STA-1380 – Elementary Statistics
Week #7

Hello! Welcome to the additional online **Weekly Resources** for the course of **STA-1380**. Following a traditional calendar semester, these will be some of the topics your professors will go over. If you do not see material your section is going over for the week, please look at the other resources listed for this course. In addition to these resources, there might be **Group Tutoring** for this course, please see our website for more details. These sessions will go over these materials in more detail as well as any questions about the material.

Any additional help or services can be found through the **Baylor Tutoring Website**. Visit to schedule a free 30-minute private tutoring session, drop-in times for your course, the Baylor Tutoring YouTube channel, or any additional tutoring resources.

Contacts: Sid Rich M-Th 9am-8pm (Fall and Spring class days)   Office Phone: 254-710-4135

Topic of the Week:
"Sampling Distributions Continued"

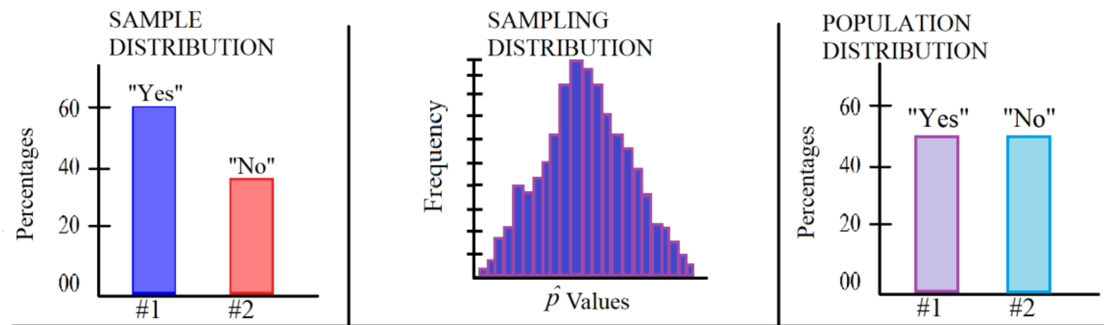
Key Points:

- Proportion Distribution Parameters
- Binomial Approximations

Last week's guides introduced what a Sampling Distribution is and how it is used for quantitative data with sample means, variances, etc. This week's guides will focus on the categorical data that is translated into proportions and ratios.

The Sampling Distribution of Proportions is unlike the Sample or Population Distributions in that it seeks to be normally distributed. A Sampling Distribution of a Proportion seeks to identify the true proportion of a population by finding the means of all proportions sampled.

In the chart below, notice how the Sample Distribution and the Population Distribution are almost identical in appearance, with only the sample size and the $\hat{p}$ vs p being the only differences. This is because there are only two options that an individual in the sample can choose, 'Yes' or 'No'. Questions will be asked to seek out a yes or no answer. "Will you vote for the Republican Candidate in the upcoming election?" Will result in a yes or no response, which can be tallied up into a proportion of those for and against that candidate. The process of polling for a response is the same regardless of the question so long as there is a binary choice.



	Sample Distribution	Sampling Distribution	Population Distribution
Proportion	$\hat{p}$		p
Mean		$\mu_{\hat{p}}$	
Std. Div.		$\sigma_{\hat{p}}$	
Sample Size	n	'XX' amount of trials with n subjects	N
X-Axis	Option Choices	$\hat{p}$ values from 'XX' trials	Option Choices
Y-Axis	% of option chosen	How often the $\hat{p}$ occurs	% of option chosen

The Sampling Distribution seen here is structured almost identically to that of the mean chart, this is because it functions the same way. The purpose of a Sampling Distribution is to normalize the samples collected in a way that data can be analyzed and interpreted. In the future, you will learn to use your Sampling Distribution to determine if the preconceived population parameters are correct or use Statistical reasoning to disprove them.

Since the proportion of a categorical poll is a ratio and a continuous variable able to take on an infinite number of possible results, the $\hat{p}$ of many samples and are averaged to find the mean and center point of the Sampling Distribution. The std. dev of this Sampling Distribution works in a similar way to the Sampling Distribution of Means, in that the population parameters are used rather than the sample statistics. This is why Mu is once again used instead of '*S*'.

Recall that the term 'frequency' is not the exact amount of how often it occurs, but the proportion of occurrences. The Y-axis will be in decimal places for the same reason the Normal Distribution's Y-axis is. The Area under the Sampling Distribution's graph will be equal to 1.

Highlight #1
"Proportion Distribution Parameters"

Definition: The set values gained from a sample of a larger population or derived from the population in order to provide inferences.

Notation: $\hat{p}$ = Sample Proportion, $\hat{q}$ or (1- $\hat{p}$)= Sample Counter-Proportion, n = Sample Size.
$\mu_{\hat{p}}$ = Sampling Mean, $\sigma_{\hat{p}}$ = Sampling Std. Dev., $\sigma_{\hat{p}}^2$ = Sampling Variance,
p = Population Proportion, q or (1- p)= Pop Counter-Proportion, N = Pop. Size

The Sample Proportion parameters can be identified by having the little hat on all of the symbols. 'P-hat' is the symbol given to the sample proportion (47%, 38%, etc.). The counter-proportion is the portion of the sample that chose 'No'. This symbol can either be written as 'Q-hat' or (1-'P-hat').

