Section 3D – Conclusions *(updated version)*

Vocabulary

Population: The collection of all people or objects to be studied.

Sample: Collecting data from a small subgroup of the population.

Statistic: A number calculated from sample data in order to understand the characteristics of the data. For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Hypothesis Test: A procedure for testing a claim about a population.

Null Hypothesis (H_0): A statement about the population that involves equality. It is often a statement about "no change", "no relationship" or "no effect".

Alternative Hypothesis (H_A or H_1): A statement about the population that does not involve equality. It is often a statement about a "significant difference", "significant change", "relationship" or "effect".

Population Claim: What someone thinks is true about a population.

Test Statistic: A number calculated in order to determine if the sample data significantly disagrees with the null hypothesis. There are a variety of different test statistics depending on the type of data.

Sampling Variability: Also called "random chance". The principle that random samples from the same population will usually be different and give very different statistics. The random samples will usually be different than the population parameter.

P-value: The probability of getting the sample data or more extreme because of sampling variability (by random chance) if the null hypothesis is true.

Significance Level (α): Also called the Alpha Level. This is the probability of making a type 1 error. The P-value is compared to this number to determine significance and sampling variability. If the P-value is lower than the significance level, then the sample data significantly disagrees with the null hypothesis and is unlikely to have happened because of sampling variability.

Critical Value: We compare a test statistic to this number to determine if the sample data significantly disagrees with the null hypothesis. If the absolute value of the test statistic is higher than the absolute value of the critical value, then the sample data significantly disagrees with the null hypothesis.

Randomized Simulation: A technique for visualizing sampling variability in a hypothesis test. The computer assumes the null hypothesis is true, and then generates random samples. If the sample data or test statistic falls in the tail, then the sample data significantly disagrees with the null hypothesis. This technique can also calculate the P-value and standard error without a formula.

Conclusion: A final statement in a hypothesis test that addresses the claim and evidence.

<u>Introduction</u>: So far, we have learned many of the key ideas used in hypothesis testing. We learned how to write a null and alternative hypothesis and how to determine if the sample data significantly disagrees with the null hypothesis by comparing the test statistic to the critical value. We also learned how to compare the P-value and significance level in order to judge if the sample data occurred by random chance and whether to reject the null hypothesis or fail to reject.

The last step in a hypothesis test is to write a formal conclusion. A conclusion is a statement about the claim the person made. Do we disagree with that claim (reject the claim) or do we agree with the claim (support the claim)? The conclusion is also a statement about evidence. Do we have evidence to back up our view about the population or not?

Conclusion: A final statement in a hypothesis test that addresses the claim and evidence.

Note: "Reject H_0" or "Fail to reject H_0" is NOT a conclusion. This is a simple statement of what the P-value tells us about the null hypothesis. It does not address evidence and the claim.

Writing conclusions can be difficult. Many statistics students struggle with the logic of the conclusion. Think of it this way.

Suppose we have a low P-value and H_0 is the claim. A low P-value means we have evidence and will reject H_0. Since H_0 is our claim, then our conclusion will be that we should "reject the claim". In other words, we have evidence that the random sample data significantly disagrees with the claim. In this case, our conclusion will be that "we have significant evidence to reject the claim".

Suppose we have a low P-value and H_A is the claim. Again, a low P-value means that we have evidence and we will reject H_0. Disagreeing with H_0 is equivalent to agreeing with or supporting H_A. So our conclusion will be that "we have evidence to support the claim". In other words, random sample data significantly agrees with the claim.

Suppose we have a high P-value and H_0 is the claim. A high P-value means that we do not have evidence and we will fail to reject H_0. Since H_0 is our claim, then we should think of our conclusion as "failing to reject the claim". So our conclusion will be that "we do NOT have evidence to reject the claim". In this case, the random sample data is close to H_0 and does not significantly disagree with the claim.

Suppose we have a high P-value and H_A is the claim. A high P-value is not evidence and means we will fail to reject H_0. You have to be able to reject H_0 in order to support H_A. Since our random sample data does not significantly disagree with H_0, it also does not significantly agree with H_A. So, our conclusion is that we do not have significant evidence to support the claim. The random sample data does not significantly agree with the claim.

