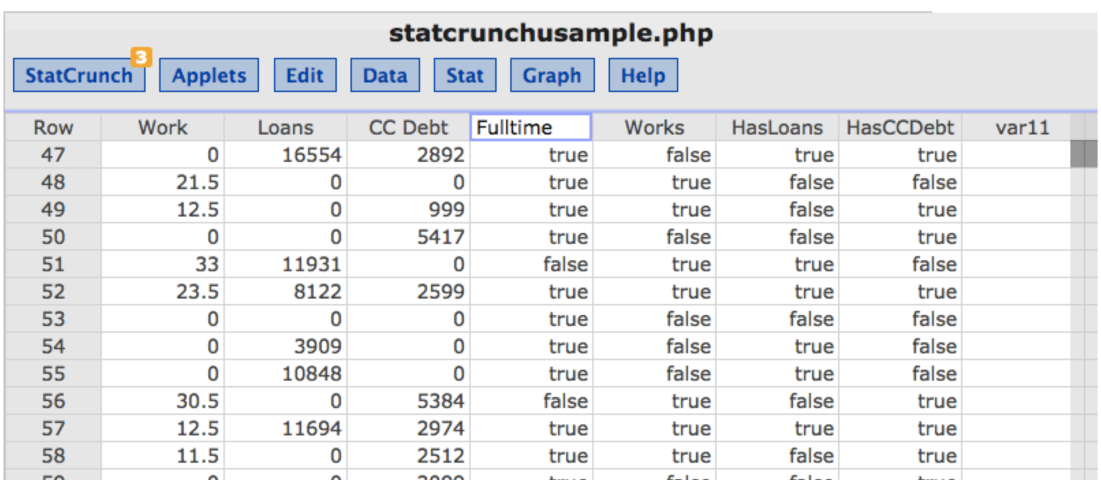
Choose `Data >> Compute >> Multiple Expressions`
The following two images show how to fill in the Multiple Expressions form for all of the 8 options in the Project. The first image shows how to create the new categorical columns for Options 1 – 4 (fulltime students and students who work), and the second image shows how to create the categorical columns for Options 5 – 8 (students with loans and students with credit card debt).





Here's what is going on in these two images from StatCrunch. Under the Expression headings, StatCrunch is creating a new temporary logical variable with values true or false according to whether the value in the corresponding spreadsheet column is positive. It then assigns the logical value (true or false) to the variable given under the Name heading. For example, in the first row of the form directly above, each row in new spreadsheet column HasLoans is assigned the value **true** if the value in Loans is positive. Otherwise HasLoans is assigned the value **false**. (note: the quantitative variable CC Debt is placed inside double quotes because it contains an embedded blank)

The resulting spreadsheet from StatCrunchU, complete with the 4 new Categorical variables, appears below. The first several columns have been scrolled off of the screen to make room for all four of the new categorical variables. Remember: the data in your sample of 1000 students will differ from that shown below.



Row	Work	Loans	CC Debt	Fulltime	Works	HasLoans	HasCCDebt	var11
47	0	16554	2892	true	false	true	true	
48	21.5	0	0	true	true	false	false	
49	12.5	0	999	true	true	false	true	
50	0	0	5417	true	false	false	true	
51	33	11931	0	false	true	true	false	
52	23.5	8122	2599	true	true	true	true	
53	0	0	0	true	false	false	false	
54	0	3909	0	true	false	true	false	
55	0	10848	0	true	false	true	false	
56	30.5	0	5384	false	true	false	true	
57	12.5	11694	2974	true	true	true	true	
58	11.5	0	2512	true	true	false	true	

Instructions for finding a confidence interval or conducting the hypothesis test

- Choose Stat, Proportion Stats, One sample, With data
- In the pop-up window titled One Sample Proportion enter the following:
    - **Values in:** choose the variable you created
    - **Success:** type the variable value, such as Job or Debt.
    - **Where:** leave this blank
    - **Group by:** leave this blank
    - **Perform:** click either Hypothesis test OR Confidence interval
    - For the hypothesis test, enter the hypothesized p and the correct Ha.
- Hit Compute!

UNIT 9

Inference for Means

Contents

Module 22 Distribution of Sample Means

Module 22.1 Introduction to the Distribution of Sample Means

Learning Goals:

- Describe the sampling distribution of sample means.
- Describe the effect of sample size on the variability of sample means.

Introduction:

Up to this point we have tested hypotheses and found confidence intervals in order to draw conclusions about a population proportion. Proportions summarize categorical data. We will now focus on quantitative data. We summarize quantitative data with means.

In this Unit we will use quantitative data to test hypotheses and find confidence intervals in order to draw conclusions about a population mean or a difference in population means.

1) Which of the following research questions involves quantitative data?

   a) What proportion of community college students has a student loan?
   b) What is the average loan amount owed by community college students?
   c) On average do college students borrow more than $5,000 before they graduate?
   d) Do more than 80% of college graduates have credit card debt?
   e) Is the average interest rate on a student loan greater than 8%?
   f) Do the majority of community college students work on average 20 hours or more each week?

Example:

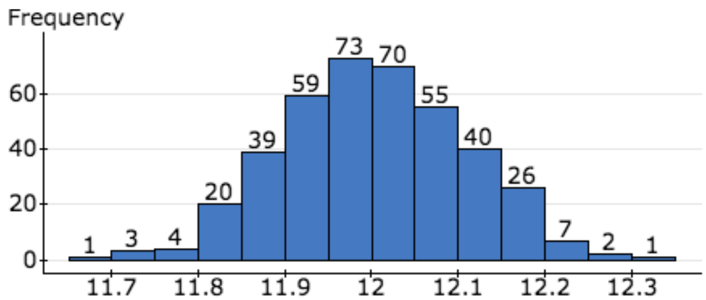
In a quality control process at a soda bottling plant, 400 bottles, filled during a 10-minute period, are pulled from the assembly line. These bottles will represent our population of soda bottles. The bottles are labeled with a volume of 12 fluid ounces, but there is variability in the amount each bottle actually contains.

2) The histogram shows the distribution of the volumes for the population of 400 bottles. The mean is 12fl.oz. and the standard deviation is 0.1fl.oz.

   a) Select a random soda from the list of 400 sodas.

b) Draw an arrow to show the bin that your soda lies in. Is the volume in your soda unusual?

DISTRIBUTION OF INDIVIDUAL SODAS
Mean = 12, SD = 0.1



c) Is it unusual for a bottle of soda to contain 11.9fl.oz.? What makes you think so?

d) What is the probability that a bottle of soda contains less than 11.9fl.oz.?

e) Do you think that the quality control engineer will be happy with this distribution? Why or why not?

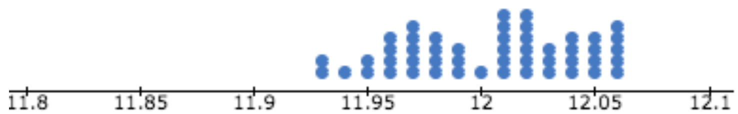
3) Now, let's think about a 6-pack of soda. If a 6-pack is randomly selected from this population of 400 sodas, is it unusual for the mean amount of soda to be 11.9floz in the 6-pack?

To answer this question, we need to examine the variability in the distribution of sample means.

a) Collect your own random sample of 6 sodas and determine the mean volume for your 6-pack.

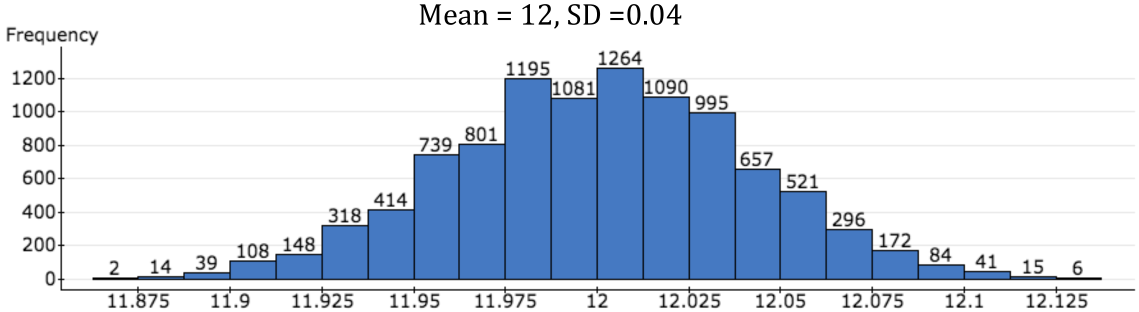
b) Here is a dot plot of the mean volumes of 50 randomly selected 6-packs of soda. Plot your mean volume in the dot plot.

DISTRIBUTION OF SAMPLE MEANS



c) Here is a histogram of the mean volumes of 10,000 6-packs of soda. The mean of the sample means is 12floz. The standard deviation of the sample means is 0.04floz.

DISTRIBUTION OF SAMPLE MEANS
Mean = 12, SD =0.04



Draw an arrow to show the bin where your sample mean lies.
Is the mean volume of your 6-pack unusual? How do you know?

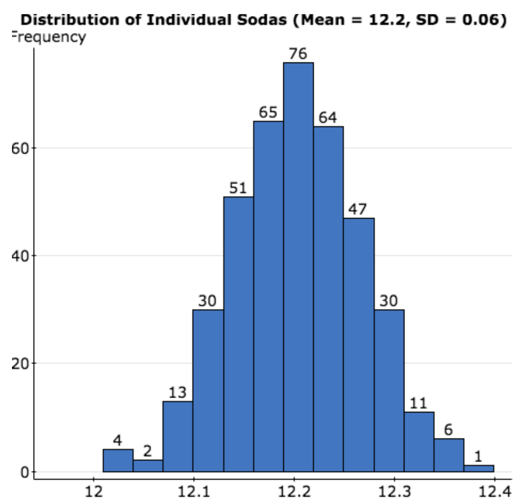
d) Is it unusual for the mean amount of soda to be 11.9floz in the 6-pack? How do you know?

e) What is the probability that the mean amount of soda in a 6-pack is 11.9floz or less?

4) Which is more unusual? Why?

- A 12-pack of soda with a mean volume of 12.2 fl.oz,
- A 6-pack of soda with a mean volume of 12.2 fl.oz, or
- A single soda with a volume of 12.2 fl.oz?

5) The quality control engineer resets the bottle-filling machine and selects another 400 bottles during a 10-minute run. He produces the following histogram. Are the new settings better or worse? Why do you think so?



Group work:

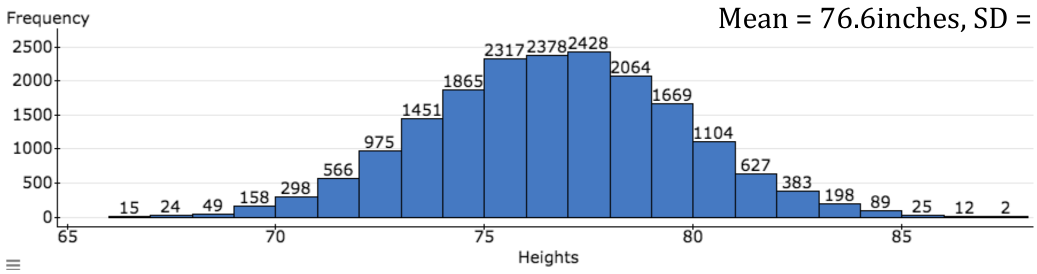
1) There are 18,697 men who play NCAA (National Collegiate Athletic Association) basketball. Their mean height is 76.6 inches with a 3-inch standard deviation.

   Which do you think is more likely? Randomly selecting one NCAA male basketball player who is 79 inches (6'7") tall or randomly selecting a team of 5 players that is 79 inches tall? Why do you think so?

2) A randomly selected NCAA male basketball player is 79.2inches tall.

a) Locate the bin that contains this player in the histogram of individual heights.

DISTRIBUTION OF INDIVIDUALS
Mean = 76.6inches, SD = 3.0inches



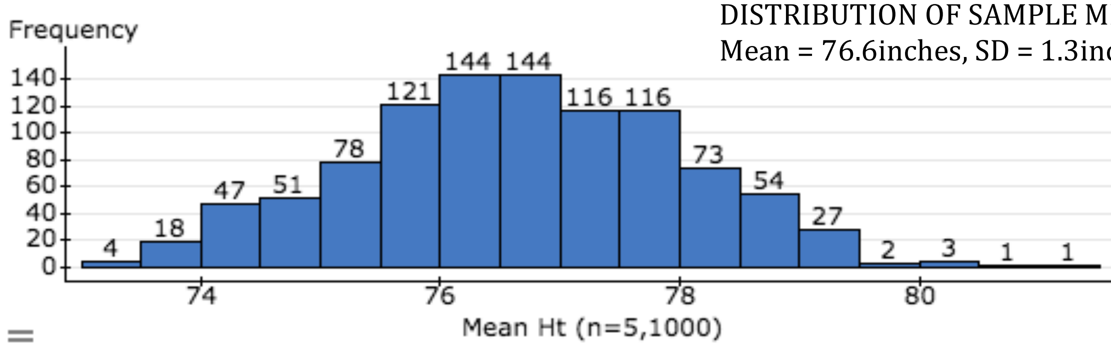
b) Give an interval of typical heights for these players.

c) What is the probability that a randomly selected player is at least 79inches tall?

3) We randomly select a team of 5 players. The team members have the following heights (in inches): 80.1, 84.6, 78.9, 77.5, 75.2.

a) Locate the bin that contains this team's mean height in the histogram of 1,000 sample means.

DISTRIBUTION OF SAMPLE MEANS
Mean = 76.6inches, SD = 1.3inches



b) Give an interval of typical mean heights for randomly selected teams of 5 players.

c) What is the probability that a randomly selected team of 5 players has a mean height of at least 79inches?

4) If we randomly chose a team of 15 players is it more likely or less likely that the mean height of a team is 79inches or more? Why do you think so?

