Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

0. Submit HW1 immediately (by Wed Jun29, 9am) to
STATGRADER@STAT.UCLA.EDU by email.
1. 2-sided tests.
2. Normal distribution, CLT, and Halloween candy example.
3. Validity conditions for testing proportions.
4. Reject the null vs. accept the alternative, wealth and echinacea examples.
5. Sampling, bias, and students example.
6. Estimating the mean, and guessing elapsed time example.

Read chapters 2 and 3.
http://www.stat.ucla.edu/~frederic/13/sum22 .
HW2 is due Wed July 6 at 9am and is problems 2.3.15, 3.3.18, and 4.1.23.
2.3.15 starts "Consider a manufacturing process that is producing hypodermic needles that will be used for blood donations...." and
3.3.18 starts "Reconsider the investigation of the manufacturing process that is producing hypodermic needles. Using the data from the most recent sample of needles, a 90% confidence interval for the average diameter of needles is...."
4.1.23 starts "In November 2010, an article titled 'Frequency of Cold Dramatically Cut with Regular Exercise' appeared in *Medical News Today*."

HW2 is due Wed July 6 at 9am and is problems 2.3.15, 3.3.18, and 4.1.23.
2.3.15 starts "Consider a manufacturing process that is producing hypodermic needles that will be used for blood donations. These needles need to have a diameter of 1.65mm – too big and they would hurt the donor (even

more than usual), too small and they would rupture the red blood cells, rendering the donated blood useless. Thus, the manufacturing process would have to be closely monitored to detect any significant departures from the desired diameter. During every shift, quality control personnel take a sample of several needles and measure their diameters. If they discover a problem, they will stop the manufacturing process until it is corrected.

a. Define the parameter of interest in the context of this study and assign an appropriate symbol to it.

b. State the appropriate null and alternative hypotheses using the symbol defined in (a).

c. Describe what a Type I error would be in this study. Also, describe the consequence of such an error in the context of this study.

d. Describe what a Type II error would be in this study. Also, describe the consequence of such an error in the context of this study.

3.3.18 starts "Reconsider the investigation of the manufacturing process that is producing hypodermic needles. Using the data from the most recent sample of needles, a 90% confidence interval for the average diameter of needles is...."
4.1.23 starts "In November 2010, an article titled 'Frequency of Cold Dramatically Cut with Regular Exercise' appeared in *Medical News Today*."

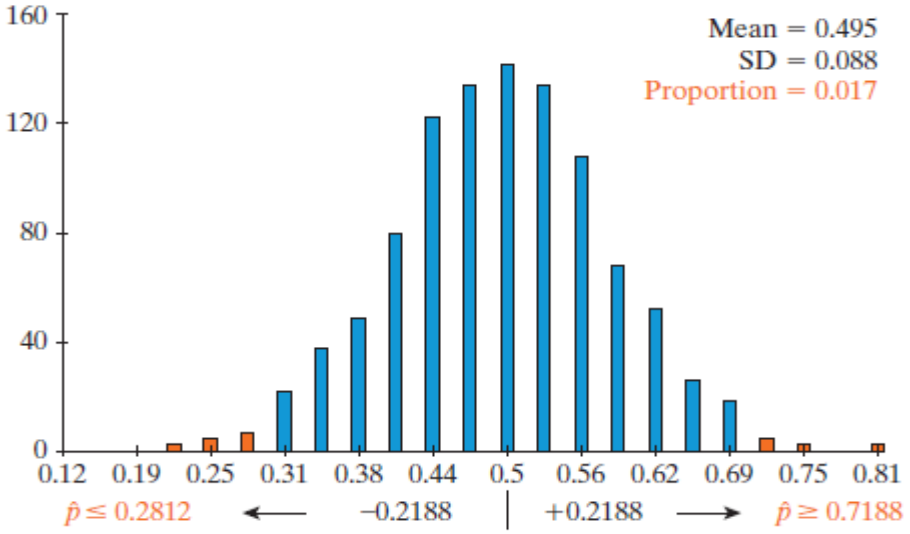
0. Submit HW1 immediately (by Wed Jun29, 9am) to STATGRADER@STAT.UCLA.EDU by email.

1. Two-Sided Tests

- The change to the alternative hypothesis affects how we compute the p-value.
- Remember that the p-value is the probability (assuming the null hypothesis is true) of obtaining a proportion that is equal to or **more extreme** than the observed statistic
- In a *two-sided test*, **more extreme** goes in both directions.

Two-Sided Tests

- Continuing with the example of predicting elections based on faces, since our sample proportion was 0.7188 and 0.7188 is 0.2188 *above* 0.5, we also need to look at 0.2188 *below* 0.5.
- The p-value will include all simulated proportions 0.7188 and above as well as those 0.2812 and below.



Two-Sided Tests

- 0.7188 or greater was obtained 9 times
- 0.2812 or less was obtained 8 times
- The p-value is (8 + 9 = 17)/1000 = 0.017.
- Two-sided tests increase the p-value (it about doubles) and hence decrease the strength of evidence.
- Two-sided tests are said to be more conservative. More evidence is needed to reject the null hypothesis.

Predicting House Elections

- Researchers also predicted the 279 races for the House of Representatives in 2004.
- They correctly predicted the winner in 189/279 ≈ 0.677, or 67.7% of the races.
- The House's sample percentage (67.7%) is a bit smaller than the Senate (71.9%), but the sample size is larger (279) than for the senate races (32).
- Do you expect the strength of evidence to be stronger, weaker, or essentially the same for the House compared to the Senate?

