

In most cases, there are two types of inferences we make about a single population parameter. Either we estimate the actual value of the parameter or we make decisions concerning the value of the population parameter. To make decisions concerning the value of the population parameter we will conduct a **hypothesis test**.

For example, the Acme Mail Service claims that their priority mail packages reach their destinations in two days, on average. The question is, is this claim true? Note there are two opposing hypotheses here:

- The average delivery time is two days (claim is true); OR
- The average delivery time is not 2 days (claim is false).

The hypothesis that says the claim is true is called the *null hypothesis*; denoted H_0 . The null hypothesis is typically based on prior research or information, or just an assumption about the population. The hypothesis that says the claim is false is called the *alternative hypothesis*; denoted H_a . Another name for alternative hypothesis is the *research hypothesis*, since a researcher who sets out to test an initial result, or claim, does so with an alternative theory, or alternative hypothesis in mind. For the above example, the hypotheses would be set up this way:

- $H_0: \mu = 2$
- $H_a: \mu \neq 2$ (where μ is the average delivery time for an Acme priority mail package.)

Note that null and alternative hypotheses always involve statements about population parameters, such as the population mean, μ , or the population proportion, p .

Example: A car manufacturer advertises that its new subcompact model gets 47 miles per gallon. From our experience with this manufacturer, we have every reason to believe the advertised mileage is too high.

- $H_0: \mu = 47$
- $H_a: \mu < 47$

Making Decisions and Considering the Consequences

The evidence provided by a random sample will lead you to one of the two alternatives. You will either decide that the evidence supports the claim (accept H_0), or that it doesn't (reject H_0 in favor of H_a .) There are two possible errors possible in this decision making process. A **type I** error occurs when we reject the null hypothesis when it is true. A **type II** error occurs when a null hypothesis when it is false. Let's look at the following examples.

Example: A parachutist is about to jump out of an airplane. He gives one last look at his parachute and considers the following two hypotheses.

H_0 : The parachute is defective.

H_a : The parachute is safe.

- State in everyday language what it means to make a Type I error in this context.
- State in everyday language what it means to make a Type II error in this context.
- What is the consequence of making a Type I error in this context?
- What is the consequence of making a Type II error in this context?

Example: A jury trial can be thought of as a hypothesis test for deciding between two alternatives:

H_0 : The defendant is not guilty.

H_a : The defendant is guilty.

We make a Type I error if the defendant is not guilty, but we say he/she is. We make a Type II error if the defendant is guilty, but we say he/she is not.

- Our jury system requires all 12 jurors to vote "guilty" in order to get a guilty verdict. In this system, do you think we would have less chance of making a Type I error or less chance of making a Type II error? Explain.
- Suppose the system of voting required only 10 out of 12 to vote guilty in order to convict. How would this affect the chances making of a Type I error and Type II error compared to the present system?

Type I and Type II Errors

	Our Decision	
Truth of H_0	Do not reject H_0	Reject H_0
If H_0 is true	Correct decision, no error $P(\text{correct}) = 1 - \alpha$	Type I error $P(\text{Type I}) = \alpha$
If H_0 is false	Type II Error $P(\text{Type II}) = \beta$	Correct decision, no error $P(\text{correct}) = 1 - \beta$

Example:

A certain type of automobile is known to sustain no visible damage 25% of the time in a 10mph crash test. A modified bumper has been proposed to increase this percentage. The hypotheses tested are $H_o : p = 0.25$ (no improvement) versus $H_a : p > 0.25$. The test will be based on an experiment involving 20 independent crashes. Intuitively, H_o should be rejected if a substantial number of crashes show no damage. Consider the following procedure:

Let X = the number of crashes with no visible damage (a binomial random variable). We will reject H_o if $X \geq 8$.

So when H_o is true X has a binomial probability distribution with $n = 20$ and $p = 0.25$.

$$\alpha = P(\text{type I error}) = P(H_o \text{ is rejected when it is true}) = P(X \geq 8) = 1 - 0.898 = 0.102$$

That is, when H_o is actually true, roughly 10% of all experiments of this type (with 20 crashed) would result in being incorrectly rejected. (Note α is called the level of significance.)

In contrast to α , there is not a single value of β . You will get a different value for β depending on how much p exceeds 0.25. For instance if $p = 0.3$, then

$$\beta = P(\text{type II error } p = .3) = P(H_o \text{ not rejected when } p = 0.3) = P(X \leq 7 | X \sim \text{Bin}(20, .3)) = .772$$

If $p = 0.5$, the $\beta = 0.132$, and if $p = 0.7$, the $\beta = 0.001$. Clearly β decreases as p moves further to the right of the null hypothesis.

