Section 3F – One-population Mean and Proportion Hypothesis Tests

Vocabulary

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

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

Random Sample: Sample data collected in such a way that everyone in the population has an equal chance to be included.

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.

One-Population Proportion Test Statistic (z): The sample proportion is this many standard errors above or below the population proportion in the null hypothesis.

One-Population Mean Test Statistic (t): The sample mean is this many standard errors above or below the population mean in the null hypothesis.

Critical Value: A number we compare our test statistic to in order to determine significance. In a sampling distribution or a theoretical distribution approximating the sampling distribution, the critical value shows us where the tail or tails are. The test statistic must fall in the tail to be significant.

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.

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.

Rejecting the null hypothesis: Random sample data significantly disagrees with the null hypothesis.

Failing to reject the null hypothesis: Random sample data does not significantly disagree with the null hypothesis.
This does not prove that the null hypothesis is correct however.

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

Introduction

So far, we have been learning all of the pieces and ideas behind a hypothesis test. In this section, we will finally start to put the pieces together and perform a hypothesis test from start to finish. While different tests have different test statistics, null and alternative hypotheses, and assumptions, the ideas and methods are very similar. Let's start by looking at the steps for performing a hypothesis test.

Steps to perform a Hypothesis Test

1. Write the null and alternative hypothesis.

Write down the population claim and the null and alternative hypothesis. Consider the type of data you will need and the number of populations to decide the appropriate test for the situation. Is this a right tailed, left tailed or two tailed test? What test statistic should we use? Consider the type 1 and type 2 errors and pick a significance level.

2. Collect random sample data and check the assumptions.

If the sample data has not been collected yet, collect random sample data that is appropriate. Once the sample data is collected, check the assumptions to make sure the data represents the population and can be used for the hypothesis test. You may need to create a histogram to check normality. If certain assumptions are not met, consider revising your hypothesis test method. You may need to move to a non-parametric test or use a randomization technique.

3. Use computer technology to calculate the test statistic, critical value, and P-value and use them to interpret significance and sampling variability.

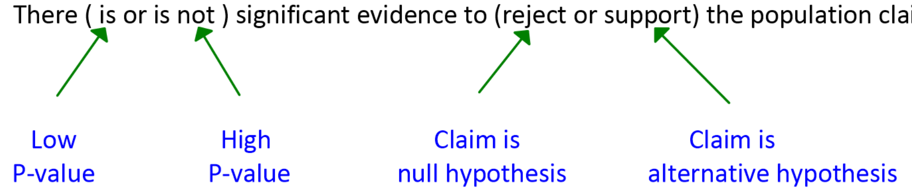
Compare the test statistic to the critical value. If the test statistic falls in the tail, then the sample data significantly disagrees with the null hypothesis. Compare the P-value to your chosen significance level. If the P-value is less than the significance level, then it is unlikely to be sampling variability.

4. Determine if we should reject or fail to reject the null hypothesis.

Remember the P-value is calculated from the assumption that the null hypothesis is true. In a sense, this is about seeing what your P-value tells you about the null hypothesis. If the P-value is less than the significance level, then we will be rejecting the null hypothesis. If the P-value is higher than the significance level, then we will be failing to reject the null hypothesis. Failing to reject does not mean that the null hypothesis is correct.

5. Write the formal conclusion addressing the claim and evidence.

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



6. Explain the conclusion in plain language that non-statisticians can understand.

Does the random sample data disagree with claim or agree with the claim? Is it a significant disagreement or a significant agreement? Do we have evidence? What kind of evidence?

Assumptions

In this chapter, we have been discussing test statistics, critical values, theoretical distributions (curves) and P-values. These are the building blocks of hypothesis tests. A key question to think about is how accurate are these building blocks? The standard error is the standard deviation of a sampling distribution. What if that sampling distribution is not normal? The standard error may not be very accurate. To deal with theoretical curves and formulas that involve standard error, the sample data must meet certain assumptions. The good news is that the assumptions for one or two population hypothesis tests are the same as the assumptions we learned for confidence intervals. In our next chapter, we will look at more advanced hypothesis tests where the assumptions get more complicated. Remember the data must meet all of the assumptions for us to use the traditional hypothesis test methods.

Notice that hypothesis tests have some assumptions in common.

Random

One is that the data should be a random sample that is unbiased and represents the population. Statistics students often ask about whether the data can be an unbiased census. The answer is no, but this does not mean that a census is not great data. It is the best. In some ways it is too good. If you have an unbiased census then you know what is going in the population. There is no need for a hypothesis test to surmise what may be true about the population, you know the population parameters. A hypothesis test uses sample data to try to understand population when a census is not possible.

Independence

Another assumption that is particularly difficult to check is that the data values within each sample must be independent of each other. This means that the individuals do not have something in common, at least not for the variable you are studying. This usually rules out the same people measured twice or family members or people from the same classroom or on the same Facebook page. If you know the characteristics of one person, it should not mean that the next person in your data set has a higher probability of that characteristic. A simple random sample from a large population will usually pass the independence criteria. Data scientists usually like the population to be at least ten times larger than the sample size for the individuals to be independent. Think about it. If we collect a simple random sample of 350 people from a population of millions, it is very unlikely that I will accidentally get family members or people that know each other.