Predicting House Elections

Distance of the observed statistic to the null hypothesis value

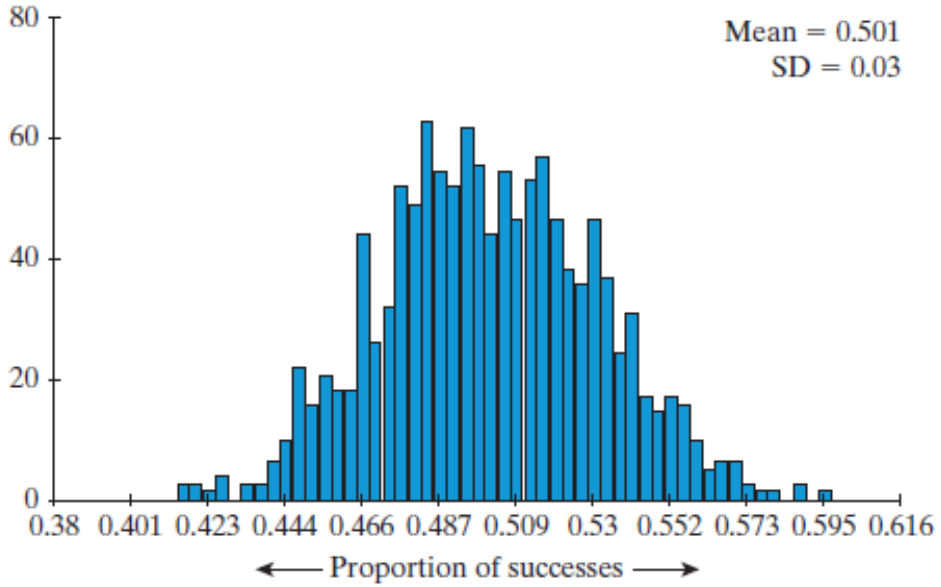
- The statistic in the House is 0.677 compared to 0.719 in the Senate
- Slight decrease in the effect size.

Sample size

- The sample size is almost 10 times as large (279 vs. 32)
- This will increase the strength of evidence.

Predicting House Elections

Null distribution of 279 sample House races



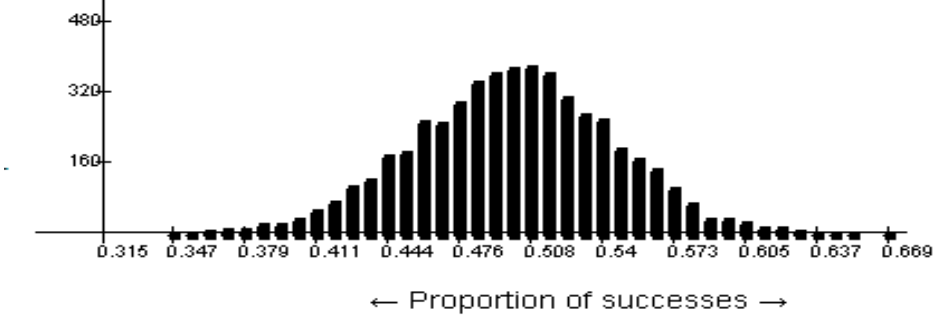
Simulated statistics ≥0.677 didn't occur at all so the p-value is 0

Predicting House Elections

- What about the standardized statistics?
  - For the Senate it was 2.43
  - For the House is 5.90.
- The larger sample size for the House outweighed the smaller effect size in this particular case. We have stronger evidence against the null using the data from the House.

2. Normal distribution, CLT, and halloween candy example.

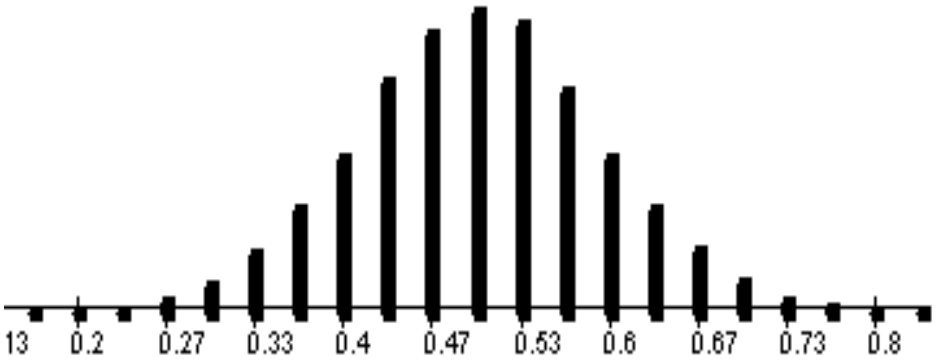
Section 1.5

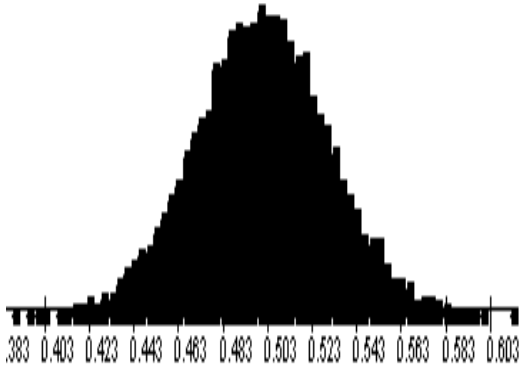


- The shape of most of our simulated null distributions always seems to be bell shaped. This shape is called the normal distribution.
- The Central Limit Theorem (CLT) dictates that, as *n* gets large, the sample mean or proportion becomes approximately normally distributed.
- When we do a test of significance using theory-based methods, only how our p-values are found will change. Everything else will stay the same.

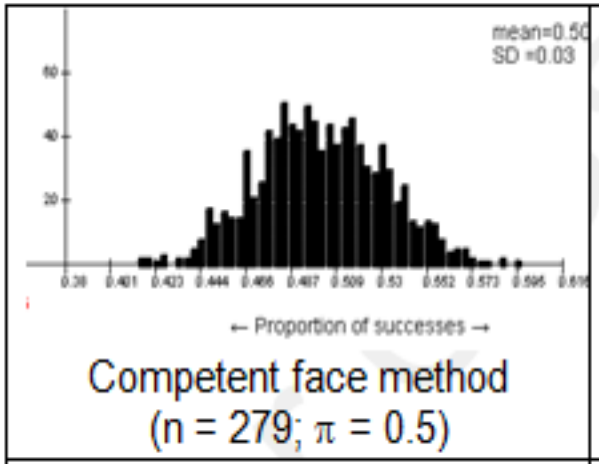
The Normal Distribution

- Both of these are centered at 0.5.
  - The one on the left represents samples of size 30.
  - The one on the right represents samples of size 300.
  - Both could be described as normal distributions.

