Hypothesis Testing

Test of Hypothesis

The test of hypothesis uses statistics that computed from samples. Given that these statistics have a sampling distribution, the decision is made in the face of random variation. This requires that we have clear decision rules for choosing between the alternatives. As for example, in a jurry trial we begin by assuming that the person is innocent and will only be convicted if there is a strong evidence against the presumption of innocence. The process has rigorous procedures for presenting and evaluating evidence, a judge to enforce the rules, and a decision mechanism.

A cereal company claim that on an average cereal packages weigh at least 16 ounces. We can test this claim by collecting a random sample of cereal packages, determining the weight of each one, and computing the sample mean package weight from the data.

Hypothesis

An assumption we make about a population parameter. A **statistical hypothesis test** is a method of making statistical decisions using experimental data.

Null hypothesis

The statistical hypothesis which is picked up for the test is known as null hypothesis. The null hypothesis is usually denoted by H_0.

Example

Let us consider a normal distribution with mean μ and variance σ^2. Then the hypothesis that the normal distribution has specified mean 5 i., e., $\mu = 5$ is known as null hypothesis.

Alternative hypothesis

Any hypothesis other than the null hypothesis is known as alternative hypothesis. It is usually denoted by H_a or H_1.

Example

Let us consider a normal distribution with mean μ and variance σ^2. We are going to test the hypothesis, $H_0 : \mu = 5$. Then the alternative hypothesis may be

$$H_a : \mu \neq 5$$

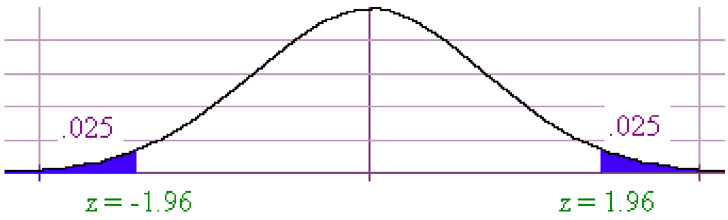
Or $H_a : \mu > 5$

Or $H_a : \mu < 5$

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

Rejection Regions

Suppose that $\alpha = 0.05$. We can draw the appropriate picture and find the Z score for -0.025 and 0.025. We call the outside regions the rejection regions.



We call the blue areas the *rejection region* since if the value of Z falls in these regions, we can say that the null hypothesis is very unlikely so we can reject the null hypothesis.

Test : A body of rules which leads to the decision regarding acceptance or rejection of the hypothesis is called a test. The statistic which is usually used to test the parameter of a population is known as test statistic.

Test may be classified as

- ➔ One tailed test
- ➔ Two tailed test

One tailed test

A test for which the entire rejection region lies in only of two tails either in the right tail or in the left tail of the sampling distribution of the test statistic is called one tailed. If we are interested to test the hypothesis $H_0 : \mu = \mu_0$ vs $H_a : \mu > \mu_0$ then we should use right tailed test. If we test $H_0 : \mu = \mu_0$ vs $H_a : \mu < \mu_0$ then we should use right tailed test.

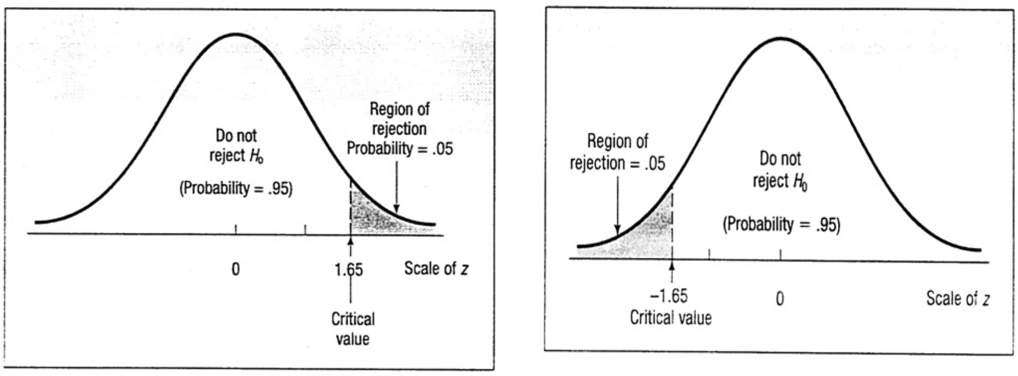


Fig: Sampling Distribution of the Statistic z, Right and Left-Tailed Test, 0.05 Level of Significance.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

Two tailed test

A test for which the rejection region is divided equally between two tails of the sampling distributions of the test statistics is called a two tailed test. If we are interested to test the $H_0: \mu = \mu_0$ vs $H_a: \mu \neq \mu_0$ then we should use two tailed test.

