

# Chapter 6: Inferential Methods for a Single Numerical Variable

As seen in Chapters 1-4, research hypotheses involving either a single categorical variable or two categorical variables require us to test claims about proportions. When a research question involves a single numerical variable, however, we end up testing claims about the mean, or average. In this chapter, we will consider inferential methods (hypothesis tests and confidence intervals) for the mean of a single numerical variable. Consider the following example.

## Example 6.1: Time Perception Impaired by Nicotine Withdrawal

A study conducted by researchers at Pennsylvania State University investigated whether time perception, a simple indication of a person's ability to concentrate, is impaired during nicotine withdrawal. The study results were presented in the paper "Smoking Abstinence Impairs Time Estimation Accuracy in Cigarette Smokers" (Psychopharmacology Bulletin [2003]: 90-95). After a 24-hour smoking abstinence, 20 smokers were asked to estimate how much time had passed during a 45-second period. Suppose the resulting data on perceived elapsed time (in seconds) were analyzed as shown below (these results are artificial but are similar to the actual findings).

We can summarize the data with:

1. What is the mean of the observed data? The standard deviation?
2. If another sample of  $n=20$  subjects were obtained, would these new subjects have a mean exactly the same as the mean from this sample? Why or why not?
3. Given your answer to the previous question, do you think it is appropriate to use only this sample mean to make inferences about the mean perceived elapsed time in the greater population of all smokers subjected to nicotine withdrawal? Explain.

## The Distribution of the Sample Mean

The sample mean is a random quantity; that is, it changes from sample to sample. Therefore, the sample mean actually has its own distribution. This distribution will tell us two things:

1. What values the sample mean can assume
2. How often it will assume these values

This distribution is referred to as the *distribution of the sample mean*. An understanding of this distribution allows us to make decisions about a population mean for a single numerical variable.

Before we discuss the procedure for inference, let's consider the next example to gain a better understanding of how the distribution of sample means works and how we use this distribution to make a decision concerning our research question.

## The Central Limit Theorem (CLT)

Note that a statistician would not necessarily carry out a simulation study such as this to answer a research question. Instead, one could use a "short-cut" known as a t-test to investigate a research question concerning a single population mean. This short-cut is a result of something known as the central limit theorem, which states the following:

Consider a random sample of  $n$  observations from ANY population with mean  $\mu$  and standard deviation  $\sigma$ . When  $n$  (the number of subjects in the sample) is sufficiently large, the distribution of sample means will be approximately a NORMAL distribution with a mean of  $\mu$  and a standard deviation of  $\frac{\sigma}{\sqrt{n}}$ . This approximation gets better as the sample size ( $n$ ) increases.

We can see the Central Limit Theorem applied to the sample means calculated from our hypothetical population from Example 6.2 as follows:

- We see that the distribution of sample means is approximately normal:
- We see that this normal distribution is centered at the true population mean,  $\mu = 45$ .
- The standard deviation of all observations in our hypothetical population is  $\sigma = 9.4$  seconds. According to the central limit theorem, then, the standard deviation of the distribution of sample means is given by

Finally, recall from the previous chapter that given the mean and standard deviation of a distribution, we can determine whether a given observation is an outlier or not based on its position on this distribution. This will help us decide whether the sample mean *actually observed* in the research study is an outlier on the distribution that assumes the null hypothesis is true.

The only question that remains is this: How large does  $n$  (the number of subjects in a study) have to be in order for us to use the short-cut provided by the Central Limit Theorem?

### **i** Sample Size for the Central Limit Theorem

For almost all populations, samples of size  $n = 30$  or  $40$  subjects will be sufficient to say that the distribution of sample means is approximately normal. However, if the distribution is very skewed, the sample size may have to be much larger than  $30$  in order for the central limit theorem to apply.

In summary, we can use the Central Limit Theorem to help us create a procedure for comparing a population mean to some hypothesized value.

This works because:

1. We know the distribution of sample means will be approximately normal if either (i) the original population is normally distributed, or (ii) our sample size is sufficiently large.
2. We know the distribution of sample means will be centered at the true population mean (which we can set to some hypothesized value in the null hypothesis).
3. We know that the variability in the distribution of sample means is given by (i.e., the variability decreases as the sample size gets larger, which we can see in the above examples).

In the next section, we put all of the pieces together to create what is known as the *t*-test.

## **The *t*-test for a Single Population Mean**

Back to **Example 6.1**: Recall that the researchers wanted to show that the mean perceived elapsed time for smokers suffering from nicotine withdrawal was greater than the actual 45 seconds that had elapsed. The data collected in the study were summarized as follows:

1. Set up the null and alternative hypotheses

$H_0$ :

$H_a$ :

2. Find the *t*-statistic and the *p*-value

To determine whether or not the distance between  $\mu$  (the hypothesized population mean) and (the mean from our observed sample) is larger than what we would expect by random chance, we will use the following statistic:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{\text{observed sample mean} - \text{mean of the distribution of sample means}}{\text{standard deviation of the distribution of sample means}}$$

Why use this statistic? Because this quantity measures the position of our observed sample mean on the null model, just like the Z-score discussed in the previous chapter.

