Chapter # 04

## ⇒ Linear Regression with one regressor

### → Regression (Simple)

Simple linear regression is a regression model that estimates the relationship blw one independent variable and one dependent variable.

Simple linear regression gets its adjective "Simple", because it concerns the study of only one predictor variable. In contrast multiple linear regression, which we will study later in this course, gets its adjective "multiple" because it concerns the study of two or more predictor variables.

In simple regression model we talked about dependent (Y) and independent variables (x). for which we have some examples such as:

| | X (independent) | | Y (dependent) |
|---|---|---|---|
| i) | Income | → | expenditure |
| ii) | Study hours | → | Marks |
| iii) | Height of parents | → | height of children |

| Income | Expenditure |
|---|---|
| 50K | 40K |
| 55K | 42K |
| 75K | 50K |
| 81K | ? |

↳ Here we can be questioned that if a professional income is 81K so what or how much will be its expendure which is surely dependent on income.

↳ So we can estimate or predict our expenditure from our income.

So from the above given examples we can define a simple regression such as in which we check the dependence of a dependent variable on another independent variable.

⇒ The Linear regression model OR regression equations

$$\Rightarrow Y_i = \beta_0 + \beta_1 X_i + U_i$$

i) $Y_i$ = Dependent variable

ii) $\beta_0$ = Intercept (is the value of a dependent variable when independent

variable is Zero) ⇒ present only theoritically. But, note that "when there is no predictor donot predict."

$$Y_i = \beta_0 + \beta_1 X_i$$

$X_i = 0$, so.

↳ independent variable.

$$Y_i = \beta_0 + \beta_1(0)$$

$Y_i = \beta_0$ → ↳ intercept.

iii) $\beta_1$ = Regression coefficient or slope. (tells the rate of change in independent with unit change in independent variable.

$$\beta_{class\,size} = \frac{\text{Change in test score}}{\text{Change in class size}} \quad \frac{\Delta Y}{\Delta X}$$

↳ $\beta_1 = \frac{\Delta Y}{\Delta X}$ → dependent. / ↳ independent.

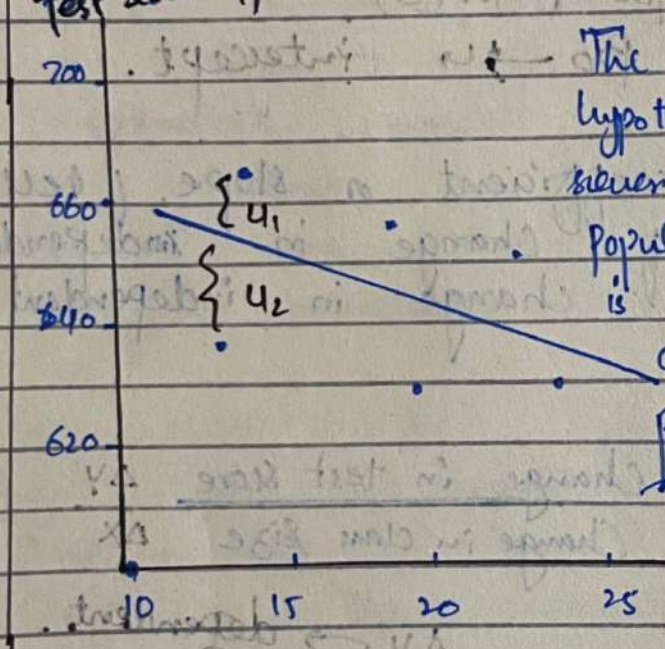
iv) $X_i$ = Independent variable

v) $U_i$ = Error term.

According to the equation, if you knew $\beta_0$ and $\beta_1$, not only would you be able to determine the change in dependent variable (test score) at a district associated with a change in independent.

variable (class size), but you would also be able to predict the average test score itself for a given class size.