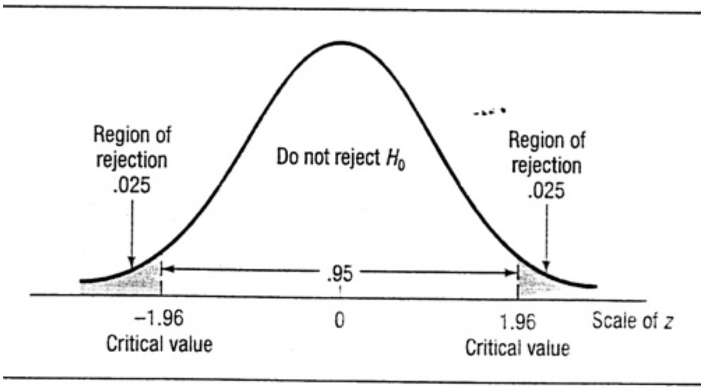


Fig: Regions of Non-rejection and Rejection for a Two-Tailed Test, 0.05 Level of Significance.

Critical value

The value of the standard statistic (Z or t) beyond which we reject the null hypothesis; the boundary between the acceptance and rejection regions.

Error

When using probability to decide whether a statistical test provides evidence for or against our predictions, there is always a chance of driving the wrong conclusions. Even when choosing a probability level of 95%, there is always a 5% chance that one rejects the null hypothesis when it was actually correct. This is called ***Type I error***, represented by the Greek letter α. The probability of a type I error (α) is called the ***significance level***.

It is possible to err in the opposite way if one fails to reject the null hypothesis when it is, in fact, incorrect. This is called ***Type II error***, represented by the Greek letter β. These two errors are represented in the following chart.

Table: Types of error

Type of decision	H_0 true	H_0 false
Reject H_0	Type I error (α)	Correct decision (1-α)
Accept H_0	Correct decision (1-β)	Type II error (β)

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

A related concept is power, which is the probability of rejecting the null hypothesis when it is actually false. Power is simply 1 minus the Type II error rate, and is usually expressed as $1 - \beta$.

When choosing the probability level of a test, it is possible to control the risk of committing a Type I error by choosing an appropriate α.

P-values

We define the p-value as the probability of obtaining a value of the test statistic as extreme as or more extreme than the actual value obtained when the null hypothesis is true. Thus the p-value is the smallest significance level at which a null hypothesis can be rejected given the observed sample statistic. Consider a random sample of size n from a population that has a normal distribution with mean μ and standard deviation σ, and resulted computed sample mean $\overline{X}$. We are ask to test the null hypothesis

$$H_0 : \mu = \mu_0$$

against the alternative hypothesis $H_1 : \mu \rangle \mu_0$

The p-value for the test is

$$\text{p-value}= P\left(\frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \geq Z_p \; H_0 : \mu = \mu_0 \right)$$

For example in cereal box problem with population mean $\mu = 15.9$, $\sigma = 0.4$, $n = 25$ and under the null hypothesis we had obtain a sample mean of 16.1 ounces. Then the p-value would be

$$P\left(\overline{X} > 16.1 \; H_0 : \mu = 15.9\right) = P\left(Z > \frac{16.1 - 15.9}{0.08} = 2.5 \right) = 1 - P(Z \leq 2.5) = 1 - 0.9968 = 0.0062$$

Steps in Hypothesis Testing

Step 1: Identify the null hypothesis H_0 and the alternate hypothesis H_A.

Step 2: Choose α. The value should be small, usually less than 10%. It is important to consider the consequences of both types of errors.

Step 3: Select the test statistic and determine its value from the sample data. This value is called the observed value of the test statistic. Remember that a t statistic is usually appropriate for a small number of samples; for larger number of samples, a z statistic can work well if data are normally distributed.

Step 4: Compare the observed value of the statistic to the critical value obtained for the chosen α .

Step 5: Make a decision.

