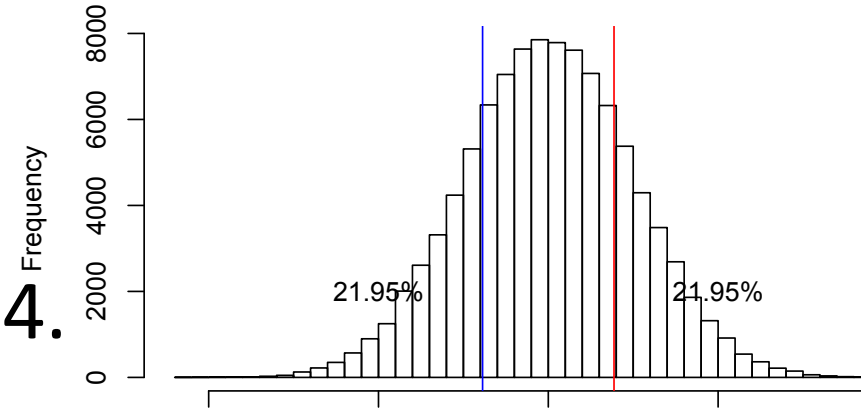
Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

0. Elections and faces testing and computing the p-value.
1. Echinacea and rejecting the null example.
2. Sampling, bias, and students example.
3. Estimating the mean, t-test, and guessing elapsed time example.
4. Sig. level, type 1 and type 2 errors, and power.

Read chapter 3.
http://www.stat.ucla.edu/~frederic/13/F18 .
HW1 is due Tue Oct16.
HW2 is due Thu Oct25 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."
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*."

Test and p-value.



The z statistic is $\frac{0.523-0.5}{.0297} = 0.774$.

Since n is large, by the central limit theorem, the z-statistic is l
a draw from the standard normal distribution. So the p-value
the probability of a standard normal being larger than .774 in
absolute value, for a 2-sided test.

In *R*, pnorm(.774) is the prob. of a std. normal being < .774.

pnorm(.774,lower=F) = .2195 is the prob. std. normal ≥ .774.

So 2 * pnorm(.774,lower=F) is the prob. it is ≥ .774 or ≤ -.774.

2 * pnorm(.774,lower=F) = 43.9%. Not stat. sig.

1. Failing to reject the null vs. accepting the null.

- 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 ($z = -0.42$; $P = .67$)."

1. Failing to reject vs. accepting

- 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".


Cold and flu on NBCNEWS.com
Got a cold? Sorry, echinacea won't help much
Study shows the popular herbal remedy may bring milder symptoms — but that could be due to chance
Recommend 7


CNN
Health » Diet + Fitness | Living Well | Parenting + Family
Echinacea fails to curb the common cold

1. Failing to reject vs. accepting.

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."

