3. Estimation and Testing II

3.2 Hypothesis Testing

Hypothesis testing or significance testing is a method for testing a claim or hypothesis about a parameter in a population, using data measured in a sample. We break up significance testing into four steps:

Step 1: Setting Null and Alternative Hypotheses

The first step of hypothesis testing is to convert the research question into null and alterative hypotheses. We start with the null hypothesis (H_0). The null hypothesis is a claim of “no difference.” The opposing hypothesis is the alternative hypothesis (H_1 or H_a). The alternative hypothesis is a claim of “a difference in the population,” and is the hypothesis the researcher often hopes to bolster. It is important to keep in mind that the null and alternative hypotheses reference population values, and not observed statistics.

Step 2: Calculating the Test statistic

We calculate a test statistic from the data. There are different types of test statistics. This chapter later describes those different types. Large test statistics indicate data are far from expected, providing evidence against the null hypothesis and in favor of the alternative hypothesis.

Step 3: p Value and Decision

The test statistic is converted to a probability called a p-value. The p- value answers the question “If the null hypothesis were true, what is the probability of observing the current data or data that is more extreme?”

Small p values provide evidence against the null hypothesis because they say the observed data are unlikely when the null hypothesis is true. We apply the following convention:

When p value $\leq \alpha$, then the null hypothesis should be rejected in favor of alternative at α % level of significance.

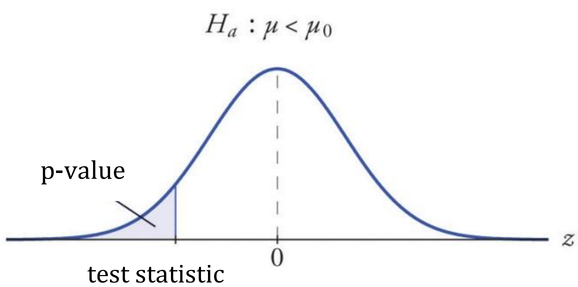
Step 4: Conclusion

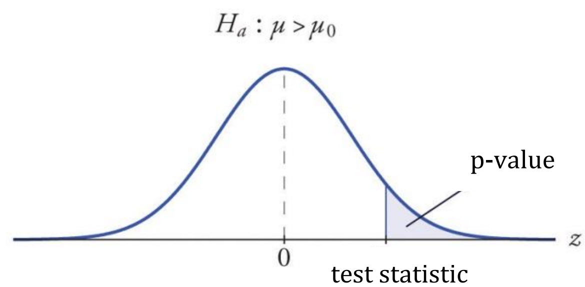
With decision made using the p-value, the conclusion should be written with non-technical form.

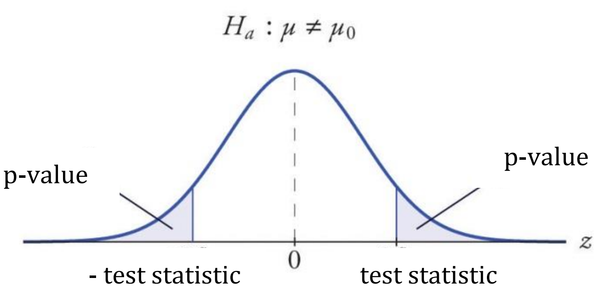
One-tailed and Two-tailed Testing

The equality part of the hypotheses always appears in the null hypothesis. In general, a hypothesis test about the value of a population mean μ must take one of the following three forms (where μ_0 is the hypothesized value of the population mean).

$H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_a: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$
One – Tailed	**One – Tailed**	**Two – Tailed**
(Lower tail test)	**(Upper tail test)**	







Depending on whether a hypothesis is one-tailed or two-tailed, the way of calculating p-value differs according to the diagram.

Type I and Type II Errors

Because hypothesis tests are based on sample data, we must allow for the possibility of errors. A Type I error is rejecting H_0 when it is true. The probability of making a Type I error when the null hypothesis is true as an equality is called the level of significance. Applications of hypothesis testing that only control the Type I error are often called significance tests.