Let's look at some steps for writing a formal conclusion to a hypothesis test.

Step 1: Address the Claim

If the null hypothesis (H_0) is the claim: *There are two possibilities.*

- Yes, we have evidence to reject the claim
  OR
- No, we do not have evidence to reject the claim.

If the alternative hypothesis (H_A) is the claim: *There are two possibilities.*

- Yes, we have evidence to support the claim
  OR
- No, we do not have evidence to support the claim.

Step 2: Address the evidence (Yes or No)

So how do we know if we have evidence or not?

We have already seen that a low P-value (less than the significance level) from unbiased random sample data passing conditions indicates significant evidence. Remember, scientists usually require a low P-value passing conditions as evidence on their reports. If we have a high P-value greater than the significance level or if the sample data was biased or did not pass conditions, then we do NOT have significant evidence.

Note: While a low P-value from good data is the usual method for determining significance evidence, a test statistic falling in the tail determined by the critical value also indicates significance evidence. Remember a test statistic in the tail will correspond to a small P-value.

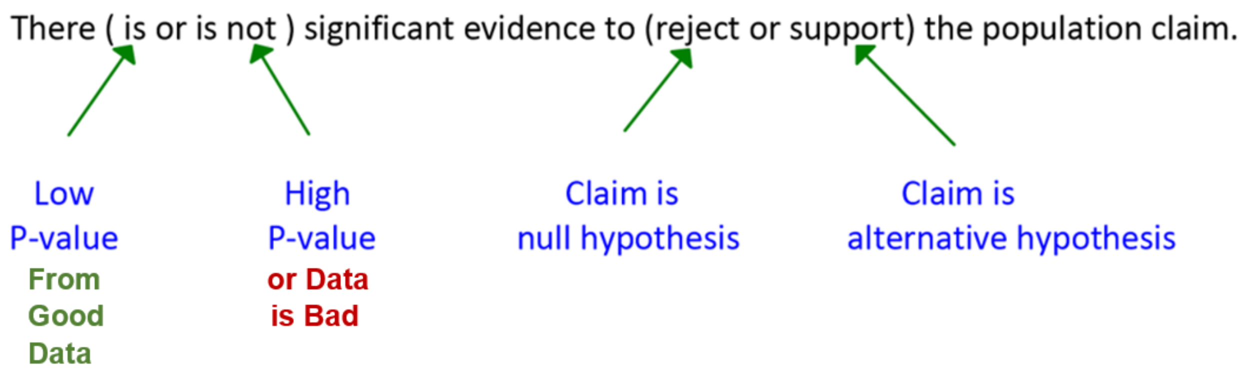
Low P-value (less than significance level) from Good Data: We have significant evidence.

Low P-value from Bad Data: We do NOT have significant evidence.

High P-value (higher than significance level) from Good or Bad data: We do NOT have significant evidence.

<u>Step 3: Write the conclusion sentence</u>

Remember a low P-value is considered significant statistical evidence but a high P-value is not evidence. When the claim is H_0, we will either be rejecting or not rejecting the claim. When the claim is H_A, we will either be supporting or not supporting the claim.



Notice there are 4 Possible Conclusions.

- If the claim is H_0 , P-value from good data is low *(Think "Yes Evidence Reject")*
  Conclusion Sentence: There <u>is</u> significant evidence to <u>reject</u> the claim.
  (Data indicates that the null hypothesis is wrong and the null hypothesis is the population claim. So we thing the population claim is wrong and we have evidence.)

- If the claim is H_0 , P-value is high or data is bad *(Think "No Evidence Reject")*
  Conclusion Sentence: There is <u>not</u> significant evidence to <u>reject</u> the claim.
  (You cannot tell if the null hypothesis is wrong and the null hypothesis was the population claim. Data is inconclusive or bad. You cannot tell if the population claim is wrong.)