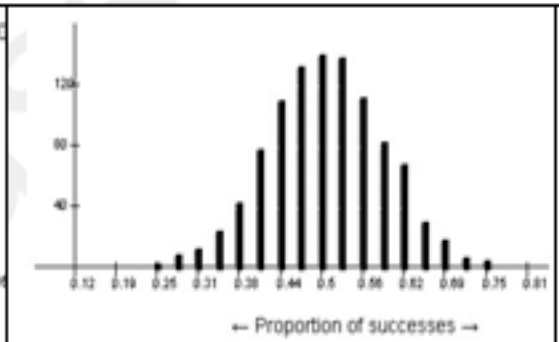




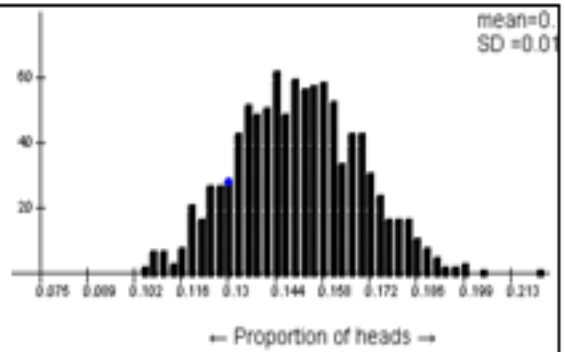
- Which ones will normal distributions fit?



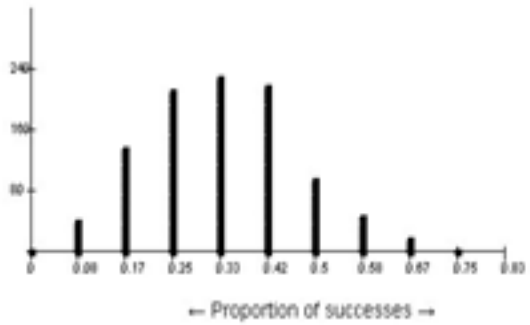
Competent face method
(n = 279; π = 0.5)



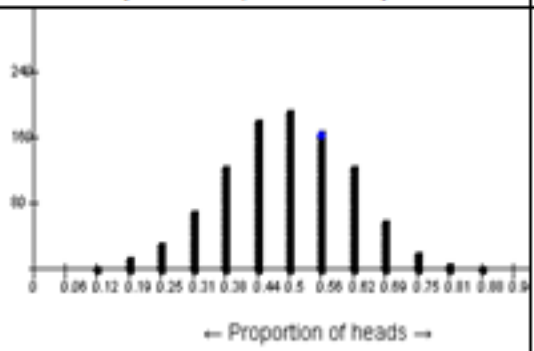
Competent face method
(n = 32; π = 0.5)



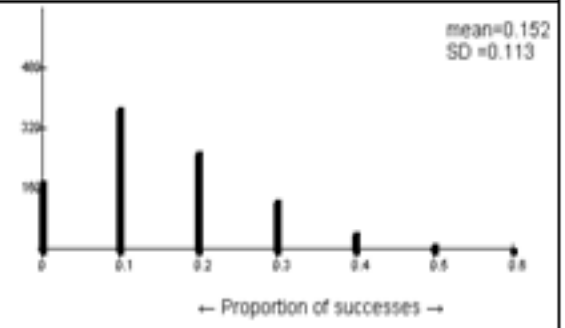
St. George's
(n = 361; π = 0.15)



Rock Paper Scissors
(n = 12; π = 0.3333)



Doris/Buzz
(n = 16; π = 0.5)



St. George's
(n = 10, π = 0.15)

When can I use a theory-based test that uses the normal distribution?

- The shape of the randomized null distribution is affected by the sample size and the proportion under the null hypothesis.
- The larger the sample size the better.
- The closer the null proportion is to 0.5 the better.
- For testing proportions, you should have at least 10 successes and 10 failures in your sample to be confident that a normal distribution will fit the simulated null distribution nicely.

Advantages and Disadvantages of Theory-Based Tests

- **Advantages of theory-based tests**
  - No need to set up some randomization method
  - Fast and Easy
  - Can be done with a wide variety of software
  - We all get the same p-value.
  - Determining confidence intervals (we will do this in chapter 3) is much easier.
- **Disadvantages of theory-based tests**
  - They all come with some validity conditions (like the number of success and failures we have for a single proportion test).

Example 1.5: Halloween Treats

- Researchers investigated whether children show a preference to toys or candy
- Test households in five Connecticut neighborhoods offered children two plates:
  - One with candy
  - One with small, inexpensive toys
- The researchers observed the selections of 283 trick-or-treaters between ages 3 and 14.

Halloween Treats

- Null: The proportion of trick-or-treaters who choose candy is 0.5.
- Alternative: The proportion of trick-or-treaters who choose candy is not 0.5.

- H_0: $\pi = 0.5$
- H_a: $\pi \neq 0.5$

- 283 children were observed
  - 148 (52.3%) chose candy
  - 135 (47.7%) chose toys

Standard Deviation of $\hat{p}$

- Under the null distribution, the standard deviation of $\hat{p}$ is $\sqrt{\pi(1-\pi)/n}$ where π is the proportion under the null and n is the sample size.
- $\sqrt{\frac{0.5(1-0.5)}{283}} = 0.0297.$

Theory-Based Inference

- The theory-based standard error works if we have a large enough sample size.
- We have 148 successes and 135 failures. Is the sample size large enough to use the theory-based method?

Standardized Statistic

- $\frac{0.523-0.5}{.0297}$ = 0.774.
- This is our Z-statistic, meaning the sample proportion is 0.774 SEs above the mean.
- Remember that a standardized statistic of more than 2 indicates that the sample result is far enough from the hypothesized value to be unlikely if the null were true.
- We had a standardized statistic that was not more than 2 (or even 1) so we don't really have strong evidence against the null.

Halloween Treats

- To compute the p-value in *R*,
  2*(1-pnorm(.774)) ~ 0.439.
- The theory-based p-value is 0.439 so if half of the population of trick-or-treaters preferred candy, then there's a **43.9%** chance that a random sample of 283 trick-or-treaters would have 148 or more, or 135 or fewer, candy choosers.
- Since 43.9% is not a small p-value, we don't have strong (or even moderate) evidence that trick-or-treaters prefer one type of treat over the other. We cannot reject the null hypothesis.

