Chapter 10
Testing Hypotheses

Once inferences are made about a population, the question becomes: are statements made about the population true based on the statistics underlying the statements?

A **hypothesis** is a statement that something is **true** and is frequently developed for the sole purpose of testing its veracity**.** In this chapter, we want to test hypotheses. That is, if a claim is made about a population, how likely is it that the claim is true? We will "test" the veracity of the statement by drawing a sample from the population about which the claim was made and performing a statistical test called a **hypothesis test.**

Example: If we flip a quarter 100 times and get 94 heads, it is very likely that the quarter favors heads. However, if we flip it 100 times and get 51 heads, it would be difficult to conclude that the quarter favors heads. We would conclude that the quarter favors heads only if there were **significantly** more heads than we would expect using an unbiased quarter. 51 is not significant while 94 is. Therefore, **our tests will be based on criteria for identifying results that are significant enough for us to make a decision about the likely truth of the statement or hypothesis. The question is: "How far away from the claim must our test results be before we disbelieve the claim?"**

Let's look at how we can test claims (hypotheses) about the **mean** of a population. First, we need to know the terminology to use in performing the test.

The Terms of Hypothesis Testing About a Mean

Null Hypothesis (H_o) : This is the hypothesis that is **directly tested** as true. It must contain the condition of **equality**. **H_o = some value; $H_o \leq$ some value; or $H_o \geq$ some value**.

The **final conclusion** of our test will be either to **reject H_o** or **fail to reject H_o**.

Alternate Hypothesis (H_1) : The statement that must be true if H_o is false. It will never contain an equality. **$H_1 \neq$ some value; $H_1 <$ some value; or $H_1 >$ some value.**

H_1 is the complement of H_o. If H_o: $\mu \leq 10$, then H_1: $\mu > 10$.

Taken together, the null and alternate hypotheses **must include all possible outcomes.**

Depending on how stated, the claim may be the null or the alternate hypothesis. However, the null hypothesis will always have the condition of equality and will always be the one tested.

Test Statistic : A value or statistic based on sample data drawn from a population. This is the **z score** based on the sample mean (how many standard deviations it is from the claimed mean) and is used to decide whether to reject or not reject the null hypothesis. We will reject the claim if the mean of our test sample is significantly far away from the claimed mean (i.e., too many standard deviations **z** away from the claim).

Critical Values: The z score that acts as the cut-off for what is significant and not significant. The value of "z" that would separate the "reject" from the "fail to reject" decision. This z-score is based on the level of confidence (Chpt 9) we want to have in rejecting or failing to reject the claim.

Critical Region (or Region of Rejection): All values of the sample test statistic that would cause us to reject the null hypothesis. If we are testing $H_o \leq$ some value, then the critical region is called **right-tailed** because we can only reject the claim if our sample mean is significantly **greater than** the claimed mean.. If we are testing $H_o \geq$ some value, then the critical region is called **left-tailed** because we can only reject the claim if our sample mean is significantly **less than** the claimed mean. If we are testing that H_o = some value, then the critical region is said to be **two-tailed** because we can reject the claim if our sample mean is significantly **above or below** the claimed mean.

Significance Level: The probability of rejecting the null hypothesis when it is true. This is α and is the same α we have previously discussed. For example, if we want to be 95% sure we do not reject a true claim or fail to reject a false one, then there is only an α = 5% probability that we will make such an error. The significance level says that we will accept a claim about a population mean unless the mean that we got from our test sample is very unlikely (5% or less) to have happened if the claim were true. (In other words, our sample mean must be **significantly** far away from the claimed mean).

Of course, based on our test, we could always get such a sample and reject a true hypothesis or accept a false one! These are called testing errors and are described as follows:

> **Type I Error** : The mistake of rejecting the null hypothesis when it is true. The probability of a Type I error is called α.
>
> **Type II Error** : The mistake of failing to reject the null hypothesis when it is false. The probability of a Type II error is called β.

Steps in Hypothesis Testing

Note: There are more steps here than in the text. The text includes more than one of these steps in each of its steps. The procedures are exactly the same.

Step 1: Identify the specific claim and decide whether it is the null or alternative hypothesis (i.e., does it contain equality?).

Step 2: Identify the opposite of the specific claim.

Step 3: Write the claims in symbolic form : H_o with equality and H_1 without.

Step 4: Select the significance level α. If $\alpha = .05$, it means that we would fail to reject a sample as proof of the null hypothesis unless the probability of getting such a sample is less than 5%. In other words, **we would reject the claim about the population mean only if the sample was so significantly different from the claim that there was only a 5% (α) chance or less of drawing this sample if the claim were true.**

Step 5: Find the critical z value from α (or α/2 if appropriate) using the z-table. This "z" is the number of standard deviations away from the claim that the sample mean must be in order for us to reject the null hypothesis.

Step 6: Calculate the test statistic (z) from the formula

$$z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$$

where $\bar{x}$ is the mean of the test sample, μ is the claimed mean, σ is the population standard deviation, and n is the size of our test sample. In other words, the test statistic is the number of standard deviations the sample mean is from the claimed mean.