A Type II error is accepting H_0 when it is false. It is difficult to control for the probability of making a Type II error. Statisticians avoid the risk of making a Type II error by using "do not reject H_0" and not "accept H_0".

3.2.1 Testing Single Population Mean (One Sample Z test)

The one sample z-test is used to test a single population mean against a given value. In addition, this test is used only when the population standard deviation σ is known from a prior source. Finally, data represent a simple random sample, and measurements that comprise the data are assumed to be accurate and meaningful.

Here, the null and alternative hypothesis can take any of the 3 forms,

$$H_0: \mu \geq \mu_0 \qquad H_0: \mu \leq \mu_0 \qquad H_0: \mu = \mu_0$$
$$H_a: \mu < \mu_0 \qquad H_a: \mu > \mu_0 \qquad H_a: \mu \neq \mu_0$$

Let us define the test statistic z in terms of the sample mean, the sample size and the population standard deviation σ :

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

Then the null hypothesis should be rejected if the p-value corresponding to the calculated z value ≤ α, where α is the desired level of significance.

Example

Suppose the manufacturer claims that the mean lifetime of a light bulb is at least 10,000 hours. In a sample of 30 light bulbs, it was found that they only last 9,900 hours on average. Assume the population standard deviation is 120 hours. At 5% significance level, can we reject the claim by the manufacturer?

The hypothesis of interest is :

$$H_0: \mu \geq 10000$$
$$H_a: \mu < 10000$$

```
> xbar <- 9900            # sample mean
>  mu0 <-10000            # hypothesized value
>  sigma <-120            # population standard deviation
>  n <-30                 # sample size
>  z <-(xbar−mu0)/(sigma/sqrt(n))
>  z                      # test statistic
[1] -4.564355
>
> pval <-pnorm(z)
> pval                   # lower tail p−value
[1] 2.505166e-06
```

As the p-value (2.505e-06) turns out to be less than .05, we reject the null hypothesis at 5% significance level. So the mean lifetime of a light bulb is not above 10,000 hours.

Example

Suppose the mean weight of King Penguins found in an Antarctic colony last year was 15.4 kg. In a sample of 35 penguins same time this year in the same colony, the mean penguin weight is 14.6 kg. Assume the population standard deviation is 2.5 kg. At 0.05 level of significance, can we reject the hypothesis that the mean penguin weight does not differ from last year?

The hypothesis of interest is :

$$H_0: \mu = 15.4$$
$$H_a: \mu \neq 15.4$$

```
> xbar <-14.6            # sample mean
>  mu0 <-15.4             # hypothesized value
>  sigma <-2.5            # population standard deviation
>  n <-35                 # sample size
>  z <-(xbar−mu0)/(sigma/sqrt(n))
>  z                      # test statistic
[1] -1.893146
>  pval <-2 * pnorm(z)
```

```
>  pval                   # two-tailed p-value
[1] 0.05833852
```

Here the p-value should be doubled, as the hypothesis is a two tailed one.
Since the final p-value turns out to be greater than the .05, we do not reject the null hypothesis at 5% significant level.

3.2.2 Testing Single Population Mean (One Sample t test)

The prior section used a z statistic to test a sample mean against an expectation. The z statistic needed population standard deviation σ (without estimating it from the data) to calculate the test statistic. To conduct a one-sample test when the population standard deviation is not known, we use a variant of the z statistic called the t statistic. The advantage of the t statistic is that it can use sample standard deviation s instead of population standard deviation σ.
The null and alternative hypotheses are identical to those used by the z test. The null hypothesis is $H_0:\mu=\mu_0$. Alternatives are

H_a: $\mu \neq \mu_0$ (two-sided)

H_a: $\mu > \mu_0$ (one-sided to right)

H_a: $\mu < \mu_0$ (one-sided to left)

Let us define the test statistic t in terms of the sample mean, the sample size and the sample standard deviation s as:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Here, the p-value corresponding to the calculated t value would be taken from the Student t distribution with n − 1 degrees of freedom. Once the p-value is found, then the null hypothesis should be rejected at α % level of significance, if the p-value ≤ α.