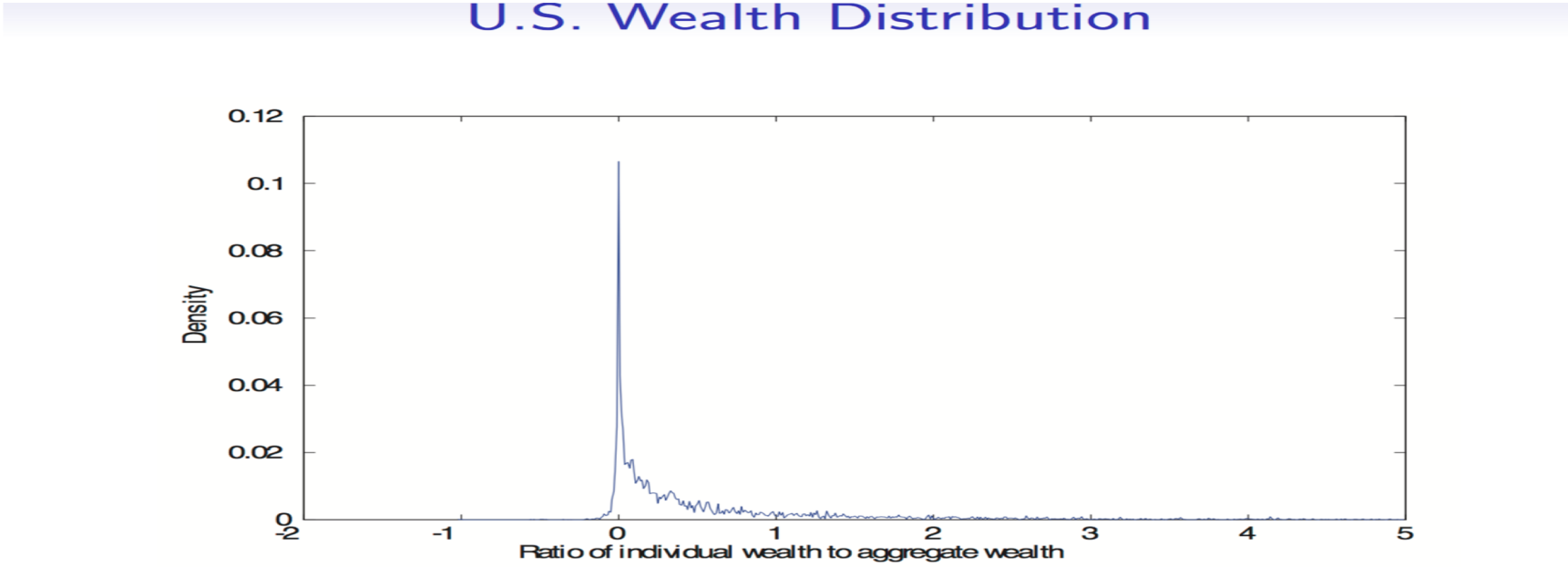
3. Validity conditions for testing proportions.

- You should have at least 10 successes and 10 failures in your sample to be confident a normal distribution will fit the simulated null distribution nicely.
- Your observations should be (at least approximately) independent. We will discuss what this means later in this lecture when we talk about sampling.

4. Rejecting the null vs. accepting the alternative.

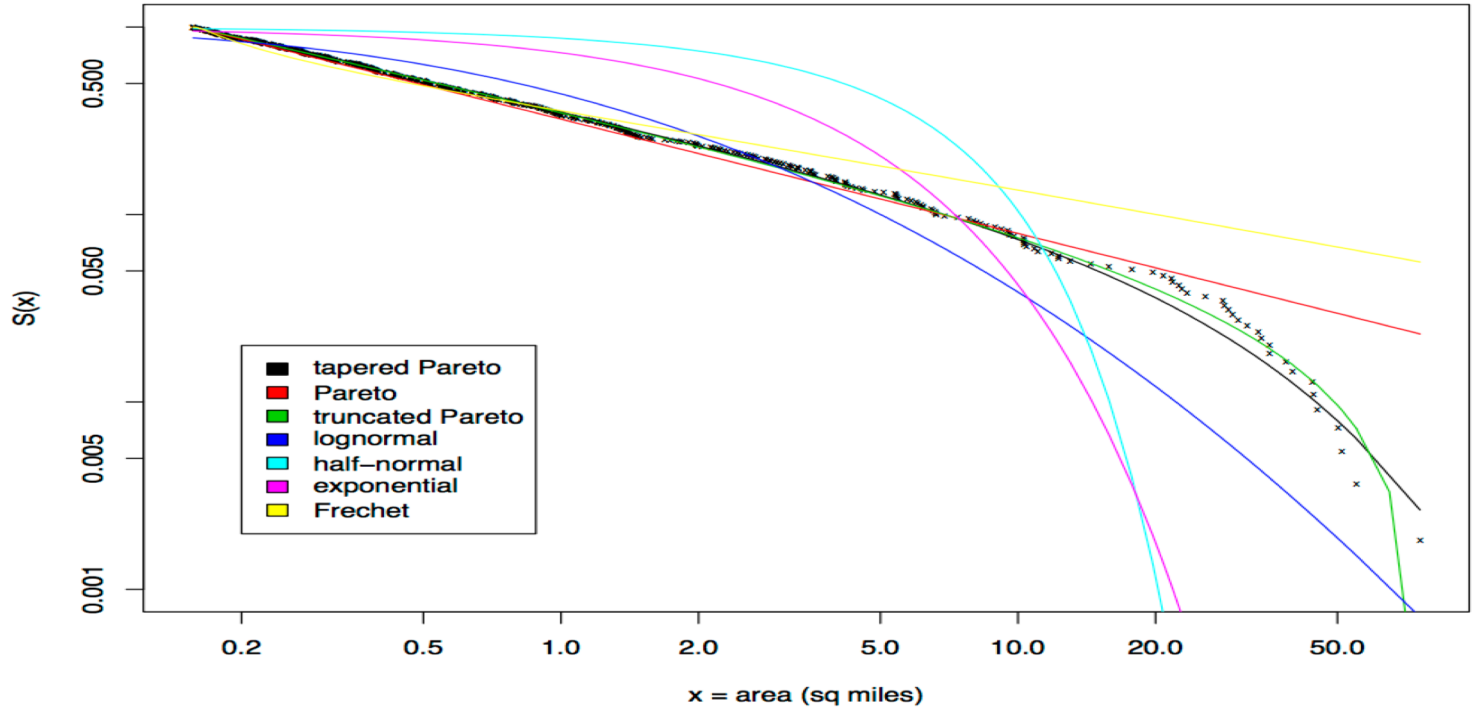
– Benoit Mandelbrot.
  We've tested it on many datasets and found the Pareto distribution "fits perfectly".

