HW 14-5

1. a. 510/1000 = 0.51
   b. 0.0316
   c. (0.4784, 0.5416)
2. a. 300/500 = 0.6
   b. 0.044
   c. (0.556, 0.644)
   d. 550/1000 = 0.55
   e. 0.032
   f. (0.518, 0.582)
3. a. True, less data you have, less accurate your results are ∴ larger spread of data
   b. True, CI is estimate ± ME, by finding our CI it estimates the true population proportion to be between 0.41 and 0.63
4. a. ME = 0.07
   b. (0.53, 0.67)
5. a. ME = 0.04
   b. (0.52, 0.6)

Name Key Algebra 2 Homework 15-5

1. The results of testing a new drug on 1000 people with a certain disease found that 510 of them improved when they used the drug. Assume these 1000 people can be regarded as a random sample from the population of all people with this disease. The standard deviation of the sampling distribution is 0.0158.
   a. Find the sample proportion of people whose condition would improve when they used the drug.

   $\hat{p} = \frac{510}{1000} = .51$

   b. Find the estimated margin of error.

   $2(.0158) = .0316$

   c. Create a 95% confidence interval to estimate the proportion of people with this disease that would improve if they used the new drug.

   $.51 \pm .0316 = (.4784, .5416)$

2. A newspaper in New York took a random sample of 500 registered voters from New York City and found that 300 favored a certain candidate for governor of the state. A second newspaper polled 1000 registered voters in upstate New York and found that 550 people favored this candidate. The standard deviation for the sampling distribution of voters in New York City is 0.022. The standard deviation for the sampling distribution of voters in upstate New York is 0.016.
   a. Find the sample proportion of registered voters from New York City that favored a certain candidate for governor of the state.

   $\hat{p} = \frac{300}{500} = .6$

   b. Find the estimated margin of error for NYC voters.

   $2(.022) = .044$

   c. Create a 95% confidence interval to estimate the proportion of registered voters from New York City that favored a certain candidate for governor of the state.

   $.6 \pm .044 = (.556, .644)$

   d. Find the sample proportion of registered voters from upstate New York that favored a certain candidate for governor of the state.

   $\hat{p} = \frac{550}{1000} = .55$

   e. Find the estimated margin of error for upstate New York voters.

   $2(.016) = .032$

   f. Create a 95% confidence interval to estimate the proportion of registered voters from upstate New York that favored a certain candidate for governor of the state.

   $.55 \pm .032 = (.518, .582)$

3. Decide if each of the following statements is true or false. Explain your reasoning in each case.

   a. The smaller the sample size, the larger spread of the data.

   True, less data you have, less accurate your results are ∴ larger spread of data

   b. If the margin of error is 0.11 and the observed proportion of red chips is 0.52, then the true population proportion is likely to be between 0.41 and 0.63.

   $.52 \pm .11 = (.41, .63)$ True. CI is est ± ME

   By finding our CI it estimates the true pop. prop. to be between 0.41 and 0.63.

4. The school newspaper at a large high school reported that 120 out of 200 randomly selected students favor assigned parking spaces. The standard deviation of the sampling distribution is 0.035.

   a. Compute the margin of error. $2(.035) = .07$

   b. Calculate the resulting 95% confidence interval.

   $\hat{p} = \frac{120}{200} = .6$

   $.6 \pm .07 = (.53, .67)$

5. A newspaper in a large city asked 500 women the following: "Do you use organic food products (such as milk, meats, vegetables, etc.)?" 280 women answered "yes." The standard deviation of the sampling distribution is 0.02.

   a. Compute the margin of error. $2(.02) = .04$

   b. Calculate the resulting 95% confidence interval.

   $\hat{p} = \frac{280}{500} = .56$

   $.56 \pm .04 = (.52, .6)$

Inference of a Population Proportion

The vast majority of the statistics that you've done so far have been **descriptive**. With descriptive statistics, we summarize how a data set "looks" with measures of central tendency, like the mean, and measures of dispersion, like the standard deviation. But, the more powerful branch of statistics is known as **inferential** where we try to **infer** properties about a population from samples that we take. We do this by using **probability** and **sampling variability** to estimate how likely the sample is given a certain population. **Inferential statistics** begins with the most basic question: How can we estimate the **population proportion/mean**, p / μ , if we know a sample proportion/mean, $\hat{p} / \bar{x}$?

- We will look at sampling distributions of the sample proportions/means. We've learned earlier in the unit that the sampling distribution of sample proportions/means will have a shape that is ~ Normal/symmetric, centered at the Population mean/proportion.
- The closer your sample is to the mean (center), the more consistent it is with your sampling distribution, which is centered at the population proportion/mean.
  - Nothing rare or unusual with your sample.
  - Result is due to chance (normal sampling variability)
- The further your sample is from the mean, the more unusual it is to have come from your sampling distribution centered at the mean.
  - Your sample is "extreme (unusual)" and **not typical of chance behavior**.
  - Result may NOT be due to chance (normal sampling variability)
  - More evidence it came from a sampling distribution with a different mean.
  - We say the results are "statistically significant"

Let's investigate the distribution of sample proportions to see how this all works.

1. A school is trying to determine the proportion of students who own cell phones. They do a survey of **all** juniors and find that 168 out of 236 of have cell phones. They then take a **sample** of freshmen and find that 30 out of 52 freshmen in the sample own cell phones.

a. Calculate the population proportion, *p*, of juniors who own cell phones. Round to the nearest hundredth.