2. 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
…	…	…

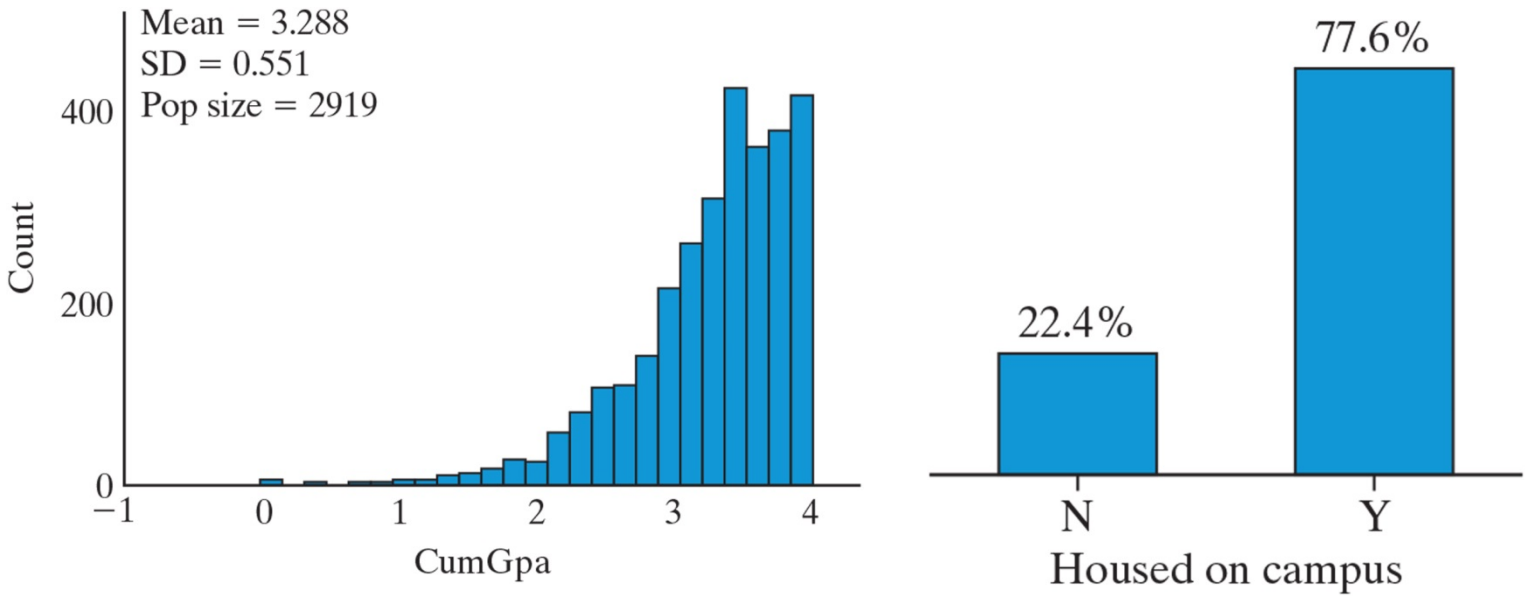
Sampling Students

- 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 do not 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 over or under-estimate the population parameter on average.

Sampling Students

- Bias is a property of a sampling *method*, not the sample
  - A method must *consistently* (on average) produce non-representative results to be considered biased
- 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.
- Do not 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

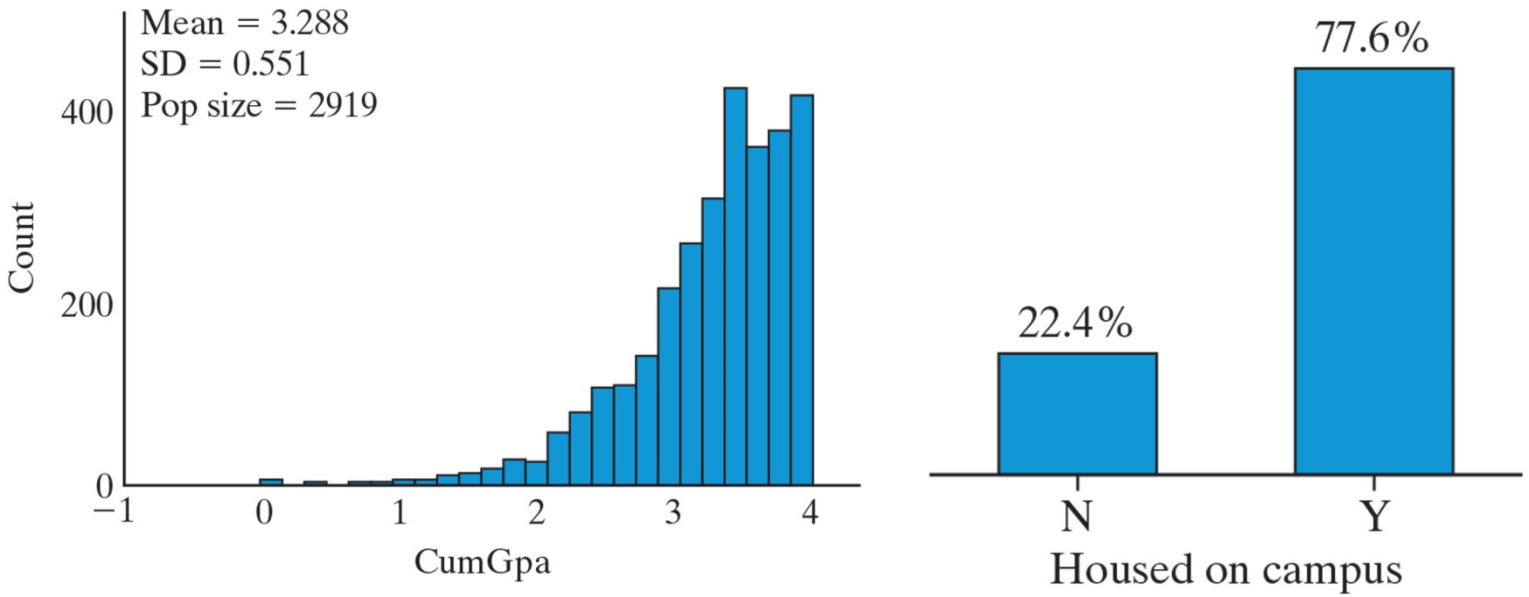
- What is the average cum GPA for these 30 students?
  - $\bar{x}$ is the sample average
  - $\bar{x}$ = 3.24
- What proportion live on campus?
  - $\hat{p}$ is the sample proportion
  - $\hat{p}$ = 0.80
- μ is the population mean.
- π is the population proportion.

