

## Day 17 Worksheet

### Definition

#### Central Limit Theorem for Proportions

Given a population with a proportion  $p$ , the set of all possible samples of size  $n$  forms a sampling distribution that approaches a normal distribution with a

mean of  $p$  and

a standard error of  $SE = \sqrt{\frac{p(1-p)}{n}}$ .

As the sample size,  $n$ , increases, the sampling distribution gets closer and closer to a normal distribution.

If we do not know the population proportion  $p$ , we can approximate the standard error using the sample proportion  $\hat{p}$  instead of  $p$ . In this case

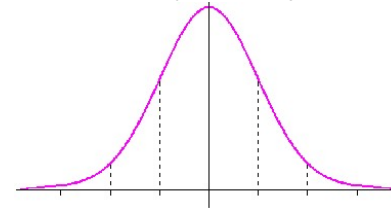
$$SE \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

#### The Fine Print - (Conditions to check)

- The sample observations must be independent and generally not more than 10% of the overall population
- There should be at least 10 successes and 10 failures in our sample. i.e. If we ask a yes/no question, we should have at least 10 yes and 10 no responses in our sample. This is often written as  $np \geq 10$  and  $n(1-p) \geq 10$ .

### Activity 1, continued :

D. Draw a picture of your sampling distribution.



E. Use Excel to find the probability of a sample as extreme as the data above, given the mean and standard error you just found.

F. What is your conclusion?

#### Activity 1: Facial prototyping, with a larger $n$

When researchers notice a phenomenon with a small sample size, they often conduct a follow-up study with a larger sample size. Let's suppose our facial prototyping researchers decide to conduct a follow-up study with 300 participants. They use the same Tim and Bob images as the prior activity and find that 192 participants select Tim as the photo on the left. Rather than building a simulation, this time we will use the formulas from the Central Limit Theorem to describe the bell curve that we would have created if we had done a simulation.

- State the null hypothesis.
- State the alternative hypothesis.
- If the null hypothesis is true, what is the mean and standard error for the sampling distribution?

**Activity 2:** Two students wanted to try this out with their own names: Heather and Jennifer. They use their own pictures and find 300 volunteer participants. They find that 162 of the 300 participants assigned the left picture to Heather and the other 138 selected Jennifer.

- State the null and alternative hypotheses.
- Will the mean and standard error change? If so, find the new mean and standard error. If not, explain why they will stay the same as the Bob and Tim example.
- Draw a picture of your sampling distribution.

**Activity 2, continued:**

D. Use Excel to find the probability of a sample as extreme as 162 out of 300, i.e.  $\hat{p} = 0.54$ .

E. How does this p-value compare with the p-value you did with the Tim and Bob version of the experiment? Why might this have happened? Which of the following sound reasonable?

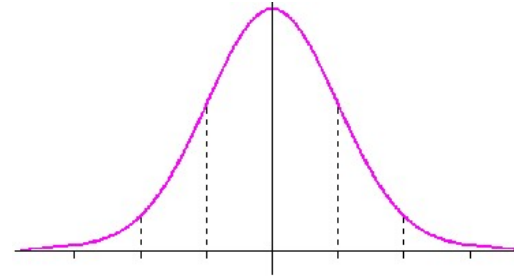
- Facial prototyping is actually not a real phenomenon and the researchers just got lucky with Tim and Bob.
- The phenomenon might be real, but it only works for the names Tim and Bob.
- The names Jennifer and Heather are associated with similar facial prototypes.
- Other reasons or confounding variables. What is your explanation?

F. Suppose the researchers decide to classify two names as "significantly different profile types" if testing them provides a certain p-value. What p-value would you select as the cut-off? Explain your reasoning in complete sentences.

**Activity 4: Finding p-values**

According to the CDC, approximately 9.3% of people in the US have diabetes. A researcher is investigating whether this rate is lower in a community of recent immigrants from Africa. She gathers a random sample of 500 recent immigrants and finds that the rate of diabetes in her sample is 8.2%.

- 1 State the null and alternate hypotheses.
- 2 Find the p-value. Use the mean and standard error you computed in the prior activity rather than creating a simulation in TinkerPlots.
- 3 State your conclusions as a complete sentence.



**Activity 3:** For each of the scenarios, determine whether the sampling distribution will be basically a normal distribution. If so, state the mean and standard error for the sampling distribution.

According to the CDC, approximately 9.3% of people in the US have diabetes. Random samples of 500 people are taken to determine the rates of diabetes.  
- Does this meet the requirements for a normal distribution? (Explain why)  
If so, state the mean and standard error for the sampling distribution.

According to the ASPCA, approximately 44% of households in the US have dogs. Random samples of 80 households are taken to determine dog ownership.  
- Does this meet the requirements for a normal distribution? (Explain why)  
If so, state the mean and standard error for the sampling distribution.

The odds of finding a four-leaf clover are quite low. While exact figures are hard to find, one researcher claims that only 1 in 1000 clovers will be four-leaf clovers. A sample 2000 clovers from one field are selected in a search for an accurate rate of the appearance of four-leaf clovers.  
- Does this meet the requirements for a normal distribution? (Explain why)  
If so, state the mean and standard error for the sampling distribution.

**Activity 5:** According to the ASPCA, approximately 44% of households in the US have dogs. A researcher wants to know if the rate of dog ownership is higher in rural areas. He takes a random samples of 80 rural households and finds that 51 had a dog.

- 1 State the null and alternate hypotheses.
- 2 Find the p-value. Use the mean and standard error you computed in the prior activity rather than creating a simulation in TinkerPlots.
- 3 State your conclusions as a complete sentence.

