Lecture (6): Hypothesis Testing for Regression Parameters

Statistical Hypothesis

It is an assumption or claim (which may be true or false) about one or more parameters for one or more populations, and accepting this assumption depends on the information we obtain from the sample.

The acceptance or rejection of the hypothesis is as follows:

1. The hypothesis is accepted when the sample data supports the hypothesis, and we reject it when the sample data contradicts it.
2. The acceptance of the hypothesis results from the lack of sufficient evidence from the sample data to reject it, so accepting it does not mean that it is correct, but rejecting it based on the sample information is because the hypothesis is wrong, so the researcher tries to formulate a hypothesis in a way that he hopes will be rejected.

There are two types of hypotheses: -

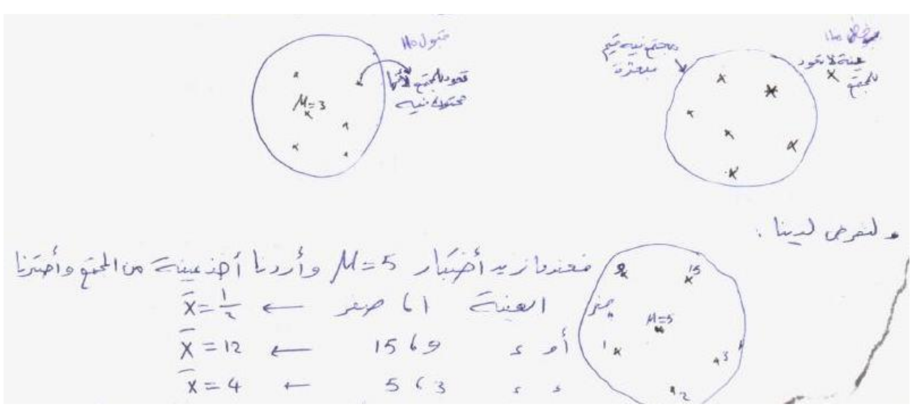
- **Null Hypothesis**

Symbolized by H_0, is related to the parameter under study and specifies the values that the researcher believes do not express the true value of the parameter.

- **Alternative Hypothesis**

It is symbolized by the symbol H_1, which determines the values of the parameter that the researcher believes to be correct, and he hopes that the sample data will lead to accepting the alternative hypothesis H_0 on the basis that it is a correct hypothesis, contrary to what he hopes for the null hypothesis H_0.

For example, testing the two hypotheses $H_0 : \mu = 3$, $H_1 : \mu \neq 3$, in fact we are testing whether our sample belongs to a community whose average is equal to 3



When we want to test $\mu = 5$ and we want to take a sample from the community and we choose the sample

$\overline{X} = \frac{1}{2} \leftarrow 0, 1$

$\overline{X} = 12 \leftarrow 15, 9$

$\overline{X} = 4 \leftarrow 5, 3$

If we test the hypothesis $H_0 : \mu = 5$, $H_1 : \mu \neq 5$ and $\overline{X} = 12, 15, 9$, then the reason for this error is in choosing a sample that does not represent the population.

Suppose we have $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$. This hypothesis means whether μ_1 and μ_2 are means of the same population, or whether the means of our two samples belong to two populations with equal means.

Suppose we have the hypothesis $H_0 : \beta_0 = 0$, $H_1 : \beta_0 \neq 0$. This means that the intercept parameter differs from zero, i.e. does the regression line pass through the origin or not (i.e. above or below the origin)?

That is, when we say $\beta_0 \neq 0$, it means either $\beta_0 > 0$, the regression line passes above the origin.

Or $\beta_0 < 0$, the regression line passes below the origin.

lso, if we have the hypothesis $H_0 : \beta_0 \leq 0$ the regression line passes through or below the origin.

$H_0 : \beta_0 \geq 0$, the regression line passes through or above the origin.

If we want to test whether the regression line passes through a certain point, we have:

$H_0 : \beta_0 = \beta_{00}$ where β_{00} is a real value.
$H_1 : \beta_0 \neq \beta_{00}$ where we test whether the regression line passes through this point or not.
For example, $H_0 : \beta_0 = 3$. The meaning of this hypothesis is whether a line passes through point 3 (3 units above the origin) or not.