5) Generalize your observations from this activity by completing the following sentences

- If a population has a mean of μ, then the mean of a distribution of sample means will be (circle one: equal to, less than, greater than) μ.

- If a population has a standard deviation of σ, then the standard deviation of a distribution of sample means will be (circle one: equal to, less than, greater than) σ.

- If we increase the sample size, the standard deviation of the distribution of sample means will (circle one: stay the same, decrease, increase).

- If the distribution of a variable (such as height) in the population has a bell-shape, then what can we say about the shape of the distribution of sample means?

Volume (fluid ounces) for 400 bottles of soda from a normal distribution with mean of 12 fl.oz. and standard deviation of 0.1 fl.oz.

Bottle	Volume (fl.oz.)								
1	12.09	41	12.01	81	12.04	121	12.13	161	11.96
2	12.03	42	11.91	82	12.02	122	11.96	162	11.87
3	12.02	43	12.31	83	12	123	11.91	163	11.98
4	11.84	44	11.98	84	11.95	124	12.17	164	11.99
5	11.95	45	11.97	85	12	125	12	165	11.99
6	12.01	46	12.14	86	11.93	126	12.02	166	12.19
7	12.1	47	12.02	87	11.95	127	12.02	167	12.17
8	11.94	48	11.74	88	12.11	128	11.95	168	12.06
9	12.09	49	12.03	89	12	129	11.97	169	12.05
10	12.1	50	11.95	90	11.87	130	11.94	170	12.04
11	11.98	51	11.93	91	12.04	131	12.13	171	11.97
12	12.3	52	11.99	92	12.03	132	11.81	172	12.12
13	11.96	53	11.76	93	12.04	133	11.92	173	11.86
14	11.79	54	12.14	94	11.91	134	11.92	174	11.94
15	12.03	55	12.01	95	12.09	135	12.05	175	11.97
16	11.85	56	12.09	96	11.97	136	12.03	176	11.88
17	12.32	57	12.06	97	11.81	137	11.88	177	12.25
18	11.99	58	12.1	98	11.89	138	12.1	178	11.98
19	11.89	59	11.94	99	11.97	139	12.09	179	12.08
20	11.95	60	11.84	100	11.95	140	11.98	180	12
21	12.01	61	12.03	101	12.22	141	11.98	181	11.84
22	12.16	62	12.03	102	11.94	142	11.82	182	11.99
23	12.03	63	12.04	103	12.11	143	11.91	183	11.94
24	12	64	12	104	12.07	144	11.96	184	12.17
25	12.06	65	11.82	105	11.9	145	12.07	185	12.02
26	11.98	66	12.01	106	12.09	146	12.06	186	12.1
27	12.2	67	11.99	107	11.88	147	12.08	187	12.08
28	12.17	68	12.18	108	11.79	148	12.01	188	12
29	11.97	69	11.98	109	11.95	149	12.02	189	11.77
30	12	70	12	110	12.11	150	12.08	190	12.03
31	12.02	71	11.96	111	12.14	151	12.26	191	11.77
32	12.21	72	12.08	112	12.17	152	12.22	192	11.89
33	11.94	73	12.05	113	11.91	153	12.11	193	12.15
34	12.04	74	12.1	114	12.23	154	12.11	194	11.99
35	12.05	75	12.11	115	11.98	155	12.18	195	12.03
36	11.98	76	11.97	116	12	156	11.98	196	11.95
37	12.04	77	11.87	117	12.1	157	11.98	197	11.97
38	12.05	78	12.1	118	11.88	158	12.03	198	12.11
39	12.07	79	12.01	119	11.95	159	12.06	199	11.99
40	12	80	11.88	120	12.02	160	11.95	200	11.78

Bottle	Volume (fl.oz.)								
201	12.06	241	11.98	281	12.11	321	11.82	361	12.03
202	12.02	242	12.02	282	12.16	322	11.85	362	11.76
203	12.15	243	12.11	283	12.2	323	11.93	363	11.98
204	11.84	244	12.05	284	11.88	324	12.09	364	11.98
205	11.93	245	11.92	285	11.85	325	12.02	365	12.01
206	11.82	246	11.93	286	12.02	326	12.02	366	11.93
207	12	247	12.04	287	12.05	327	12	367	11.94
208	12.02	248	11.96	288	11.95	328	11.99	368	12.01
209	12.12	249	11.98	289	11.96	329	11.81	369	12.04
210	11.92	250	12.08	290	12.14	330	11.84	370	12.17
211	12.09	251	12.06	291	12.13	331	12.15	371	11.98
212	12.06	252	11.96	292	11.99	332	12	372	12.03
213	12.08	253	11.85	293	12.14	333	11.82	373	11.94
214	12.07	254	11.88	294	11.94	334	12.04	374	12.11
215	12.05	255	11.96	295	11.95	335	11.82	375	12.03
216	12.18	256	12.01	296	11.85	336	11.89	376	11.9
217	11.83	257	12.19	297	11.99	337	12.11	377	12.01
218	12.01	258	11.92	298	11.99	338	12.12	378	12.22
219	12.08	259	11.97	299	11.99	339	12.06	379	11.93
220	12.09	260	11.93	300	12.01	340	12.01	380	12.04
221	11.89	261	11.99	301	12.05	341	11.87	381	12.04
222	12.04	262	11.86	302	12.08	342	12.11	382	11.95
223	11.93	263	11.97	303	11.91	343	11.99	383	11.94
224	11.98	264	11.87	304	11.91	344	12.08	384	11.91
225	12.04	265	12.22	305	12.08	345	12.28	385	11.95
226	12.16	266	12.17	306	11.99	346	11.97	386	11.98
227	11.98	267	11.98	307	12.03	347	12.2	387	11.96
228	11.89	268	12.07	308	12.1	348	12.2	388	12.04
229	11.98	269	12.13	309	12.08	349	12.01	389	12.11
230	12.02	270	12.03	310	11.92	350	11.98	390	11.79
231	11.99	271	12.05	311	12.14	351	12.07	391	12.01
232	11.98	272	12.03	312	12.08	352	11.97	392	12.24
233	12.12	273	12.06	313	11.93	353	11.92	393	12.03
234	12.28	274	12.05	314	12.07	354	11.94	394	12.08
235	11.85	275	11.99	315	11.87	355	11.88	395	12.1
236	11.94	276	11.87	316	12.04	356	12.06	396	11.92
237	12.02	277	12.16	317	12	357	12.02	397	11.84
238	12.05	278	12.22	318	11.88	358	12.18	398	12.03
239	11.94	279	12	319	12.02	359	12.06	399	11.9
240	12.16	280	11.92	320	11.95	360	12.11	400	11.95

Module 22.2 Modeling the Distribution of Sample Means

Learning Goals:

- Determine the mean and standard deviation of a distribution of sample means.
- Describe conditions necessary for use of the normal curve to model a distribution of sample means.
- Estimate probabilities using a normal model of the sampling distribution.

Introduction:

We bet you can guess what comes next … we just finished running simulations to observe how the sample means from random samples behave, so now we will create a mathematical model to summarize what we observed.

When can a normal curve be used to model the distribution of sample means?

It really depends on the population's distribution. For small samples, the distribution of sample means will have a similar shape to the population's distribution. Therefore, if a normal curve models the population's distribution well, a normal curve can also be used to model the distribution of sample means, regardless of sample size. But if the population's distribution is skewed, a normal curve is a good fit for the distribution of sample means only for large samples.

The general guideline is that samples of size 30 will have a fairly normal distribution regardless of the shape of the population's distribution.

What is the mean and standard deviation of the distribution of sample means?

If the population has a mean of μ and a standard deviation of σ, then the theoretical distribution of sample means has a mean of μ and a smaller standard deviation given by $\frac{\sigma}{\sqrt{n}}$, where n is the sample size. In theory, this is always true (regardless of the shape of the sampling distribution).

How do we calculate z-scores?

Recall that the z-score is $\frac{sample\ statistic - population\ parameter}{standard\ error}$.

The z-score for an individual measurement is $z = \frac{x-\mu}{\sigma}$

The z-score for a sample mean is $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

1) **Example:** The World Health Organization (WHO) monitors many variables to assess a population's overall health. One of these variables is birth weight. A low birth weight is defined as 2500 grams or less.

   Suppose that babies in a town had a mean birth weight of 3,500 grams with a standard deviation of 500 grams in 2005. This year, a random sample of 25 babies has a mean weight of 3,400 grams. Obviously, this sample weighs less on average than the population of babies in the town in 2005. A decrease in the town's mean birth weight could indicate a decline in overall health of the town.

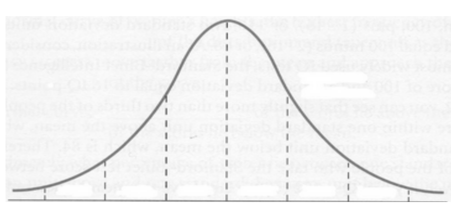
   Are differences this large expected in random sampling from a population with a mean birth weight of 3,500 grams? What is the probability that a random sample of 25 babies will have a mean birth weight of 3,400 grams or less?

   We assume that the variability in individual birth weights is the same this year as it was in 2005. In general, body measurements in a large population can be modeled by a normal curve.

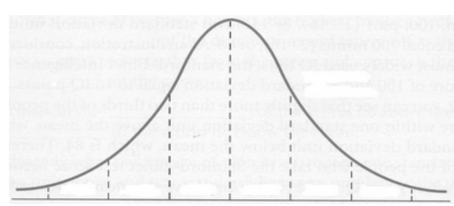
   a) Identify the following: μ , σ , $\bar{x}$, and n from the information given above.

   b) Verify that a normal curve can be used to model the distribution of sample means for this situation.

   c) Label the mean and mean±SD in the normal model for the population and for the sampling distribution.

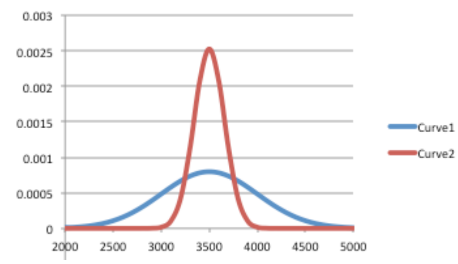


Birth Weights Individual Babies (grams)



Sample mean birth weight
Random samples of 25 babies

Here are the two normal models drawn on the same scale. Which is the population and which is the sampling distribution? How do you know?



d) What is the z-score for a baby that weighs 3400 grams? What is the z-score for a sample of babies with a mean birth weight of 3400 grams? Why do your answers make sense when you look at the normal curves in (c)?

e) What is the probability that a random sample of 25 babies weighs 3,400 grams or less? (Shade the area representing the probability in the appropriate normal curve in (c) and give your estimate.)

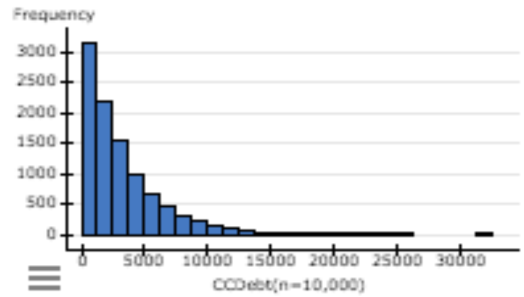
f) Is the difference between 3,400g and 3,500g statistically significant? Or is this difference what we expect to see in random sampling when the population has a mean of 3,500g? How do you know?

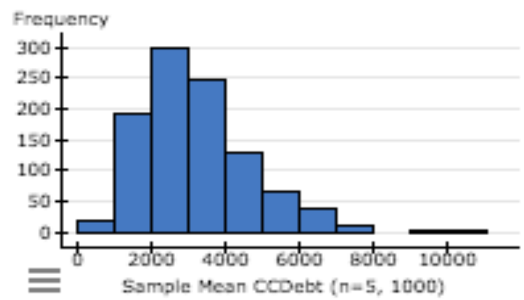
Group work

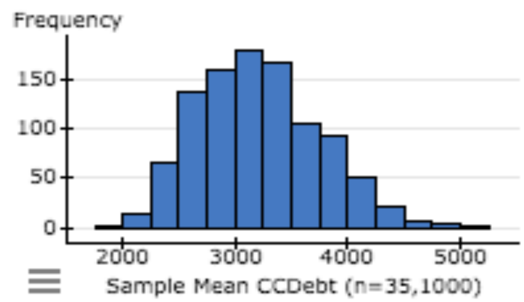
2) According to Debt.org, college students (undergraduates) have an average of \$3,200 in credit card debt. https://www.debt.org/students/debt/

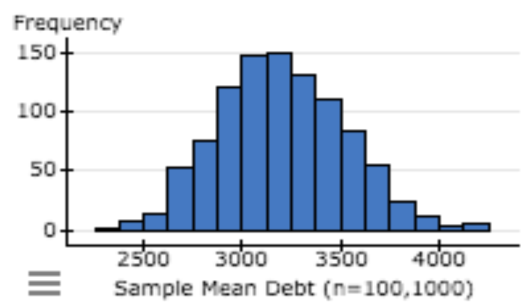
   The top histogram is a hypothetical distribution of credit card debt for 10,000 college students with μ = \$3200 and $\sigma = \$3000$. The distribution is strongly skewed to the right with a few students who have large amounts of credit card debt.

   The other three histograms are sampling distributions. Each sampling distribution contains the sample means from 1,000 samples. One histogram is composed of samples of 5 college students. The samples sizes are 35 and 100 for the other two histograms.









   a) Locate the bin containing an individual student who has \$4766 in credit card debt.

   b) Five college students are randomly selected. Their credit card debts are \$1321, \$5951, \$782, \$3365, \$1504. Locate the bin in the appropriate histogram containing the mean debt for these 5 students.

   c) How do these four histograms illustrate the conditions for the use of a normal curve to model a distribution of sample means?

d) We want to determine the probability that a random sample of 20 college students from this population has an average credit card debit of $3500 or more. Which can we use to find the probability? Circle all that will work.

