Chapter 4: Framework for inference
Learning Objectives

Reading: Section 4.1 of OpenIntro Statistics

Video: Sample vs. sampling distribution, Dr. Çetinkaya-Rundel (9:30)

Video: Working with the CLT for means, Dr. Çetinkaya-Rundel (3:16)

Video: Distribution of the sample mean, YouTube (6:17)

1. We estimate a population parameter by calculating a point estimate based on a sample. For example, the sample mean is used to estimate the population mean. Note that the terms *point estimate* and *sample statistic* are often used interchangeably.

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

3. Calculate the sampling variability of the mean, called the standard error, as

$$SE = \frac{\sigma}{\sqrt{n}},$$

where σ is the population standard deviation.

   - Note that when the population standard deviation σ is unknown (almost always), this can be estimated using the sample standard deviation, $SE = \frac{s}{\sqrt{n}}$.

4. Distinguish standard deviation (σ or s) and standard error (SE): standard deviation measures the variability in the data, while standard error measures the variability in point estimates from different samples of the same size and from the same population, i.e. measures the sampling variability.

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 means will be much more consistent across samples than when the sample sizes are small.
   - Mathematically: $SE = \frac{\sigma}{\sqrt{n}}$, when n increases, SE will decrease since n is in the denominator.

Test yourself:

1. For each of the following situations, state whether the variable is categorical or numerical, and whether the parameter of interest is a mean or a proportion.
   (a) In a survey, college students are asked whether they agree with their parents' political ideology.
   (b) In a survey, college students are asked what percentage of their non-class time they spend studying.
2. Suppose heights of all women in the US have a mean of 63.7 inches, and a random sample of 100 women's heights yield a sample mean of 65.2 inches. Which one is the population parameter and which one is the point estimate? Which one is μ and which one is $\bar{x}$?
3. Suppose heights of all women in the US have a standard deviation of 2.7 inches, and a random sample of 100 women's heights yields a standard deviation of 4 inches. Which one is the population parameter and which one is the point estimate? Which one is σ and which one is s?
4. Explain, in plain English, the difference between standard deviation and standard error.
5. Suppose heights of all men in the US have a mean of 69.1 inches and a standard deviation of 2.9 inches. Would a randomly selected man who is 72 inches tall be considered unusually tall? Would it be unusual to have a random sample of 100 men where the sample average is 72 inches?

Reading: Section 4.2 of OpenIntro Statistics

Video: Central Limit Theorem, Khan Academy (9:49)

Video: Sampling Distribution of the Mean, Khan Academy (10:52)

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

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

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

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.

   - In the case of the mean the CLT tells us that if

     (1a) the sample size is sufficiently large ($n \geq 30$) and the data are not extremely skewed or

     (1b) the population is known to have a normal distribution, and

     (2) the observations in the sample are independent,

     then the distribution of the sample mean will be nearly normal, centered at the true population mean and with a standard error of $\frac{\sigma}{\sqrt{n}}$.

     $$\bar{x} \sim N\left(mean = \mu, SE = \frac{\sigma}{\sqrt{n}}\right)$$

   - When the population distribution is unknown, condition (1a) can be checked using a histogram or some other visualization of the distribution of the observed data in the sample.
   - The larger the sample size (n), the less important the shape of the distribution becomes, i.e. when n is very large the sampling distribution will be nearly normal regardless of the shape of the population distribution.

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).

    - Our methods also assume the sample is no more than 10% of the population. If the sample is larger than 10% of the population, we should use slightly more advanced techniques.

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

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

    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.

12. 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.

13. Generic interpretation of a confidence interval: "We are XX% confident that the true population parameter is in this interval", where XX% is the desired confidence level.

- In general, it is necessary to put the interpretation into the context of the problem. Example: We are 95% confident that the Wingspan stent by Boston Scientific, when implanted in the brain of stroke patients, increases the patient mortality rate by 3% to 15%. (This example relies on the data from Section 1.1 and methods in Section 6.1 of OpenIntro Statistics.)