$H_1 : \beta_0 \neq 3$ → $\beta_{00} = 3$

Thus, in the above examples, the value of $\beta_{00} = 0$, and there are other formulas such as

Does the regression line pass through point β_{00}? — $H_0 : \beta_0 = \beta_{00}$
Does the regression line pass above the point β_{00}? — $H_1 : \beta_0 > \beta_{00}$
Does the regression line pass under the point β_{00}? — or $H_1 : \beta_0 < \beta_{00}$

To test the hypotheses, we discussed above, the t-test is used, as:

$$\frac{\hat{\beta}_0 - \beta_{00}}{S(\hat{\beta}_0)} \sim t\left(\frac{\alpha}{2}, n-2\right)$$

That is, the sampling distribution for this quantity is a t-distribution with a significance level of $\frac{\alpha}{2}$ and $n-2$ degrees of freedom, meaning that the calculated t value (Cal.t) will be

$$Cal.t = \frac{\hat{\beta}_0 - \beta_{00}}{S(\hat{\beta}_0)}$$

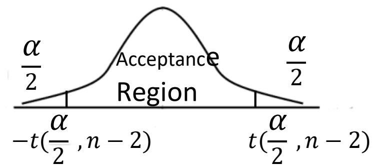
where $\hat{\beta}_0$ is the estimated value in the regression equation

β_{00} is the value for β_0 in the null hypothesis

and β_0 is the standard deviation of the estimated parameter $\hat{\beta}_0$

The calculated value of t is compared with the tabular value of t, where we have the following cases:

$$H_0 : \beta_0 = 0$$
$$H_1 : \beta_0 \neq 0$$

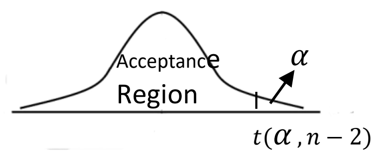


1) $\mathbf{tab.t = t(\frac{\alpha}{2}, n-2)}$

In this case, H_0 is rejected when $|\text{cal.t}| \geq \text{tab.t} = t(\frac{\alpha}{2}, n-2)$

This type of test is called a Two-sided test , Two-tails test.

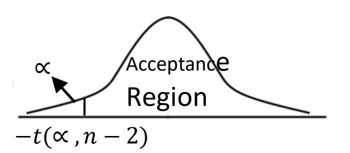
$$H_0 : \beta_0 = 0$$
$$H_1 : \beta_0 > 0$$



2) $\mathbf{tab.t = t(\alpha, n-2)}$

In this case, the null hypothesis is rejected when $\text{cal.t} \geq \text{tab.t} = t(\alpha, n-2)$

$$H_0 : \beta_0 = 0$$
$$H_1 : \beta_0 < 0$$



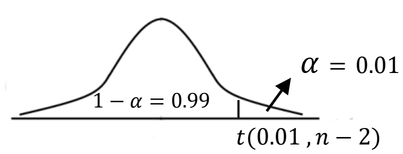
3) $\mathbf{tab.t = -t(\alpha, n-2)}$

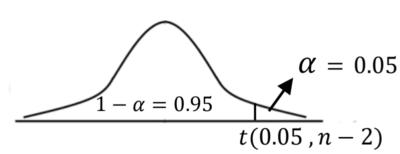
In this case, H_0 is rejected when $\text{cal.t} \leq -\text{tab.t} = -t(\alpha, n-2)$

The last two tests above are called one-sided test, one-tails test and we have α representing the level of significance which represents:

$$\alpha = \Pr(\text{Reiject } H_0/H_0\, H_0 \text{ is true})$$

It is the area of the H_0 rejection region. For example:





Regression Coefficient Test $(\hat{\beta}_1)$ $^{Y}/_{X}$

In the same way, if we have the following hypothesis:

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$ — either : $\beta_1 > 0$ / Or : $\beta_1 < 0$

Here, accepting H_0 means that our $\hat{\beta}_1$ belongs to a population in which β_1 equals zero, and the resulting $\hat{\beta}_1$ in the regression line equation came as a result of repeated errors, the most important of which is the error in selecting samples, and it has no statistical meaning.

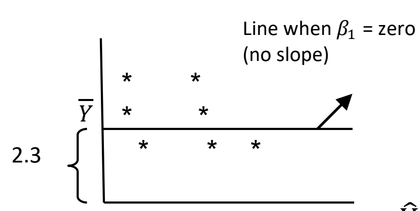
مثلا:-

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = 2.3 + 4.6 X_i$$

Before H_0, it means one of the following two things: -

First: That X has no relationship with Y through this model and that X does not explain the changes that occur in Y through it (through the model). In this case, the regression line will be drawn in the following form: -

We find (2.3), which is the intersection point $\hat{\beta}_0$, then we take a high value for X and connect the two points, so it will be in the following form: -



$$\hat{Y}_i = \bar{y} + \hat{\beta}_1 (X_i - \bar{X})$$

For our case, $\hat{Y}_i = \bar{y} = \bar{\hat{y}}$

$$\bar{y} = \bar{y} + \hat{\beta}_1 (X_i - \bar{X})$$

Second: There may be a relationship between X and Y through another model that explains the changes in Y (a non-linear model). That is, we accept ($\beta_1 = 0$) H_0 and say that there is no linear relationship between X and Y, but there may be a non-linear model that explains the relationship between X and Y.

When we accept $(\beta_1 \neq 0)H_1$, we say that there is importance for X in explaining the changes that occur in Y through the linear model we have, and it also gives us two results:

First: X actually explains the changes in Y within the relationship shown by the linear model.

Second: There may be another non-linear model that explains the relationship between X and Y.