Scatterplot of Test Scores vs. Student-Teacher Ratio




The scatterplot shows hypothetical observations for seven school districts. The population regression line is $\beta_0 + \beta_1 X$. The vertical distance from the $i$th point to the population regression line is $(Y_i - \beta_0 - \beta_1 X)$, which is the population error term $u_i$ for the $i^{th}$ observation.

In the about scatterplot the error term $u_i$ is positive because the test score in district # 1 is better than the predicted by the population regression line.

And $u_2$ is below it is below the population regression line.

So the test score for that district were worse than predicted, and $u_2 < 0$ means it is negative.

→ The linear regression model general equation is

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

↳ parameters.

↳ It is a general equation model it is based on population data and when we apply this equation on sample data then we will have fitted model

Before that we should know the difference b/w parameter and estimation.

↳ When we calculate any value from population it is called parameter.

↳ where as when we calculate any value from sample then it is called a estimator.

Fitted model:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{u}_i$$

↳ Y on X equation.

$\hat{Y}$ = Estimated value of $Y$.

↳ observed value.

$\hat{B}_0$ = Estimator of $B_0$
$\hat{B}_1$ = Estimator of $B_1$
X = Independent variable

So we have dependent and independent variables in the equation so if we know the values of $B_0$ and $B_1$ then we can easily calculate the value of dependent so for that we should have the values of $B_0$ and $B_1$

⇒ Estimating the coefficients of the linear regression models

In a practical situation such as the application to class size and test scores, the intercept $B_0$ and slope $B_1$ of the population regression line are unknown. Therefor, we must use data to estimate the unknown slope and intercept of the population regression line.

This estimation problem is similar to others you have faced in statistics, the same idea extends to the linear regression model. We donot know the population value of $B_1$, the

slope of the unknown population regression line relating X (class size) and Y (test scores). But just as it was possible to learn about the population mean using a simple of data drawn from that population, so is it possible to learn about the population slope ($\beta_1$) using a sample of data

As we know that if we have the values of ($\beta_0$) intercept and slope ($\beta_1$) then we can easily calculate the value of dependent variable using the simple linear regression model.

So,

For the $\beta_0$ and $\beta_1$ we can have the following formulas or we can use the OLS estimator to estimate the values of coefficient ($\beta_0$ and $\beta_1$)

(ii) $\beta_1 = \dfrac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$

(iii) $\beta_1 = \dfrac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$

(ii) $\beta_1 = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$

and for $\beta_0$ we have formula.

(ii) $\beta_0 = \frac{\sum y - b \sum x}{n}$

(iii) $\beta_0 = \bar{y} - b\bar{x}$

To understand these formulaes we have solve a Question and we will come to know that how they are calculated:

| X | Y | XY | $X^2$ | $\hat{y} = 2.001 + 0.387(x)$ | $e = y - \hat{y}$ |
|---|---|---|---|---|---|
| 1 | 2.1 | 2.1 | 1 | 2.388 | 2.001 − 2.388 = −0.288 |
| 1 | 2.5 | 2.5 | 1 | 2.388 | 0.112 |
| 2 | 3.1 | 6.2 | 4 | 2.775 | 0.325 |
| 3 | 3 | 9 | 9 | 3.162 | −0.162 |
| 4 | 3.8 | 15.2 | 16 | 3.540 | 0.251 |
| 4 | 3.2 | 12.8 | 16 | 3.549 | −0.349 |
| 5 | 4.3 | 21.5 | 25 | 3.936 | 0.364 |
| 6 | 3.9 | 23.4 | 36 | 4.323 | −0.423 |
| 6 | 4.4 | 26.4 | 36 | 4.323 | 0.077 |
| 7 | 4.8 | 33.6 | 49 | 4.71 | 0.09 |
| $\Sigma X =$ 39 | $\Sigma Y =$ 35.1 | $\Sigma XY =$ 152.7 | $\Sigma X^2 =$ 193 | | $\Sigma e = \Sigma(y-\hat{y}) =$ 0.003 ≈ 0 |

As we have the simple regression model such as.