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

- 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

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 cum 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

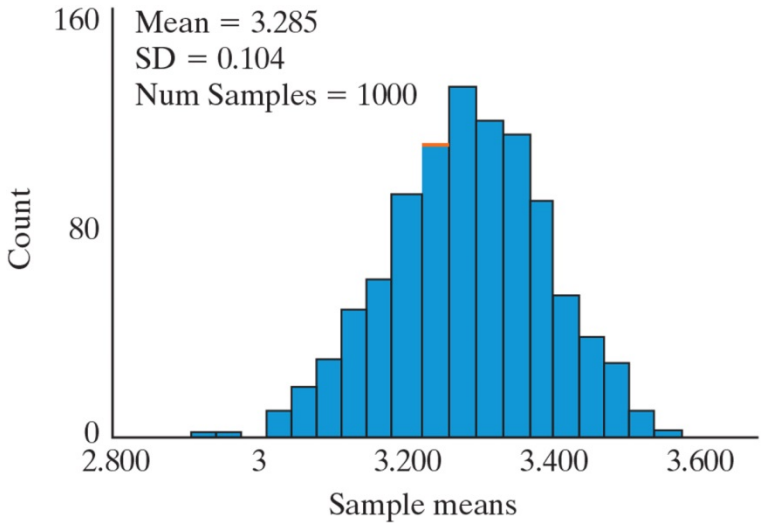
Sampling Students

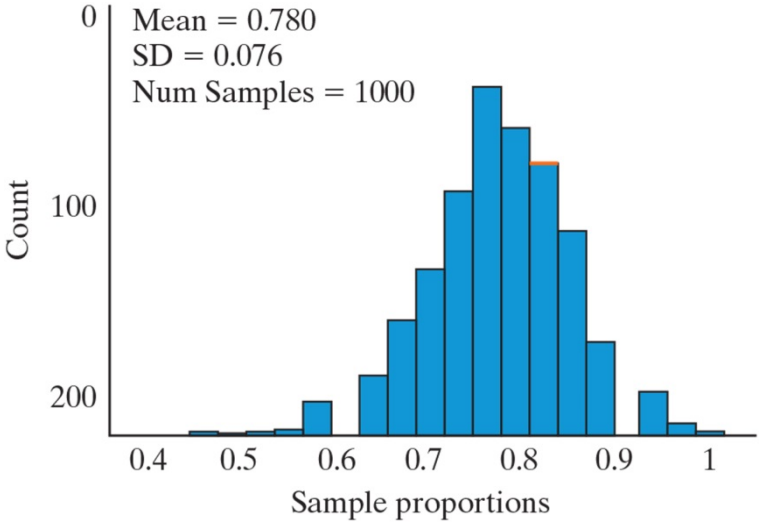
- Population parameters:
  - $\mu = 3.288$
  - $\pi \approx 0.776$ (2265/2919).
- What do the parameters describe?
  - The true average cumulative GPA and the true proportion on campus of the 2919 students
- The statistics tend to be close to the parameters.

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?
- 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.
  - Statistics from different samples cluster more closely around the center of the distribution.

3. Inference for a Single Quantitative Variable

Section 2.2

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?