- Empirical Rule
- StatCrunch Normal Calculator
- OLI Z-score calculator
- Simulation
- none of these

e) Is it unusual for a random sample of 100 college students to have an average credit card debit of $3500 or more? What is the probability? (Show your work.)

3) Which of the following questions can be answered based on a normal model? Explain. (You do not need to answer the questions, just determine if a normal model is appropriate.)

a) According to the growth charts produced by the World Health Organization, one-month-old girls have a mean head circumference of 36.55cm and a standard deviation of 1.17cm. In general, body measurements in a large population can be modeled by a normal curve.

In a study of health conditions in a county with a high poverty rate, researchers find that a random sample of 25 one-month-old girls have a mean head circumference of 36cm. Does this sample provide strong evidence that the mean head circumference for the population of one-month-old girls in this county is unusually small?

b) According to the US Census Bureau 2014 Annual Social and Economic Supplement, the mean household income in the United States was $72,641. Previous studies suggest that the standard deviation is about $35,000. Income data is skewed strongly to the right.

We are interested in determining whether the mean household income is higher in our county. We randomly sample 25 households and determine that the mean income is $65,000. Does this sample provide strong evidence that the mean income is lower in our county?

4) At Starbucks the advertised nutrition facts say that a Tall Chai Latte with nonfat milk has 32 grams of sugar, which is equivalent to 8 sugar cubes. Of course, there will be some variability in sugar content. Suppose that the standard deviation is about 0.1 grams. Engineers monitoring production processes assume that measurements are normally distributed.

   A quality control engineer randomly samples 10 Tall Chai Lattes and finds a mean sugar content of 32.07 grams. Is the difference between 32.07 and 32 statistically significant for a sample of 10?

   a) Explain why we can use a normal curve to model the distribution of sample means despite the fact that the samples only contain 10 lattes.

   b) What is the z-score for the engineer's sample? What does the z-score tell us?

   c) Is the sample statistically significant? Support your answer using probability.

Module 23 The Confidence Interval for a Population Mean

Learning Goals:

- Construct a confidence interval for the population mean when conditions are met. Interpret the confidence interval in context.
- Interpret the meaning of a confidence level associated with a confidence interval.
- Explain how the margin of error is affected by changes in sample size and by changes in confidence level.

Introduction:

In this activity we will learn to construct a confidence interval that provides a range of estimates for the population mean. As before, the confidence interval is based on a normal model for the distribution of sample statistics, in this case, sample means.

1) Recall the general formula for a 95% confidence interval is…

    Sample statistic ± margin of error
    Sample statistic ± 2(standard error)

    Given what we have learned about the distribution of sample means, this becomes

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}}$$

    where z_c is the tabulated z-value for the desired confidence level, for example 1.96 (rounded to 2 by our rule-of-thumb) for 95% confidence.

2) What conditions have to be met to use this formula to estimate a population mean?

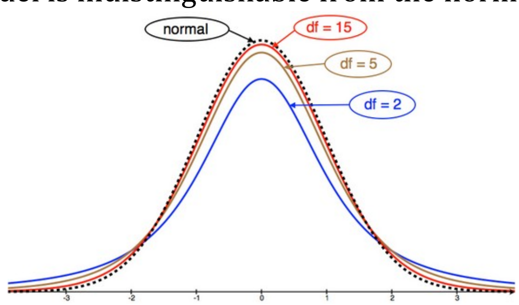
For this confidence interval, we have to know the population standard deviation, σ, based on previous studies. It may occur to you that if you do not know μ, it is unlikely that you know σ. So, we are forced to take a different approach. **We estimate σ using the sample standard deviation *s*.**

We will use the sample standard deviation s, instead of population standard deviation σ, in the calculation of the standard error.

This is the same type of adjustment we used in when substituting the sample proportion $\hat{p}$ for population proportion p in estimating the standard error in sample proportions.

But this adjustment is not as straightforward as our work with proportions. This estimate for σ introduces more uncertainty in the process. The problem is worse with smaller samples because the sample standard deviations vary more. Therefore, for small samples, s is a worse approximation for σ. Unfortunately, this makes the normal model a bad fit and inappropriate for determining critical values. For this reason, we use what is called a *T-model* instead of the normal model.

The T-model, like the normal model, is bell-shaped. However, the shape of a normal model is affected by its mean and standard deviation while a T-model is affected by the sample size, or more specifically by the *degrees of freedom*. The degrees of freedom are one less than the sample size (n-1). Therefore, each sample size has its unique T-model. As sample sizes increase, the T-model is indistinguishable from the normal model.



What is the formula for the confidence interval if we use the sample standard deviation to estimate the population standard deviation?

The confidence interval formula changes from

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}} \text{ to } \bar{x} \pm t_c \frac{s}{\sqrt{n}}$$

So how do we determine the T-score we need for our confidence interval?
We will use an applet or we will let StatCrunch do this for us.

What are the conditions for use of the T-model?
The conditions for use of a T-model are similar to the conditions for use of a normal model; however, we can use the T-model for smaller samples in some situations.

- Must have a random sample (as with all statistical inference procedures.)
- If a normal model fits the population's distribution, then use the T-model.
- If a normal model does not fit the population's distribution (or you can't determine if this), then, with sample sizes of 30 or greater, use the T-model.
- For a sample of less than 30, plot the data. If the distribution of data in the sample is not heavily skewed, then we assume the population's distribution is also not heavily skewed and we use the T-model.

1) **Example:** Community college students often work and have family responsibilities in addition to attending college. Busy schedules can result in less sleep. In a project for her statistics class, a student randomly selects 20 students and determines that this sample sleeps on average 6.2 hours a night with a standard deviation of 0.7 hours.

   Here is the distribution of sleep hours in her sample.

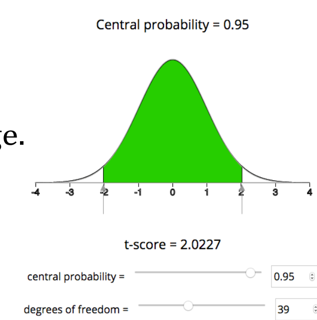
   Estimate the mean hours of sleep for students at this college with a confidence interval.

   a) Verify that conditions are met for use of the confidence interval formula involving the T-model.

   b) Why does this student need to use a T-score instead of a Z-score in her confidence interval?

   c) Determine the t-score using the OLI applet image.

   d) Find and interpret the confidence interval.



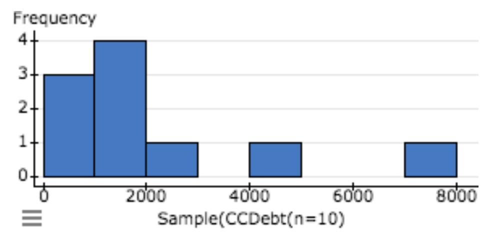
Group work:

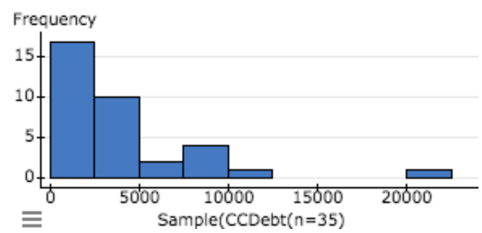
1) We want to estimate the mean amount of credit card debt ($) owed by students at our college.

   Here are two different samples randomly selected from the population of students at our college. Can we use the T-model to find a confidence interval using the sample mean from either of these samples? Why or why not?

Summary statistics:

Column	n	Mean	Std. dev.
Sample(CCDebt(n=10)	10	2239.8	2077.2873
Sample(CCDebt(n=35)	35	3910.8	4338.716



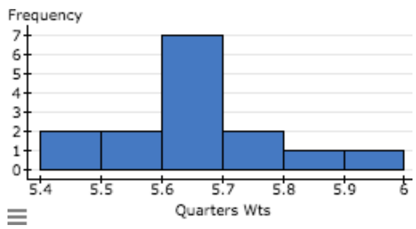


2) To discourage the use of counterfeit coins, vending machines can be set to reject unusually light coins and unusually heavy coins. Such settings are based on an estimated mean weight of the coins. What is the mean weight of quarters?

   As with most manufacturing processes, we can assume that a normal model is a good fit for the distribution of weights.

   We collect 15 quarters from various drawers in our house and car and weigh them.

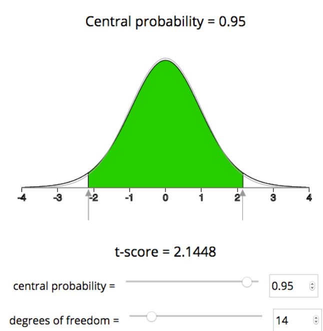
Here is information about the sample:

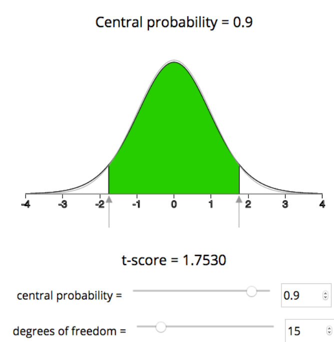


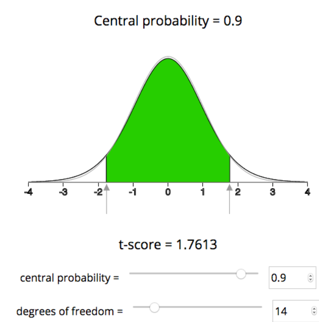
Column	n	Mean	Std. dev.
Quarters Weights (grams)	15	5.65	0.12

a) Explain why we can use the T-model despite the small sample size and slight skew in the data.

b) Pick the image with the correct T-score for a 90% confidence interval for this situation and calculate the 90% confidence interval.







c) According to Wikipedia, quarters are manufactured with a weight of 5.67 grams. Does this seem reasonable given your confidence interval? Why or why not?

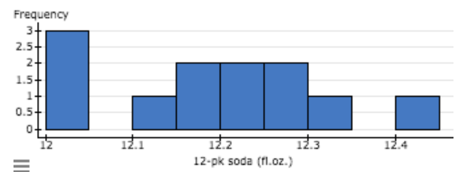
3) A soda is labeled with a volume of 12 fluid ounces. Quality control processes are set to slightly overfill bottles so that customers do not receive less than 12-fl.oz. There is inevitable variability in the filling process; but, as with most manufacturing processes, we can assume that a normal model is a good fit for the distribution of volumes.

   What is the mean volume of soda in a can marked 12 fluid ounces?

   To answer this question, suppose that we measured the volume of 12 sodas in a 12-pack and got a mean of 12.19 fl.oz. Below is information about the sample of 12 sodas.

   Summary statistics:

Column	n	Mean	Std. dev.
12-pk soda (fl.oz.)	12	12.190833	0.11988315



   Below is the StatCrunch print-out for the confidence interval based on the T- model.

   One sample T confidence interval:
   μ : Mean of variable

   95% confidence interval results:

Variable	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
12-pk soda (fl.oz.)	12.190833	0.034607285	11	12.114663	12.267003

   a) The StatCrunch T confidence interval is based on the T-model. Explain why use of the T-model is appropriate here despite the small sample size and the slight skew in the data.

   b) What does the standard error tell us?

   c) Interpret the confidence interval in the context of soda volume.

4) Below are two T confidence intervals based on the same sample of 12 sodas. One is the 90% confidence interval and the other is the 99% confidence interval. Which is which? How do you know?

Variable	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
12-pk soda (fl.oz.)	12.190833	0.034607285	11	12.128683	12.252984

Variable	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
12-pk soda (fl.oz.)	12.190833	0.034607285	11	12.08335	12.298317

5) Below are two 95% confidence T intervals based two different samples. One sample was a 6-pack, the other a 12-pack. Which is which? How do you know?

Variable	Sample Mean	Std. Err.	L. Limit	U. Limit
Sample 1	12.190833	0.034607285	12.114663	12.267003

Variable	Sample Mean	Std. Err.	L. Limit	U. Limit
Sample 2	12.106667	0.037475918	12.010332	12.203002

Module 24 Hypothesis Test for a Population Mean

Module 24.1 Hypothesis Test for a Population Mean

Learning Goals:

- Under appropriate conditions, conduct a hypothesis test about a population mean. State a conclusion in context.
- Interpret the P-value as a probability.

Introduction:

Up to this point, we have had quite a bit of experience conducting hypothesis tests. But we have always worked with categorical data and tested hypotheses about proportions. In this activity we learn to conduct a hypothesis test about a population mean. We will see that the process and the logic of the hypothesis test are very similar to what we did previously.

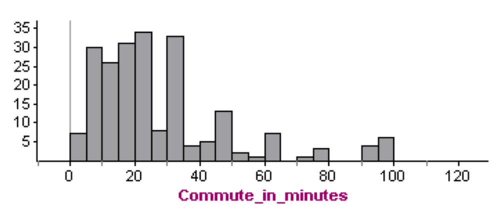
Example:

According to a 2015 study, the average American worker spends about 25 minutes a day commuting to (or from) work. Using U.S. Census data, we will test the hypothesis that commute times are no different now than in the year 2000.

Source: http://www.cheatsheet.com/business/5-cities-with-the-most-brutal-commutes.html/?a=viewall

State the hypotheses:

Collect the data:
A random sample of 215 adults is selected from the 2000 Census. The mean commute time is 25.7 minutes with a standard deviation of 21.78 minutes. Here is a histogram of the data.



- The sample mean of 25.7-minutes from the year 2000 is obviously very close to the 25-minute commute time in 2015. Why do we need to do a hypothesis test? Can't we just say that there is no statistically significant difference in the commute time in the years 2000 and 2015?

Assess the data:

- The skew in this data suggests that the distribution of commute times in the population in the year 2000 is also skewed.

  Verify that we can use a T-curve to model the distribution of sample means despite the likely skew in the population's commute times.