$$\hat{Y} = \beta_0 + \beta_1 X_i$$

so know calculating $\beta_1$ such as using the formula

$$\hat{\beta}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$\hat{\beta}_1 = \frac{10(1527) - (39)(35.1)}{10(193) - (39)^2}$$

$$\hat{\beta}_1 = \frac{10(152.7) - (39)(35.1)}{1930 - (39)^2}$$

$$\hat{\beta}_1 = \frac{158.1}{409}$$

$$\boxed{\hat{\beta}_1 = 0.387}$$

$$\hat{\beta}_0 = \frac{\sum y - \hat{\beta}_1 \sum x}{n}$$

$$\hat{\beta}_0 = \frac{35.1 - 0.387(39)}{10}$$

$$\boxed{\hat{\beta}_0 = 2.001}$$

So now we calcalate the value of dependent variables as in column 5.

of the table using different values of X.

→ If $X = 10$ then

$$\hat{y} = 2.001 + 0.387(10)$$

$$\boxed{\hat{y} = 5.871}$$

→ if $X = 0$

then

$$\boxed{\hat{y} = 2.001}$$

⇒ Ordinary least Squares Estimator (OLS Estimator)

In data analysis, we use OLS estimator for estimating the unknown parameters in linear regression model. The goal is to minimizing the differences between blw the calculated observations in some arbitrary dataset and the responses predicted by the linear approximation of the data.

The OLS estimator chooses the regression coefficients so that the estimated regression line is as close as possible to the

observed data, where closeness is measured by the sum of squared mistakes made in predicting Y given X.

The sum of these squared prediction mistakes over all n observations is.

$$Y_i - (\hat{B}_0 + \hat{B}_1 X_i)$$

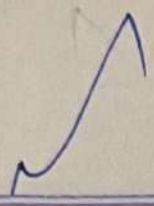
$$Y_i - \hat{B}_0 - \hat{B}_1 X_i$$

so it becomes as.

$$\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \rightarrow (i)$$

The estimators of the intercept and slope that minimize the sum of squared mistakes in equation (i) are called the ordinary least squares (OLS) estimators of $\beta_0$ and $\beta_1$.

You could compute the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ by trying different values of $\beta_0$ and $\beta_1$ repeatedly until you find those that minimize the total squared mistakes in equation or expression (i) They are the least squares estimators



because it is BLUE and consistent

⇒. Why Use the OLS estimators?

There are both practical and theoritical seasons to use the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$. Because OLS is the dominant method used in practice, it has become the common language for regression analysis throughout economics, finance, and the social science more generally. The OLS formulas are built into virtually all spreadsheet and statistical software packages, making OLS easy to use.

→ Under the assumptions of the simple linear regression model the OLS estimator is unbaised and consistant. The OLS estimator is also efficient among a certain class of unbaised estimators.

⇒ Measures of Fit:

Having estimated a linear regression, you might wonder how well that regression line describes the data. Does the regressor account for much

or for little of the variation in the dependent variable? Are the observations tightly clustered around the regression line, or are they spread out.

The $R^2$ and the standard error of regression measures how well the OLS regression line fits the data. The $R^2$ ranges blw 0 and 1 and measures the fraction of the variance of $Y_i$ that is explained by $X_i$. The standard error of regression measures how far is the $Y_i$ from its predicted value.

⇒ The $R^2$ &

The $R^2$ is a statistical measure that represents the proportion of a variance for a dependent variable that is explained by an independent variable or variables in a regression model.

Mathematically, the $R^2$ can be written as the ratio of the explained sum of squares (ESS) to the total sum.

of squares (TSS). The explained sum of squares (ESS) is the sum of squared deviations of the predicted value $\hat{Y}_i$, from its average, and the total sum of squares (TSS) is the sum of squared deviations of $Y_i$ from its average.

