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

1. echinacea example, continued.
2. Sampling, bias, and students example.

Read chapters 2 and 3.

1. Rejecting Ho vs. accepting Ha.

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."

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
…	…	…

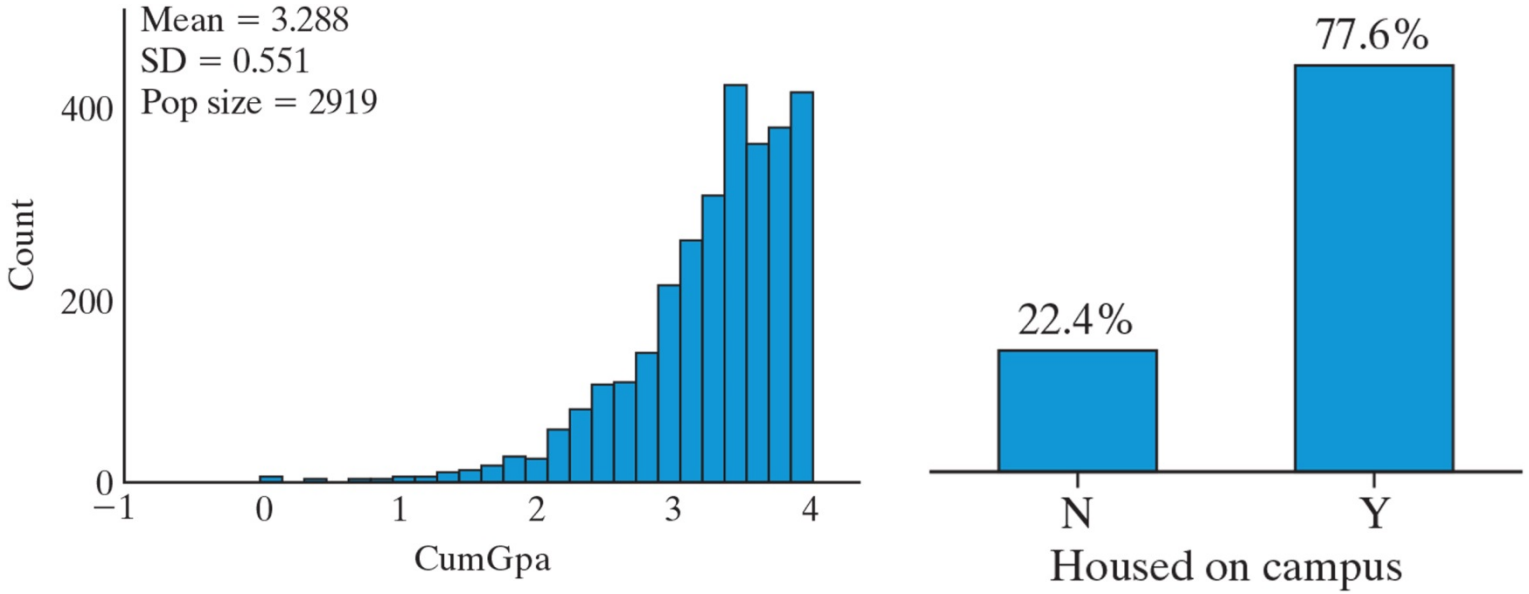
Sampling Students

- 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 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

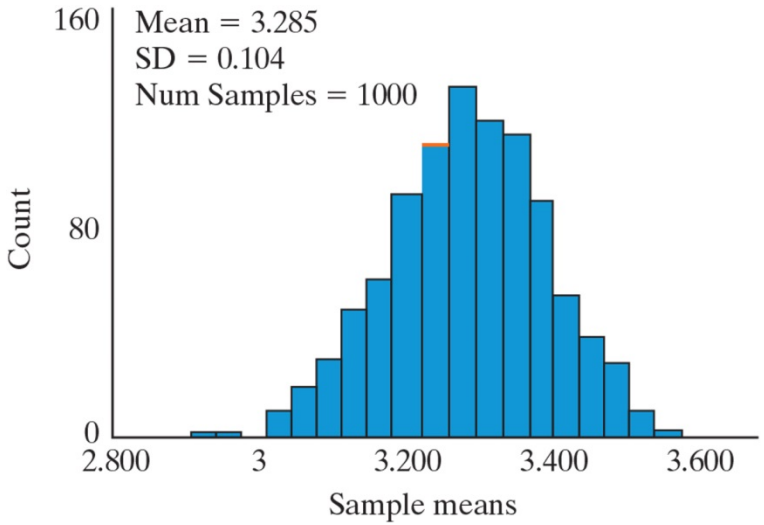
Sampling Students

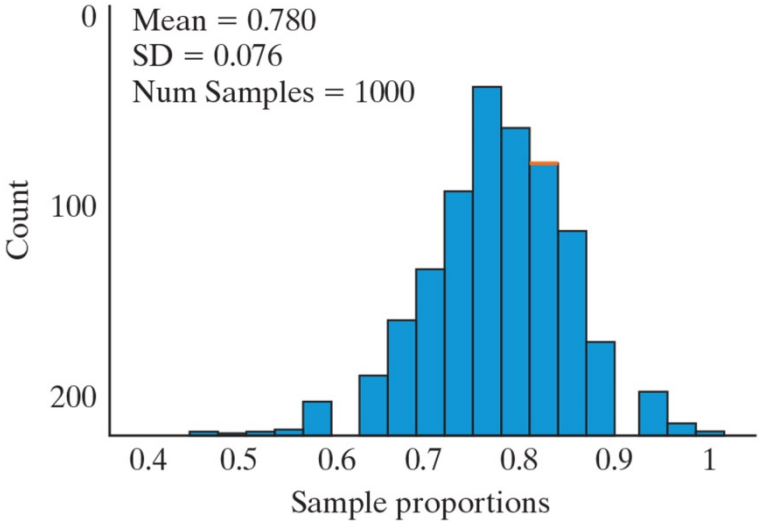
- Population parameters:
  - $\mu = 3.288$
  - $\pi \approx 0.776$ (2265/2919).
- What do the parameters describe?
  - The true average cumulative GPA and the true proportion on campus of the 2919 students
- The statistics tend to be close to the parameters.

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.