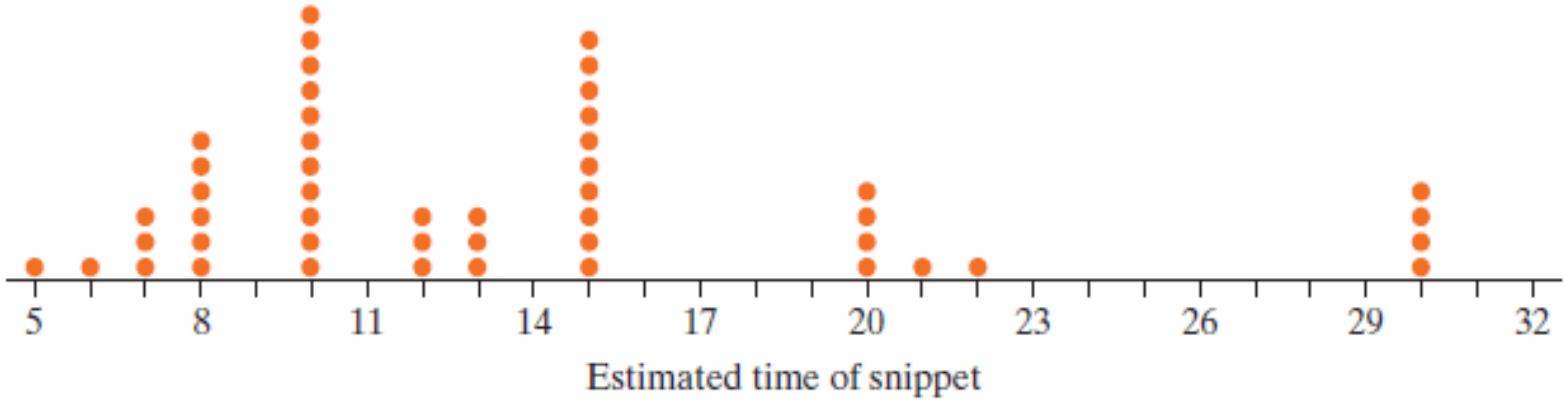
Hypotheses

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)

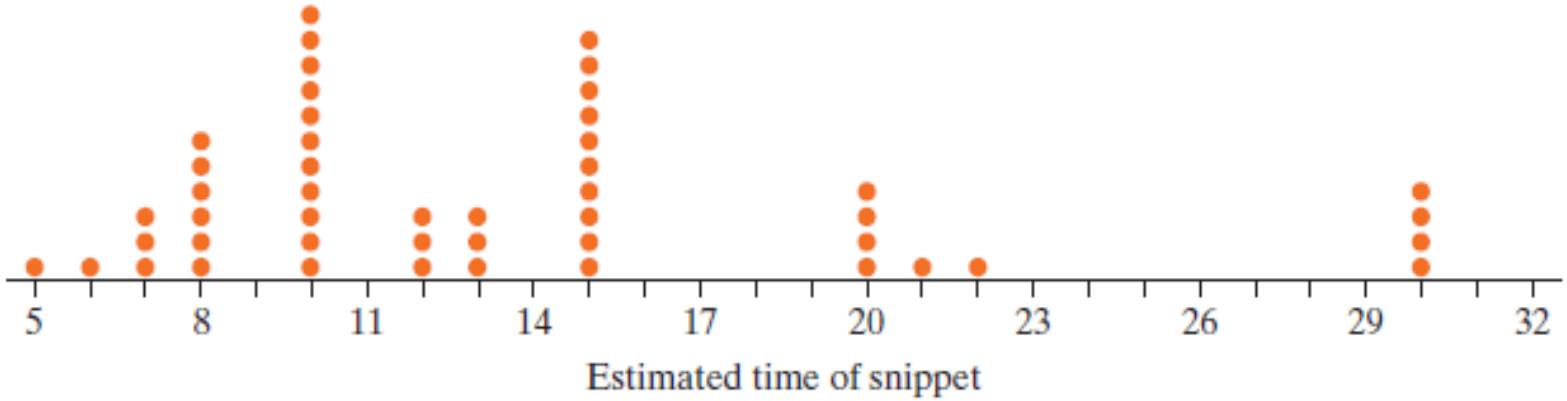
Estimating Time

- 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?