Sample Size Requirements

Most assumptions also have some sort of way of checking that the data set is large enough to ensure that it matches the theoretical curve. For proportions, this is usually having at least ten successes and at least ten failures. So ten or more people or objects in the sample data have the categorical criterial you are studying and ten or more do not. If the sample data has at least ten successes and failures, then the sampling distribution for sample proportions is likely to look normal, the standard error will be roughly accurate and the Z-distribution will be a good fit. For mean averages, this is usually having a sample size greater than or equal to 30 or nearly normal. If one of these two criterial is met, then the sampling distribution for sample means is likely to look normal, the standard error will be roughly accurate, and the T-distribution will be a good fit.

Key Question: If we have not collected the data yet, is there a way to know if the sample size is large enough for a given hypothesis test?

Absolutely. If the random sample data has not been collected yet, we can determine what sample sizes should be appropriate. It is generally at least 30 for mean average hypothesis tests, but is more difficult for proportions. We need to collect enough data to ensure that we will get at least ten successes and at least ten failures. Since we have an approximate population proportion (π) in the null hypothesis, we can use this to calculate how many success and failures I am likely to get.

Expected number of successes = $n \times \pi$

Expected number of failures = $n \times (1 - \pi)$

For example, suppose the null hypothesis is π = 0.20. Will a sample size of 120 randomly selected people be enough?

Expected number of successes = $n \times \pi$ = 120 × 0.2 = 24

Expected number of failures = $n \times (1 - \pi)$ = 120 × (1 – 0.2) = 120 × 0.8 = 96

Since both of these are greater than ten, a random sample of 120 should be enough.

You can also turn these formulas around and calculate the minimum sample size to get ten. I don't like these too much because you want your expected to be significantly greater than ten. If you calculate for an expected number of success and failures of exactly ten, you may get 8. (We don't always get what we expect in random data.)

In the previous example, here are the minimum sample size calculations.

Minimum sample size (n) based on number of successes and population proportion = $\frac{10}{\pi}$ = $\frac{10}{0.2}$ = 50

Minimum sample size (n) based on number of failures and population proportion = $\frac{10}{(1-\pi)}$ = $\frac{10}{0.8}$ = 12.5

Always take the larger of the two calculations. In this example, the population proportion was 0.2. We will need a random sample of above 50 if we expect to get at least ten successes and failures.

One-Population Hypothesis Test Assumptions

One-population Mean Assumptions

- The quantitative sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- The sample size should be at least 30 or have a nearly normal shape.

One-population Proportion Assumptions

- The categorical sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- There should be at least ten successes and at least ten failures.

Example 1

Tim needed to collect some data to test a claim that the population mean average monthly cost of living in California is higher than in Nevada. He put a survey on Facebook. He received data from 217 people living in California and 168 people living in Nevada. The shape of both of the data sets was skewed right. Can we use this data to test perform a two-population mean T-test?

Are the data sets random or representative? *No. This was voluntary response data.*

Are both samples at least 30 or Normal? *Yes. The data sets were not bell shaped (skewed right) but since both of the sample sizes were over 30 (217 and 168) it does pass the 30 or normal requirement.*

Are the data values within the samples independent? *No. These people came from the same Facebook page. They may know each other and even have similar jobs, socioeconomic levels and salaries.*

Are the data values between the samples independent? *No. These California and Nevada samples came from the same Facebook page. They may know each other and even have similar jobs, socioeconomic levels and salaries.*

Since this data did not pass all of the assumptions, we should not use it to judge a population claim with a hypothesis test.

Example 2

Julie needs to test a claim about the percent of people in her hometown that will vote. She took a simple random sample of 45 people and found that 32 said they would vote and 13 said they would not. Can we use this data to test a claim about the population proportion (percentage)?

Is the sample data random or representative? *Yes. This was a simple random sample.*

Are there at least 10 success? *Yes. There are at least 10 success (32 said they will vote)*

Are there at least 10 failures? *Yes. There are at least 10 failures (13 said they will not vote)*

Are the data values within the samples independent? *Yes. There were 45 people out of a relatively large population. So it is unlikely that they would be family members. One person saying they vote will not affect the probability of other people voting. They did come from the same town, but that is the population of interest, so will not be a factor in independence.*

Since the data met the assumptions, we can use this data to test the claim about the population proportion with a hypothesis test and the Z-test statistic. We just need the number of events (voters) = 32 and the total number of trials = 45.

What if the data does not meet all of the traditional assumptions?

Are there ways of working with sample data that does not meet the traditional assumptions?

Absolutely. The one assumption that really needs to be there for any population hypothesis test is that the sample data should be representative, unbiased, and collected randomly. If we are conducting an experiment to prove cause and effect, we will need to separate our groups with random assignment. Besides that, there are many options.