– from B. Moll (2012).



4. Rejecting the null vs. accepting the alternative.

– Benoit Mandelbrot.
We've tested it on many datasets and found the Pareto distribution "fits perfectly".



4. Rejecting the null vs. accepting the alternative.

- Benoit Mandelbrot.
  We've tested it on many datasets and found the Pareto distribution "fits perfectly".
- Think about it. What is the null hypothesis of the test. Is it possible to show that the model fits perfectly?
- You might not reject the null with a certain n, and then as n grows, you reject it.
- Nowadays people are using the tapered Pareto distribution instead of the Pareto.
- Echinacea vs. placebo. n = 58. Oneil et al. 2008.

4. Rejecting the null vs. accepting the alternative.

- 28 in echinacea group and 30 in placebo group.
- "[V]olunteers recruited from hospital personnel were randomly assigned to receive 3 capsules twice daily of either placebo (parsley) or E. purpurea [echinacea] for 8 weeks during the winter months. Upper respiratory tract symptoms were reported weekly during this period.
- "Individuals in the echinacea group reported 9 sick days per person during the 8-week period, whereas the placebo group reported 14 sick days ($z = -0.42$; $P = .67$)."

4. Rejecting Ho vs. accepting Ha.

- conclusion in Oneil et al. (2008), "commercially available E. purpurea capsules did not significantly alter the frequency of upper respiratory tract symptoms compared with placebo use."
- From sciencebasedmedicine.org, "[The study] added to the evidence that *Echinacea* is not useful for prevention of colds or flus. They found no difference in incidence of cold symptoms."
- ABC News headline "Study: Echinacea no help for colds".

4. Rejecting Ho vs. accepting Ha.

4. Rejecting Ho vs. accepting Ha.

Today, most of the evidence seems to indicate that echinacea does boost the immune system a little bit and help to fight colds. From WebMD: "Extracts of echinacea do seem to have an effect on the immune system, your body's defense against germs. Research shows it increases the number of white blood cells, which fight infections. A review of more than a dozen studies, published in 2014, found the herbal remedy had a very slight benefit in preventing colds."

5. Sampling Students

Example 2.1A

Sampling Students

- We will look at data collected from the registrar's office from the College of the Midwest for ALL students for Spring 2011

Student ID	Cumulative GPA	On campus?
1	3.92	Yes
2	2.80	Yes
3	3.08	Yes
4	2.71	No
5	3.31	Yes
6	3.83	Yes
7	3.80	No
8	3.58	Yes
…	…	…

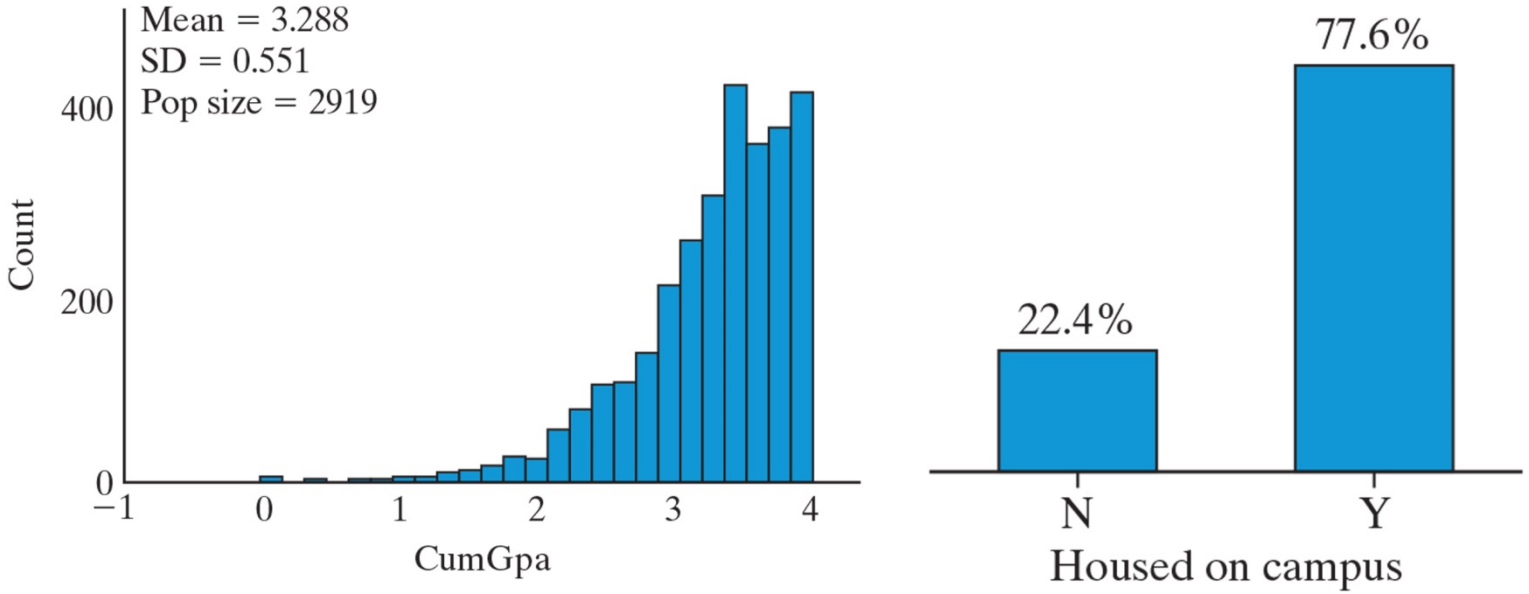
Sampling Students

- What type of variable is “On campus”?
- What type is Cumulative GPA?