- If the claim is H_A , P-value from good data is low *(Think "Yes Evidence Support")*
  Conclusion Sentence: There <u>is</u> significant evidence to <u>support</u> the claim.
  (Data indicates that the null hypothesis is wrong and the alternative hypothesis is correct. Since the alternative hypothesis is the population claim, we thing the population claim is correct and we have evidence.)

- If the claim is H_A , P-value is high or data is bad *(Think "No Evidence Support")*
  Conclusion Sentence: There is <u>not</u> significant evidence to <u>support</u> the claim.
  (You cannot tell if the alternative hypothesis is correct and the alternative hypothesis was the population claim. Data is inconclusive or bad. You cannot tell if the claim is correct.)

Step 4: Explain the conclusion sentence

Our job as data scientists, statisticians and data analysists is to explain. People rarely understand the language and difficult ideas in statistics. The conclusion is a summary of the hypothesis test, but is rarely understood. It is always good to explain the conclusion in plain language. Here is a summary table to help.

Hypothesis Test Conclusion Table

	Reject Ho or Fail to reject Ho?	Conclusion if Claim is Null Hypothesis (Ho)	Conclusion if Claim is Alternative Hypothesis (Ha)	Explain: What does the hypothesis test tell us?
HIGH P-VALUE + GOOD DATA (P-value higher than the significance level AND random sample data DOES meet the conditions / assumptions for the hypothesis test and is relatively unbiased.)	Fail to Reject Ho	There is NOT significant evidence to REJECT the claim that …	There is NOT significant evidence to SUPPORT the claim that…	**Sample Statistic from good unbiased random sample data does NOT significantly disagree with the null hypothesis and so cannot support the alternative hypothesis.**
HIGH P-VALUE + BAD DATA (P-value higher than significance level BUT the data does NOT meet the conditions / assumptions for the hypothesis test OR has other sources of bias.)	Fail to Reject Ho	There is NOT significant evidence to REJECT the claim that …	There is NOT significant evidence to SUPPORT the claim that…	**Sample Statistic from bad biased sample data does NOT significantly disagree with null hypothesis.** P-values calculated from bad biased data should not be taken as evidence to make decisions about a population claim.
LOW P-VALUE + GOOD DATA (Low P-value AND random sample data DOES meet the conditions / assumptions for the hypothesis test and is relatively unbiased.)	Reject Ho	There IS significant evidence to REJECT the claim that…	There IS significant evidence to SUPPORT the claim that…	**Sample Statistic from good unbiased random sample data significantly disagrees with the null hypothesis and supports the alternative hypothesis.**
LOW P-VALUE + BAD DATA (P-value lower than significance level BUT the data does NOT meet the conditions / assumptions for the hypothesis test OR has other sources of bias.)	Fail to Reject Ho	There is NOT significant evidence to REJECT the claim that …	There is NOT significant evidence to SUPPORT the claim that…	**Sample Statistic from bad biased sample data significantly disagrees with null hypothesis.** P-values calculated from bad biased data should not be taken as evidence to make decisions about a population claim.

Example 1

A nursing magazine recently claimed that the population mean average amount of a particular medicine that is being given to patients is about 100 milligrams. Looking at a large unbiased random sample passing conditions, we found a P-value of 0.0041 and a 5% significance level (α = 0.05) was used in the study. What would be the conclusion?

H_0: μ = 100 milligrams (Claim)
H_A: μ ≠ 100 milligrams

Step 1: Address the Claim

The claim is H_0 , so there are two possible conclusions.

- Yes, we have evidence to reject the claim
  OR
- No, we do not have evidence to reject the claim.

Step 2: Address the Evidence

The P-value = 0.0041 = 0.41% is less than the 5% significance level. So, this is a low P-value close to zero. This means that sampling variability (random chance) is very unlikely and tells us the sample data significantly disagrees with the null hypothesis. We would reject the null hypothesis. Since the claim is the null hypothesis, we have evidence to reject the claim.

Step 3: Writing the conclusion sentence:

"There (*is or is not*) significant evidence to (*reject or support*) the claim."