- Calculate the T-score:

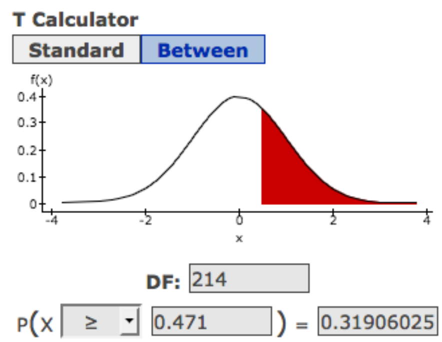
$$T = \frac{statistic - parameter}{standard\ error} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{25.7 - 25}{21.78/\sqrt{215}} = 0.471.$$

- Use the StatCrunch T-calculator to find the P-value. Note that the T-calculator requires you to calculate and enter degrees of freedom and a T-score.

  What is the P-value for this test?

  How does it compare with the StatCrunch print-out?

  Shade the entire P-value in the T-model.



One sample T hypothesis test:
μ : Mean of population
$H_0 : \mu = 25$
$H_A : \mu \neq 25$

Hypothesis test results:

Mean	Sample Mean	Std. Err.	DF	T-Stat	P-value
μ	25.7	1.4853837	214	0.47125871	0.6379

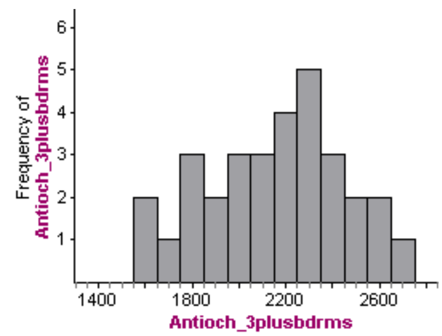
State a conclusion:

Group work

1) For her statistics project, a student wants to determine whether rental rates are lower in Antioch than the county average. According to her research, the average monthly rent paid for a 3+bedroom house in Contra Costa County is $2500. (Source: http://www.zilpy.com/US/California/Contra_Costa_County/)

   She randomly selects thirty 3+bedroom houses in Antioch by randomly selecting 6 rentals from the tulia.com list each week for 5 weeks.

   The sample mean is $2,137.52 with a standard deviation of $299.83



   a) Explain why she can conduct a hypothesis test using this data.

   b) Her hypotheses are Ho:μ=2500, Ha:μ<2500. What does μ represent in terms of rents? Be specific!

   c) Here is the StatCrunch output for her hypothesis test. What can she conclude about 3+bedroom rentals in Antioch?

One sample T hypothesis test:
μ : Mean of population
H_0 : μ = 2500
H_A : μ < 2500

Hypothesis test results:

Mean	Sample Mean	Std. Err.	DF	T-Stat	P-value
μ	2137.52	54.741218	29	-6.6217014	<0.0001

   d) What does a standard error of 54.74 mean in this context?

   e) What does the T-Stat tell us in this context?

2) For their project, a group of statistics students investigate the features that distinguish a large egg from an extra large egg. With some internet research, they learn that USDA guidelines are based on a carton of 12 eggs. The minimum net weight per dozen for large eggs is set at 24 ounces.

Source: http://www.fsis.usda.gov/wps/portal/fsis/topics/food-safety-education/get-answers/food-safety-fact-sheets/egg-products-preparation/shell-eggs-from-farm-to-table/ct_index

Each team member buys two cartons of large eggs from different grocery stores in the area, for a total of 6 cartons. Because there are so many brands, they choose the third brand from the left on each grocery store shelf.

They assume that weights of dozens of large eggs will be normally distributed and they feel that their process for selecting cartons will produce a random sample. Therefore, they proceed with a T-test.

Here is the StatCrunch printout for their project.

One sample T hypothesis test:
μ : Mean of population
$H_0 : \mu = 24$
$H_A : \mu > 24$

Hypothesis test results:

Mean	Sample Mean	Std. Err.	DF	T-Stat	P-value
μ	24.25	0.12247449	5	2.0412415	0.0484

a) Which of the following are correct ways for them to state their conclusions?

- At a 5% level of significance, our sample gives strong evidence that the mean weight of a dozen large eggs is greater than 24 ounces.
- At the 5% level of significance, the amount that our sample differs from a mean weight of 24 ounces is statistically significant. Therefore, we reject the null hypothesis in favor of the alternative.
- The 0.25-difference between 24.25 and 24 is too small to be statistically significant at the 5% level. It is not surprising to see random samples vary from 24 ounces as much as our sample.
- We fail to reject the null hypothesis because our sample is only off by 0.25 ounces, which is less than 5%.

b) Which if the following is an accurate interpretation of the P-value?

- If it is true that a random sample of a dozen large eggs weighs 24.25 ounces or more, then there is a 4.8% chance that the mean weight of all cartons of a dozen large eggs is 24 ounces.

- If it is true that all cartons of a dozen large eggs weigh 24 ounces on average, then there is a 4.8% chance that a random carton of a dozen large eggs will weigh 24.25 ounces or more.

- There is a 4.8% chance that the mean weight of all cartons of a dozen large eggs is greater than 24 ounces.

- There is a 4.8% chance that the mean weight of a random sample of a dozen large eggs will weigh 24 ounces.

3) A group of statistics students wants to determine if LMC students work more than 20 hours a week on average. They ask 20 randomly selected students who are working in the math lab on a Tuesday afternoon at 2:15, "Do you work more than 20 hours a week?"

   Will these students be able to run a T-test to determine if LMC students work more than 20 hours a week? Why or why not?

4) A group of statistics students wants to investigate how fast on average they can download a 4-minute song from iTunes using an iPhone and an AT&T phone plan that advertises up to 6Mbps download speed.

   They want to test the hypothesis that the mean download time is greater than 6 seconds. They plan to use a T-test.

   What advice do you have for them about designing their study?

5) The following is an excerpt from the Food and Agriculture Organization of the United Nations (highly summarized).

   In testing meat contamination, the indicator bacteria most commonly used in testing should not exceed 100 per cm for the criterion "good microbiological standard."

   Source: http://www.fao.org/docrep/010/ai407e/AI407E24.htm

   In a T-Test of the following hypotheses,

   Ho: The meat is not spoiled (mean bacteria count is 100 per cm)
   Ha: The meat is spoiled (mean bacteria count is greater than 100 per cm)

   a) Describe Type I and Type II errors for this situation.

   b) Which is the more serious error? Why?

   c) What can be done to avoid making Type I and Type II errors?

Module 24.2 Hypothesis Test for a Mean with Matched Pairs

Learning Goals:

- Under appropriate conditions, conduct a hypothesis test about a mean for a matched pairs design. State a conclusion in context.
- Interpret the P-value as a conditional probability.

Introduction:

In this activity we will again be testing a hypothesis about a population mean, but this time we will have data from a matched pairs design. For our purposes, *matched pair* means that we have two measurements for each individual, such as a pre- and post-exam score for a set of students.

It can also mean that two groups of subjects are matched based on demographic characteristics. One person in each pair is exposed to a treatment and the other person is not. We will not work with this type of matched pairs design in this course, but it is handled the same way.

A matched pairs model is really a straightforward extension of what we already have done. The new population is simply the difference in measurements for each individual. For example, $x_{post} - x_{pre}$ in the pre- and post-exam scenario. And the mean of the population of differences is indicated by the new notation μ_d, read "mu sub d." Here, *d* stands for difference. The analysis itself just uses the differences as the raw data.

Example:

The following post was pulled from allnurses.com

"Hey everyone,
So I am in midst of an argument at work between a CNA and a new grad RN over which temp is more accurate Oral or Tympanic [ear] now I know different variables come into play using both i.e. with oral the patient may have drank something cold or hot or with tympanic laying on one side on pillow can cause false highs etc. etc., so lets assume no such variables are in play - which temp would be most accurate?"

Source: http://allnurses.com/general-nursing-discussion/which-temp-is-615783.html

Suppose that we have 12 randomly selected patients and a nurse measures each patient's oral temperature and then tympanic (ear) temperature.

Here are the (hypothetical) results: Complete the table.

	1	2	3	4	5	6	7	8	9	10	11	12
Oral (°F)	98.4	97	96.5	97.1	98.9	98.2	96.8	99	97.9	100	99.6	98.7
Ear (°F)	98.9	98.2	96	97.8	99	98.7	98.6	98	98.6	101	100	99.6
Difference (Ear-Oral)	0.5	1.2	-0.5	0.7	0.1	0.5	1.8	-1	0.7			

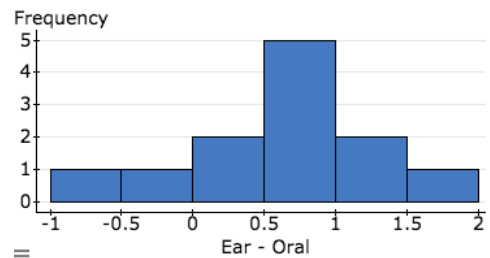
Let's test the claim that oral and tympanic temperatures are different using a 5% significance level.

State the hypotheses:

Collect the data:

Here is a histogram of the 12 differences.

- Draw an arrow to locate Patient #2 in the histogram.
- What does a positive difference indicate?
- What does the mean of 0.525 represent?



Summary statistics:

Column	n	Mean	Std. dev.
Ear - Oral	12	0.525	0.74360913

Assess the data:

- Verify that the conditions for use of the T-model are met.
- Verify that the T-score is about 2.45 as shown in the StatCrunch print-out.

Paired T hypothesis test:
$\mu_D = \mu_1 - \mu_2$: Mean of the difference between Ear and Oral
$H_0 : \mu_D = 0$
$H_A : \mu_D \neq 0$
Hypothesis test results:

Difference	Mean	Std. Err.	DF	T-Stat	P-value
Ear - Oral	0.525	0.21466147	11	2.4457114	0.0325

Differences stored in column, Differences.

- What is the P-value?

State a conclusion:

Group work:

1) Determining the alternative hypothesis requires some careful thinking with a matched pairs design. For each of the following situations, state the alternative hypothesis as $\mu_d > 0, \mu_d < 0, or\ \mu_d \neq 0$. Briefly explain your reasoning.

    a) In a study of a drug designed to reduce cholesterol levels, patients take the drug for six weeks. Researchers record the cholesterol levels of patients before and after taking the drug. The difference (before minus after) is used to test the claim that the drug reduces cholesterol levels.

    b) In a study of a treatment designed to increase red blood cell counts, researchers use the difference in red blood cell count (before minus after) to test the claim that the treatment has the desired effect.

    c) In an automobile safety study, researchers use a matched pairs design to examine the effect of cruise control on reaction time to brake. They test the claim that the use of cruise control is less safe because it increases reaction time to brake. For the T-test the difference in reaction times is defined as "cruise control" minus "no cruise control."

2) Students from a statistics class perform an experiment to assess the effect of music on concentration. They measure "concentration" by recording the time it takes a student to complete a simple word puzzle. Each student does a word puzzle while listening to classical music and in silence. A coin flip determines the order of treatments.

   They test the claim that listening to classical music will improve concentration; therefore, the time for completing a word puzzle will be faster with music (a shorter completion time) and the mean of the difference in completion times ("no music" minus "music") will be positive. (H_0: $\mu_d = 0$, H_a: $\mu_d > 0$).

   They use a sample of 15 student volunteers; therefore, the sample is not randomly chosen, but the treatments are randomly ordered, which allows them to conduct a hypothesis test if the conditions are met. The distribution of the difference in puzzle completion times is not strongly skewed and there are no outliers, so the conditions are met for use of a T-test.

   For a sample mean of 3 seconds and a standard deviation of 5 seconds, the P-value is approximately 0.024.

   What can we conclude? Explain why you chose, or do not choose, each option.

   a) For the population of college students, this study suggests that listening to classical music produces a 3-second improvement in puzzle completion time.

   b) Classical music is associated with statistically significant improvements in concentration as measured by time to complete a simple word puzzle.

   c) College students listening to classical music completed word puzzles significantly faster than students not listening to music.

   d) For the population of college students, this study provides fairly strong evidence that listening to classical music improves word puzzle completion time, which in turn suggests that listening to classical music may improve concentration.

3) William S. Gosset invented the T-model when he was an employee of the Guinness Brewing Company in Dublin. The Guinness Co. was interested in increasing grain yields to lower the cost of brewing beer. Samples from these experiments were often small and statistical methods of the day were inadequate for analyzing small data sets.

   In his famous 1908 paper titled "The Probable Error of a Mean," Gossett illustrates his T-model using the results of an experiment published by Dr. Voelcker in the *Journal of Agricultural Society*. Voelcker's experiment was designed to determine if kiln-dried seed produced greater yields. In this experiment 11 different plots of land were planted with two different types of seed: regular or kiln-dried.

   Grain grown in different plots may experience different growing conditions, such as differences in soil fertility or amount of light. To reduce the potential influence of these confounding variables, the experiment used a matched pairs design and both types of seed were grown in each of the 11 plots.

   The table shows the data from this experiment as reported in Gossett's paper. The variable is corn yield in pounds per acre.

   We used the data from Gossett's paper to run a matched pairs T-test.

	lbs. head corn per acre		
	N. K. D.	K. D.	Diff.
1899	1903	2009	+106
	1935	1915	− 20
	1910	2011	+101
	2496	2463	− 33
	2108	2180	+ 72
	1961	1925	− 36
	2060	2122	+ 62
1900	1444	1482	+ 38
	1612	1542	− 70
	1316	1443	+127
	1511	1535	+ 24
Average	1841·5	1875·2	+33·7

Paired T hypothesis test:
$\mu_D = \mu_1 - \mu_2$: Mean of the difference between Kiln-dried seed and Regular seed
$H_0 : \mu_D = 0$
$H_A : \mu_D > 0$
Hypothesis test results:

Difference	Mean	Std. Err.	DF	T-Stat	P-value
Kiln-dried seed - Regular seed	33.727273	19.951346	10	1.6904761	0.0609

Differences stored in column, Differences.

a) What does H_0: $\mu_d = 0$ mean in the context of corn yields and seed types?

b) Explain why the alternative hypothesis is $\mu_d > 0$ instead of $\mu_d < 0$.

c) The standard error is about 20. What are the units for the standard error? What does this number tell us?

d) What does the T-stat of 1.69 tell us in this context?

e) In his paper Gossett quotes Voelcker's conclusion, "In such seasons as 1899 and 1900 there is no particular advantage to kiln-drying before sowing." He adds that his own examination justifies this conclusion. Our T-test is consistent with this conclusion at a 5% level of significance.