Student ID	Cumulative GPA	On campus?
1	3.92	Yes
2	2.80	Yes
3	3.08	Yes
4	2.71	No
5	3.31	Yes
6	3.83	Yes
7	3.80	No
8	3.58	Yes
…	…	…

Sampling Students

- Here are graphs (a histogram and a bar graph) representing all of the 2919 students at the College of the Midwest for our two variables of interest.



Sampling Students

- We usually don’t have information on an entire population like we do here.
- We usually need to make inferences about a population based on a sample.
- Suppose a researcher asks the first 30 students she finds on campus one morning whether they live on campus. This would be a quick and convenient way to get a sample.

Sampling Students

For this scenario:

- What is the population?
- What is the sample?
- What is the parameter
- What is the statistic?
- Do you think this quick and convenient sampling method will result in a similar sample proportion to the population proportion?

Sampling Students

- The researcher's sampling method might overestimate the proportion of students that live on campus because if it is taken early in the morning most of those that live off campus might not have arrived yet.
- We call this sampling method ***biased.***
- A sampling method is ***biased*** if statistics from samples *consistently* over or under-estimate the population parameter.

Sampling Students

- Bias is a property of a sampling *method*, not the sample
  - A method is biased if *on average* its results are not representative.
- Sampling bias also depends on what is measured.
  - Would the morning sampling method be biased in estimating the average GPA of students at the college?
  - What about estimating the proportion of students wearing orange shirts?

Sampling Students

- What's a better way of selecting a representative sample?
- Use a ***random*** mechanism to select the observational units
- Don't rely on ***convenience samples***
- A ***Simple Random Sample (SRS)*** is where every collection of size *n* is equally likely to be the sample selected from the population.

Sampling Students

- How could we take a Simple Random Sample of 30 students from the College of the Midwest?
- Represent each student by ID numbers 1 to 2919
- Have the computer randomly select 30 numbers between 1 and 2919

Sampling Students

IDs of the 30 people selected, along with their cumulative GPA and residential status

ID	Cum GPA	On campus?	ID	Cum GPA	On campus?	ID	Cum GPA	On campus?
827	3.44	Y	844	3.59	N	825	3.94	Y
1355	2.15	Y	90	3.30	Y	2339	3.07	N
1455	3.08	Y	1611	3.08	Y	2064	3.48	Y
2391	2.91	Y	2550	3.41	Y	2604	3.10	Y
575	3.94	Y	2632	2.61	Y	2147	2.84	Y
2049	3.64	N	2325	3.36	Y	2590	3.39	Y
895	2.29	N	2563	3.02	Y	1718	3.01	Y
1732	3.17	Y	1819	3.55	N	168	3.04	Y
2790	2.88	Y	968	3.86	Y	1777	3.83	Y
2237	3.25	Y	566	3.60	N	2077	3.46	Y

Sampling Students

- What is the average cumulative GPA for these 30 students?
  - $\bar{x}$ is the sample average
  - $\bar{x}$ = 3.24
- What proportion live on campus?
  - $\hat{p}$ is the sample proportion
  - $\hat{p}$ = 0.80
- μ is the population mean.
- π is the population proportion.

Sampling Students

- How do we know if $\bar{x}$ and $\hat{p}$ are close to the population values, μ and π?
- A different sample of 30 students would probably have had different values.
- How are these statistics useful in estimating the population parameter values?
- Let's take more simple random samples of 30 students to examine the null distribution of the statistics from other samples.

Sampling Students

- We took 5 different SRSs of 30 students
- Each sample gives different statistics
- This is ***sampling variability***.
- The values don't change much:
  - Average GPAs range from 3.22 to 3.40
  - Sample proportions range from 0.63 to 0.83

Random sample	1	2	3	4	5
Average GPA ()	3.22	3.29	3.40	3.26	3.25
proportion on campus ( )	0.80	0.83	0.77	0.63	0.83