If the test statistic falls in the critical region: Reject H_0 in favour of H_A .

If the test statistic does not fall in the critical region: Conclude that there is not enough evidence to reject H_0 .

Important tests of significance

The important tests of significance in statistics can be classified broadly as

a) Normal test : Normal tests are widely used in testing hypotheses regarding means, proportions and correlation coefficients.

b) t test: The t -tests are used for testing hypotheses regarding means, regression coefficients and correlation coefficients.

c) χ^2 test: The χ^2 -tests used for testing the equality of a set of variances and correlation coefficients.

d) F test: The F -tests is used in the analysis of variances for comparing the equality of more than two means.

Test of significance about mean

Here we consider the following cases:

- Comparison of a sample mean with an assigned population mean.
- Comparison of two independent sample means.
- Comparison of two correlated sample means.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

Comparison of a sample mean with an assigned population mean

Case 1: σ is known or estimated from a large sample $(n \geq 30)$.

Case 2 σ is unknown and the sample is small.

Case 1

Hypothesis

$$H_0 : \mu = \mu_0$$
$$H_A : \mu \neq \mu_0$$

Test statistic

$$|Z| = \left| \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right|$$

Case 2

Hypothesis

$$H_0 : \mu = \mu_0$$
$$H_A : \mu \neq \mu_0$$

Test statistic

$$|t| = \left| \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \right| \text{ with } n-1 \ d.f.$$

Where, $\bar{X} = \frac{1}{n}\sum x$ and

$$s^2 = \frac{1}{n-1}\sum\left(\sum x^2 - n\bar{x}^2\right)$$

Example: The mean life time of a sample of 100 light tubes produced by a company is found to be 1570 hours with standard deviation of 80 hours. Test the hypothesis that the mean life time of the tubes produced by the company is 1600 hours (consider 5% significance level).

Solution

1. **Hypothesis**

   We consider the following hypothesis

$$H_0 : \mu = 1600 \text{ vs } H_A : \mu \neq 1600$$

2. Significance level

Given that the significance level $\alpha = 5\% = 0.05$

3. Test statistic

In order to test the hypothesis we consider the following test statistic

$$|Z| = \left| \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right|$$

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

We have, $\bar{X}=1570$, $\sigma=80$, $n=100$.

$$\therefore |Z|=\left|\frac{1570-1600}{\frac{80}{\sqrt{100}}}\right|=3.75$$

4. Critical value

The critical value or tabulated value is 1.96.

5. Making decision

Since the calculate value is greater than the tabulated value, so we reject the null hypothesis. So that, the mean life time of the tubes produced by the company is not 1600 hours.

Example: A sample of 400 male students is found to have a mean height 67.47 inches. Can it be reasonably regarded as a sample from a large population with mean height 67.39 inches and standard deviation 1.30 inches? Test at 5% level of significance.

Solution

1. **Hypothesis**

We consider the following hypothesis

$$H_0: \mu = 67.39'' \text{ vs } H_A: \mu \neq 67.39''$$

2. Significance level

Given that the significance level $\alpha = 5\% = 0.05$

3. Test statistic

In order to test the hypothesis we consider the following test statistic

$$|Z|=\left|\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right|$$

We have, $\bar{X}=67.47''$, $\sigma=1.30''$, $n=400$.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

$$\therefore |Z| = \left| \frac{67.47 - 67.39}{\frac{1.30}{\sqrt{400}}} \right| = 1.231$$

4. Critical value

The critical value or tabulated value is 1.96.

5. Making decision

Since the calculate value is less than the tabulated value, so we accept the null hypothesis. So that, we may conclude that the given sample (with mean height 67.47 inches) can be regarded to have been taken from a population with mean height 67.39 inches and standard deviation 1.30 inches at 5% level of significance.

Example: Suppose that it is known from experience that the standard deviation of the weight of 8 ounces packages of cookies made by a certain bakery is 0.16 ounces. To check its production is under control on a given day, the true average of the packages is 8 ounces, they select a random sample of 40 packages and find their mean weight is 8.122 ounces. Test whether the production is under control or not at 5% level of significance.