Notice this problem has a low P-value ("is evidence") and claim is null ("reject").

Formal Statistics Conclusion: "There is significant evidence to reject the claim that the mean average amount of this medicine given to patients is 100 mg."

Step 4: Explain the conclusion in plain language?

Significant statistical evidence disagrees with the claim that the population mean average is 100 mg.

<u>Example 2</u>

An online article is currently estimating that more than 35% of people in the U.S. voted in the last election. Looking at a large unbiased random sample passing conditions, we found a P-value of 0.267 and a 10% significance level (α = 0.10) was used in the study. What would be the conclusion?

H_0: p = 0.35
H_A: p > 0.35 (claim)

Step 1: Address the Claim

The claim is H_A , so there are two possible conclusions.

- Yes, we have evidence to support the claim
  OR
- No, we do not have evidence to support the claim.

Step 2: Address the Evidence

The P-value = 0.267 = 26.7% is more than the 10% significance level. So this is a <u>high</u> P-value and is <u>not</u> close to zero. This data could have happened because of sampling variability (random chance) and tells us the sample data does <u>not</u> significantly disagrees with the null hypothesis. We would fail to reject the null hypothesis. We do not have evidence to reject the null hypothesis. You have to reject the null hypothesis to be able to support the alternative hypothesis. So, we do <u>not</u> have evidence to <u>support</u> the claim.

Step 3: Writing the conclusion sentence:

"There (*is or is not*) significant evidence to (*reject or support*) the claim."

Notice this problem has a high P-value ("NOT evidence") and claim is null ("support").

Formal Statistics Conclusion: "There is <u>not</u> significant evidence to <u>support</u> the claim that more than 35% of people in the U.S. voted in the last election."

Step 4: Explain the conclusion in plain language?

The hypothesis test was inconclusive. Random unbiased sample data does not significantly agree with the population claim. We do not have significant statistical evidence to agree with the population claim.

Important Notes:

- The claim being the null hypothesis does not in itself mean that you will reject the claim. It may be a "not reject" situation. You would only reject the claim if the claim is H_0 and the P-value was low and came from good unbiased data.
- The claim being the alternative hypothesis does not in itself mean that you will support the claim. It may be a "not support" situation. You would only support the claim if the claim is H_A and the P-value was low and came from good unbiased data.
- Always address the "claim" in a conclusion. Never say that you support H_A or you reject H_0. That is not a conclusion.

Conclusions from Bad Biased Data

So far, we have looked at conclusions based on P-values calculated from data that was relatively unbiased and met the conditions for the hypothesis test. The sample size was not too small and the data was collected in a manner that reflects the population.

However, what do we do with P-values calculated from biased data where the data does NOT meet the conditions for the hypothesis test? The sample size may be too small or maybe the data was collected with a poor method like convenience or voluntary response.

Usually when data is biased and does not pass the conditions for the hypothesis test, we will NOT do the hypothesis test at all. Let us say a person did do the test though. What could we say based on a P-value or Test Statistic calculated from bad biased data?

The key is that you do NOT have evidence. We should NOT make a decision about a population based on bad biased data no matter what the P-value is.

- A low P-value from unbiased data that meets the conditions of the hypothesis test will be significant evidence. The sample statistic (sample data) significantly disagrees with the null hypothesis. Therefore, we feel pretty confident that we should reject the null hypothesis and support the alternative hypothesis.
- A low P-value from bad biased data that does NOT meet the conditions of the hypothesis test should NOT be considered significant evidence. The sample statistic (sample data) significantly disagrees with the null hypothesis but this data does not reflect the population. Therefore, we cannot be confident that the null hypothesis is wrong. Therefore, we would fail to reject the null hypothesis. And would not have significant evidence to support the alternative hypothesis.

Note about Bias: Remember in section 1C we learned that data can be biased for all sorts of reasons. Hypothesis test conditions help us judge sample size requirements and the method of collecting data but do not rule out all bias. It is always good to think about possible sources of bias even if the data passes all of the conditions for the hypothesis test.
