AP Statistics – Classwork 1/12/15

1. Identify the population, the parameter, the sample, and the statistic:

(a) A pediatrician wants to know the 75th percentile for the distribution of heights of 10-year-old boys, so she takes a sample of 50 patients and calculates Q_3 = 56 inches.

Population: 10-year-old boys; Parameter: 75th percentile of heights; Sample: 50 patients; Statistic: Q3 of the sample, or 56 inches.

(b) A Pew Research Center Poll asked 1102 12- to 17-year-olds in the United States if they have a cell phone. Of the respondents, 71% said "Yes."

Population: 12-to-17-year-olds in the US; Parameter: proportion who own a cell phone; Sample: 1102 12-to-17-year-olds polled; Statistic: proportion fot he sample who said "Yes", or 71%.

2. What is sampling variability? Why do we care?

Sampling variability refers to the fact that a statistic will take on different values from sample to sample.
We need to estimate sampling variability so we know how close our estimates are to the truth—the margin of error.

3. What is a sampling distribution? Why do we care?

A sampling distribution of a statistic is the values the statistic can take, and the frequency that each value occurs. It is the distribution of the calculated statistic from all possible samples of size n from a given population.
We need to know what counts as unusual.

4. What is the difference between the distribution of the population, the distribution of the sample, and the sampling distribution of a sample statistic?

The distribution of the population is a distribution of measurements of all individuals in the population.
The distribution of the sample is the distribution of measurements of the n individuals in the sample.
The sampling distribution of a sample statistic is the distribution of the calculated statistic from all possible samples of size n from a given population.

5. What is an unbiased estimator? What is a biased estimator? Provide some examples.

Unbiased doesn't mean perfect! Unbiased means not consistently too high or consistently too low when taking <u>many</u> random samples. Biased means statistic is consistently higher or lower than the parameter.

If sampling process is biased (undercoverage, response, non-response), there are no guarantees, even if a statistic is an unbiased estimator.

For example, maximum is a biased estimator because the center of a sampling distribution of sample maximums will always be below the true maximum of the population.

6. How can you reduce the variability of a statistic?

By taking a larger sample.

Also by better design, such as stratified sampling.

7. What effect does the size of the population have on the variability of a statistic?

Not much, assuming the population is at least 10 times the sample – the 10% condition

8. What is the difference between accuracy and precision? How does this relate to bias and variability?

Accurate = unbiased, precise = low variability

9. What are the mean and the standard deviation of the sampling distribution of a sample proportion? Are these formulas on the formula sheet? Are there conditions that need to be met for these formulas to work?

Yes the formulas are on the sheet, find them. We need the 10% condition to use the standard deviation formula, but the mean will always be p.

10. The superintendent of a large school district wants to know what proportion of middle school students in her district are planning to attend a four-year college or university. Suppose that 80% of all middle school students in her district are planning to attend a four-year college or university. What is the probability that an SRS of size 125 will give a sample proportion of at most 75%?

$\mu_{\hat{p}}$ = 0.80. *Because the school district is large, we can assume that there are more than*

10(125) = 1250 middle school students, so $\sigma_{\hat{p}} = \sqrt{\dfrac{0.80(0.20)}{125}} = 0.036$. *We can consider the distribution of $\hat{p}$ to be approximately Normal since np = 125(0.80) = 100 $\geq$ 10 and n(1 − p) = 125(0.20) = 25 $\geq$ 10.*

$P(\hat{p} \leq 0.75) \approx$ `normalcdf(lower: -9999, upper: 0.75, mean =` `0.80, SD = 0.036)` $= 0.0824$.

About 8% of all SRSs of size 125 will give a sample proportion of middle school students who are planning to attend a four-year college or university of at most 0.75.

11. What are the mean and standard deviation of the sampling distribution of a sample mean? Are these formulas on the formula sheet? Are there any conditions for using these formulas?

Yes the formulas are on the sheet, find them. We need the 10% condition to use the standard deviation formula, but the mean will always be mu.

12. What is the shape of the sampling distribution of a sample mean when the sample is taken from a Normally distributed population? Does the sample size matter?

Normal. No.

13. At the P. Nutty Peanut Company, dry-roasted, shelled peanuts are placed in jars by a machine. The distribution of weights in the jars is approximately Normal, with a mean of 16.1 ounces and a standard deviation of 0.15 ounces.
(a) Without doing any calculations, explain which outcome is more likely: randomly selecting a single jar and finding that the contents weigh less than 16 ounces or randomly selecting 10 jars and finding that the average contents weigh less than 16 ounces.

(a) Because averages are less variable than individual measurements, you would expect the sample mean of 10 jars to be closer, on average, to the true mean of 16.1 ounces. Thus, it is more likely that a single jar would weigh less than 16 ounces than for the average of 10 jars to be less than 16 ounces.

(b) Find the probability of each event described above.

(b) Let X = weight of the contents of a randomly selected jar of peanuts. X has the N(16.1, 0.15) distribution. `P(X < 16) = normalcdf(lower: −100, upper: 16, mean: 16.1, SD: 0.15) = 0.2525.`
Let $\bar{x}$ = average weight of the contents of a random sample of 10 jars. $\bar{x}$ has the N(16.1, $0.15/\sqrt{10}$) distribution. `P(`$\bar{x}$` < 16) = normalcdf(lower: −100, upper: 16, mean: 16.1, SD: `$0.15/\sqrt{10}$`) = 0.0175.` *This answer agrees with the answer to part (a) because this probability is much smaller than 0.2525.*

14. What is the shape of the sampling distribution of a sample mean when the sample is NOT taken from a Normally distributed population? Does the sample size matter? Does this concept have a name?

As the sample size gets larger, the shape of the sampling distribution gets closer to a Normal distribution. This concept is called the Central Limit Theorem

15. Suppose that the number of texts sent during a typical day by a randomly selected high school student follows a right-skewed distribution with a mean of 15 and a standard deviation of 35. Assuming that students at your school are typical texters, how likely is it that a random sample of 50 students will have sent more than a total of 1000 texts in the last 24 hours?

A total of 1000 texts among 50 students is the same as an average number of texts of 1000/50 = 20. We want to find P($\bar{x}$ > 20), where $\bar{x}$ = sample mean number of texts. Since n is large (50 > 30), the distribution of $\bar{x}$ is approximately N(15, $35/\sqrt{50}$).

P($\bar{x}$ > 20) ≈ `normalcdf (lower: 20, upper: 9999, mean: 15, SD: `$35/\sqrt{50}$`) = 0.1562.`

There is about a 16% chance that a random sample of 50 high school students will send more than 1000 texts in a day.

For Homework, do Chapter Review exercise from Chapter 7, R7.1 to R7.7 in the online text book.
Test is moved to Wednesday, 1/14.