Note that this is very much like the Z-score, with one minor exception. We don't know the true population standard deviation,  $\sigma$ , so we estimate it with the standard deviation calculated from the 20 observed subjects in the study (this estimate is commonly denoted by  $s$ ).

This t-statistic comes from what is called a t-distribution. The amount of variability in a t-distribution depends on the sample size  $n$  (the greater the sample size, the smaller the standard deviation of the distribution of sample means). Therefore, this distribution is indexed by its *degrees of freedom* (df).

#### **i** Degrees of Freedom for a Single Mean

For inference on a single mean,  $df = n - 1$ .

To find the p-value associated with this test statistic, we must remember that this is an upper-tailed test (we are trying to find evidence that the mean is greater than 45 seconds). So, the p-value will be the probability we would observe a sample mean (or a t-statistic) greater than that obtained in the actual study by chance alone, assuming the null hypothesis is true:

3. Write a conclusion in context of the problem.

#### **i** Checking the Normality Assumption:

For the t-test to be valid, at least one of the following conditions must be met:

- Either the sample size is sufficiently large (greater than 30 or so), or
- The distribution of the observed data is approximately normal (which would indicate that the population is normally distributed so that the Central Limit Theorem would apply even with a small sample size)

For Example 6.1, we have a sample size of 20 subjects, which is not sufficiently large. So we must check whether the data seem to come from a normal distribution. The histogram seems to indicate that this is a reasonable assumption:

### **Example 6.3: Time Perception for Smokers**

For the data given in Example 6.1, we found evidence that the mean perceived elapsed was in fact greater than the actual 45 seconds that had elapsed. This study alone, however, doesn't really prove that the nicotine withdrawal was what impaired one's perception of time. Why not?

Suppose that the researchers also studied 20 subjects who were smokers that did NOT abstain from smoking prior to the data collection (so, they were not suffering from nicotine withdrawal). Data on their perceived elapsed times are given in the file Nicotine2.JMP.

**Research Question:** Is there evidence the mean perceived elapsed time for all smokers not suffering from nicotine withdrawal is significantly greater than the actual 45 seconds?

Carry out the formal t-test to address this research question.

1. Set up the null and alternative hypotheses

Ho:

Ha:

2. Check normality assumptions.
3. Find the t-statistic and the p-value
4. Write a conclusion in the context of the problem.

### Confidence Interval for a Single Population Mean

In **Example 6.1**, we found evidence that the mean perceived elapsed time for smokers suffering from nicotine withdrawal significantly differed from the actual 45 seconds of time that had elapsed. Our next question is obvious: HOW MUCH does it differ? To answer this question, we must construct a confidence interval.

Recall our discussion of confidence intervals from earlier in the quarter:

This procedure does NOT require any hypotheses concerning our population parameter of interest (the mean, in this case). We will use both our sample data (in particular, the observed mean) and what we know about the distribution of sample means to obtain a range of likely values for our population mean.

#### Warning

1. A confidence interval allows us to estimate the population parameter of interest (recall that the hypothesis test does NOT allow us to do this). Therefore, when available, a confidence interval should always accompany the hypothesis test.
2. The confidence interval does not require any hypothesized value for the population parameter. Instead, we center the confidence interval on the sample mean. Consider the following example.

### Example 6.4: Estimated Perceived Time from Nicotine Withdrawl

Our goal is to construct a 95% confidence interval for the mean perceived elapsed time for smokers suffering from nicotine withdrawal. To do this, we will center our distribution of sample means on the observed mean. Then, we will find the lower and upper endpoints that separate the middle 95% of the distribution from the rest (since we are constructing a 95% confidence interval).

The formula for calculating the endpoints of this confidence interval is given as follows:

The appropriate t-quantile can be found using JMP. To find this value, you need the following information:

- confidence level =
  - df =
1. Interpret the meaning of this interval. What does this interval tell us about the true mean perceived elapsed time for all smokers that are suffering from nicotine withdrawal?
  2. Does this interval agree with what you learned from the hypothesis test? Explain.
  3. How would your calculations change if you wanted to obtain a 90% confidence interval, instead?

### More On The Interpretation Of Confidence Intervals

The 95% refers to the process of constructing the confidence interval. This means that if we were to take many samples of size 20 from the population, constructing a confidence interval each time, we would expect 95% of them to capture the true population mean. Consider the following example:

### Example 6.5: Confidence Interval Coverage

Our goal is to take samples from a population in order to estimate the true population mean. Shown below are 10 random samples of size  $n = 5$ . Construct a confidence interval for each of the samples.

A graphical representation of the intervals is presented below:

1. Why are some of the 90% confidence intervals wider than others?
2. In truth, these 10 random samples were generated from a population with a mean of 10. How many of the confidence intervals captured this true mean? What does it mean to say that we are 90% confident?