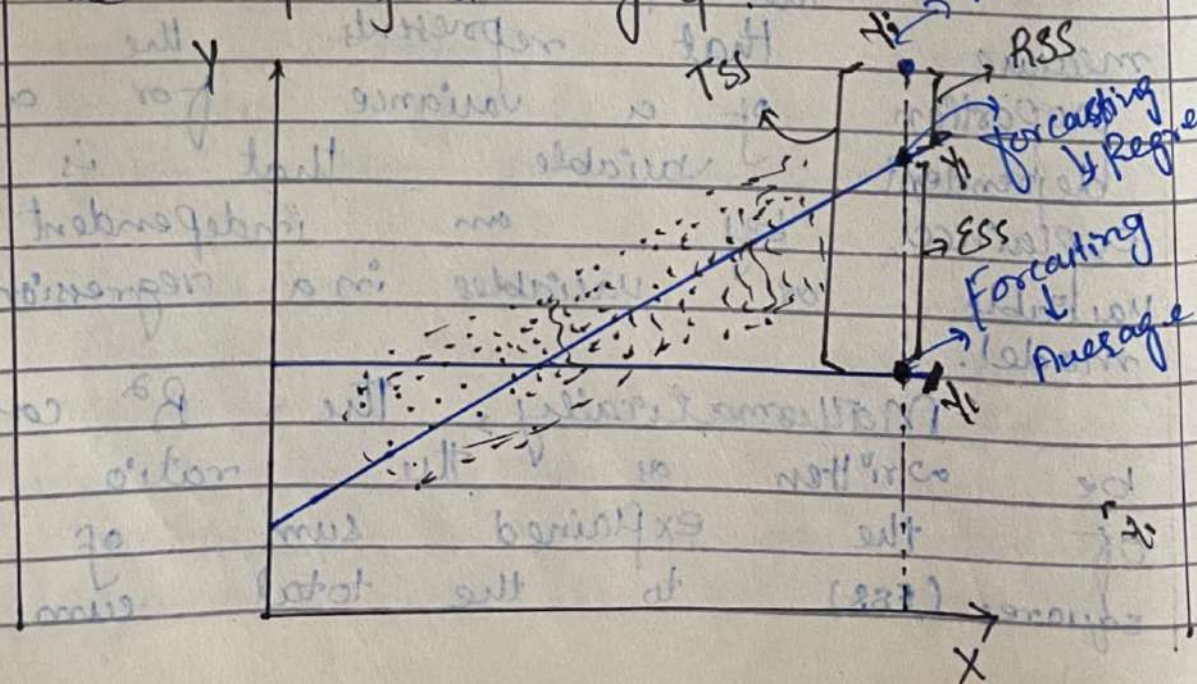
$$ESS = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

$$TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

→ forcasting from average

~~forcasting~~ from ~~average~~ actual value.

Now we will explain with the help of a graph.



Predicting

In the graph we are finding the value of Y given X. And here we are predicting the value of Y through forcasting, (i) forcasting by average. (ii) and then forcasting through a regression line.

When we forcasted through average then the difference b/w the predicted value and the actual value is called Total sum of Square.

$$TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

And after that we forcasted through the regression line and we found that there is a little variation from the $Y_i$ that -that in forcasting through average. so some of the portion of the TSS is explained by the regression line so we said it Explained sum of square.

$$ESS = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2.$$

And the difference b/w regression line and the observed value is called Residual sum of squares. (RSS)

So,

$$TSS = ESS + RSS$$

So, when we called about the ratio of ESS in TSS, it is basically called $R^2$.

$$R^2 = \frac{ESS}{TSS}$$

So as $R^2$ is greater the model will be considered fit

$R^2$ ranges b/w 0 and 1

$$0 \leq R^2 \leq 1$$

↳ $R^2$ is also called Coefficient of determination

$$\Rightarrow R^2 = 1 - \frac{RSS}{TSS}$$

As $TSS = ESS + RSS$

$\therefore ESS = TSS - RSS$

So equation becomes

$$R^2 = 1 - \frac{TSS - RSS}{TSS}$$

$$R^2 = \frac{TSS}{TSS} - \frac{RSS}{TSS}$$

$$R^2 = 1 - \frac{RSS}{TSS}$$

& $R^2 = \frac{ESS}{TSS} \Rightarrow R^2 = 1 - \frac{RSS}{TSS}$

⇒ The standard Error of the Regression
The standard error of the regression basically measures the magnitude of a typical regression residual in the units of y.