Many advanced techniques in statistics were developed to deal with not meeting all of the traditional assumptions. Non-parametric hypothesis tests were invented for just this purpose. They generally have less assumptions and can be used in a variety of situations. I am a big fan of the non-parametric "Mann-Whitney" test when comparing population averages. Non-parametric hypothesis tests are usually covered in more advanced statistics courses.

One technique that we have discussed is randomized simulation or a randomization test. In some ways, this falls under the umbrella of non-parametric tests. In the section on P-value, we saw that we can create a simulated sampling distribution based on the premise that the null hypothesis is true. We can then calculate the P-value directly from the simulation. This technique does not require the data to match up with a theoretical distribution curve, so randomization often does not have as many requirements as traditional approaches to hypothesis testing. Notice that there is no requirements for sample size or a normal shape.

Assumptions for a Randomization (Randomized Simulation) Hypothesis Test

- The sample data should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- If multiple samples were collected that were not matched pair, then the data values between the samples should be independent of each other.

Using Technology to Calculate Test Statistics, Critical Values and P-value

Always use technology to calculate test statistics, critical values and P-values. We may need to calculate something from time to time, but always use a computer to do the bulk of the calculations. Your job as a data analyst is to analyze and explain data, not to calculate.

In this book, we have been using two free programs, StatKey and Statcato. Here are some basic directions for using Statcato. Statcato printouts will usually be provided for you to analyze in the problems.

Statcato One-population percentage (proportion) Hypothesis Test with Z-test statistic

Statistics => Hypothesis Tests => 1-Population Proportion

- "Samples in Column" (If you have raw categorical data, enter the column. For example "C1".)
- "Summarized Sample Data" (Use this option if you have the number of events, total number of trails).
- Put in significance level as a decimal proportion.
- "Hypothesized proportion" (This is asking for the population proportion (π) in H_0)
- "Alternative Hypothesis" (Less than, greater than, not equal. This tells the computer whether to do a left tail, right tail or two-tail hypothesis test.)

Statcato One-population mean average Hypothesis Test with T-test statistic

Statistics => Hypothesis Tests => 1-Population Mean

- "Samples in Column" (If you have raw quantitative data, enter the column. For example "C1".)
- "Summarized Sample Data" (Use this option if you have the sample size, sample mean and sample standard deviation).
- Put in significance level as a decimal proportion.
- "Hypothesized mean" (This is asking for the population mean (μ) in H_0)
- "Alternative Hypothesis" (Less than, greater than, not equal. This tells the computer whether to do a left-tail, right-tail or two-tail hypothesis test.)
- Leave population standard deviation as "unknown".

Example 1 (One Population Proportion Z-test)

Ex) A doctor thinks that the population percent of people in in his city that have a certain infection is about 6%. He took a simple random sample of 175 people in the city and found that 13 of them had the infection. Use the following Statcato printout to test the doctor's claim that exactly 6% have the infection. (Use a 5% significance level.)

Hypothesis Test - One population proportion: confidence level = 0.95
Input: Summary data
Null hypothesis: p = 0.06
Alternative hypothesis: p ≠ 0.06

N	Sample Proportion	Significance Level	Critical Value	Test Statistic Z	p-Value
175	0.074	0.05	-1.96, 1.96	0.796	0.4262

Null and alternative hypothesis?

H_0 : π = 0.06 (claim)
H_A : π ≠ 0.06

Type of hypothesis test? One population proportion Z-test (two tail)

Assumptions? The data did pass all of the assumptions, so we can proceed with the hypothesis test.

- The sample data was collected randomly.
- The sample data was sufficiently large since it had at least ten success and failures. The data had 13 success and 175-13 = 162 failures.
- Individuals were likely to be independent since it was a simple random sample out of a large population. It is unlikely that the individuals in the sample data will be related.

Test stat Z = 0.796

Test Stat Sentence: The sample percent 7.4% was only 0.796 standard errors above the population value 6%.

- Z - Test Statistic does not fall in the tail and so is not significant. (The sample proportion needs to be higher than +1.96 or more to be in the right tail (significantly higher) or −1.96 or less to be in the left tail (significant lower.)
- This tells us that the sample value (7.4%) was not significantly different from the population value (6%)
- The sample data does not significantly disagree with the null hypothesis.

P-value = 0.426

P-Value Sentence: If Ho is true, and the population percent really is 6%, we had a 42.6% probability of getting the sample percentage of 7.4% or more extreme by random chance.

- This is a high P-value. (The P-value of 42.6% is much larger than the 5% significance level.)
- If H_0 is true, this tells us that the sample data could have happened by random chance (sampling variability).
- Sample value of 7.4% is not significantly different from the population value of 6%.
- There is not a significant disagreement between the sample data and the null hypothesis.

Reject H_0 or Fail to reject H_0? Fail to reject H_0 since the P-value is larger than the significance level.

Conclusion?

There is not significant evidence to reject the claim that 6% of the population have the infection.

(The random sample data does not significantly disagree with the claim that 6% of the city is infected. The doctor might be correct, but we do not have evidence.)