Step 8: Draw a normal distribution showing the mean, critical value, critical region, and test statistic.

Step 9: If the test statistic falls within the critical region (area above or below the critical value), reject the null hypothesis. This would tell us that the sample mean is not only different than the population mean, but that it is significantly different.

If the test statistic does not fall within the critical region, fail to reject the null hypothesis.

Step 10: State the decision in terms of “reject”” or “fail to reject”. If we fail to reject, it does not mean that we are accepting the hypothesis, it only means that the evidence provided by the sample is not strong enough to warrant rejection of the null hypothesis.

Note that because we always test the null hypothesis, in some cases we will be testing the original claim and in some cases the opposite of the original claim.

As with the other statistical concepts, the larger the sample size, the more accurate our results. In this case, that means the less likely we are to commit a Type I or Type II error.

Example:

The engineering department of a car manufacturer claims that the fuel consumption rate of one model is equal to 35 mi/gal. The advertising department wants to test this claim to see whether the announced figure should be higher or lower than 35 mi/gal. The quality control group suggests that $\sigma = 4$ mi/gal, and a sample of 50 cars yields $\bar{x} = 33.6$ mi/gal. Using the data from the sample, test the claim of the engineering department about mean miles per gallon. Use the 0.05 level of significance.

The claim:

The opposite of the claim:

The null hypothesis (contains the equality): H_o :

The alternate hypothesis: H_1 :

Significance level α:

The critical value $z_{\alpha/2}$:

The test statistic:

The graph:

Where does the test statistic fall on the graph relative to the critical value?

What is your decision about the engineering department's claim?

Alternate approach using P-values.

The P-value is defined as the probability of having gotten the sample, or one *even farther away* from the claim, if the claim were indeed true. It is found by going into the z-table with the test statistic as "z", and pulling out the corresponding probability of that "z". The probability so determined is called the **P-value**.

Compare the P-value to α.

> If P-value $\leq \alpha$ (or $\alpha/2$ for a two-tailed test), reject the null hypothesis.
>
> If P-value $> \alpha$ (or $\alpha/2$ for a two-tailed test), fail to reject the null hypothesis.

For the example above, calculate the P-value associated with the sample and decide whether to reject or fail to reject the null hypothesis.

Homework: pp. 291-292, Exercises 1, 2, 3, 4, 5, 7

Testing Claims about a Proportion

Review the material about proportions (percentages) in Chapter 9.

In this case, the critical value is found from z-table just like before, and the test statistic is found from the formula

$$\mathbf{z = (p - \pi) / \sqrt{\pi(1-\pi)/n}}$$

based on claim

based on sample

where **p** = x/n
n = sample size
x = number of successes in the sample of size n
π = claimed percentage of successes in the population
(1-π) = claimed percentage of failures in the population

Example:

An auditor for the US Postal Service wants to examine its special Two Day Priority mail handling to determine the proportion of parcels that actually require longer than two days for delivery. A randomly selected sample of 150 such parcels is found to contain nine that required longer than two days. Use a 0.05 level of significance to test the auditor's claim that the percentage of late parcels exceeds 4%.

The claim:

The opposite of the claim:

The null hypothesis (contains the equality): H_o :

The alternate hypothesis: H_1 :

Significance level α:

The critical value z :

The test statistic:

The graph:

Where does the test statistic fall on the graph relative to the critical value?
What is your decision about the auditor's claim?

What is the p-value of the test statistic?

Homework: p. 295, Exercises 9, 11, 13

Testing Hypotheses with the t-Distribution

All of the above is fine if the sample size n is large, i.e. $n \geq 30$, or if we know the standard deviation of the population σ.

However, if the sample size is small ($n < 30$), the value of σ is unknown, and the parent population is essentially normal, then we have to use the t-distribution to test a hypothesis.

Therefore, the critical values are found using Appendix F, (degree of freedom = n-1) and the test statistic (t) is given by

$$\mathbf{t = (\bar{x} - \mu) / (s/\sqrt{n})}$$

where "s" is the **sample** standard deviation.

Example:

The Carolina Tobacco Company claims that its best selling cigarettes contain at most an average of 40 mg of nicotine per cigarette. Test this claim at the 0.01 significance level by using the results of 15 randomly selected cigarettes for which $x = 42.6$ mg and $s = 3.7$ mg. Other evidence suggests that the distribution of nicotine contents is normally distributed.

The claim:

The opposite of the claim:

The null hypothesis (contains the equality): H_o :

The alternate hypothesis: H_1 :

Significance level α:

The critical value t :

The test statistic:

The graph:

Where does the test statistic fall on the graph relative to the critical value?

What is your decision about the tobacco company's claim?

Homework: **pp. 300-301, Exercises 15, 17, 19**
pp. 303-304, Exercises 21, 22, 23, 25
pp. 305-307, Exercises 27, 29, 33, 35, 37, 39