The SER measures the spread of the distribution of u. The standard error of regression is almost the sample standard deviation of the OLS residuals.

$$SER = \frac{SSR}{n-2}$$

OR it can also be defined as

The standard error of regression also known as the standard error of the estimate, represents the average distance that the

Observed values fall from the regression line. Conveniently, it tells you how wrong the regression model is on average using the units of the response variable.

⇒ What should we make of low $R^2$ and large SER

The fact that if the $R^2$ of any regression is low and the SER is large does not, by itself, imply that this regression is either "good" or "bad". What the low $R^2$ does tell us is that other important factors influence test score. These factors could include differences in the student body across districts, differences in school quality unrelated to the student teacher ratio, or luck on the test.

The low $R^2$ and high SER donot tell us what these factors are, but they do indicate that the student teacher ratio alone explains only a small part

of the variation in test scores in these data.

⇒ The Least Squares Assumptions:

This section presents a set of three assumptions on the linear regression model and the sampling scheme under which OLS provides an appropriate estimator of the unknown regression coefficients $\beta_0$ and $\beta_1$.

(1)→ Assumption # 01: The independent variable and the error term should not correlated

(*)⇒ The conditional distribution of $u_i$ given $X_i$ has a mean of zero.

This assumption is a formal mathematical statement about the "other factors" contained in $u_i$ and asserts that these other factors are unrelated to $X_i$ in the sense that, given a value of $X_i$, the mean of the distribution of these other factors is zero.

So, the error term $u_i$ has a conditional mean zero given $X_i$: $E(u_i|X_i) = 0$

So, if the conditional mean of one random variable given another is zero, then the two random variables have zero covariance and thus are uncorrelated. Thus the conditional mean assumption $E(u_i|X_i) = 0$ implies that $X_i$ and $u_i$ are uncorrelated or $corr(X_i, u_i) = 0$

However, if $X_i$ and $u_i$ are correlated, then it must be the case that $E(u_i|X_i) \neq 0$

So, if $X_i$ and $u_i$ are correlated, then the conditional mean assumption is violated.

(2)→ Assumption # 02:

$(X_i, Y_i)$, $i = 1, \ldots, n$, are independently and identically distributed.

This assumption is a statement about how the sample is drawn. If the observation are drawn by simple random sampling from a single large population, then $(X_i, Y_i)$, $i = 1, \ldots, n$, are i.i.d.

For example, let X be the ~~large population~~ age of the workers and Y be his or

her earnings, and imagine drawing a person at random from the population of workers. That randomly drawn person will have a certain age and earnings. If a sample of n workers is drawn from this population, then $(X_i, Y_i)$, $i = 1, \ldots, n$, necessarily have the same distribution. If they are drawn at random they are also distributed independently from one observation to the next; that is, they are i.i.d.

The i.i.d. assumption is a reasonable one for many data collection schemes. For example, survey data from a randomly chosen subset of the population typically can be treated as i.i.d.

(3)→ **Large Outliers are Unlikely**

The third least squares assumption is that large outliers — that is, observations with values of $X_i$, $Y_i$, or that both are far outside the usual range of data — are unlikely. Large outliers can make OLS regression

results misleading.

Another way to state this assumption is that X and Y have finite Kurtosis.

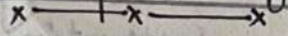
⇒ Sampling distributions of the OLS Estimators :

Because the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are computed from a randomly drawn samples, the estimators themselves are random variables with a probability distribution — the sampling distribution — that describes the values they could take over different possible random samples.

This section presents these sampling distributions.

In small samples, these distributions are complicated, but in large samples, they are approximately normal.