Of course, Voelcker did not perform any statistical inference procedures. After all, statistical inference procedures for analyzing small data sets did not exist, which is why Gossett wrote his paper.

Voelcker just observed that the kiln-dried seeds produced 33.7 more pounds of corn per acre than the regular seeds. Why do statisticians want to run a hypothesis test instead of just relying on descriptive summaries of the data?

Historical note excerpted from Wikipedia: Another researcher at Guinness had previously published a paper containing trade secrets of the Guinness brewery. To prevent further disclosure of confidential information, Guinness prohibited its employees from publishing any papers regardless of the contained information. However, after pleading with the brewery and explaining that his mathematical and philosophical conclusions were of no possible practical use to competing brewers, he was allowed to publish them, but under a pseudonym ("Student"), to avoid difficulties with the rest of the staff. Thus his most noteworthy achievement is now called Student's, rather than Gosset's, t-distribution. Source: https://en.wikipedia.org/wiki/William_Sealy_Gosset

Module 25 Inference for a Difference between Population Means

Module 25.1 Inference for a Difference in Population Means

Learning Goals:

- Under appropriate conditions, conduct a hypothesis test about a difference between two population means. State a conclusion in context.
- Construct a confidence interval to estimate a difference in two population means (when conditions are met). Interpret the confidence interval in context.

Introduction:

In this activity we will use StatCrunch to conduct inference procedures (hypothesis test and confidence interval) for a difference in population means.

To use T-procedures, samples must be randomly selected and independent, or, in the case of an experiment, randomization must be used to assign treatments.

Conditions for use of the t-test that we learned earlier in Unit 10 must be met for both samples.

The hypotheses for two population means are similar to those for two population proportions. The null hypothesis, H_0, is a statement of "no effect" or "no difference."

H_0: $\mu_1 - \mu_2 = 0$, which is the same as H_0: $\mu_1 = \mu_2$

The alternative hypothesis, H_a, takes one of the following three forms:

- H_a: $\mu_1 - \mu_2 < 0$, which is the same as H_a: $\mu_1 < \mu_2$
- H_a: $\mu_1 - \mu_2 > 0$, which is the same as H_a: $\mu_1 > \mu_2$
- H_a: $\mu_1 - \mu_2 \neq 0$, which is the same as H_a: $\mu_1 \neq \mu_2$

If conditions are met, we will use StatCrunch to find a P-value based on a T-model determined by degrees of freedom $(n_1 - 1) + (n_2 - 1)$.

The confidence interval estimates the difference between two population means or the difference between means coming from two treatments. The confidence interval has the usual form: *sample statistic ± margin of error*, where the margin of error is based on the standard error. For those who like formulas, here is the formula for the confidence interval:

$$(\bar{x}_1 - \bar{x}_2) \pm T_c \cdot \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}$$

If conditions are met, we will use StatCrunch, instead of this formula, to find confidence intervals in this module based on a T-model determined by degrees of freedom $(n_1 - 1) + (n_2 - 1)$.

Swimming With Dolphins Lifts Depression: Study

Taking a dip with dolphins can be a tremendous therapy for people with depression, according to a study published in the weekly British Medical Journal (BMJ).

Nature lovers - biophiles, to give them their scientific name - have long argued that interaction with animals can soothe a troubled mind but this claim has always been anecdotal, lacking the scientific data to back it up.

Seeking to find out more, psychiatrists Christian Antonioli and Michael Reveley at Britain's University of Leicester, recruited 30 people in the United States and Honduras who had been diagnosed with mild or moderate depression.

The severity of their symptoms was calculated according to established yardsticks for mental health, the Hamilton and Beck scales, which are based on interviews and questionnaires with the patient.

The volunteers were required to stop taking any anti-depressant drugs and psychotherapy for four weeks.

Half of the group was then randomly selected to play, snorkel and take care of dolphins each day at an institute for marine sciences in Honduras.

The other half was assigned to a programme of outdoor activities, also at the institute, that included swimming and snorkelling at a coral reef, but without the dolphins.

Two weeks later, both groups had improved, but especially so among patients who had been swimming with the dolphins.

Measurable symptoms of depression in the dolphin group had fallen by half and by two-thirds according to the two scales - twice as much as in the non-dolphin group.

In addition, a self-rating measurement of anxiety symptoms, the Zung scale, found a fall of more than 20 percent among the dolphin group, compared with a decline of 11 percent among the non-dolphin groups.

"To the best of our knowledge, this is the first randomized, single blind, controlled trial of animal-facilitated therapy with dolphins," say Antonioli and Reveley.

Source: http://www.banderasnews.com/0511/hb-swimwithdolphins.htm

The news article summarizes the actual results from the 30 subjects and the 15 in the treatment group that swam with dolphins had greater improvements than the 15 in the control group that swam and snorkeled. But, are the differences statistically significant? Or would we expect to see the differences described when comparing two identical treatments? In other words, could the differences be due to expected variability arising naturally when comparing two groups? We will investigate these questions in groups.

Each group will investigate **one** option and present a short report to the class using the document camera. Details are on the next page.

AFTER the class presentations, in the box below you will write an overall conclusion summarizing the findings of this study for a general audience … something that could be added to newspaper article about this study.

The researchers used three different tools for assessing depression: Hamilton scale, Beck inventory, and Zung self-rating. Higher scores on these scales indicate more severe depression or anxiety.

We do not have access to the raw data, so we will assume that the conditions for use of the T-model are met despite the small sample sizes.

Options:

Option 1: Based on the Hamilton depression scale, is the difference in improvement between the treatment group and control group statistically significant? Conduct a hypothesis test to answer this question.

Option 2: How big is the treatment effect based on the Hamilton depression scale? Estimate the size of the treatment effect with a confidence interval.

Option 3: Based on the Beck depression inventory, is the difference in improvement between the treatment group and control group statistically significant? Conduct a hypothesis test to answer this question.

Option 4: How big is the treatment effect based on the Beck depression inventory? Estimate the size of the treatment effect with a confidence interval.

Option 5: Based on the Zung self-rating anxiety scale, is the difference in improvement between the treatment group and control group statistically significant? Conduct a hypothesis test to answer this question.

Option 6: How big is the treatment effect based on the Zung self-rating anxiety scale? Estimate the size of the treatment effect with a confidence interval.

Group work:

1) Psychologists have developed a variety of tools to measure depression or anxiety levels. As you can imagine, measuring a mental state is harder and less precise than measuring physical conditions, like height or weight.

   For this reason, in this study, the researchers used three different surveys that are commonly used by the mental health professionals.

   Review the survey associated with the scale or inventory that is relevant to your research question. Your instructor will provide an easy way to access these links.

   DO NOT complete the survey because this will take too much time. Just review the survey and discuss in your group how to best summarize the survey during your presentation to the class. You might also want to pull a few survey questions as examples.

   Hamilton depression scale (0-38):
   http://healthnet.umassmed.edu/mhealth/HAMD.pdf

   Beck depression inventory (0-60):
   http://www.hr.ucdavis.edu/asap/pdf_files/Beck_Depression_Inventory.pdf

   Zung self-rating anxiety scale (20-80):
   https://www.mnsu.edu/comdis/isad16/papers/therapy16/sugarmanzunganxiety.pdf

2) Here are the results of the study excerpted from the British Medical Journal. Source : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1289317/

	Number of participants	Change = mean of differences (before – after)	SD of differences (before – after)
Hamilton rating scale for depression			
Treatment group	15	7.27	3.47
Control group	15	3.60	3.36
Beck depression inventory			
Treatment group	15	13.40	7.58
Control group	15	6.07	7.28
Zung self-rating anxiety scale			
Treatment group	15	9.80	7.32
Control group	15	5.73	5.76

The researchers used both matched-pairs and two-sample T-procedures.

- Using each of the 3 scales, they rated each subject both before and after the experiment. They then calculated the difference (before minus after) to represent the effect of the treatment on the subject according to each scale; this is the matched-pairs part.
- The researchers then calculated the mean and standard deviation of these differences for the control group and for the treatment group; this is the two-sample part. These are the means and standard deviations in the table.

For example, suppose subject #1 is randomly assigned to the treatment group. His Hamilton rating was 14 before the experiment and 8 after the experiment. His change (an improvement) is 6 (14-8=6). His change score is one of the 15 numbers used to calculate the mean of 7.27 and SD of 3.47 shown in the table.

Subject #1 will also have a Beck "before" and "after" score. The difference between these two scores will be one of the 15 numbers used to calculate the mean of 13.40 and SD of 7.58 shown in the table, etc.

a) StatCrunch requires you to enter a mean, standard deviation and sample size for each of the two groups. Circle the information in the table you will use.

b) Write a sentence to explain what each of these numbers tells us.

3) Conduct a two-sample T-procedure in StatCrunch that addresses your research question. (Open a StatCrunch spreadsheet. Choose Stat, T Stats, Two sample, With summary)

Group presentation:

- State your research question and briefly explain the scale used in your investigation (e.g. Hamilton).
- Identify the appropriate mean improvement for the treatment group and for the control group from the table of results. Explain what these numbers tell us. Do the same for the standard deviations.
- Complete the StatCrunch template below that matches your inference procedure (hypothesis test or confidence interval).
- State a conclusion that references the context of the study.
- Be prepared to explain the meaning of each of the numbers in the StatCrunch print-out.

Research question:

Sample results and their meaning:

Two sample T-procedure:

μ_1 :

μ_2 :

$\mu_1 - \mu_2$: Difference between two means

Inference Results:

For a hypothesis test:

$H_0 : \mu_1 - \mu_2$

$H_A : \mu_1 - \mu_2$

(with pooled variances)

Difference	**Sample Diff.**	**Std. Err.**	**DF**	**T-Stat**	**P-value**
$\mu_1 - \mu_2$					

For a 95% confidence interval:

Difference	**Sample Diff.**	**Std. Err.**	**DF**	**L. Limit**	**U. Limit**
$\mu_1 - \mu_2$					

Conclusion:

Module 25.2 Unit 9 Lab

The Unit 9 Lab is a series of questions about a study conducted by a professor of kinesiology at Michigan State University. Three different perspectives on this study are provided in this document:

- An excerpt from a Huffington Post article about the study;
- A short abstract of the official study published in the *Recreational Sports Journal*, which is the official journal of the NIRSA (National Intramural-Recreational Sports Association);
- The flyer produced by the researchers to advertise the benefits of a Campus Recreation and Fitness Facilities membership (CRFF) to Michigan State University students.

Carefully read these three excerpts and then answer the questions that follow at the end of this document.

If You Want A Higher GPA, Study Shows You Should Join A Gym

The Huffington Post | By Nina Friend
Posted: 07/11/2014 2:57 pm EDT Updated: 07/11/2014 2:59 pm EDT

Can pumping iron pump your grade point average? New research from Michigan State University says yes.

An MSU study released this week compares 4,843 freshman and sophomores on the basis of whether they had membership to the school's recreational sports and fitness centers. The findings are significant: the students who belonged to the gym had higher GPAs than those who didn't.

James Pivarnik, the lead researcher of the study and a professor of kinesiology and epidemiology at MSU, explained students' cumulative GPAs were 0.13 points higher. Although the number may not appear significant, it could "mean the difference to those students on the cusp of getting into graduate school or even advancing to the next academic year."

Pivarnik found that students with a gym membership, in addition to having higher GPAs, had more credits completed by the end of their freshman year and stayed in school longer, helping to boost the school's retention rate. …

http://www.huffingtonpost.com/2014/07/11/gym-higher-gpa-msu-study_n_5575054.html

Recreational Sports Journal: The Official Journal of the NIRSA
RSJ Volume 38, Issue 1, April
Original Research

Academic Success and Retention: The Role of Recreational Sports Fitness Facilities 2014, 38, 14 – 22

http://dx.doi.org/10.1123/rsj.2013-0010

This study evaluated the role of a university recreational sports and fitness center, in students' academic success. Study participants included freshmen at a large Midwestern university (n = 4,843; 56% women; 67% white). Recreational sports fitness facility members (students who purchased a recreational sports fitness facilities membership in their first semester; n = 1,138) were compared with nonmembers (students who did not purchase a recreational sports fitness facility membership in their first semester; n = 3,705). $M \pm SD$ and percentages were calculated for all variables of interest. Differences between groups were analyzed using t tests and percentages. Members had significantly higher high school grade point averages (GPA) ($p = .002$). After four consecutive semesters, members had significantly higher cumulative college GPA ($p \leq .0001$) and cumulative credits completed ($p \leq .0001$). Significantly more members than nonmembers were enrolled in school after two completed years, 89% and 85%, respectively. Results show recreational sports fitness facility membership is associated with, and may be beneficial to, college students' academic success.

Authors: Samantha J. Danbert, James M. Pivarnik, Richard N. McNeil, Ira J. Washington

http://journals.humankinetics.com/rsj-back-issues/rsj-volume-38-issue-1-april/academic-success-and-retention-the-role-of-recreational-sports-fitness-facilities

ACADEMIC SUCCESS:
THE ROLE OF CAMPUS RECREATION FITNESS FACILITIES

Samanta J. Danbert, James M. Pivarnik, Richard N. McNeil, & Ira J. Washington
Recreation Sports and Fitness Services, Michigan State University

INTRODUCTION

Engaging in fitness activities makes an educational difference! Campus Recreation and Fitness Facilities (CRFF) give college students the means to participate in daily physical activity, which is important to current health and chronic disease prevention. However, as demonstrated in recent research jointly conducted between the Department of Recreational Sports and Fitness Services and the Department of Kinesiology, participating in recreational activities also is related to student success, specifically academic success (e.g., grade point average, credits completed, first to second and first to third year retention, class standing).