Example:

Suppose the manufacturer claims that the mean lifetime of a light bulb is at least 10,000 hours. In a sample of 30 light bulbs, it was found that they only last 9,900 hours on average. Assume the sample standard deviation is 125 hours. At .05 significant level, can we reject the claim by the manufacturer?

The hypothesis of interest is :

$$H_0: \mu \geq 10000$$
$$H_a: \mu < 10000$$

```
> xbar <-9900             # sample mean
> mu0 <-10000             # hypothesized value
> s <-125                 # sample standard deviation
> n <-30                  # sample size
> t <-(xbar−mu0)/(s/sqrt(n))
> t                       # test statistic
[1] -4.38178
> pval <- pt(t, df=n−1)
> pval                    # lower tail p−value
[1] 7.035026e-05
```

The lower tail p-value of the test statistic (7.035e-05) turns out to be less than 0.05. We reject the null hypothesis at 5% significance level. So the mean lifetime of a light bulb is not above 10,000 hours.

3.2.3 Testing Single Population Proportion

A z test can be used to test a proportion from a single population. To carry out the test usually the sample should include at least 10 successes and 10 failures (Some texts say that 5 successes and 5 failures are enough.) and the population size is at least 10 times as big as the sample size.

Here, the null and alternative hypothesis can take any of the 3 forms,

$$H_0: p \geq p_0 \qquad H_0: p \leq p_0 \qquad H_0: p = p_0$$
$$H_a: p < p_0 \qquad H_a: p > p_0 \qquad H_a: p \neq p_0$$

Let us define the test statistic z in terms of the sample proportion, sample size and the hypothesized value:

$$z = \frac{\bar{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

p-value can be found using the z distribution as in one sample z-test and the null hypothesis should be rejected if the p-value ≤ α, where α is the given level of significance.

Example

Suppose that 12% of apples harvested in an orchard last year was rotten. 30 out of 214 apples in a harvest sample this year turns out to be rotten. At .05 significance level, test the hypothesis that the proportion of rotten apples in harvest is more than 12% this year?

The hypothesis of interest is :

$$H_0\colon p \leq 0.12$$
$$H_a\colon p > 0.12$$

```
> pbar = 30/214          # sample proportion
> p0 = .12               # hypothesized value
> n = 214                # sample size
> z = (pbar-p0)/sqrt(p0*(1-p0)/n)
> z                      # test statistic
[1] 0.908751
> pval = pnorm(z, lower.tail=FALSE)
> pval                   # upper tail p-value
[1] 0.1817408
```

The upper tail p-value of the test statistic (0.18) turns out to be greater than 0.05. We do not reject the null hypothesis at 5% significance level. So the proportion of rotten apples in harvest is not more than 12% this year.

3.2.4 Testing the Difference between Two Independent Population Means : σ_1 and σ_2 unknown

We compare two independent samples to determine if the means for each sample are statistically significantly different. Either the two samples come from the same population and the populations mean difference is zero, or the two samples come from different populations and the population mean difference is not zero. The two samples do not have to be the same size.

Here, the null and alternative hypothesis can take any of the 3 forms,

$$H_0\colon \mu_1 - \mu_2 \geq 0 \qquad H_0\colon \mu_1 - \mu_2 \leq 0 \qquad H_0\colon \mu_1 - \mu_2 = 0$$
$$H_a\colon \mu_1 - \mu_2 < 0 \qquad H_a\colon \mu_1 - \mu_2 > 0 \qquad H_a\colon \mu_1 - \mu_2 \neq 0$$

Let us define the test statistic t in terms of the two sample means $\bar{x}_1$, $\bar{x}_2$, the sample sizes n_1, n_2 and the sample standard deviations s_1^2, s_2^2 :

$$\boldsymbol{t} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Here the degree of freedom to be used for t value is,

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

Here, the p-value would be taken from the Student t distribution with above degrees of freedom. Once the p-value corresponding to the calculated t value is found, then the null hypothesis should be rejected at α % level of significance, if the p-value ≤ α.