Solution

1. **Hypothesis**

We consider the following hypothesis

$$H_0 : \mu = 8 \text{ vs } H_A : \mu \neq 8$$

2. Significance level

Given that the significance level $\alpha = 5\% = 0.05$

3. Test statistic

In order to test the hypothesis we consider the following test statistic

$$|Z| = \left| \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right|$$

We have, $\bar{X} = 8.112$, $\sigma = 0.16$, $n = 40$.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

$$\therefore |Z| = \left| \frac{8.112-8}{\frac{0.16}{\sqrt{40}}} \right| = 4.427$$

4. Critical value

The critical value or tabulated value is 1.96.

5. Making decision

Since the calculate value is greater than the tabulated value, so we reject the null hypothesis. So that, the production is not under control.

Case 2: σ is unknown and the sample is small

Example

A random sample of 10 boys had the following I. Q's:

70, 120, 110, 101, 88, 83, 95, 98, 107, 100

Do these data support the assumption of a population mean I. Q. of 100?

Solution

1. **Hypothesis**

We consider the following hypothesis

$$H_0 : \mu = 100 \text{ vs } H_A : \mu \neq 100$$

2. Significance level

Given that the significance level $\alpha = 5\% = 0.05$

3. Test statistic

In order to test the hypothesis we consider the following test statistic

$$|t| = \left| \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \right|$$

We have, $\bar{X} = \frac{1}{n}\sum x = \frac{1}{10} \times 972 = 97.2$,

$s = \sqrt{\frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right)} = \sqrt{\frac{1}{9}\left(96312 - 10 \times (97.2)^2\right)} = 14.27$, $n = 10$.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

$$\therefore |t| = \left| \frac{97.2 - 100}{\frac{14.27}{\sqrt{10}}} \right| = 0.62$$

4. Critical value

The critical value or tabulated value of t for $(10-1) = 9\ d.f$ is 2.2622.

5. Making decision

Since the calculate value is less than the tabulated value, so we accept the null hypothesis. So that, the data support that the population mean I. Q is 100.

Example: Is the temperature required to damage a computer on the average less than 110 degrees? Because of the price of testing, twenty computers were tested to see what minimum temperature will damage the computer. The damaging temperature averaged 109 degrees with a standard deviation of 3 degrees. (Use $\alpha = 0.05$)

Solution

1. **Hypothesis**

We consider the following hypothesis

$$H_0 : \mu = 110 \text{ vs } H_A : \mu < 110$$

2. Significance level

Given that the significance level $\alpha = 5\% = 0.05$

3. Test statistic

In order to test the hypothesis we consider the following test statistic

$$|t| = \left| \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \right|$$

Here, $n = 20$, $\bar{X} = 109$ and $\sigma = 3$

$$\therefore |t| = \left| \frac{109 - 110}{\frac{3}{\sqrt{20}}} \right| = 1.49$$

4. Critical value

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

This is a one tailed test, so we can go to our t-table with 19 degrees of freedom to find the critical value or tabulated value. The critical value is 1.73.

5. **Making decision**

Since the calculate value is less than the tabulated value, so we accept the null hypothesis and conclude that there is insufficient evidence to suggest that the temperature required to damage a computer on the average less than 110 degrees.

Example: The specimen of copper wires drawn from a large lot has the following breaking strength (in Kg. weigth):

578, 572, 570, 568, 572, 578, 570, 572, 596, 544

Test whether the mean breaking strength of the lot may be taking to be 578 Kg. weights by using 10% level of significance.

Solution

1. **Hypothesis**

We consider the following hypothesis

$$H_0: \mu = 578 \text{ vs } H_A: \mu \neq 578$$

2. **Significance level**

Given that the significance level $\alpha = 10\% = 0.1$

3. **Test statistic**

In order to test the hypothesis we consider the following test statistic

$$|t| = \left| \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \right|$$

We have, $\bar{X} = \frac{1}{n}\sum x = \frac{1}{10} \times 5720 = 572$,

$s = \sqrt{\frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right)} = \sqrt{\frac{1}{9}\left(3273296 - 10 \times (572)^2\right)} = 12.72$, $n = 10$.

$$\therefore |t| = \left| \frac{572 - 578}{\frac{12.72}{\sqrt{10}}} \right| = 1.49$$

4. Critical value

The critical value or tabulated value of t for $(10-1)=9\ d.f$ is 1.8331.

5. Making decision

Since the calculate value is less than the tabulated value, so we accept the null hypothesis at 10% level of significance. Hence we may conclude that the mean breaking strength of copper wires lot may be taken as 578 Kg. weight.