More than 75 group fitness classes are offered each week !

THE STUDENTS

4843 Michigan State first year students participated

- 56% women; 44% men
- 67% Caucasian; 33% Students of Color

Sample split into two groups:

- Purchased a CRFF membership (N=1138)
- Did not purchase a CRFF membership (N=3705)

More than 9,000 students purchase fitness memberships

RESULTS

	CRFF Members	CRFF Non-Members
Number	1138	3705
Gender (% Female)	41.2%	60.1%
High school GPA	3.55 (± .38) *	3.52 (± .30)
Pre-college GPA	2.99 (± .52) *	2.9 (± .36)
Collegiate GPA	3.13 (± .52) *	3.00 (± .59)
College credits completed	56.6 (± 8.9) *	54.1 (± 11.3)
1-year retention	91% *	88%
2-year retention	89% *	85%
Class standing	74% *	60%

* significant difference between groups ($p<0.002$)

WHAT WE LEARNED

- After four consecutive semesters, CRFF members completed more college credits than non-members.
- CRFF members had higher 1-year and 2-year retention rates than non-members.
- After two consecutive semesters, more CRFF members achieved sophomore status than non-members.
- CRFF members began college with a higher precollege GPA and maintained a higher collegiate GPA while in college.

Questions to answer:

1) Jessica Utts in her article "What Educated Citizens Should Know About Statistics and Probability"[1] writes,

   Probably the most common misinterpretation of statistical studies in the news is to conclude that when a relationship is statistically significant, a change in an explanatory variable is the *cause* of a change in the response variable. This conclusion is appropriate only under very restricted conditions, such as for large randomized experiments. For single observational studies, it is rarely appropriate to conclude that one variable caused a change in another. Therefore, it is important for students of statistics to understand the distinction between randomized experiments and observational studies, and to understand how the potential for confounding variables limits the conclusions that can be made from observational studies.

   a) How do the MSU flyer and the Huffington Post article illustrate the common misinterpretation that Utts describes? How does the abstract of the study published in the Recreational Sports Journal do a better job?

      Be specific. In your answer discuss the study's design and cite language from the MSU flyer and the Huffington Post article that is problematic given the study's design; also cite language from the RSJ abstract that is more accurate and explain why.

   b) In this study gym membership is the explanatory variable. College GPA is one of the response variables. However, there is one serious confounding variable that probably does a better job explaining the higher college GPA for gym members. Read carefully and identify this confounding variable. Explain why it is probably confounding the relationship between gym membership and college GPA.

2) The researchers conducted several t-tests.

   a) Are the conditions for use of a t-test met by this data? Explain.

   b) Why is it important to verify that conditions are met?

3) In the RSJ summary of the original research, it says, "After four consecutive semesters, members had significantly higher cumulative college GPA ($p \leq .0001$) and cumulative credits completed ($p \leq .0001$)."

---

[1] Utts, J. What Educated Citizens Should Know About Statistics and Probability. *The American Statistician, May 2003, Vol. 57, No. 2*

a) What does the term “significantly higher” mean to a statistician?

b) In the Huffington Post article, it says, “James Pivarnik, the lead researcher of the study and a professor of kinesiology and epidemiology at MSU, explained students' cumulative GPAs were 0.13 points higher.” In this study, why is such a small difference in GPAs statistically significant?

4) In the article “What Educated Citizens Should Know About Statistics and Probability”, Jessica Utts writes, “Students need to understand that a statistically significant finding may not have much *practical* importance. This is especially likely to be a problem when the sample size is large, so it is easy to reject the null hypothesis even if there is a very small effect.”

In the flyer the results from this study are being used to encourage students to buy a membership to the Recreational Sports and Fitness Center. Membership fees range from $70 to $260.
(Source: http://recsports.msu.edu/Fitness/Memberships.html)

a) How does Pivarnik, the study’s author, address the issue of practical importance?

b) Do you think the GPA difference found in this study is of practical importance? Why or why not?

Module 25.3 Unit 9 Project

Overview: This project is different from others. Instead of working in groups, we will start with individual work using StatCrunch, followed by speed-dating, and then you will have a short amount of time to revise your work. We will repeat this cycle 3 times with 3 separate problems. Your instructor may require you to turn in a written report.

Instructions:

1) Log into StatCrunch, go to Resources. Click on StatCrunchU and select a random sample of 100 students. You will use this sample for all 3 problems.
2) Open a Word document. Your instructor will project a problem onto the overhead screen. Type the research question into your Word document.
3) Conduct the appropriate T-procedure in StatCrunch and paste your table into the Word document.
4) Write a conclusion about StatCrunchU and answers the research question.

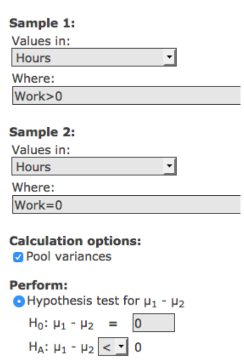
Warm-up: We will do a quick warm-up before we get started.

1) You and your classmate will have different random samples. When you discuss your work, which of the following should be the same?

>StatCrunch procedure	>Sample mean	>P-value
>Hypotheses	>Standard error	>L. and U. limits of confidence interval
>Degrees of freedom	>T-stat	>Conclusion

2) Do StatCrunchU students who work take fewer units than those without a job?

Choose Stat, T-Stats, Two-sample, With Data.
Values in: choose the response variable Hours
Where: Use the short-cut shown for creating the two comparison groups: those who work (Work>0) and those who do not work (Work=0)



Two sample T hypothesis test:
μ_1 : Mean of Hours where Work>0
μ_2 : Mean of Hours where Work=0
$\mu_1 - \mu_2$: Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 < 0$
(with pooled variances)

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	-2.0116877	0.62267192	98	-3.2307346	0.0008

Your classmate writes the following conclusion: "Fail to reject Ho in favor of Ha." What feedback do you have for him?

The Three Problems.

1. Do StatCrunchU students who work have greater loan debt than those who do not work?

2. According to Debt.org, college students have an average of $3,200 in credit card debt. Do StatCrunchU students have less credit card debt on average?

3. Who has more credit card debt on average, female or male StatCrunchU students? How much more?

UNIT 10

ANOVA

Module 26 ANOVA

Learning Goal:
Conduct a hypothesis test for differences in three or more population means using an ANOVA F-test.

Introduction

In this Unit we learn the ANOVA One Way F-test. It allows us to compare means from three or more populations. Alternatively, we can think of this test as examining the relationship between two variables. The explanatory variable is categorical and has 3 or more values. The response variable is quantitative.

For example, we could compare mean sleep hours for four populations: college freshmen, sophomores, juniors, and seniors. Here the explanatory variable is College Class; it has four values (freshmen, sophomore, etc.). The response variable is quantitative: sleep hours.

1) Which of the following scenarios are a candidate for use of the ANOVA?

- We compare student loan debt for male and female college students.
- We compare the proportion of college students receiving student loans based on their employment status: not employed, employed part-time, employed full-time.
- We compare student loan debt for college students based on their academic standing: satisfactory academic progress, academic warning, suspension, reinstatement.

Stating hypotheses

As before, the null hypothesis is a statement of "no difference" or "no relationship between explanatory and response variables" or "no treatment effect." (These all mean the same thing.). We can write the null hypothesis using symbols:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$

Unlike previous tests, where we had three options for the alternative hypothesis (less than, greater than or not equal to), in the ANOVA F-test the alternative hypothesis cannot specify the way in which the means differ. The alternative hypothesis just says, "not all of the population means are equal" or equivalently "at least one population mean differs from the others" or "there is no relationship between the explanatory and response variable" or in the case of an experiment "no treatment effect." (These all mean that same thing.)

2) It is not possible to write the ANOVA's alternative hypothesis concisely with symbols. Why not?

The F-statistic

We need a way to judge the variation in the sample means to determine if the differences are statistically significant.

- If we find that the sample means are not close together (or at least one deviates substantially from the others), we'll say that we have evidence against H_o and the population means are not equal.
- Otherwise, if the sample means are close together, we'll say that we do not have evidence against H_o , i.e. the differences in the sample means are not statistically significant and the sample means could come from populations with equal means.

The F-statistic measures the variation in sample means relative to the variation in the data within each sample. The F-statistic is very complicated and we will not compute it by hand, but, in essence, it is a ratio of variations (hence the name Analysis of Variance – ANOVA)

$$F = \frac{variation\ in\ sample\ means}{variation\ in\ each\ sample}$$

3) Do elementary school children carry too much weight in their backpacks?

   The explanatory variable is Grade (3^{rd}, 5^{th} or 7^{th}) and the response variable is percent of body weight carried in a school backpack. For example, if a child weighs 40 pounds and carries a 10-pound backpack, PercentWt = 25 because 10 ÷ 40 = 0.25.

   We use ANOVA to test the hypotheses:

   H_0: Mean percent of bodyweight carried in a school backpack is the same for the populations of students in the three grades. (3^{rd}, 5^{th} or 7^{th})

   H_a: At least one population has a mean percent that differs from the others.

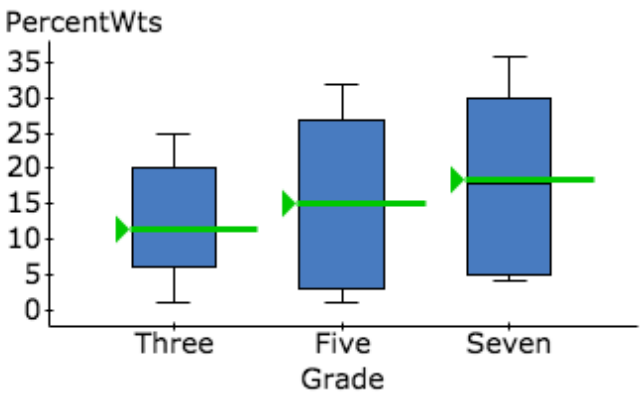
   a) Imagine that we run the ANOVA test as part of two different studies using data from random samples of 3^{rd}, 5^{th} and 7^{th} graders. Examine the boxplots and summary statistics for the hypothetical data from the two studies on the next page.

      For one of these studies F = 6.1 with a P-value of 0.0038; for the other study F = 2.3 with a P-value of 0.1098.

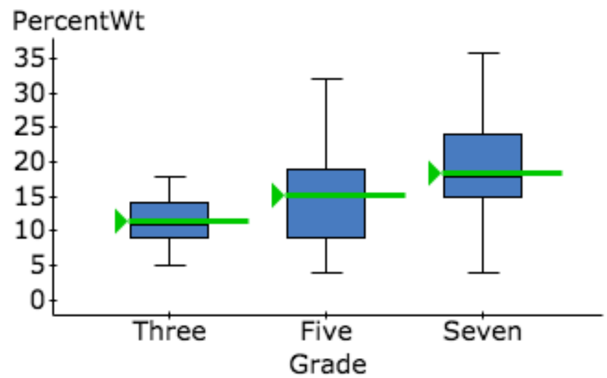
      Which is which? Why do you think so?

   b) Which study suggests that the samples come from populations with different means? Why do you think so?

STUDY #1



STUDY #2



Sample means are marked in each boxplot.

Grade	n	Mean	Std. Dev.
Three	21	11.4	7.7
Five	22	15.1	12.4
Seven	19	18.5	10.4

Grade	n	Mean	Std. Dev.
Three	21	11.4	3.7
Five	22	15.1	7.1
Seven	19	18.5	7.8

4) This ratio of variations is the idea behind the comparing more than two means, hence the name Analysis of Variance (ANOVA).

- When the variation within the sample data is large relative to the difference in the sample means, like in (*circle one: Study #1 or Study #2*), the data provide very little evidence against H_0 . In this case, the F-statistic is small and, as we have seen before, a small test statistic has a *(circle one: large or small)* P-value. We (*circle one: fail to reject the null hypothesis or reject the null hypothesis.*) There is *(circle one: weak or strong)* evidence that the population means differ.

- When the variation within the sample data is small relative to the differences in the sample means like in (*circle one: Study #1 or Study #2*), the data give stronger evidence against H_0. In this case, the F-statistic is large and, as we have seen before, a large test statistic has a *(circle one: large or small)* P-value. We (*circle one: fail to reject the null hypothesis or reject the null hypothesis.*) There is *(circle one: weak or strong)* evidence that the population means differ.

Conditions for use of the ANOVA F-test

- Samples are randomly selected from each population and are independent.
- The response variable varies normally within each of the populations. If we cannot verify the shape of the distribution of the response variable in the population, then we can still use the ANOVA in the following situations:
  - If the sample sizes are large (n > 30 for each sample), use the ANOVA regardless of the shape of the variable's distribution in the populations.
  - If a sample is small (n < 30), examine the distribution of the variable in that sample. If the distribution of the data in each small sample is not heavily skewed and without outliers, this suggests that the variable may be normally distributed in the associated population and we use the ANOVA test.
- This is the new condition: the populations all have the same standard deviation.
  - We will often have difficulty verifying that the population standard deviations are the same; therefore, we check this condition by examining the *sample* standard deviations to see if they are approximately equal. A common rule of thumb is the ratio between the largest sample standard deviation and the smallest is less than 2. If that's the case, we check off this condition as satisfied.

5) Are conditions met for use of the ANOVA F-test in our two fictitious studies of children's backpack weights? Explain.

6) When ANOVA F-test suggests that the population means differ, we can examine confidence intervals estimating each population mean to try to determine which population means account for the difference. These are the same one-sample T-intervals we learned in Unit 9.

