

April 13-24

*Instructor: Brian Powers*

## 19.1 Statistical Hypotheses

**Definition 19.1.** A **statistical hypothesis** is an assertion or conjecture concerning one or more populations.

We never know with certainty the validity of a hypothesis unless we examine the entire population. What we can say is the probability of observing our sample data if the hypothesis was correct. If this probability is extremely small we would argue that it is unlikely that the hypothesis is valid. We then “reject the hypothesis.”

The **null hypothesis** (denoted  $H_0$ ) is the hypothesis we test, the rejection of which leads us to acceptance of the **alternative hypothesis** (Denoted  $H_1$  or  $H_A$  or  $H_a$  depending on the book). If there is sufficient evidence in the data in favor of the alternative hypothesis, we **reject  $H_0$** , otherwise we say that we **fail to reject  $H_0$** . We do not say that we “accept  $H_0$ ”. The null hypothesis is the “status quo”, it is what we assume to be true unless we can prove otherwise.

## 19.2 The Test Statistic

A **test statistic** is a statistic on which we base our decision. The **critical region** Are the values of the test statistic for which we will reject the null hypothesis, often an interval. The boundary of the non-critical region is known as the **critical value**.

**Example 19.2.** Say we want to test whether  $p > .10$  for some Binomial experiment where  $n = 20$ .  $X$ , the value of the binomial random variable can be our test statistic, and we may decide to reject  $H_0$  if  $X > 8$ , so 8 is the critical value.

## 19.3 Type I and Type II error

There are two errors we can make when doing a hypothesis test.

**Definition 19.3.** Rejection of the null hypothesis when in fact it is true is a **type I error**.

**Definition 19.4.** Failure to reject the null hypothesis when in fact it is false is a **type II error**.

So we can see our possible outcomes in a table:

	$H_0$ is true	$H_0$ is false
Fail to reject $H_0$	Correct	Type II error
Reject $H_0$	Type I error	Correct

The probability of committing a Type I error is known as the **level of significance** and is denoted by  $\alpha$ . The probability of committing a Type II error is denoted by  $\beta$ , and can only be computed when the alternative hypothesis is “simple.”

You can always decrease  $\alpha$  by decreasing the size of the critical region - at the extreme, if you never reject  $H_0$ , then  $\alpha = 0$ , but then  $\beta = 1$ . You can reduce  $\beta$  by increasing the size of the critical region, but  $\alpha$  will increase. In general, one way to reduce both  $\alpha$  and  $\beta$  is to increase the sample size.

**Example 19.5.** In general, people are able to get 70% correct on a certain memory quiz. An herbal supplement is supposed to increase memory so that people get 90% correct. Suppose we test the herbal supplement on someone, and we decide to believe the claim if the person can get 80 / 100 correct or more. Find  $\alpha$  and  $\beta$ . What if we reject if they get 90/100 correct or more?

**Example 19.6.** Suppose we know that a population is normal with  $\sigma = 5$ . We want to test whether or not  $\mu = 35$ . Say we take a sample size of 40, and we decide to reject  $H_0$  if  $\bar{X} < 34$  or  $\bar{X} > 36$ . What is  $\alpha$ ? What if we increase the sample size to 60?

**Example 19.7.** For a normal population with  $\sigma = 5$  and a sample size of 40, test  $H_0 : \mu = 35$  vs  $H_1 : \mu = 45$ , where the critical region is the same as the previous example. Find  $\beta$ . What is  $\beta$  if  $H_1 : \mu = 36$ ?

This emphasizes that as the distance between the true value of the parameter and the value in the null hypothesis increases, then  $\beta$  will decrease.

**Definition 19.8.** The **power** of a test is the probability of rejecting  $H_0$  given that a specific alternative is true.

In practice, the level of significance is chosen first and ideally the researcher will choose the test which has the maximum power, (although this is not always possible). Since power is dependent on the specific value of the parameter, it can be viewed as a function of  $\theta$

**Example 19.9.** Suppose we know that a population is normal with  $\sigma = 1$ , but  $\mu$  is unknown. If we want to test whether  $\mu > 0$  by observing just a single sample, what critical region should we use? What is the power if  $\mu = .5$ ? What about for  $\mu = 2$  or  $\mu = 5$ ?