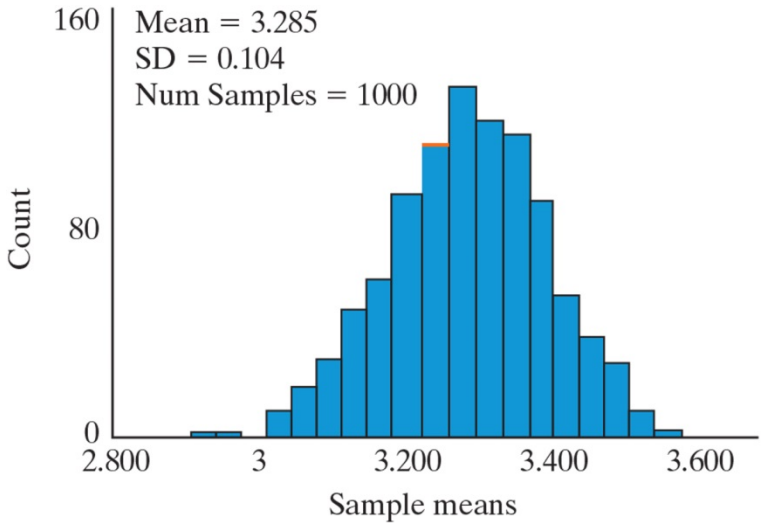
Sampling Students

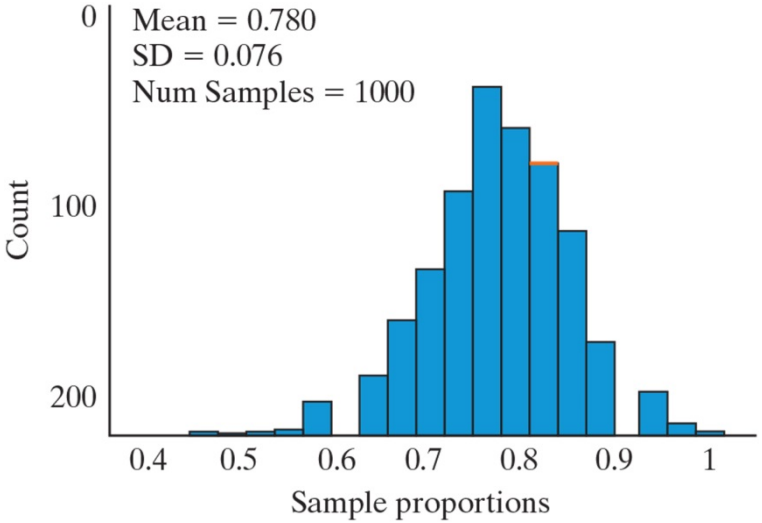
- Population parameters:
  - $\mu = 3.288$
  - $\pi \approx 0.776$ (2265/2919).
- What do the parameters describe?
  - The true average cumulative GPA and the true proportion on campus of the 2919 students
- The statistics tend to be close to the parameters.

Random sample	1	2	3	4	5
Average GPA ()	3.22	3.29	3.40	3.26	3.25
proportion on campus ( )	0.80	0.83	0.77	0.63	0.83

Sampling Students

- We took 1000 SRSs and have graphs of the 1000 sample means (for the GPAs) and 1000 sample proportions (for living on campus).
- The mean of each distribution falls near the population parameter.





Sampling Students

- What would happen if we took all possible random samples of 30 students from this population?
  - The averages of the statistics would match the parameters exactly
- Statistics computed from SRSs cluster around the parameter.
- Why is this an unbiased sampling method?
  - There is no tendency to over or underestimate the parameter.
- The sampling method and statistic you choose determine if a sampling method is biased.
- A sample mean of a simple random sample is an unbiased estimate of the population mean. Same for proportions instead of means.

Sampling Students

- We can *generalize* when we use simple random sampling because it creates:
  - A sample that is representative of the population.
  - A sample statistic that is unbiased and thus close to the parameter for large *n*.

Sampling Students

- If the researcher at the College of the Midwest uses 75 students instead of 30 with the same early morning sampling method will it be less biased?
- No. Selecting more students *in the same manner* doesn't fix the tendency to oversample students who live on campus.
- A smaller sample that is random is actually more accurate.

Sampling Students

- What is an advantage of a larger sample size?
  - Less sample to sample variability.

6. Inference for a Single Quantitative Variable

Section 2.2

https://www.youtube.com/watch?v=ho7796-au8U

Example 2.2: Estimating Elapsed Time

- Students in a stats class (for their final project) collected data on students' perception of time
- Subjects were told that they'd listen to music and be asked questions when it was over.
- 10 seconds of the Jackson 5's "ABC" and subjects were asked how long they thought it lasted
- Can students accurately estimate the length?

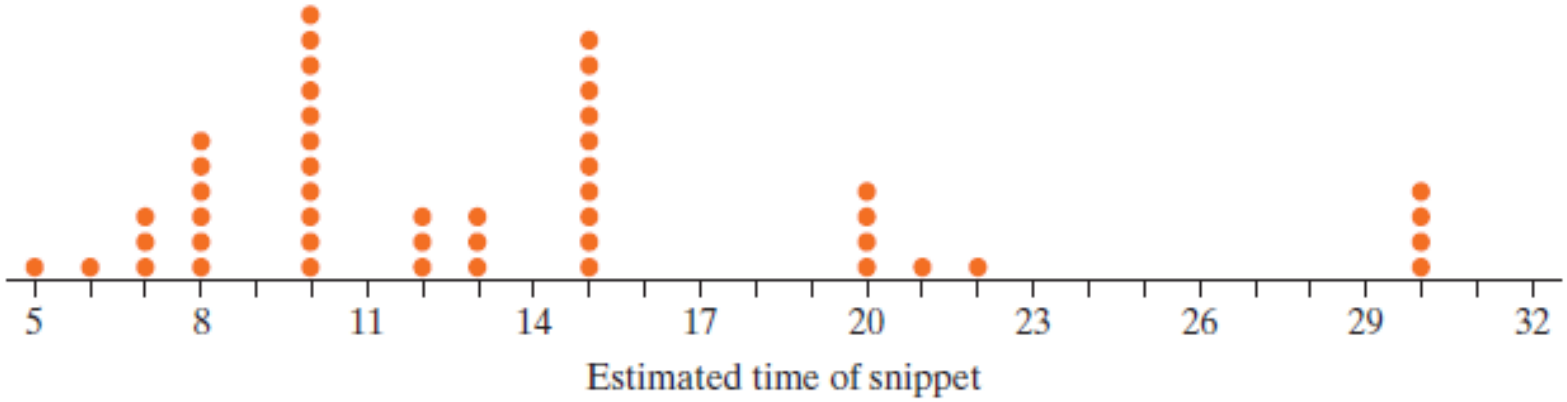
Hypotheses

Null Hypothesis: People will accurately estimate the length of a 10 second-song snippet, on average. ($\mu = 10$ seconds)

Alternative Hypothesis: People will not accurately estimate the length of a 10 second-song snippet, on average. ($\mu \neq 10$ seconds)