Note the quantity $1 - \beta$ is called the **power of the test**.

In real life, we don't know whether the H_o (the claim) is actually true or not; therefore we won't know which error might be committing when we conduct our hypothesis test. We would like to guard against both types of errors.

It is difficult to prevent both a Type I and a Type II error. Look at the above example, what would happen if the rejection region changed from $X \geq 8$ to $X \geq 10$? What would happen to α ? What would happen to β ?

In general, all researchers have to decide which error is more important to avoid, and take it from there. The power of the test increases as the level of significance increases. However increasing the significance level also increases the probability of a type I error.

Despite these facts most researchers use a small level of significance. They are more willing to make an error by failing to reject a claim (i.e. H_0) than to make an error by accepting another claim H_a . Because of this, instead of accepting H_0 we will write that we *fail to reject* H_0 meaning that the evidence was not strong enough to favor rejection.

Hypothesis Test for the μ when σ is known:

In the “real world”, the hypotheses being tested would be determined when the researcher determines his/her question that they want answered. So before the data is ever collected, the hypotheses are determined along with; the sample size, the level of significance, how the sample will be randomly selected, what tool will be used for measurements, what the plan is to deal with non-response etc.

Possible hypotheses to consider:

$$H_0 : \mu = \mu_0$$

$$H_a : \mu \neq \mu_0$$

$$H_0 : \mu = \mu_0$$

$$H_a : \mu < \mu_0$$

$$H_0 : \mu = \mu_0$$

$$H_a : \mu > \mu_0$$

Where μ_0 is some known value for the population mean. Note, the null hypothesis in the last two cases could be greater than or equal to or less than or equal to. The significance level would be its largest for equality and less for the other possible values of the mean.

Once the data is collected we determine the sample mean, \bar{x} and look to see if there is evidence against the null hypothesis. This is accomplished by calculating the number of standard errors the sample mean, \bar{x} , from the population mean μ_0 .

$$Z_c = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Z_c is called the test statistic. Note if the sample size, n , is large we can use s instead of σ . We can determine whether to reject the null hypothesis by calculating the p-value.

P-value =

$$2P(Z > |Z_c|)$$

$$P(Z < Z_c)$$

$$P(Z > Z_c)$$

We will reject H_0 if $p\text{-value} \leq \alpha$, otherwise we fail to reject H_0 .

Another way to draw a conclusion from the test statistic is to simply compare it to values of the standard normal distribution.

$$\text{Reject } H_0 \text{ if } |Z_c| \geq Z_{\alpha/2}$$

$$\text{Reject } H_0 \text{ if } Z_c \leq -Z_\alpha$$

$$\text{Reject } H_0 \text{ if } Z_c \geq Z_\alpha$$

Example: A real estate agent believes that the average closing cost of purchasing a new home is \$6500 over the purchase price. She selects 40 new home sales at random and finds that the average closing costs are \$6600. The standard deviation of the population is \$120. Test her belief at $\alpha = 0.05$.

Example: A biologist knows that the average length of a leaf of a certain full-grown plant is 4 inches. The standard deviation of the population is 0.6 inch. A sample of 45 leaves of that type of plant were given a new type of plant food have an average length of 4.2 inches. Is there reason to believe that the new food is responsible for change in the average growth of leaves? Use $\alpha = 0.01$. Find a 99% confidence interval for the mean. Do the results concur?

Hypothesis Test for the μ when σ is unknown:

Possible hypotheses to consider:

$$H_o : \mu = \mu_o$$

$$H_a : \mu \neq \mu_o$$

$$H_o : \mu = \mu_o$$

$$H_a : \mu < \mu_o$$

$$H_o : \mu = \mu_o$$

$$H_a : \mu > \mu_o$$

Where μ_o is some known value for the population mean. Assuming that the data is normally distributed, we determine the sample mean, \bar{x} , sample standard deviation, s , and look to see if there is evidence against the null hypothesis. This is accomplished by calculating the number of standard errors the sample mean, \bar{x} , from the population mean μ_o .

$$t_c = \frac{\bar{x} - \mu_o}{s / \sqrt{n}}$$

t_c is called the test statistic. We can determine whether to reject the null hypothesis by calculating the p-value.

P-value =

$$2P(t > |t_c|)$$

$$P(t < t_c)$$

$$P(t > t_c)$$

We will reject H_o if $p\text{-value} \leq \alpha$, otherwise we fail to reject H_o .

Another way to draw a conclusion from the test statistic is to simply compare it to values of the standard normal distribution.