Test yourself:

1. Explain, in plain English, what is going on in Figure 4.8 of the book (page 166).
2. List the conditions necessary for the CLT to hold. Make sure to list alternative conditions for when we know the population distribution is normal vs. when we don't know what the population distribution is, and when the sample size is barely over 30 vs. when it's very large.
3. Confirm that $z^\star$ for a 98% confidence level is 2.33. (Include a sketch of the normal curve in your response.)
4. Calculate a 95% confidence interval for the average height of US women using a random sample of 100 women where the sample mean is 63 inches and the sample standard deviation is 3 inches, and interpret this interval in context of the data.
5. Explain, in plain English, the difference between standard error and margin of error.
6. A little more challenging: Suppose heights of all men in the US have a mean of 69.1 inches and a standard deviation of 2.9 inches. What is the probability that a random sample of 100 men will yield a sample average less than 70 inches?
   Hint: First check if we should expect the sample mean to be distributed nearly normally, i.e. if the CLT holds. If so, sketch a normal curve with mean μ and the appropriate standard error. Shade the area you're interested in, and calculate it using methods we learned in the previous unit.

Reading: Section 4.3 of OpenIntro Statistics

Video: Null and alternative hypotheses, YouTube (2:42)

14. Explain how the hypothesis testing framework resembles a court trial.

15. 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.

16. Construction of hypotheses:

- Always construct hypotheses about population parameters (e.g. population mean, μ) and not the sample statistics (e.g. sample mean, $\bar{x}$). 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 (μ < or > the null value) or two-sided ($\mu \neq$ the null value), and the choice depends on the research question.

17. 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.

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

18. Calculate a p-value as the area under the normal curve beyond the observed sample mean (either in one tail or both, depending on the alternative hypothesis). Note that in doing so you can use a Z score, where

$$Z = \frac{sample\ statistic - null\ value}{SE} = \frac{\bar{x} - \mu_0}{SE}$$

- 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.

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

20. 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 strong evidence for the alternative hypothesis. The smaller the p-value, the stronger the evidence.
- If the p-value > the significance level, do not reject the null hypothesis since this means that obtaining a sample statistics at least as extreme as the observed data is reasonably likely to happen by chance, and conclude that the data does not provide strong 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.

21. 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.

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

- Use a smaller α to reduce the prospect of a Type 1 error (but raise the prospect of a Type 2 error)
- Use a larger α to reduce the prospect of a Type 2 error (but raise the prospect of a Type 1 error)

Test yourself:

1. List errors in the following hypotheses: $H_0 : \bar{x} > 20$ and $H_A : \bar{x} \geq 25$
2. What is wrong with the following statement?
   "If the 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 "The average height of adult males in the US is 69.1 inches", 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 40 adult males, and finds a sample average of 70.2 inches, and a p-value of 0.0082. What should she conclude?
   (d) Interpret this p-value (as a conditional probability) in context of the question.
   (e) Suppose that the true average is in fact 69.1 inches, 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?
4. Go back to Section 1.8 and describe the differences and similarities between the hypothesis testing procedure using simulation and using theory. Especially discuss how the calculation of the p-value changes while the definition stays the same.

Reading: Section 4.4 of OpenIntro Statistics

23. Notice that sampling distributions of point estimates coming from samples that don't meet the required conditions for the CLT (about sample size, skew, and independence) will not be normal.

Test yourself: explain what is going on in Figure 4.20 of the book (page 186).

Reading: Section 4.5 of OpenIntro Statistics

24. Formulate the framework for statistical inference using hypothesis testing and nearly normal point estimates:

(1) Set up the hypotheses first in plain language and then using appropriate notation.
(2) Identify the appropriate sample statistic that can be used as a point estimate for the parameter of interest.
(3) Verify that the conditions for the CLT holds.
(4) Compute the SE, sketch the sampling distribution, and shade area(s) representing the p-value.
(5) Using the sketch and the normal model, calculate the p-value and determine if the null hypothesis should be rejected or not, and state your conclusion in context of the data and the research question.

25. If the conditions necessary for the CLT to hold are not met, note this and do not go forward with the analysis. (We will later learn about methods to use in these situations.)

Test yourself: 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).

Reading: Section 4.6 of OpenIntro Statistics

Video, slow but thorough: Calculating power, YouTube (29:18)

26. Calculate the required sample size to obtain a given margin of error at a given confidence level by working backwards from the given margin of error.

27. Distinguish statistical significance vs. practical significance.

28. Define power as the probability of correctly rejecting the null hypothesis (complement of Type 2 error), and calculate power of a hypothesis test using a two step process:

i. Find the cutoff value(s) for the point estimate in order to be able to reject the null hypothesis at the given significance level.
ii. Find the probability of getting a random sample of size n that yield the cutoff point estimate calculated in the previous step for a desired effect size.

Test yourself: 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?