## 19.4 One- and Two-Tailed Tests

A hypothesis of this form:

$$H : \theta = \theta^*$$

is called a **simple hypothesis**. A hypothesis of the form

$$H : \theta > \theta^* \quad \text{or} \quad H : \theta < \theta^*$$

is a **one-sided hypothesis**. A hypothesis of the form

$$H : \theta \neq \theta^*$$

is known as a **two-sided hypothesis**.

For this course we will be performing hypothesis tests with a simple null hypothesis. The alternative will usually be one-sided or two-sided. If the alternative hypothesis is one-sided, we call this a **one-tailed test** because the critical region (rejection region) will be of the form  $T > t^*$  or  $T < t^*$  where  $T$  is the test statistic and  $t^*$  is the critical value. If the alternative hypothesis is two-sided, then we call this a **two-sided test** because the critical region will be of the form  $(-\infty, t_1^*) \cup (t_2^*, \infty)$ .

## 19.5 Using P-values

Simply rejecting the null hypothesis if the test statistic is in the critical region does not tell us how significant our observations are. The technique of using p-values addresses this concern. We calculate a  $P$ -value, which measures how significantly our observations differ from what we would expect under the null hypothesis. Formally

**Definition 19.10.** A **P-value** is the lowest level of significance at which the observed value of the test statistic is significant.

We can intuitively think of a  $P$ -value as the probability of observing a test statistic as extreme (or more extreme) than we observe under the null hypothesis. Or more simply put, it is a measure of evidence in support of the null hypothesis. We can then use our judgement to determine whether to reject the null hypothesis based on the  $P$ -value

Alternative Hypothesis	P-value calculation
"<" (Lower-tailed)	$P(T < t)$ under $H_0$
">" (Upper-tailed)	$P(T > t)$ under $H_0$
" $\neq$ " (Two-tailed)	$2P(T >  t )$ under $H_0$

Table 19.1: P-value calculation wher  $t$  is the value of the test statistic,  $T$  is the random variable test statistic.

## 19.6 Hypothesis Tests Concerning a Single Population Mean

### 19.6.1 Test on Single Mean, Variance Known

Suppose we sample  $X_1, X_2, \dots, X_n$  independently from a population with unknown mean  $\mu$  and known variance  $\sigma^2$ . We want to test the hypothesis

$$H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu \neq \mu_0$$

The test statistic should be based on  $\bar{X}$ . By the central limit theorem, the test statistic

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

approximately follows a standard normal distribution. we can determine the critical region for a test with significance level  $\alpha$ . As this is a two-tailed test, our critical values are found by

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

Or in other words, we reject  $H_0$  if

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2} \text{ or } z < -z_{\alpha/2}$$

We could also convert these critical values of  $Z$  to critical values for  $\bar{X}$ .

**Example 19.11.** A random sample of 100 recorded deaths in the United States during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use  $\alpha = .05$ .

**Example 19.12.** A manufacturer of sports equipment has developed a new synthetic fishing line that the company claims has a mean breaking strength of 8 kilograms with a standard deviation of .5 kilograms. Test the hypothesis that  $\mu = 8$  kilograms against the alternative that  $\mu \neq 8$  if a random sample of 50 lines is tested and found to have a mean breaking strength of 7.8 kilograms. Use  $\alpha = 0.01$ .

### 19.6.2 Test on Single Mean, Variance Unknown

In the case of a small sample size or  $\sigma^2$  unknown, we will naturally use as our test statistic

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Which approximately follows a  $t$ -distribution with  $n - 1$  degrees of freedom. The procedure for finding critical regions and rejection of  $H_0$  is the same as the previous case, though.

**Example 19.13.** A sample of measurements of data were taken:

$$7.07, 7.00, 7.10, 6.97, 7.00, 7.03, 7.01, 7.01, 7.01, 6.98, 7.08$$

Assuming the population is normal, we want to test if  $\mu = 7.0$  or not. Find the  $p$ -value for  $T$ .