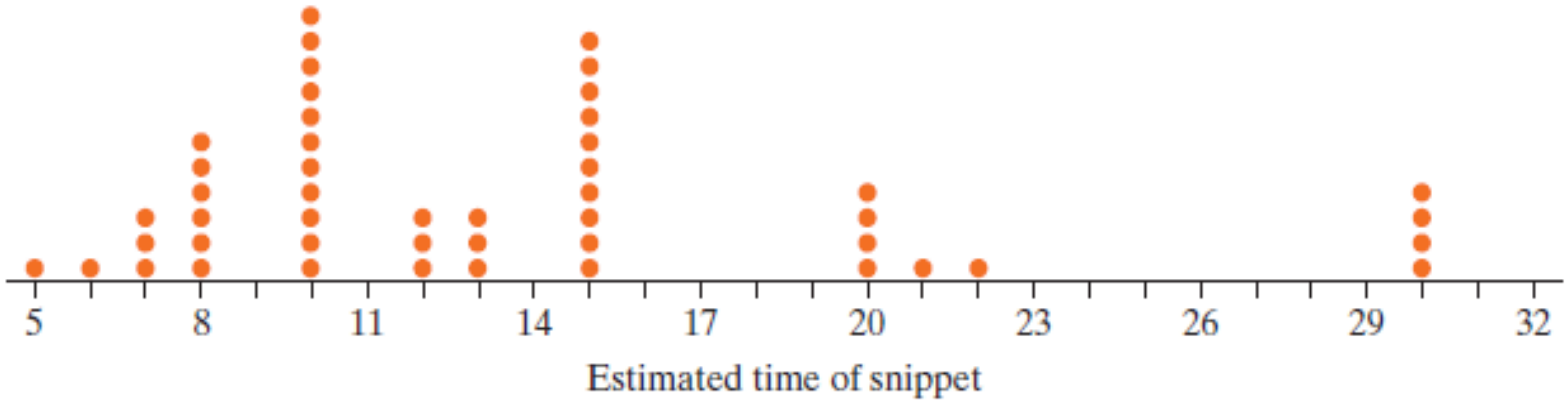
Skewed, mean, median

- 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?
- Flipping coins will not work to model what would happen under a true null hypothesis.
  - 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 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 and 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*pt(3.95,df=47,lower=F) = 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 right. However, 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 people don't accurately estimate the length of a 10-second song snippet and in fact they significantly overestimate it.
- To what larger population can we make our inference?

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 will 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 those from a test for a single proportion.

4. Significance level, Type 1 and Type 2 errors

Section 2.3

Significance Level

- We think of a p-value as telling us something about the strength of evidence from a test of significance.
- The lower the p-value the stronger the evidence.
- Some people think of this in more black and white terms. Either we reject the null or not.

Significance Level

- The value that we use to determine how small a p-value needs to be to provide convincing evidence whether or not to reject the null hypothesis is called the **significance level**.
- We reject the null when the p-value is less than or equal to ($\leq$) the significance level.
- The significance level is often represented by the Greek letter alpha, α.

Significance Level

- Typically we use 0.05 for our significance level. There is nothing magical about 0.05. We could set up our test to make it
  - harder to reject the null (smaller significance level say 0.01) or
  - easier (larger significance level say 0.10).

Type I and Type II errors

- In medical tests:
  - A type I error is a false positive. (They conclude someone has a disease when they don't.)
  - A type II error is a false negative. (They conclude someone does not have a disease then they actually do.)
- These types of errors can have very different consequences.

Type I and Type II Errors

TABLE 2.9 A summary of Type I and Type II errors

		What is true (unknown to us)	
		Null hypothesis is true	Null hypothesis is false
What we decide (based on data)	Reject null hypothesis	Type I error (false alarm)	Correct decision
	Do not reject null hypothesis	Correct decision	Type II error (missed opportunity)

•

Type I and Type II errors

TABLE 2.10 Type I and Type II errors summarized in context of jury trial

		What is true (unknown to the jury)	
		Null hypothesis is true (defendant is innocent)	Null hypothesis is false (defendant is guilty)
What jury decides (based on evidence)	Reject null hypothesis (Jury finds defendant guilty)	Type I error (false alarm)	Correct decision
	Do not reject null hypothesis (Jury finds defendant not guilty)	Correct decision	Type II error (missed opportunity)

The probability of a Type I error

- The probability of a type I error is the significance level.
- Suppose the significance level is 0.05. If the null is true we would reject it 5% of the time and thus make a type I error 5% of the time.
- If you make the significance level lower, you have reduced the probability of making a type I error, but have increased the probability of making a type II error.

The probability of a Type II error

- The probability of a type II error is more difficult to calculate.
- In fact, the probability of a type II error is not even a fixed number. It depends on the value of the true parameter.
- The probability of a type II error can be very high if:
  - The true value of the parameter and the value you are testing are close.
  - The sample size is small.

Power

- The probability of rejecting the null hypothesis when it is false is called the **power** of a test.
- Power can also be thought of as 1 minus the probability of a type II error.
- We want a test with high power and this is aided by
  - A large effect size, i.e. a sample mean far from the parameter in the null hypothesis.
  - A large sample size.
  - A small standard deviation.