

IX. Inference with a Single Sample Mean

- 1) Logic of single-sample inference
- 2) Steps in a hypothesis test
- 3) 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 μ , where does the sample mean lie, or how many standard error units does the sample mean lie from μ ? We then ask whether our assumption about μ 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 μ , or is our sample a subset of an entirely different population with mean μ ? 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 μ 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 - Review

- In order to illustrate the steps in a hypothesis test, we will walk through an example. It is well established that people who offend tend to have lower intelligence than people who do not offend. The research question guiding our study is this: Do offenders have lower IQ than the population? We know that the mean IQ in the population is 100 with a standard deviation of 15. Suppose that we draw a sample of 100 incarcerated individuals, and find that the mean IQ is 94.1.
 - **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 this example, $H_1 : \mu < 100$ and $H_0 : \mu \geq 100$.
- **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 σ , so our probability distribution is the z-distribution.
- **Step 3: Make decision rules.** Once we have our theoretical probability distribution, we need to select the level of significance to 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 α . 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, α . 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.
- **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 σ . So the test statistic is computed as:

$$TS = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{94.1 - 100}{15/\sqrt{100}} = \frac{-5.9}{1.5} = -3.933$$

This value means that our sample mean is 3.933 standard error units less than 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 instead lies in the acceptance region, we will “accept H_0 .” In our example, we reject H_0 because -3.933 is less than -1.65.
 - * By rejecting the null hypothesis, we accept the alternative in its place. Thus, we conclude that offenders have significantly lower IQ than the general population.
- Here is a useful table to remember critical values for hypothesis tests using the z -distribution.

Significance Level	One-Tailed Critical Value	Two-Tailed Critical Value
.10	1.28	1.65
.05	1.65	1.96
.01	2.33	2.58
.001	3.08	3.27

- 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.

Reality	Decision	
	Accept H_0	Reject H_0
H_0 is true	Correct decision	Type I error (α)
H_0 is false	Type II error (β)	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.

- Let’s work through some more examples.
 - 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. The research question is: Do high-risk neighborhoods make more calls to police than other neighborhoods ($\alpha = .01$)?
 - * 1) $H_1 : \mu > 2.82$; $H_0 : \mu \leq 2.82$
 - * 2) z -distribution
 - * 3) $\alpha = .01$ (one-tailed); $z_{\text{crit}} = 2.33$; reject H_0 if $TS > 2.33$
 - * 4) $TS = \frac{43.03 - 2.82}{2.31/\sqrt{3795}} = \frac{40.21}{.37} = 1072.329$
 - * 5) Reject H_0 , there are significantly more calls to police in high-risk places.
 - Homicide rates in the U.S. since 1950 have a mean of 6.66 with a standard deviation of 1.97. In 2017, the homicide rate among the 50 states was 5.3. The research question is: Has there been a change from the average homicide rate over the last 50 years ($\alpha = .05$)?
 - * 1) $H_1 : \mu \neq 6.66$; $H_0 : \mu = 6.66$
 - * 2) z -distribution
 - * 3) $\alpha = .05$ (two-tailed); $z_{\text{crit}} = 1.96$; reject H_0 if $|TS| > 1.96$
 - * 4) $TS = \frac{5.33 - 6.66}{1.97/\sqrt{50}} = \frac{-1.33}{.2786} = -4.77$
 - * 5) Reject H_0 , there was a significant change in the homicide rate in 2017.
- 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 don’t know σ , we must estimate it using s , the sample standard deviation. Since we have to estimate this additional parameter (μ in addition to σ), 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 σ is unknown, instead of the z -distribution.

- Let’s see this with some examples.
 - 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, among whom the mean number of subsequent offenses is 2.25, with a variance of 12.51 offenses. Let’s also say, based on past research, that the mean number of offenses among all youths is 0.75. We will use $\alpha = .01$.
 - * 1) $H_1 : \mu > 0.75$; $H_0 : \mu \leq 0.75$
 - * 2) t -distribution
 - * 3) $\alpha = .01$ (one-tailed); $t_{\text{crit}} = 2.358$; reject H_0 if $TS > 2.358$
 - * 4) $TS = \frac{\bar{x} - \mu}{\sqrt{s^2/n-1}} = \frac{2.25 - 0.75}{\sqrt{12.51/(121-1)}} = \frac{1.5}{.323} = 4.646$
 - * 5) Reject H_0 ; training school youths have significantly more offenses than the population of youths.
 - We want to know if boot camp youths engage in fewer prosocial activities than other youths. Suppose that we have a sample of 10 boot camp youths and find that the mean number of prosocial activities is 1.53, with standard deviation 2.21. Suppose also that the population mean is 2.21. Use $\alpha = .10$.
 - * 1) $H_1 : \mu > 0.75$; $H_0 : \mu \leq 0.75$
 - * 2) t -distribution
 - * 3) $\alpha = .01$ (one-tailed); $t_{\text{crit}} = 2.358$; reject H_0 if $TS > 2.358$
 - * 4) $TS = \frac{\bar{x} - \mu}{\sqrt{s^2/n-1}} = \frac{2.25 - 0.75}{\sqrt{12.51/(121-1)}} = \frac{1.5}{.323} = 4.646$
 - * 5) Reject H_0 ; training school youths have significantly more offenses than the population of youths.
 - Suppose we want to know if there is an association between child maltreatment and arrest. We have a sample of 61 young adults with a history of child abuse and find that the mean number of arrests is 2.57 with a standard deviation of 2.81. Suppose that the mean number of arrests in the population is 1.25. Use $\alpha = .05$.
 - * 1) $H_1 : \mu \neq 1.25$; $H_0 : \mu = 1.25$
 - * 2) t -distribution
 - * 3) $\alpha = .05$ (two-tailed); $t_{\text{crit}} = 2.000$; reject H_0 if $TS > 2.000$
 - * 4) $TS = \frac{\bar{x} - \mu}{\sqrt{s^2/n-1}} = \frac{2.57 - 1.25}{2.81/\sqrt{(61-1)}} = \frac{1.32}{.363} = 3.639$
 - * 5) Reject H_0 ; child abuse victims are arrested a significantly *different* number of times than individuals who were not abused as children.

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.
 - Let’s compute the p -values associated with our large-sample test statistics earlier.
 - * In the IQ example, the test statistic was -2.94. We conducted a one-tailed significance test, since $H_1 : \mu < 100$. We find that the probability in the tail of the standard normal distribution at this value is 0.0016.
 - * In the training-school example, the test statistic was 3.07. We conducted a two-tailed significance test, since $H_1 : \mu \neq 0.35$. The combined probability in the tail is $0.0011 + 0.0011 = 0.0022$.
 - You might notice that if we reject the null hypothesis, the p -value will always be less than α . If, on the other hand, we accept the null hypothesis, the p -value will always be greater than α .