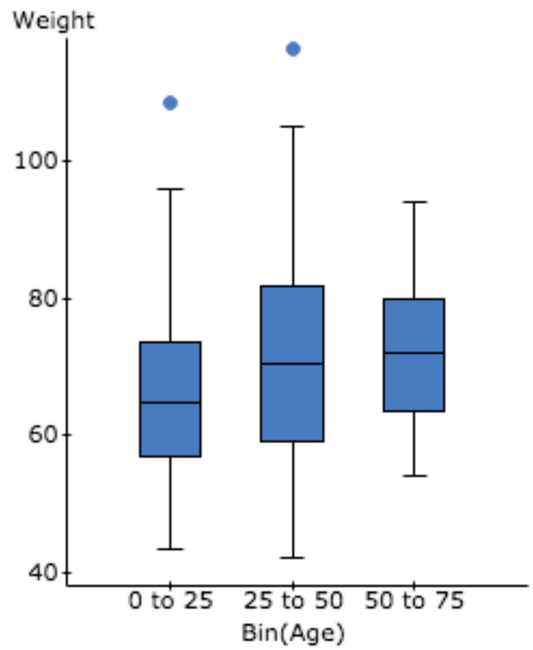
   The fictitious data from Study #2 give these 95% confidence intervals. Which population means appear to differ? Which might be the same?

Grade	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
Three	11.4	0.8	20	9.7	13.1
Five	15.1	1.5	21	11.9	18.3
Seven	18.5	1.8	18	14.8	22.3

(Note: Statisticians actually use more sophisticated statistical techniques called "multiple comparisons" to identify differences, but we will not study those techniques in this introductory statistics course.)

Group Work

1) In a nationwide study of the weights of adults with gym memberships, researchers obtain a random sample of gym members and group them into three age categories. They conduct an ANOVA test and provide the StatCrunch outputs shown here.



Analysis of Variance results:
Responses: Weight (kg)
Factors: Bin(Age)

Response statistics by factor

Bin(Age)	n	Mean	Std. Dev.	Std. Error
0 to 25	180	66.060556	12.163718	0.90663
25 to 50	304	70.750329	13.880549	0.79610408
50 to 75	23	72.121739	10.78846	2.2495494

ANOVA table

Source	DF	SS	MS	F-Stat	P-value
Bin(Age)	2	2699.7154	1349.8577	7.7819725	0.0005
Error	504	87423.629	173.45958		
Total	506	90123.344			

Which conclusions are appropriate at the 5% level of significance? Check all that apply.

- ☐ We cannot draw a conclusion because conditions for the ANOVA F-test are not met.
- ☐ The differences observed in sample mean weights do not provide strong evidence of a difference in population mean weights for the adult gym members in these three age groups.
- ☐ There are statistically significant differences in sample mean weights for the three age groups.
- ☐ This study provides strong evidence that the mean weights differ for the three populations of gym members defined by these age groups.
- ☐ This study provides strong evidence that the mean weight of all gym members who are under 25 is less than the mean weight for all members who are in the 50-75 age group.
- ☐ This study suggests that the mean weights are the same for the three populations of gym members defined by these three age categories.

2) A Statistics student designs a survey for her project. Here are her survey questions: (You do not need to answer the survey questions.)

   How do you get to LMC? (car, bike, bus, walk, other)
   What is your gender? (male, female)
   What is your employment status? (unemployed, employed part-time, employed full-time)
   How many units are you taking this semester?
   What was your high school GPA?
   What is your college GPA?
   How many hours of sleep do you get in a typical night?
   How old are you? (under 20, 20 up to 25, 25 up to 30, 30+)
   Do you have children? (yes, no)
   How many hours do you exercise in a typical week?

   a) From the list of survey questions identify an explanatory variable and a response variable that could be used for an ANOVA test.

   b) State your research question and the hypotheses you would test in the context of the variables you chose.

3) Conduct an ANOVA test using StatCrunch and data from StatCrunchU.

   Use the ANOVA F-test to compare freshmen, sophomores, juniors and seniors (Class) for a quantitative response variable of your choice. You can choose between credit hours, work hours, student loan debt or credit card debt.

   State the hypotheses, check conditions, give summary statistics (means, SDs, n's), give F and the P-value, state your conclusion in context.

   <u>Instructions for accessing the data</u>: Log into StatCrunch. Under **Resources**, under **Take a sample from StatCrunchU**, click on **StatCrunchU**. Scroll down underneath the survey to **Sample size** and select 200. Click Survey.

   You will see a spreadsheet with the survey responses for your random sample of 200 students. To understand the variable defining each column, refer to the survey questions above the spreadsheet.

   StatCrunch Instructions for the ANOVA F-test

   - With the data spreadsheet open, select **Stat**, **ANOVA**, **One Way**
   - Under **Compare,** select **Values in a single column.** For **Responses in** choose the column for response variable you chose. For **Factors in** choose the column with the explanatory variable (Class).
   - Press **Compute!**

UNIT 11

Chi-Square

Contents

Module 27 Chi-Square

Module 27.1 Chi-Square Test for Independence

Learning Goal: Conduct a chi-square test for independence between two categorical variables in a contingency table.

Introduction:

Previously, we learned the ANOVA test, which determines if two or more population means differ. You can also think of the ANOVA test as examining the relationship between two variables. The explanatory variable is categorical and has 3 or more values. The response variable is quantitative.

In this activity we work with categorical variables. We will now determine if there is a relationship between two categorical variables. Another way to say this is: Are the variables independent of each other or dependent on each other?

This new statistical test is called a *Chi-Square Test for Independence.*

In this Chi-Square test we select one sample and for each individual in the sample we collect two pieces of categorical data. The Chi-square test of independence determines whether explanatory variable impacts the distribution of values for the response variable.

Check your understanding:

1) For each research question, to identify the appropriate hypothesis test: z-test for one proportion, t-test for one mean, t-test for a difference in two means, ANOVA, Chi-square.

   a) Is there a relationship between amount of credit card debt and employment status?

   b) Is there a relationship between whether or not a student owns a credit card and employment status (unemployed, employed part-time, employed full-time)?

   c) Is gender associated with hybrid car ownership?

   d) Do the majority of community college students drive to their college?

   e) Do women drive more on average than men?

Stating hypotheses for Chi-Square Test for Independence

The null hypothesis can be stated in several equivalent ways:

- There is no relationship between the variables.
- There is no association between the variables.
- The variables are independent.

The alternative hypothesis says there is a relationship (association), which means the variables are dependent.

Check your understanding:

2) In a study of marketing to children, researchers examine cereal placement on grocery store shelves. The explanatory variable is cereal type (child vs. adult). The response variable is shelf placement (bottom shelf, middle shelf, top shelf)

   State the null and alternative hypotheses.

Analyzing the data for a Chi-square test

Here is the two-way table for the 77 cereals in this random sample of cereals. In a Chi-Square setting, we call the data observed counts.

	Bottom	Middle	Top	Total
Adult	11	5	36	52
Child	9	16	0	25
Total	20	21	36	77

What does independent mean? Two categorical variables are independent if the distribution of the response variable does not change when we take the explanatory variable into account. In other words, Shelf and Target are independent if shelf placement looks the same for child and adult cereals and matches the distribution of cereals across shelves when we ignore the cereal type.

Check your understanding:

3) Fill in the missing percentages to complete the distribution of shelf placement for adult cereals, child cereals, and all cereals together.

	Bottom	Middle	Top	Total
Adult	11	5	36 36/52=69%	52
Child	9 9/25=36%	16	0 0/25=0%	25
Total	20 20/77=26%	21	36	77

If the Target variable and Shelf variable are independent, then the distribution of shelf placement will be the same (or very close to the same) for Adult and Child cereals. But we see differences in these distributions. For example, there is a larger percentage of adult cereals on the top shelf. Therefore, we need to determine if these differences are small enough to attribute to the variation we expect to see in random sampling, or not. Surprise, surprise, we need a test statistic and a P-value.

The Chi-Square test statistic

The Chi-Square test statistic (written χ^2) measures how much the observed data in our sample differs from what we expect to happen when the null hypothesis is true.

Here is the formula:

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

We will not have you calculate the χ^2 test statistic by hand. However, we will spend a little time doing some calculations to help you understand the expected counts.

The expected counts are the counts we expect if the null hypothesis is true (the variables are not related; they are independent.)

If Shelf and Target are independent, then the placement distribution for Adult cereals and Child cereals will be the same as the placement distribution overall. Note that we are looking at the distribution of the response variable.

Previously, we found the overall placement distribution (bottom row of the previous table.) Record those percentages below:

Bottom shelf: 20/77 = 0.26 = 26%   Middle shelf: ________   Top shelf: __________

Let's start by finding the expected counts for ADULT cereals. We use these overall placement percentages to calculate the number of adult cereals we expect to see on each shelf:

- Bottom shelf: If the target consumer is not affecting the location (which is what the null hypothesis says), we expect 26% of 52 adult cereals to be on the bottom Therefore, the expected count is 0.26(52) = 13.52

Check your understanding:

4) Find the expected counts for adult cereals for the middle shelf and for the top shelf.

   a) Middle shelf: What is the expected count of adult cereals for the middle shelf?

   b) Top shelf: What is the expected count of adult cereals for the top shelf?

(To check your work, the expected counts for adult cereals across the shelves should add to 52.)

   c) Now find the expected counts for CHILD cereals:

Bottom: Middle: Top:

We calculated the expected counts by hand in this example to help you understand where the expected counts come from. But, in general, we will use StatCrunch to find expected counts.

Before we can use StatCrunch to find the χ^2 test statistic and the P-value, we need to make sure conditions are met for use of the X^2 density curve. Here are the conditions:

- There is a random sample from the population or random assignment of treatments.
- Each expected count is at least 5 (this is same idea as requiring success and failures to be at least 10 when we used the normal approximation for inference on proportions).

Check your understanding:

5) Verify that the conditions are met for use of the X^2 model for the cereal study.

Assessing the evidence

StatCrunch gives the results in an expanded two-way table. You can choose to show the expected counts or the distribution of the response variable (e.g. row percents if the explanatory variable defines the rows.)

When you have the data spreadsheet in StatCrunch, choose Stats, Tables, Contingency, with Data.

Contingency table results:
Rows: Target
Columns: Shelf

Cell format
Count (Expected count)

	bottom	middle	top	Total
adult	11 (13.51)	5 (14.18)	36 (24.31)	52
child	9 (6.49)	16 (6.82)	0 (11.69)	25
Total	20	21	36	77

Chi-Square test:

Statistic	DF	Value	P-value
Chi-square	2	37.049833	<0.0001

Check your understanding:

6) What is the X^2 statistic?

What is the P-value?

7) State a conclusion.

Group Work.

8) For movies, is there a relationship between studio type (Big 6, Other) and the movie genre (Action/Adventure, Other)? Or are they independent?

To investigate this question, students want to use the movie data that we analyzed earlier in the course. The movies in the data set are the top 75 USA box office sales earners of all time. Data was taken from IMDb.com.

Contingency table results:
Rows: Studio
Columns: Genre

	Action/Adventure	Other	Total
Big6	25	15	40
Other	18	17	35
Total	43	32	75

Can they use this data to conduct a chi-square test for independence? Why or why not?

9) We will use a study conducted during the 1980's to practice the chi-square test of independence.

   Background: Clinical depression is a recurrent illness requiring treatment and often hospitalization. Nearly 50% of people who have an episode of major depression will have a recurrence within 2-3 years.

   The Study: During the 1980's the federal government, through the National Institutes of Health (NIH), sponsored a multi-centered, randomized, controlled, clinical trial to evaluate two drugs to prevent the recurrence of depression in patients who have had at least one previous episode of the illness (Prien et al., *Archives of General Psychiatry*, 1984).

   The Study Design: Patients suffering from depression were recruited from 5 medical clinics in 5 large cities. They were randomly assigned to one of the 3 treatment groups: Imipramine, Lithium, or a Placebo. Patients were followed for 2-4 years to see whether or not they had a recurrence of depression. If they did not have a recurrence within this time frame, then their treatment was considered a success. The study was double-blinded.

   a) Select all the appropriate pairs of hypotheses to investigate the association between treatment and recurrence of depression.

H_0: There is a relationship between treatment and recurrence of depression.
H_a: There is no relationship between treatment and recurrence of depression.

H_0: Recurrence of depression is independent of treatment
H_a: Treatment and recurrence are dependent

H_0: There is no association between treatment and recurrence of depression.
H_a: There is an association between treatment and recurrence of depression.

b) In the StatCrunch print-out, what does the 18.13 tell us?

Contingency table results:
Rows: Treat
Columns: Outcome

Cell format
Count (Expected count)

	No Recurrence	Recurrence	Total
Imipramine	27 (18.13)	11 (19.87)	38
Lithium	14 (17.65)	23 (19.35)	37
Placebo	11 (16.22)	23 (17.78)	34
Total	52	57	109

Chi-Square test:

Statistic	DF	Value	P-value
Chi-square	2	12.959236	0.0015

c) Which of the following statements explains the expected counts? (Choose all of the statements that are correct)

- the number of patients expected to relapse or not if all 3 treatments are similarly effective
- the number of patients expected to to relapse or not if there is a relationship between treatment type and recurrence
- the number of patients expected to relapse or not if there is a no association between treatment type and recurrence
- the expected number to relapse or not if the null hypothesis of independence is true.

d) The P-value is 0.0015. Choose all interpretations that are correct.

a. The P-value is statistically significant at a significance level of 1%.
b. The P-value is not statistically significant at a significance level of 1%.
c. If outcome is independent of treatment type, we can expect a chi-square value of 12.96 or greater to occur less than 1% of the time.
d. The results are so rare we conclude that these results are due to fluctuations expected when randomly assigning subjects to treatments when the treatments have no association with the patient's outcome.
e. We can expect a chi-square value of 12.96 or greater about 0.15% of the time if there is no relationship between treatment type and patient outcome.

e) What can we conclude?

10) Is there a relationship between political affiliation and willingness to participate in political surveys?