The proportion, '$\hat{p}$' can be computed by the number of 'Yes' responses over the sample size.

$$\hat{p} = \frac{x}{n}$$

Note that 'x' is the number of people that chose the response of interest, and that 'Q-hat' can be calculated by the number of people who answered other than the response of interest.

The Sampling Distribution as mentioned above, functions the same as the Sampling Distribution of Means. As such, the equations are structured the same, only using the parameters associated with proportions.

The Sampling Mean of proportions is found by the sum of '$\hat{p}$' over number of trials.

$$\mu_{\hat{p}} = \frac{\sum(\hat{p})}{\#\ \text{trials}} \quad or \quad \mu_{\hat{p}} = p$$

The Sampling Variance is found by multiplying the Population proportion and counter proportion over the sample size.

$$\sigma_{\hat{p}}^{2} = \frac{pq}{n}$$

Because the Sampling parameters seek to be unbiased estimators, it can be assumed that the summation of all $\hat{p}$ will equal the population proportion of p. Remember that the variance is the Shape Parameter that can be altered and calculated, and that the square root of the variance is the Std. Dev.

Highlight #2
"Normal Approximation to the Binomial"

Definition: When 'n' is sufficiently large, a Binomial Distribution can be solved with a Sampling Distribution of Proportions.

This approximation is one of the most well-known crossovers between Discrete Distributions and Continuous Distributions. The very design of a sample proportion is a yes or no question with a binary choice. Because of this Binary characteristic, the Sampling Distribution's properties will mimic the ways a Binomial Distribution is constructed.

The approximation also has a similar function to the CLT. This rule of thumb is often referred to as the "NP" or "10 rule" and is used to determine if a Binomial Distribution can be approximated and use the Sampling Distribution for Proportions for inference. The rule is that the sample size multiplied by the proportion, and its counter proportion must both be equal to or greater than 5. This rule is to ensure that the Binomial Distribution is not heavily skewed and has enough of a sample size to appear as a bell-shaped curve.

$$np \geq 5 \text{ and } n(1-p) \geq 5$$

When a Binomial has been found to meet all of the requirements, you will find that the $\hat{p}$ will equal *np* (*p* being the innate probability of the Binomial). If you go back to the previous guides, you will find that this is the same formula used for the means of Binomial Distributions. For simplicity, use the Mean and Std. Dev. formulas for the Binomial. In future courses you will have to account for something called the 'continuity correction factor,' which allows for some 'leeway' in the transition from a Discrete to a Continuous Distribution. You will not need to account for these in STA-1380.

Check Your Learning

1. A Baylor student wishes to figure out whether Common Grounds or Dutch Bros is the most popular coffee shop for students. It was previously assumed that this question split the populus evenly. With a sample of 15 students, 66% of them chose Common Grounds as their preferred coffee shop.

   a. What is the Standard Deviation of the Sampling Distribution?

   b. What is the probability of having less than result occurring?

2. A Shampoo Factory makes high-end color-safe products. Because of their expensive merchandise, they have to ensure that all of their products are filled correctly. The shampoo is sold in crates of 520 bottles each supposed to be filled to 38 fl oz. A random sample of 50 of these bottles are weighed to see if they are properly filled. If more than 3% of the bottles sampled are improperly filled, the entire crate is deemed unsellable and not sold to the salons and vendors.

   a. Which numbers are the Population Parameters? Which are the Test Statistics?

   b. Find the probability that a crate is discarded if the true proportion of bottles in the crate tested that are improperly filled is 7%.

Things Students Struggle With

1. Is this a Normal or Sampling question?:

   a. Most of the time, with all of the block text that Statistics deals with, it can be difficult to discern what the question is truly asking for. In some cases, there might be too much information, with certain points being used as red herrings to distract you. The best way to find out what the question is asking for is to see if there is a sample conducted, or if a single statistic is given. If it is just the statistic, it's a normal distribution question. If the question contains a sample, then the Sampling Distribution will be used.

2. When is '*n*' large enough?:

   a. It seems like a trick answer to say that it depends, but for statistics, it truly does. In some cases, a bell-shaped, normally distributed set of data will not need a larger mean. Meanwhile, a Binomial Distribution might need a sample larger than 30 in order for it to be approximated via the Normal Distribution. The truth comes down to the numbers and convenience. If the sample size is large and one computation is easier and more reliable than another, statisticians will take a short cut and use the approximation. However, if a data set is skewed, or there are not enough individuals in a sample for the Binomial to equal its Normal Approximation, then the short cut cannot be taken. Each question and its circumstances are different. Use your knowledge and intuition to decide whether or not Sampling Distributions and approximations can be used.

Concluding Comments

That's it for this week! Please reach out if you have any questions and don't forget to visit the Tutoring Center website for further information at https://www.baylor.edu/tutoring.

Answers to CYL

1. **a.** (.5*.5÷15)^ ½ =1.291
   b. Z-score = 1.239 Probability = .8923

2. **a.** 520 = N, 50 = n, 7% = p. (Note that 3% is the Factory Standard, $\hat{p}$ is never given)

   b. Z-score = -1.109 Probability = .8663 (Since $\hat{p}$ is not given, this solves for the Probability of $\hat{p}$ > Factory Standard)