Test of significance about Variance

In addition to the need for test based on the sample mean there are a no. of situations where we want to determine it the population variance is a particular value or set of values. In modern quality control this need is particularly important because a process that has an excessively large variance can produce many defective items. Here we will develop procedures for testing the population variance σ^2, based on a random sample of n observations from a normally distributed population.

Here we consider the following cases:

i) Comparison of a sample variance with a specified value of population variance.
ii) Comparison of two independent sample variances
iii) Comparison of k (k > 2) independent sample variances.

Comparison of a sample variance with a specified population variance

Let us consider a random sample of size n from a normal population with variance σ^2 and $s^2 = \frac{\sum(x_i - \overline{x})^2}{n-1}$ represent the estimate of the unknown population variance σ^2. Then our null hypothesis is

$$H_0 : \sigma^2 = \sigma_0^2 \ \ vs\ H_1 : \sigma^2 \neq \sigma_0^2$$

and the test statistic is

$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ is distributed as χ^2 with (n – 1) degrees of freedom. This is an exact test and usually two-tailed.

Example: The quality control manager of stonehead chemicals has asked you to determine if the variance of impurities in its shipments of fertilizer is within the establish standard. This standard states that for 100-pound bags of fertilizer the variance in the pounds of impurities cannot exceed 4.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

Solution: A random sample of 20 bags is obtained and pounds of impurities are measured for each bag the sample variance was commuted to be 6.62.

Here $H_0 : \sigma^2 \leq \sigma_0^2 = 4$
Against the alternative $H_1 : \sigma^2 > 4$

Test statistics is

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(20-1)(6.62)}{4} = 31.45$$

$\therefore \chi_{19}^2\ (cal) = 31.45$
$\chi_{(19,0.05)}^2\ lab = 30.14$

Therefore we reject the null hypothesis and conclude that the variability of the impurities exceeds the standard. As a result we recommended that the production process be studied and improvements made to reduce the variability of the product components.

The p-value for this test is the probability of obtaining a chi-square statistics with 19 d.f. that is greater then the observed 31.45

p-value $= p\left(\frac{19 \times s^2}{\sigma_0^2} > \chi_{19}^2 = 31.45\right) = 0.036$

Equality of variances between two normally distributed populations

Let two random samples of sizes n_1 *and* n_2 are drawn from normal population. We wish to determine whether these normal population have the same variance σ^2.

The null hypothesis is

$H_0 : \sigma_1^2 = \sigma_2^2 \quad H_1 : \sigma_1^2 \neq \sigma_2^2$

The test statistics is

$$F_{(v_1, v_2)} = \frac{s_1^2}{s_2^2}$$

where s_1^2 and s_2^2 are the estimate of σ_1^2 and σ_2^2 from the two samples and is distributed as F with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

Example: The research staff of an on-line financial trading firm was interested in determining if there is a difference in the variance of the maturities of AAA-rated industrial bonds compared to CCC-rated industrial bonds.

Salutation: We design a study that compares the population variances of maturities for the two different bonds. We will test the null hypothesis

$$H_0 : \sigma_x^2 = \sigma_y^2$$
$$H_1 : \sigma_x^2 \neq \sigma_y^2$$

where σ_x^2 is the variance in maturities for AAA-rated bonds and σ_y^2 is the variance in maturities for CCC-rated bonds.

The significance level of the test was chooses as $\alpha = 0.02$
The test statistics

$$F_{(v_1, v_2)} = \frac{s_1^2}{s_2^2}$$

It is noted that among s_1^2 and s_2^2, the larger sample variance is in the numerator and smallest one in the denominator.

Now a random sample of 17 AAA rated bonds resulted in a sample variance $s_x^2 = 123.35$ and an independent random sample of 11 CCC rated bonds resulted in a sample variance $s_y^2 = 8.02$. The test statistic is thus

$$\frac{s_x^2}{s_y^2} = \frac{123.35}{8.02} = 15.38$$

Given a significance level of $\alpha = 0.02$ we find that the critical value of F from interpolation of table 7 of the Appendix is

$$F_{16,10,0.01} = 4.53$$

It is clearly observed that the computed value of F (15.38) exceeds the critical value (4.55) and we reject H_0 in favor of H_1. Thus there is strong evidence that the variances in maturities are different for test two types of bonds.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU