

# Lecture 05 Part 03 - Single Sample Inference

Data Analysis in CJ (CJUS 6103)

## Outline

- I. Logic of Single-Sample Inference
- II. Steps in a Hypothesis Test Redux
- III. On p-Values

## Logic of Single-Sample Inference

Interval estimation and hypothesis testing represent two sides of the same coin. In both, we want to be able to say something about the population based on data that we gather from a sample of that population. Recall that the problem with a sample statistic is that, although it is known, it varies randomly from one sample to the next as a result of sampling error.

When we apply techniques of interval estimation, we begin with the sample mean as our point estimate in order to estimate a plausible range of values for the population mean.

When we apply techniques of hypothesis testing, we make an assumption about what the value of the population parameter is (and sometimes, we actually know what the parameter is). In other words, we ask: In the sampling distribution centered around  $\mu$ , where does the sample mean lie, or how many standard error units does the sample mean lie from  $\mu$ ? We then ask whether our assumption about  $\mu$  is realistic (in a probabilistic sense), given the sample statistic that we obtain from our data.

Thus our question with single-sample inference is this: Is it likely that our sample was drawn from a population with mean  $\mu$ , or is our sample a subset of an entirely different population with different mean  $\mu$ ? We know from sampling theory that it is very unlikely that our sample mean will be identical to the proposed population mean. However, we want to use what we have learned from probability theory to provide what we believe to be a reasonably good answer. There are two possible answers to this question.

Yes, our sample was drawn from the population of interest. In this case, our sample mean was drawn from a sampling distribution of means centered around the population mean that we have assumed to be true. The observed difference between the sample mean and the population mean is due to sampling error, or chance.

No, our sample was drawn from a different population. In this case, our sample mean was drawn from a different sampling distribution of means centered around a different population mean. The observed difference between the sample mean and the population mean is thus due to the fact that there are actually two different populations with two different population means, reflecting the fact that there are systematic differences.

Note that the above question is an inherently probabilistic one (i.e., what is the probability that we would obtain the observed sample mean if  $\mu$  was actually true?). We are trying to determine whether the observed difference is random or systematic. Ultimately, then, we want to use a test statistic that allows us to get probabilities.

# Steps in a Hypothesis Test Redux

In order to illustrate the steps in a hypothesis test, we will walk through an example. Let's consider the Minneapolis crime hot spot example. Among all "places" (addresses and intersections) in Minneapolis in 1986, the mean number of calls to police was 2.82, with standard deviation 2.31. Among a sample of 3795 high-risk places, the mean was 43.03.

## Step 1. Formally State Hypotheses

There are two hypotheses that you must state formally, a null or statistical hypothesis ( $H_0$ ) and an alternative or research hypothesis ( $H_1$ ). These hypotheses must always be stated with respect to the mean of the sampling distribution (or the population mean), not the sample mean. This is the value that we want to compare the sample mean to. It is often useful to start with the alternative hypothesis.

The alternative hypothesis is our research hypothesis, and is stated in such a way that we answer our research question in the affirmative. The alternative hypothesis can be non-directional or directional.

The null hypothesis is what we are interested in testing, and is always stated in such a way that we answer our research question in the negative. If we conduct our hypothesis test and find that there is not much support for the null hypothesis, we must be prepared to accept the alternative hypothesis in its place.

In our example,  $H_1: \mu > 2.82$  and  $H_0: \mu \leq 2.82$ .

## Step 2. Obtain a Probability Distribution

Every test statistic is associated with a probability distribution, or what we also have been referring to as a sampling distribution. Remember that this is a theoretical distribution that provides us the results from an infinite number of trials or samples. It takes on a particular shape (usually normal) and is associated with a range of probabilities, which help us make informed decisions about the null hypothesis. If the null hypothesis is true, we expect our sample statistic to lie relatively close to the hypothesized mean of the sampling distribution. If the alternative hypothesis is true, we expect the sample statistic to lie in the tails of the sampling distribution (or one specific tail in a directional hypothesis test). In order to determine whether to use the  $z$ - or  $t$ -distribution, you must ask yourself whether or not you have the population standard deviation.

In our example, we know  $\sigma$ , so our probability distribution is the  $z$ .

## Step 3. Make Decision Rules

Once we have our theoretical probability distribution, we need to select the level of significance and determine the critical region.

First, we must select a level of significance by stating in advance how often we are willing to falsely reject the null hypothesis. By convention, researchers select a significance level of 0.05 or 0.01, which we refer to as  $\alpha$ . This indicates that we are willing to make a Type I error 5 times out of 100 (or 1 time out of 100). In other words, we are willing to falsely reject the null hypothesis 5% of the time.

Second, we must establish the critical region of our probability distribution. This is the area of our theoretical distribution that will lead us to reject the null hypothesis. The size of this area is determined by our significance level,  $\alpha$ . If we were to state a non-directional alternative hypothesis, we would be conducting what is called a two-tailed significance test. In other words, we would have to split the significance level into two parts, one half (0.025) of it in each tail of the distribution. We would thus have two critical regions. If we were to state a directional alternative hypothesis, we would be conducting what we call a one-tailed significance test. In other words, we would have only one critical region.

There is an important reason that we select a significance level and critical region for our probability distribution. We know that since we have data from a sample, and there is sample-to-sample variation, there is bound to be some error that is due to chance and sampling. So the question is this: How far away must we be from the null hypothesis before we decide that our data are not simply the result of chance? The critical region tells us how many standard error units our sample statistic must be from the hypothesized mean of the sampling distribution for us to decide that it is not simply the result of chance. If our statistic falls within the critical region, we are pretty certain (95% certain, in fact) that this reflects a true difference, and so we are willing to reject the null hypothesis and accept the alternative hypothesis in its place.

In this example, we will use an  $\alpha$ -level of .01 and conduct a one-tailed test (we hypothesized that our sample mean comes from a population with a mean that is greater than 2.82). This means that our critical  $z$ -score is 2.33. We will reject  $H_0$  if our test statistic is greater than 2.33.

#### Step 4. Calculate the Test Statistic

This is the actual quantity that we obtain from our sample data and use to test the null hypothesis. We want to use the value that we obtain from our test statistic and compare it to the probability distribution to determine how likely it is that we obtained the value by chance. In most cases, there will be a formula that we must calculate by hand to derive the test statistic.

The first thing we want to know is whether the population standard deviation is known or unknown. This will help us determine exactly which formula we must apply. In our example, we know  $\sigma$ . So the test statistic is computed as

$$TS = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{43.03 - 2.82}{2.31/\sqrt{3795}} = \frac{40.21}{.037} = 1086.76$$

This value means that our sample mean is 1086.76 standard error units above the hypothesized population mean.

#### Step 5. Make a Decision About the Null Hypothesis

If our test statistic falls within the critical region, we will reject  $H_0$  and accept the alternative hypothesis in its place. If our test statistic falls outside the critical region and lies instead in the acceptance region, we will accept  $H_0$ . In our example, we reject  $H_0$  because 1086.76 is greater than 2.33.

By rejecting the null hypothesis, we accept the alternative in its place. Thus, we conclude that our sample of high-risk places has a significantly higher number of calls for service, on average.

#### Critical Scores and Decision Errors

Here is a useful table to remember critical values for hypothesis tests using the  $z$ -distribution. Note that there is no such table for the  $t$ -distribution, as those values are dependent upon degrees of freedom (which can vary).

Significance Level	One-Tailed Critical Value	Two-Tailed Critical Value
.10	1.282	1.645
.05	1.645	1.960
.01	2.326	2.576
.001	3.090	3.291

In hypothesis testing, there are two types of error that we must be aware of. A type I error (*false positive*) is when we reject the null hypothesis, but in fact it is true. A type II error (*false negative*) is when we

accept the null hypothesis, but it in fact is false. Note that the *false positive* and *false negative* labels can be swapped - it depends on what hypothesis you use to apply the label. For example, we could consider a Type I error to be a *false negative* if the hypothesis we use to apply the label is the null (i.e., we falsely reject (negative) a true null) as opposed to using the alternative hypothesis (i.e., we falsely accept (positive) a false alternative hypothesis). Overall, this terminology is confusing. I find that simplifying it to *Type I Error = to falsely reject a true null* and *Type II error = to falsely accept a false null* is easier to remember.

Reality	Decision	
	Accept $H_0$	Reject $H_0$
$H_0$ is true	Correct decision	Type I error ( $\alpha$ )
$H_0$ is false	Type II error ( $\beta$ )	Correct decision

We can reduce the probability of a type I error by decreasing level of significance. You should recognize, however, that there is a hydraulic relationship between type I and type II errors. If we want to ensure a small probability of falsely rejecting the null hypothesis (say 0.001), we necessarily increase the probability of falsely accepting the null hypothesis. Thus, these two errors are inversely related.

## Additional Example

### Homicide Rates in US

Homicide rates in the U.S. since 1950 have mean ( $\mu$ ) 6.66 with standard deviation ( $\sigma$ ) of 2.03. In 2017, the mean homicide rate among the 50 states was 5.33. The research question is: Has there been a change from the average homicide rate over the last 70 years ( $\alpha = .05$ )? Note that the terminology I use - *change* - does not indicate a direction (higher/lower) so this is a two-tailed hypothesis.

#### Step 1: Formally state hypotheses

$$H_1 : \mu \neq 6.66; H_0 : \mu = 6.66$$

#### Step 2: Obtain a probability distribution

Because the population standard deviation  $\sigma$  is known, we will use the  $z$ -distribution.

#### Step 3: Make decision rules

$$\alpha = .05 \text{ (two-tailed); } z_{\text{crit}} = 1.96; \text{ reject } H_0 \text{ if } |TS| > 1.96$$

#### Step 4: Calculate the test statistic

$$TS = \frac{5.33 - 6.66}{1.97/\sqrt{50}} = \frac{-1.33}{.2786} = -4.77$$

#### Step 5: Make a decision about the null hypothesis

Reject  $H_0$ ; there was a significant change in the homicide rate in 2017 compared to the 70 years prior.

## Unknown Population Standard Deviation

In all of the previous examples, the population standard deviation was known. Thus, we relied on the  $z$ -distribution as our probability distribution. What do we do differently if we do not know the population standard deviation? Recall that when we do not know  $\sigma$ , we must estimate it using  $s$ , the sample standard deviation. Since we have to estimate this additional parameter ( $\mu$  in addition to  $\sigma$ ), we have additional sampling error that we must take into account. We take this additional error into account by adjusting for the **degrees of freedom**, which we determine by  $df = n - 1$ . We also rely on the  $t$ -distribution as our probability distribution when  $\sigma$  is unknown, instead of the  $z$ -distribution.

### Training School Youth Example

We want to know if training school youths are *more likely* to offend than non-training school youths. Let's say that we have a random sample of 121 training-school youths ( $n = 121$ ), among whom the mean number ( $\bar{x}$ ) of subsequent offenses is 2.25, with a variance ( $s^2$ ) of 12.51 offenses. Let's also say, based on past research, that the mean number ( $\mu$ ) of offenses among all youths is 0.75. We will use  $\alpha = .01$ .

Note that the use of **more likely** implies a directional hypothesis - you should take this to mean that we are conducting a one-tailed hypothesis test.

#### Step 1: Formally state hypotheses

$$H_1 : \mu > 0.75; H_0 : \mu \leq 0.75$$

#### Step 2: Obtain a probability distribution

We will use the  $t$ -distribution because we do not know the population standard deviation. Our degrees of freedom are equal to  $df = n - 1 = 121 - 1 = 120$

#### Step 3: Make decision rules

$$\alpha = .01 \text{ (one-tailed)}; t_{crit} = 2.36; \text{reject } H_0 \text{ if } TS > 2.36$$

#### Step 4: Calculate the test statistic

$$TS = \frac{\bar{x} - \mu}{\sqrt{s^2/n-1}} = \frac{2.25 - 0.75}{\sqrt{12.51/(121-1)}} = \frac{1.5}{.323} = 4.64$$

#### Step 5: Make a decision about the null hypothesis

Reject  $H_0$ ; training school youths commit significantly more offenses than the overall population of youths.

## Childhood Maltreatment Example

- Research Question
  - Is there an association between childhood maltreatment and arrest as an adult?
  - Note that I am asking if there is an **association** - no implied direction (higher/lower) so this is a two-tailed question.
- Sample details
  - $\mu = 1.25$
  - $n = 61$
  - $\bar{x} = 2.57, s = 2.81$

### Step 1: Formally state hypotheses

$$H_1 : \mu \neq 1.25; H_0 : \mu = 1.25$$

### Step 2: Obtain probability distribution

Because we are using  $s$  as a **point estimate** for  $\sigma$ , we need to use the  $t$ -distribution.

### Step 3: Make decision rules

$\alpha = .05$  (two-tailed);  $t_{crit} = 2.0003$ ; reject  $H_0$  if  $TS > 2.0003$

*Note:* I go to the 4th decimal place (in these notes) because otherwise it's just a value of 2. By and large, two or three decimal places is fine for this class.

### Step 4: Calculate test statistic

$$TS = \frac{\bar{x} - \mu}{\sqrt{s^2/n-1}} = \frac{2.57 - 1.25}{2.81/\sqrt{(61-1)}} = \frac{1.32}{.363} = 3.636$$

### Step 5: Make a decision about the null hypothesis

Reject  $H_0$ ; child abuse victims are arrested a significantly *different* number of times than individuals who were not abused as children. Alternatively, we can conclude that there is some *association* that exists between child maltreatment and adult arrest.

## On p-Values

One additional piece of information we may obtain is the p-value, or the probability of obtaining our test statistic or a more extreme test statistic. For a one-tailed significance test, this is simply calculated as the probability that lies in the tail from the value of the test statistic. For a two-tailed significance test, we have to take twice the probability that lies in the tail from the value of the test statistic.

In the training school example, our TS was 4.64. We conducted a one-tailed significance test. This gives us a probability value of  $4 \times 10^{-6}$ .

In the childhood maltreatment example, the TS was 3.64. We conducted a two-tailed significance test so we must multiply our probability by two (because the critical regions are in opposite tails of the distribution). This gives us a probability value of  $6 \times 10^{-4}$ .

What you may also notice is that, if we reject  $H_0$  we will **always** have a probability value lower than our selected  $\alpha$  level. If we fail to reject, it will be equal to or greater than  $\alpha$ .