Example

In the data frame column 'mpg' of the data set 'mtcars', there are gas mileage data of various 1974 U.S. automobiles. Meanwhile, another data column in 'mtcars', named 'am', indicates the transmission type of the automobile model (0 = automatic, 1 = manual). In particular, the gas mileages for manual and automatic transmissions are two independent data populations. Assuming that the data in 'mtcars' follows the normal distribution, test the hypothesis that the difference between the mean gas mileage of manual and automatic transmissions are same or not. Use 5% level of significance.

The hypothesis of interest is :

$$H_0: \mu_a - \mu_m = 0$$
$$H_a: \mu_a - \mu_m \neq 0$$

```
> attach(mtcars)
> auto<-mpg[am==0]
> manual<-mpg[am==1]
> t.test(auto,manual)
```

```
Welch Two Sample t-test

data:  auto and manual
t = -3.7671, df = 18.332, p-value = 0.001374
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.280194  -3.209684
sample estimates:
mean of x mean of y
 17.14737  24.39231
```

Here the p-value of the test is 0.00137, which is <0.05. So at 5% level of significance we reject the null hypothesis of equal means of two populations. Hence the millage for auto transmission and manual transmission are not the same.

Further, this `t.test` command gives the 95% confidence for the mean difference also as (`-11.280194,  -3.209684`)

3.2.4 Testing the Difference between Two Independent Population Means : σ_1 and σ_2 unknown but equal

Here the situation in similar to the 3.2.3 case, and the only difference is the additional information that the two variances are equal. So the hypotheses of interest are the same. The test statistic and the degree of freedom are bit different and given as follows.

Test Statistic :

$$\boldsymbol{t} = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where s_p is called the pooled sample variance and can be calculated using the formula :

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

Here the degree of freedom for the t value is, $n_1 - n_2 - 2$.

Here, the p-value would be taken from the Student t distribution with $n_1 - n_2 - 2$ degrees of freedom. Then the null hypothesis should be rejected at α % level of significance, if the p-value ≤ α.

Example

Consider the same example given in section 3.2.3 and assume that the population variances for manual transmission car mileages and automatic transmission car mileages are equal. Test the hypothesis that the difference between the mean gas mileage of manual and automatic transmissions are same. Use 5% level of significance.

The hypothesis of interest is :

$$H_0: \mu_a - \mu_m = 0$$
$$H_a: \mu_a - \mu_m \neq 0$$

```
> auto<-mpg[am==0]
> manual<-mpg[am==1]
> t.test(auto,manual,var.equal = T)

        Two Sample t-test

data:  auto and manual
t = -4.1061, df = 30, p-value = 0.000285
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -10.84837  -3.64151
sample estimates:
mean of x mean of y
 17.14737  24.39231
```

Here the p-value of the test is 0.00028, which is <0.05. So at 5% level of significance we reject the null hypothesis of equal means of two populations. Hence the millage for auto transmission and manual transmission are not the same.

3.2.5 Testing the Difference between Two Population Means: Paired Samples

Paired data involves taking two measurements on the same subjects, called repeated sampling. Think of studying the effectiveness of a diet plan. You would weigh yourself prior to starting the diet and again following some time on the diet. Depending on how much weight you lost, you would determine if the diet was effective. What you might be interested in is estimating the true difference of the original weight and the weight lost after a certain period.