Reject H_o if $|t_c| \geq t_{\alpha/2, n-1}$

Reject H_o if $t_c \leq -t_{\alpha, n-1}$

Reject H_o if $t_c \geq t_{\alpha, n-1}$

Example: A taxi company claims that its drives have an average of at least 12.4 years of experience. In a study of 15 taxi drivers, that average experience was 11.2 years with a sample standard deviation of 2 years. At $\alpha = 0.10$, is the average number of years' experience less than the company's claim? Assume the data are from a normal distribution.

Example: To see whether people are keeping their car tires inflated to the correct level of 35 psi, a tire company manager selects a sample of 25 tires and checks their pressure. The mean of the sample is 33.5 psi with a sample standard deviation of 3 psi. Are the tires inflated properly? Use $\alpha = 0.10$.

Hypothesis Test for the p :

Possible hypotheses to consider:

$$\begin{aligned} H_o &: p = p_o \\ H_a &: p \neq p_o \end{aligned}$$

$$\begin{aligned} H_o &: p = p_o \\ H_a &: p < p_o \end{aligned}$$

$$\begin{aligned} H_o &: p = p_o \\ H_a &: p > p_o \end{aligned}$$

Where p_o is some known value for the population proportion.

Once the data is collected we determine the sample proportion, \hat{p} and look to see if there is evidence against the null hypothesis. This is accomplished by calculating the number of standard errors the sample mean, \hat{p} , from the population mean p_o .

$$Z_c = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$$

Z_c is called the test statistic. We can determine whether to reject the null hypothesis by calculating the p-value.

P-value =

$$2P(Z > |Z_c|)$$

$$P(Z < Z_c)$$

$$P(Z > Z_c)$$

We will reject H_o if $p\text{-value} \leq \alpha$, otherwise we fail to reject H_o .

Another way to draw a conclusion from the test statistic is to simply compare it to values of the standard normal distribution.

Reject H_o if $|Z_c| \geq Z_{\alpha/2}$ Reject H_o if $Z_c \leq -Z_\alpha$ Reject H_o if $Z_c \geq Z_\alpha$

Example: A Harris Poll found that 35% of people said that they drink a caffeinated beverage to combat midday drowsiness. A recent survey found that 19 out of 48 people stated that they drank a caffeinated beverage to combat midday drowsiness. At $\alpha = 0.02$ is the claim of percentage found in the Harris Poll believable?

Example: Nationally 60.2% of federal prisoners are serving time for drug offenses. A warden feels that in his prison, the percentage is higher. Her surveys 400 inmates' records and finds that 260 of the inmates are drug offenders. At $\alpha = 0.05$, is he correct?

Homework:

For each of the following, stated the null and alternative hypotheses, compute the test statistic, determine the p-value, make the decision and summarize the results.

1. A report in *USA Today* stated that the average age of commercial jets in the United States is 14 years. An executive of a large airline company selects a sample of 36 planes and finds a sample mean of 11.8 years with a sample standard deviation of 2.7 years. At $\alpha = 0.01$, can it be concluded that the average age of the planes in his company is less than the national average?
2. An item in *USA Today* reported that 63% of Americans owned an answering machine. A survey of 143 employees at a large school showed that 85 owned an answering machine. At $\alpha = 0.05$, test the claim that the percentage is the same as stated in *USA Today*?
3. Average undergraduate cost for tuition, fees, and room and board for all institutions last year was \$19,410. A random sample of costs this year for 40 institutions of higher learning indicated a sample mean of \$22,098 and the sample standard deviation was \$6050. At the 0.05 level of significance, is there sufficient evidence to conclude that the cost of attendance has increased?
4. A job placement director claims that mean starting salary for nurses is \$34,000. A random sample of 10 nurses' salaries has a mean \$33,450 and a standard deviation of \$4000. Is there enough evidence to reject the director's claim at $\alpha = 0.05$?
5. A physician claims that joggers' maximal volume of oxygen uptake is greater than the average of all adults. A sample of 15 joggers has a mean of 40.6 ml/kg and a standard deviation of 6 ml/kg. If the average of all adults is 36.7 ml/kg, is there enough evidence to support the physician's claim at $\alpha = 0.01$?
6. Of families, 48.8% have stock holdings. A random sample of 250 families indicated that 142 owned some type of stock. At what level of significance would you conclude that the true proportion is different from the claim?
7. A manager states that in his factory, the average number of days per year missed by employees due to illness is less than the national average of 10. The data for a random sample of 40 workers is given below. Is there evidence to believe the manager's statement at $\alpha = 0.05$?

0	6	12	3	3	5	4	1
3	9	6	0	7	6	3	4
7	4	7	1	0	8	12	3
2	5	10	5	15	3	2	5
3	11	8	2	2	4	1	9