	Survey Participation Yes	Survey Participation No
Democrat	49	47
Independent	15	27
Republican	32	30
None or will not say	8	10

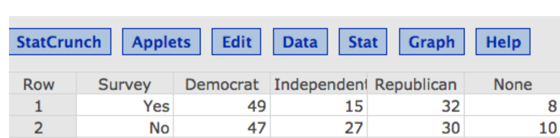
a) State your hypotheses.

b) Use StatCrunch to find expected counts and to test for independence (See instructions below.) Fill in the expected counts in the table.

c) Explain why we can use the X^2 model.

d) Give the P-value and state your conclusion.

StatCrunch instructions:
Open StatCrunch and enter the data as shown.
Choose **Stat, Tables, Contingency, With Summary**



Row	Survey	Democrat	Independent	Republican	None
1	Yes	49	15	32	8
2	No	47	27	30	10

Select Columns: select the response categories (Democrat, Independent, Republican, None)
Row labels: select the explanatory variable (Survey)
Display: choose expected count (if you want to see this)
Hypothesis tests: choose chi-square test for independence

Module 27.2 Unit 11 Lab

Instructions: For each question, submit a file or hardcopy (as directed by your instructor) containing the following…

- Your name.
- A statement of the research question.
- Confirmation that the conditions for the chi-square test are satisfied (this will be easier if you first run the StatCrunch analysis).
- A screenshot of your StatCrunch results, that must contain both the observed and expected counts, and the P-value for the test.
- A hypothesis test in our four-step format that addresses the question. Your conclusion must include an interpretation of the P-value.
- A sentence interpreting the conclusion in terms a member of the public could understand.

1) Is there an association between taking prenatal vitamins and autism? A New York Times article in 2011 reported on a study that compared mothers of autistic children with those whose children showed typical development. The table shows the number of mothers in each group who did and did not use prenatal vitamins during the three months before pregnancy.
Is there an association?

	Autism	Typical Development
Vitamin	111	70
No Vitamin	143	159

Module 27.3 Unit 11 Project

Instructions: This is the last project before your final project is due. This is one last chance to practice making a great project.

In this project, you will perform a chi-square test for independence using StatCrunch on a sample of students from StatCrunchU.

Spend 5 minutes analyzing the question, and preparing a StatCrunch analysis. Then you will speed date about your analysis two times, record the suggestions from the speed dating and rewrite your analysis. Your instructor will then call on individuals to present their final analyses to the class.

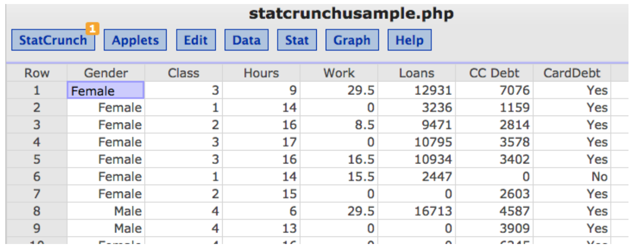
Your analysis should include the following…

- Your name.
- A statement of the research question.
- Confirmation that the conditions for the chi-square test are satisfied (this will be easier if you first run the StatCrunch analysis).
- A screenshot of your StatCrunch results that contains both the observed and expected counts, and the P-value for the test.
- A hypothesis test in our four-step format that addresses the question. Your conclusion must include an interpretation of the P-value.
- A sentence interpreting the conclusion in terms a member of the public could understand.
- A summary of the feedback you received during speed-dating.

Project Question

For students, is having credit card debt related to having student loan debt, or are they independent? Define the necessary categorical variables using ifelse statements and conduct a chi-square test on a sample of 500 StatCrunch U students.

StatCrunch for Contingency Tables with Details: An example.
Is carrying a student load independent of academic class? We run a chi-square test on a sample of 500 StatCrunchU students, and add a categorical data column for credit card debt.

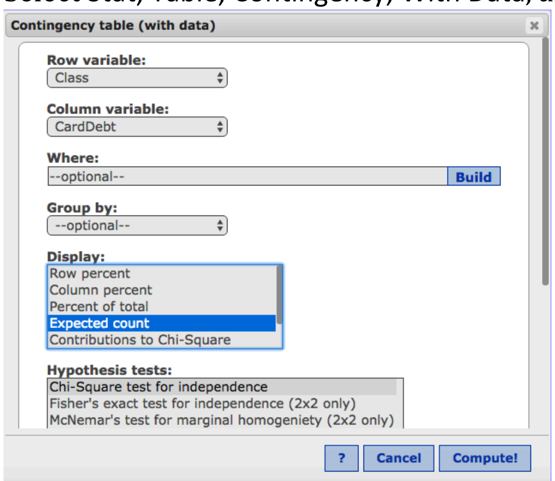


statcrunchusample.php

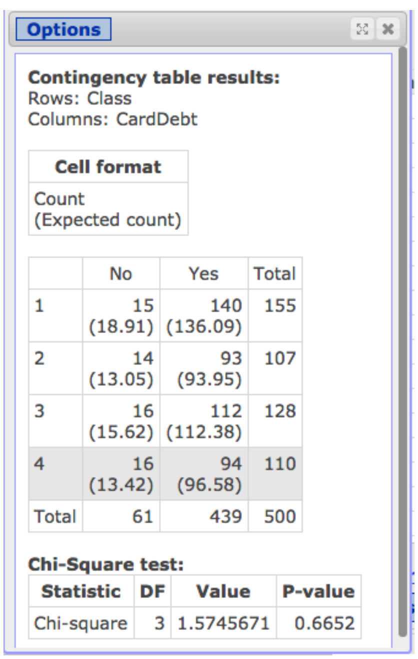
StatCrunch | Applets | Edit | Data | Stat | Graph | Help

Row	Gender	Class	Hours	Work	Loans	CC Debt	CardDebt
1	Female	3	9	29.5	12931	7076	Yes
2	Female	1	14	0	3236	1159	Yes
3	Female	2	16	8.5	9471	2814	Yes
4	Female	3	17	0	10795	3578	Yes
5	Female	3	16	16.5	10934	3402	Yes
6	Female	1	14	15.5	2447	0	No
7	Female	2	15	0	0	2603	Yes
8	Male	4	6	29.5	16713	4587	Yes
9	Male	4	13	0	0	3909	Yes

Select Stat, Table, Contingency, With Data, and complete the form as shown below.



Click Compute!



Options

Contingency table results:
Rows: Class
Columns: CardDebt

Cell format
Count (Expected count)

	No	Yes	Total
1	15 (18.91)	140 (136.09)	155
2	14 (13.05)	93 (93.95)	107
3	16 (15.62)	112 (112.38)	128
4	16 (13.42)	94 (96.58)	110
Total	61	439	500

Chi-Square test:

Statistic	DF	Value	P-value
Chi-square	3	1.5745671	0.6652

UNIT 12

Vocabulary

Module 28 Vocabulary

Distributions for Quantitative Data: Vocabulary

Individuals	Individuals are the objects described by a set of data. Individuals may be people, but they may also be animals or things.
Variable	A statistical variable is a characteristic that describes individuals in the data. A variable can take different values for different individuals.
Categorical variable	A categorical variable places an individual into one of several groups or categories. Example: a person's college major.
Quantitative variable	A quantitative variable takes a number value, which you can add or average. Example: number of people in household.
Distribution of a variable	A statistical distribution is an arrangement of the values of a variable (often a graph) showing their observed frequency of occurrence. Here is an other definition: "a representation that shows the possible values of a variable and how often the variable takes those values."
Describing the distribution of a quantitative variable	To describe a distribution, describe the shape, center, spread and outliers. **Shape:** describe the overall trends in the data. Typical ways to describe shape include symmetric, left or right skew, uniform. **Center:** give a single number that represents the data; a typical value or average (We will make this more precise in Topic 2.2) **Spread:** give a single number that measures how much the data varies. Range (maximum value minus minimum value) is one way to measure spread. (We will learn other ways to measure spread in Topic 2.3 and 2.4) **Outliers:** unusual data values.
Shape	**Symmetric:** The right and left sides of the graph are mirror images of each other **Skewed to the right:** The graph has more spread in the upper half than the lower half. There may be outliers to the right, so the graph has a longer tail to the right. **Skewed to the left:** The graph has more spread in the lower half than the upper half. There may be outliers to the left, so the graph has a longer tail to the left. **Uniform:** The graph is shaped like a rectangle. Each value occurs with the same frequency.

Experiments: Vocabulary

Population	The population is the entire group of individuals or objects that we want to study. Usually, it is not possible to study the whole population, so we collect data from a part of the population, called a *sample.*
Sample	A sample is a part of the population from which we actually collect information or data. In college-level Statistics we will use a sample to draw conclusions about the entire population.
Individual	The individuals are the members of the population, so a sample contains some of the individuals. Individuals may be people, but they may also be animals or things. Individuals can also be participants in an experiment.
Variable	A variable is the information we collect from individuals in the study. A variable can be a characteristic of an individual (such as gender) or a measurement (such as height).
Observational Study	An observational study collects information about individuals without intentionally manipulating variables. The main purpose of an observational study is to describe a group of individuals or to investigate an association between two variables. All of the research questions about a population in Activity 1.1 Exercise 1 could be investigated with an observational study.
Experiment	An experiment intentionally manipulates one variable in an attempt to cause an effect on another variable. The primary goal of an experiment is to provide evidence for a cause-and-effect relationship between two variables.
Response variable	The response variable measures the outcome of a study.
Explanatory variable	The explanatory variable may explain or even cause changes in the response variable.
Confounding variables	A confounding variable (also known as a lurking variable) is not defined as an explanatory variable in the study, but it may influence the response variable. Confounding variables can cloud or confuse the link between explanatory and response variables.

Experiments: Additional Vocabulary

Experiment	An experiment intentionally manipulates one variable in an attempt to cause an effect on another variable. The primary goal of an experiment is to provide evidence for a cause and effect relationship between two variables.
Explanatory variable	If we think a variable may explain or even cause changes, we call it an explanatory variable. In an experiment the explanatory variable is also called the treatment variable because the researcher manipulates the explanatory variable by assigning subjects to different treatments.
Response variable	The response variable is the outcome we measure for each individual.
Confounding variable	A confounding variable is a variable that is not measured in the study. It has an influence on the response variable, and its effects cannot be separated from the explanatory variable. It is related to both the explanatory and response variables and may explain the apparent association between the two. So it confounds our ability to draw a cause+ and+effect relationship between explanatory and response variables.
Direct Control	When confounding variables are directly addressed up front in the experiment design, this is called direct control.
Random Assignment	In an experiment when subjects are randomly assigned to treatments, the goal is to create treatment groups that are similar. This equalizes the effects of confounding variables across the treatment groups so that any differences in the response are not due to confounding variables.
Control Group	In an experiment a control group is the group that receives no treatment or a placebo. The control group provides a baseline for comparison.
Placebo	A placebo is a treatment with no active ingredients, such as a sugar pill.
Placebo effect	When placebo treatments produce positive improvements, we call this the placebo effect. Placebo effects arise from the power of the body to heal itself when the subject receives a placebo but believes that they are receiving treatment.
Blinding	When the researcher does not know which subjects received which treatment in an experiment, we say the researcher is "blind." Blinding the researcher controls the researcher's intentional or unintentional bias. The subjects can also be blind when they do not know which treatment they receive. Double blinding is when both researcher and subjects are blind.

Inference: Vocabulary

Parameter	A number that describes an entire **population**. Example of a parameter: At LMC 70% of all students are eligible for financial aid.
Statistic	A number that we calculate from a **sample**. Example of statistics: In a randomly selected Math 34 class, 62% of students are eligible for financial aid, but in a randomly selected English 100 class, 74% of students are eligible for financial aid.

Notation for parameters and statistics:

	(Population) Parameter	**(Sample) Statistic**
Proportion (e.g. probability) *Categorical data*	p	$\hat{p}$
Mean *Quantitative data*	μ	$\bar{x}$
Standard Deviation	σ	s

Sampling distribution	A mathematical model that describes the sample proportions from all possible random samples of size n from the population.
Variability in samples	When we make a calculation from a sample, the proportions, percents and means will differ from parameters.
Law of large numbers	When we have a large sample, the statistics are closer to the actual parameters. In general, the larger the sample, the closer to the actual parameters. A sampling distribution for large samples has less variability.
Patterns in large numbers of repetitions	There is less variability in a large number of repetitions. This means that in the long run, we will see a pattern, so we are more confident about estimating the probability of an event using empirical probability with a large number of repetitions.
Sample size n	We use the letter n to represent how many people, animals or objects are in the sample.
Center	We use the mean as the measure of center in a Normal distribution graph (bell curve). The mean of the sample proportions is the population proportion.

Standard Deviation SD of population proportion (also Standard Error of Sample Proportion)	The standard deviation tells us how much the graph is stretched or compressed. The formula for the standard deviation of the sample proportions is: $$\sqrt{\frac{p(1-p)}{n}}$$
Estimated Standard Error SE of sample proportion	When the population proportion is unknown, we estimate it using p. This is known as the standard error of the sample proportion. The formula is: $$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
Normal Approximation	If samples are repeatedly drawn from a population, the distribution of p will be approximately Normal if the number of "success" and "failure" responses is at least 10. We calculate: $np \geq 10$ and $n(1-p) \geq 10$
Success	Note that a *success* is the category of interest. It is what we are counting. For example, if we are counting blue M&Ms, a success is a blue M&M. A red M&M would be considered a failure.
Expected outcome of success	We multiply the number of the sample size times the probability of success. The formula is: $Expected\ success = np$
Expected outcome of failure	We multiple the number of the sample size times the probability of failure, 1-p. The formula is: $Expected\ failure = n(1-p)$
Z-score	The z-score measures how many standard deviations a particular statistic is away from the population proportion in a Normal distribution.
Z-score formula	$$z = \frac{statistic - parameter}{standard\ deviation} = \frac{\hat{p} - p}{SD}$$
Margin of Error	The error distance from the center of confidence interval. The z-score multiplied by the standard deviation (or standard error).
Confidence Interval	$sample\ statistic\ \pm margin\ of\ error$