This test is testing the null hypothesis that there are no differences between the means of the two related groups. If we can reject the null hypothesis, then there is a significant difference between the means of the two groups. So the hypotheses of interest is,

$$H_0: \mu_d \geq 0 \qquad\qquad H_0: \mu_d \leq 0 \qquad\qquad H_0: \mu_d = 0$$
$$H_a: \mu_d < 0 \qquad\qquad H_a: \mu_d > 0 \qquad\qquad H_a: \mu_d \neq 0$$

where μ_d is the population mean difference of the two groups. Then the test statistic of interest is,

$$t = \frac{\bar{x}_d}{s_d / \sqrt{n}}$$

Where $\bar{x}_d$ is the mean of the differences in the two samples s_d is the standard deviation of the differences in the two samples and n is the sample size (number of pairs). This test statistic has a student t distribution with n-1 degree of freedom.

If the p-value corresponding to the calculated t value is ≤ α, then the null hypothesis should be rejected at α % level of significance.

Example

In the built-in data set named 'immer', the barley yield in years 1931 and 1932 of the same field are recorded. Assuming that the data in 'immer' follows the normal distribution, test the hypothesis that the mean barley yield in year 1931 (Y1) is larger than the mean barley yield in year 1932 (Y2).

The hypothesis of interest is,

$$H_0: \mu_d \leq 0 \qquad\qquad d = 1931 \text{ yield} - 1932 \text{ yield}$$
$$H_a: \mu_d > 0$$

```
> attach(immer)
> t.test(Y1,Y2,paired=T, alternative="greater")

        Paired t-test

data:  Y1 and Y2
t = 3.324, df = 29, p-value = 0.001206
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 7.778893      Inf
sample estimates:
mean of the differences
               15.91333
```

The p-value of the test is 0.0012, which is <0.05. So at 5% level of significance we reject the null hypothesis. Hence the yield in 1931 is higher than the yield in 1932.

3.2.6 Testing the Difference between Two Population Proportions

For statistical purposes, you can compare two populations or groups when the variable is categorical (for example, smoker/nonsmoker, Democrat/Republican, support/oppose an opinion, and so on). In order to make this comparison, two independent (separate) random samples need to be selected, one from each population. The interest here is to compare the response proportion between the two populations.

The null and alternative hypothesis can take any of the 3 forms,

$$H_0: p_1 - p_2 \geq 0 \qquad H_0: p_1 - p_2 \leq 0 \qquad H_0: p_1 - p_2 = 0$$
$$H_a: p_1 - p_2 < 0 \qquad H_a: p_1 - p_2 > 0 \qquad H_a: p_1 - p_2 \neq 0$$

where p_1 and p_2 are the population proportions for the group 1 and group 2 respectively.

The test statistic for this scenario is,

$$z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

Where $\bar{p}$ is the pooled proportion, which can be calculated using,

$$\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2}$$

Then the null hypothesis should be rejected if the p-value corresponding to the calculated z value ≤ α, where α is the desired level of significance.

Example

In the built-in data set named ‘quine’, children from an Australian town is classified by ethnic background, gender, age, learning status and the number of days absent from school. Assuming that the data in ‘quine’ follows the normal distribution, test the hypothesis that the difference between the female proportion of Aboriginal students and the female proportion of Non-Aboriginal students are the same.

The hypothesis of interest is,

$$H_0: p_A - p_N = 0$$
$$H_a: p_A - p_N \neq 0$$

```
> attach(quine)
> table(Eth,Sex)
   Sex
Eth  F  M
  A 38 31
  N 42 35
> prop.test(table(Eth,Sex),correct=F)

        2-sample test for equality of proportions without continuity
        correction
```

```
data:  table(Eth, Sex)
X-squared = 0.0041, df = 1, p-value = 0.9491
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1564218  0.1669620
sample estimates:
   prop 1    prop 2
0.5507246 0.5454545
```

The p-value of the test is 0.949, which is > 0.05. So at 5% level of significance we do not reject the null hypothesis. Hence the female proportion of Aboriginal students and the female proportion of Non-Aboriginal students are the same.