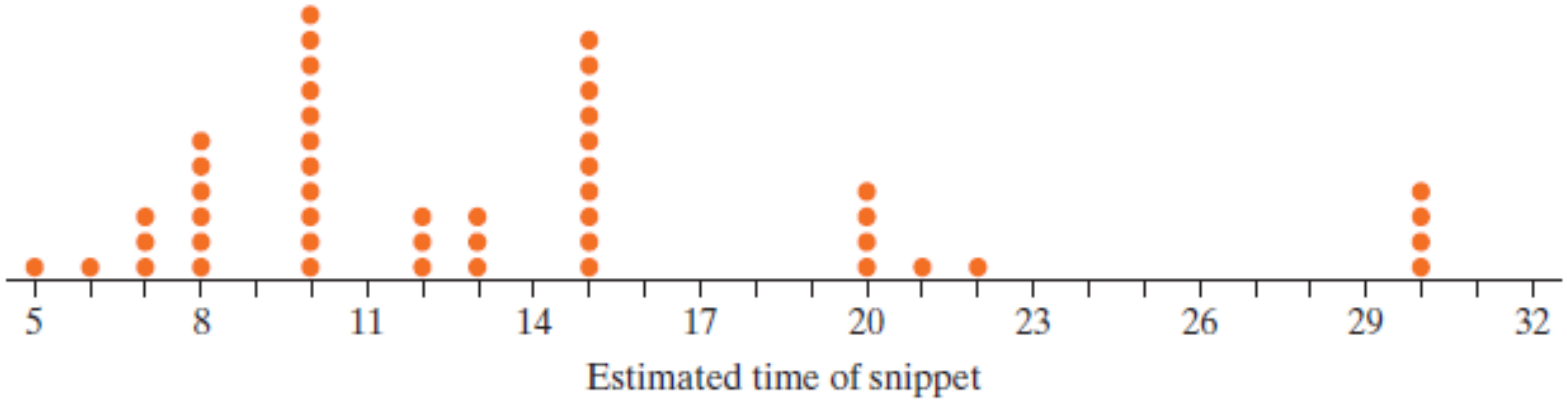
Estimating Time

- A sample of 48 students on campus were subjects and song length estimates were recorded.
- What does a single dot represent?
- What are the observational units? Variable?



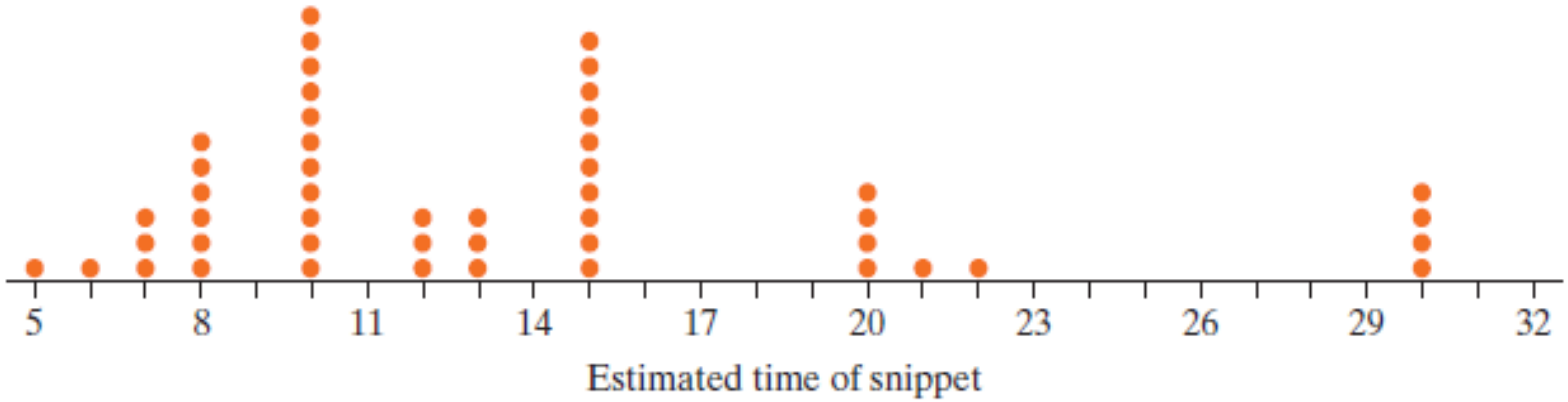
Skewed, mean, median

- The distribution obtained is not symmetric, but is **right skewed.**
- When data are skewed right, the **mean** gets pulled out to the right while the **median** is more resistant to this.



Mean vs Median

- The mean is 13.71 and the median is 12.
- How would these numbers change if one of the people that gave an answer of 30 seconds actually said 300 seconds?
- The standard deviation is 6.5 sec. Also not resistant to outliers.



Inference

- $\mathbf{H_0}$: $\mu = 10$ seconds
- $\mathbf{H_a}$: $\mu \neq 10$ seconds
- Our problem now is, how do we develop a null distribution?
  - Here we don't have population data that reflects our null hypothesis where $\mu = 10$ seconds.
  - All we have is our sample of 48.

Population?

- We need to come up with a large data set that we think our population of time estimates might look like **under a true null.**
- We might assume the population is skewed (like our sample) and has a standard deviation similar to what we found in our sample, but has a mean of 10 seconds.
- The book recommends using an applet for this. We could use *R*, or do a (theory-based) t-test.

Theory-Based Test

- Using simulations to create a population each time we want to run a test of significance is extremely time consuming and cumbersome.
- The null distribution that we developed can be predicted with theory-based methods.
- We know it will be centered on the mean given in the null hypothesis.
- We can also predict its shape and its standard deviation.

t-distribution

- The shape is very much like a normal distribution, but slightly wider in the tails and is called a t-distribution.
- The t-statistic is the standardized statistic we use with a single quantitative variable that looks approximately normal, when the sample size is small, and the statistic can be found using the formula:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

The $s/\sqrt{n}$ (standard deviation of our sample divided by the square root of the sample size) is called the standard error and is an estimate for the standard deviation of the null distribution.

Here $t = \frac{13.71 - 10.0}{6.5/\sqrt{48}} = 3.95.$

p-value = 2*(1-pt(3.95,df=47)) = 0.000261.

Validity Conditions

- The observations must be independent.
- The population must be normally distributed!
- The book says you need the sample size to be at least 20 for the t-test, but this is not technically true. The whole point of the t-test is you can use it even when your sample size is small, provided the two assumptions above hold.

But it is often hard to have any idea if the population is normal without having at least 20 observations.

Estimating Time

Formulate Conclusions.

- Based on our small p-value, we can conclude that our subjects did not accurately estimate the length of a 10-second song snippet and in fact they significantly overestimated it.
- How far can we generalize this?

Summary

- When we test a single quantitative variable, our hypothesis has the following form:
  - $\mathbf{H_0}$: μ = some number
  - $\mathbf{H_a}$: $\mu \neq$ some number, $\mu <$ something or $\mu >$ something.
- We can get our data (or mean, sample size, and SD for our data) and use the Theory-Based Inference to determine the p-value.
- The p-value we get with this test has the same general meaning as from a test for a single proportion.