### 19.6.3 Two Samples: Tests on Two Means

If the two populations are both normal with known standard deviations  $\sigma_1$  and  $\sigma_2$ , then the test statistic

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

follows a standard normal distribution. If the populations are approximately normal or if the sample size is large enough, then we can still use this result by the central limit theorem.

If the two populations have unknown standard deviations, but we can **assume them to be equal**,  $\sigma_1 = \sigma_2 = \sigma$ , then we can again use pooled variance as an estimate of  $\sigma_2$ . Of course, in this case

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{pooled} \sqrt{1/n_1 + 1/n_2}}$$

follows a  $t$ -distribution with  $n_1 + n_2 - 2$  degrees of freedom.

And finally if the variances are unknown and we do not assume them to be equal

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

Will follow a  $t$ -distribution with

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$$

degrees of freedom.

### 19.6.4 Paired Observations

When there is a natural reason to pair observations between the two samples (of equal size) then we wish to measure not the difference in the means of the two samples, but the mean of the differences between the pairs. So if the samples are  $x_1, \dots, x_n$  and  $y_1, \dots, y_n$ , we use  $d_i = x_i - y_i$  to get a list of differences. Then we can simply reduce this to the 1-sample case.

## 19.7 Test on Single Population Proportion

When testing whether the proportion  $p$  in a population is equal to some hypothesized value or not, we will base the test statistic on  $\hat{p}$ , as you may guess. Actually,  $X = \hat{p}n$ , the count of individuals in the sample with the characteristic in question, will follow a *Binomial*( $n, p_0$ ) under the null hypothesis. We could certainly use  $X$  as our test statistic, and base our inference on this distribution, including calculation of a p-value (especially with technology today, calculation of Binomial probabilities from the cdf are very fast). It is very common to use the Central Limit Theorem to approximate using the Normal distribution.

Under the null hypothesis of the form

$$H_0 : p = p_0$$

the test statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$$

follows a standard normal distribution.

## 19.8 Test on Difference of Population Proportions

If we want to see if the proportion of a certain characteristic among two populations differs, we want to do a hypothesis test on  $p_1 - p_2$ . To test the null hypothesis

$$H_0 : p_1 - p_2 = 0$$

The test statistic we want to use is

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0 \hat{q}_0 (1/n_1 + 1/n_2)}}$$

follows a standard normal distribution, where  $\hat{p}_0 = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$ .

Test Type	Null Hypothesis	Test Statistic	Notes
Single Pop. Mean			
$\sigma$ known	$H_0 : \mu = \mu_0$	$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	
$\sigma$ unknown or $n$ small	$H_0 : \mu = \mu_0$	$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$df = n - 1$
Difference of Two Pop. Means			
$\sigma_1, \sigma_2$ known	$H_0 : \mu_1 - \mu_2 = 0$	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$	
$\sigma_1 = \sigma_2$ unknown	$H_0 : \mu_1 - \mu_2 = 0$	$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{pooled} \sqrt{1/n_1 + 1/n_2}}$	$df = n_1 + n_2 - 2$
$\sigma_1 \neq \sigma_2$ unknown	$H_0 : \mu_1 - \mu_2 = 0$	$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$	$df = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{\frac{n_1 + n_2 - 2}{(s_1^2/n_1 + s_2^2/n_2)^2}}$
Paired Samples			
	$H_0 : \mu_D = \mu_{D0}$	$T = \frac{\bar{D} - \mu_{D0}}{S_D/\sqrt{n}}$	$df = n - 1$
Single Pop. Proportion	$H_0 : p = p_0$	$Z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0/n}}$	
Difference of Proportions	$H_0 : p_1 - p_2 = 0$	$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0 \hat{q}_0 (1/n_1 + 1/n_2)}}$	

Table 19.2: Test Summary of Hypothesis testing procedures. A  $Z$  test statistic follows a standard normal distribution, and a  $T$  test statistic follows a  $t$ -distribution.