$p = \frac{168}{236} = .71$

b. Calculate the sample proportion, $\hat{p}$, of freshmen who own cell phones. Round to the nearest hundredth.

$\hat{p} = \frac{30}{52} = .58$

Clearly, in the last example, the sample proportion of freshmen who own cell phones is less than the population proportion of juniors who own cell phones. But, can we attribute that variability to the two "treatments", i.e. juniors versus freshmen, or could the variability be due to **sampling variability**, i.e. the random chance that we just picked a group of freshmen who have an unusually low rate of cell phone ownership? We can establish how likely this is to happen by using simulation.

2. We would like to determine how likely it is that a sample of 52 out of a population with a proportion of cell phone ownership of 71% or 0.71 would result in a sample proportion of 0.58.

a. On your phone/tablet, run the program PSIMUL (goo.gl/9Fk40h) with a *p* value of 0.71 and a sample size of 52 for 100 simulations. How many of the 100 simulations had a proportion less than or equal to 0.58?

2

b. Based on your answer to (a), how likely is it that a sample of 52 from a population with a cell phone ownership of 71% would result in a sample proportion of only 0.58 or less?

2/100 = 2% very unlikely to get a sample of .58 or less from a known population of 71%.

c. What conclusion can you make about freshmen cell phone ownership compared to ownership by juniors? Explain.

It's likely that the population of freshman had a lower

overall population proportion of cell phone ownership than that of juniors.

(Far away from the center on the histogram ∴ likely to come

from a diff. mean proportion than juniors)

lower

Inferential statistics is never about proving beyond **any doubt** that a sample either can or cannot come from a certain population. It is about **quantifying how likely it is that it could come from a given population**.

3. Historically, the proportion of emperor penguins with adult weights above 60 pounds is 0.64. $= p$
   Take this to be the population proportion for this characteristic.

   a. A sample of 26 emperor penguins in a zoo found that 20 of the penguins had adult weights above 60 pounds. Calculate the sample proportion for this sample.

   $$\hat{p} = \frac{20}{26} = .77$$

   b. Run PSIMUL with a population proportion of p = 0.64 with a sample size of 26. Do 100 simulations. What percent of these simulations resulted in a sample proportion at or above what you found in part a?

   13/100 ~ 13% of samples had a sample prop. of .77 or more

   c. Do you have enough evidence from (b) to conclude that penguins raised in a zoo have a significantly higher proportion of weights above 60 pounds? Why or why not?

   Not very unusual. Zoo sample could have come from a pop prop. of .64.

4. Let's stick with our emperor penguins from #3. Out of a sample of 56 penguins from a zoo, it was found that 43 penguins had weights over 60 pounds. Run PSIMUL again, but now with a sample size of 56. Continue to use p = 0.64 and 100 simulations. Do you now have stronger evidence that penguins raised in zoos have a higher proportion with weights over 60 pounds? Explain.

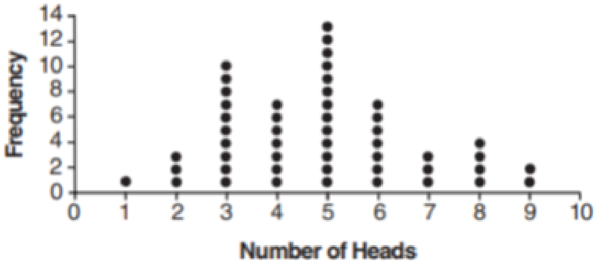
$\hat{p} = \frac{43}{56} = .77$ (bigger sample)

$2/100 = 2\%$ of our samples of S.S. 56, from a known .64 population, ~~came~~ had a sample prop. of .77 or more.

$\therefore$ It is unlikely that our sample came from a .64 population.

The zoo penguins are most likely from a larger population of fat penguins.

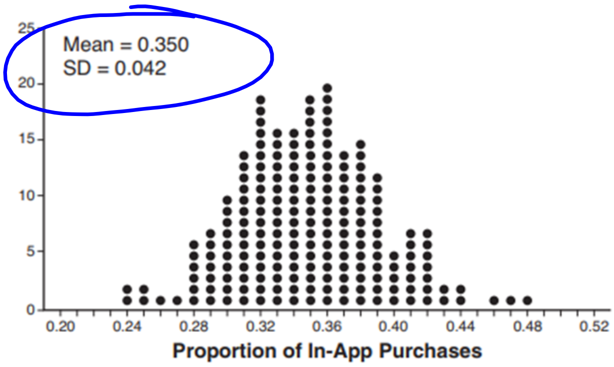
5. The results of simulating tossing a coin 10 times, recording the number of heads, and repeating this 50 times are shown in the graph below.



Based on the results of the simulation, which statement is false?

T 1) Five heads occurred most often, which is consistent with the theoretical probability of obtaining a heads.
2) Eight heads is unusual, as it falls outside the middle 95% of the data.
T 3) Obtaining three heads or fewer occurred 28% of the time. 14/50 = 28%
T 4) Seven heads is not unusual, as it falls within the middle 95% of the data.

6. Some smart-phone applications contain "in-app" purchases, which allow users to purchase special content within the application. A random sample of 140 users found that 35 percent made in-app purchases. A simulation was conducted with 200 samples of 140 users assuming 35 percent of the samples make in-app purchases. The approximately normal results are shown below.



Considering the middle 95% of the data, determine the margin of error, to the nearest hundredth, for the simulated results. In the given context, explain what this value represents.

$ME = 2SD = 2(.042) = .084$

We would expect the sample means to fall within .084 of the actual prop. of .350.