LO 1. Define sample statistic as a point estimate for a population parameter, for example, the sample proportion is used to estimate the population proportion, and note that point estimate and sample statistic are synonymous.

LO 2. Recognize that point estimates (such as the sample proportion or mean) will vary from one sample to another, and define this variability as sampling variation.

LO 3. Calculate the sampling variability of the proportion, the standard deviation of the sample proportion, as $SD = \sqrt{\frac{p(1-p)}{n}}$, where p is the population proportion.

LO 4. Note that when the population proportion p is not known (almost always), the sample proportion is used, yielding the Standard Error, $SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

LO 5. Recognize that when the sample size increases we would expect the sampling variability to decrease.

- Conceptually: Imagine taking many samples from the population. When sample sizes are large the sample proportions or means will be much more consistent across samples than when the sample sizes are small.
- Mathematically: $SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, when n increases, SE will decrease since n is in the denominator.

∗ *Reading: Section 5.1 of Advanced High School Statistics*

∗ *Test yourself:*

1. *Explain, in plain English, the difference between standard deviation of the sample proportion and standard error of the sample proportion.*

LO 6. Define a confidence interval as the plausible range of values for a population parameter.

LO 7. Define the confidence level as the percentage of random samples which yield confidence intervals that capture the true population parameter.

LO 8. Calculate an approximate 95% confidence interval by adding and subtracting 2 standard errors to the point estimate: $\textit{point estimate} \pm 2 \times SE$.

LO 9. Recognize that the Central Limit Theorem (CLT) is about the distribution of point estimates, and that given certain conditions, this distribution will be nearly normal.

LO 10. Recall that independence of observations in a sample is provided by random sampling (in the case of observational studies) or random assignment (in the case of experiments).

LO 11. Know that if the sample is large compared to the population, or more precisely, is greater than 10% of the population, the SE estimate will be inaccurate (it will, in fact, overestimate, causing the confidence interval to be wider than necessary).

LO 12. Recognize that the nearly normal distribution of the point estimate (as suggested by the CLT) implies that a more general confidence interval can be calculated as

$$\textit{point estimate} \pm z^{\star} \times SE,$$

where $z^{\star}$ corresponds to the cutoff points in the standard normal distribution to capture the middle XX% of the data, where XX% is the desired confidence level.

- Note that $z^\star$ is always positive.

LO 13. Define margin of error as the distance required to travel in either direction away from the point estimate when constructing a confidence interval, i.e. $z^\star \times SE$.

- Notice that this corresponds to half the width of the confidence interval.

LO 14. Interpret a confidence interval as "We are XX% confident that the true population parameter is in this interval", where XX% corresponds to the confidence level.

- Note that your interpretation must always be in context of the data – mention what the population is and what the parameter is (mean or proportion).

∗ *Reading: Section 5.2 of Advanced High School Statistics*

∗ *Test yourself:*

1. *Confirm that $z^\star$ for a 98% confidence level is 2.33. (Include a sketch of the normal curve in your response.)*
2. *If we want to decrease the margin of error, and hence have a more precise confidence interval, should we increase or decrease the sample size?*
3. *Explain, in plain English, the difference between standard error and margin of error.*
4. *In a random sample of 1,017 Americans 60% said they do not trust the mass media when it comes to reporting the news fully, accurately, and fairly. The standard error associated with this estimate is 0.015 (1.5%). What is the margin of error ay 95% confidence level? Calculate a 95% confidence interval and interpret it in context. You may assume that the point estimate is normally distributed (we'll learn how to check this later)*

LO 15. Explain how the hypothesis testing framework resembles a court trial.

LO 16. Recognize that in hypothesis testing we evaluate two competing claims:

- the null hypothesis, which represents a skeptical perspective or the status quo, and
- the alternative hypothesis, which represents an alternative under consideration and is often represented by a range of possible parameter values.

LO 17. Construction of hypotheses:

- Always construct hypotheses about population parameters (e.g. population proportion, p) and not the sample statistics (e.g. sample proportion, $\hat{p}$). Note that the population parameter is unknown while the sample statistic is measured using the observed data and hence there is no point in hypothesizing about it.
- Define the null value as the value the parameter is set to equal in the null hypothesis.
- Note that the alternative hypothesis might be one-sided ($p <$ or $>$ the null value) or two-sided ($p \neq$ the null value), and the choice depends on the research question.

LO 18. Define a p-value as the conditional probability of obtaining a sample statistic at least as extreme as the one observed given that the null hypothesis is true.

$$\text{p-value} = \text{P(observed or more extreme sample statistic} \mid H_0 \text{ true)}$$

LO 19. Calculate a p-value as the area under the normal curve beyond the calculated test statistic (either in one tail or both, depending on the alternative hypothesis).

- Always sketch the normal curve when calculating the p-value, and shade the appropriate area(s) depending on whether the alternative hypothesis is one- or two-sided.

LO 20. Infer that if a confidence interval does not contain the null value the null hypothesis should be rejected in favor of the alternative.

LO 21. Compare the p-value to the significance level to make a decision between the hypotheses:

- If the p-value $<$ the significance level, reject the null hypothesis since this means that obtaining a sample statistics at least as extreme as the observed data is extremely unlikely to happen just by chance, and conclude that the data provides evidence for the alternative hypothesis.
- If the p-value $>$ the significance level, fail to reject the null hypothesis since this means that obtaining a sample statistics at least as extreme as the observed data is quite likely to happen by chance, and conclude that the data does not provide evidence for the alternative hypothesis.
- Note that we can never "accept" the null hypothesis since the hypothesis testing framework does not allow us to confirm it.

LO 22. Note that the conclusion of a hypothesis test might be erroneous regardless of the decision we make.

- Define a Type 1 error as rejecting the null hypothesis when the null hypothesis is actually true.
- Define a Type 2 error as failing to reject the null hypothesis when the alternative hypothesis is actually true.

LO 23. Choose a significance level depending on the risks associated with Type 1 and Type 2 errors.

- Use a smaller α is Type 1 error is relatively riskier.
- Use a larger α is Type 2 error is relatively riskier.

LO 24. Define power as the probability of correctly rejecting the null hypothesis (complement of Type 2 error).

LO 25. Explain how power changes for changes in effect size, sample size, significance level, and standard error.

∗ *Reading: Section 5.3 of Advanced High School Statistics*

∗ *Videos:*

   – *Null and alternative hypotheses, YouTube (2:42)*

∗ *Test yourself:*

1. *List errors in the following hypotheses: $H_0 : \hat{p} > 20$ and $H_A : \hat{p} \geq 25$*
2. *What is wrong with the following statement?*
   "If p-value is large we accept the null hypothesis since a large p-value implies that the observed difference between the null value and the sample statistic is quite likely to happen just by chance."
3. *Suppose a researcher is interested in evaluating the following claim "In my county, the proportion of eligible voters who actually voted in the last election was 50%", and that she believes this is an underestimate.*
   (a) *How should she set up her hypotheses?*
   (b) *Explain to her, in plain language, how she should collect data and carry out a hypothesis test.*
   (c) *Suppose she collects a sample of eligible voters from her county and finds a sample proportion of 0.65, and a p-value of 0.029. What should she conclude?*
   (d) *Interpret this p-value (as a conditional probability) in context of the question.*

(e) *Suppose that the true proportion is in fact 0.49, what type of an error has this researcher made? In order to avoid making such an error should she have used a smaller or a large significance level?*

LO 26. Distinguish statistical significance vs. practical significance.

∗ *Reading: Section 5.4 of Advanced High School Statistics*