⇒ Completed :



⇒ Exercises :

Chapter No 5

⇒ Regression with a single regressor: Hypothesis Tests and Confidence Intervals:

This chapter continues the treatment of a linear regression with a single regressor.

⇒ Testing hypotheses about one of the regression Coefficients:

Your client, the superintendent, calls you with a problem. She has an angry taxpayer who claim that cutting class size will not help boost test scores, so reducing them is a waste of money. Class size, the taxpayer claims, has no effect on test scores.

The taxpayer is asserting that the population regression line is flat — that is, the slope '$\beta_1$' of the population regression line is zero, so the superintendent asks: Is there any evidence that you can reject the taxpayer's hypothesis that $\beta_1 = 0$ or should you accept it, at least tentatively

pending further new evidence?

So, this section discusses tests of hypotheses about the slope $\beta_1$ and intercept $\beta_0$ of the population regression line. We start by discussing two-sided tests of the slope $\beta_1$ in detail, then turn to one-sided tests and to tests of hypotheses regarding the intercept $\beta_0$

⇒ Two-sided hypothesis test concerning $\beta_1$ :-

The general approach to test hypotheses about the coefficient $\beta_1$ is the same as to test hypotheses about the population mean.

Testing hypotheses about the population mean:-

First of all we should know about the null and alternative hypotheses.

A null hypotheses is a statement, in which there is no relationship between two variables.

An alternative hypotheses is a statement in which there is some statistical significance

between two measured phenomena.

So,

Null :- The effect equals zero.

Alternative :- The effect doesn't equal zero

Recall that the null hypothesis that the mean of Y is a specific value $\mu_{Y,0}$ can be written as $H_0 : E(Y) = \mu_{Y,0}$, and the two-sided alternative is $H_1 : E(Y) \neq \mu_{Y,0}$.

To test the null hypotheses $H_0$ against the two-sided alternative proceeds as in three steps.

(i) The first is two compute the standard error of $\bar{Y}$, $SE(\bar{Y})$.

(ii) The second step is to compute the t-Statistic.

$$\text{The t-Statistic} = \frac{(\bar{Y} - \mu_{Y,0})}{SE(\bar{Y})}$$

($\bar{Y}$: estimator; $\mu_{Y,0}$: hypothesized value; $SE(\bar{Y})$: SE of estimator)

(iii) The third step is to compute the p-value, which is the smallest significance level at which the null hypotheses could be rejected, based on

test statistic actually observed.

Alternatively the third step can be replaced by simply comparing the t-statistic to the critical value, appropriate for the test with the desired significance level. For example, a two-sided test with a 5% significance level would reject the null hypotheses if $t > 1.96$. In this case, the population mean is said to be statistically significant different from the hypothesized value at the 5% significance level.

→ Testing hypotheses about the slope $\beta_1$

The null and alternative hypotheses need to be tested precisely before they can be test.

The angry texpayer hypotheses is that $\beta_1 = 0$. More generally, under the null hypotheses the true population slope $\beta_1$ takes on some specific value, $\beta_{1,0}$. Under the two sided alternative, $\beta_1$ does not equal $\beta_{1,0}$.

That is, the null hypotheses and the two - sided alternative hypothesis are..

$H_0 = \beta_1 = \beta_{1,0}$ VS $H_1 = \beta_1 \neq \beta_{1,0}$

To test the null hypotheses $H_0$, we follow the same three steps as for the population mean.

(i) Compute the standard error of $\hat{\beta}_1$

In applications the $SE(\hat{\beta}_1)$ is computed by regression software so that it is easy to use in practice.

(ii) The 2nd step is to compute the t-statistic

$$t\text{-Statistic} = \frac{\text{Estimator} - \text{hypothesize value}}{SE(\text{estimater})}$$

$$t\text{-Statistics} = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}$$

(iii) The 3rd step is to compute the p-value.

Reject the hypothesis at 5% Significance level if the p-value is less than 5%, or

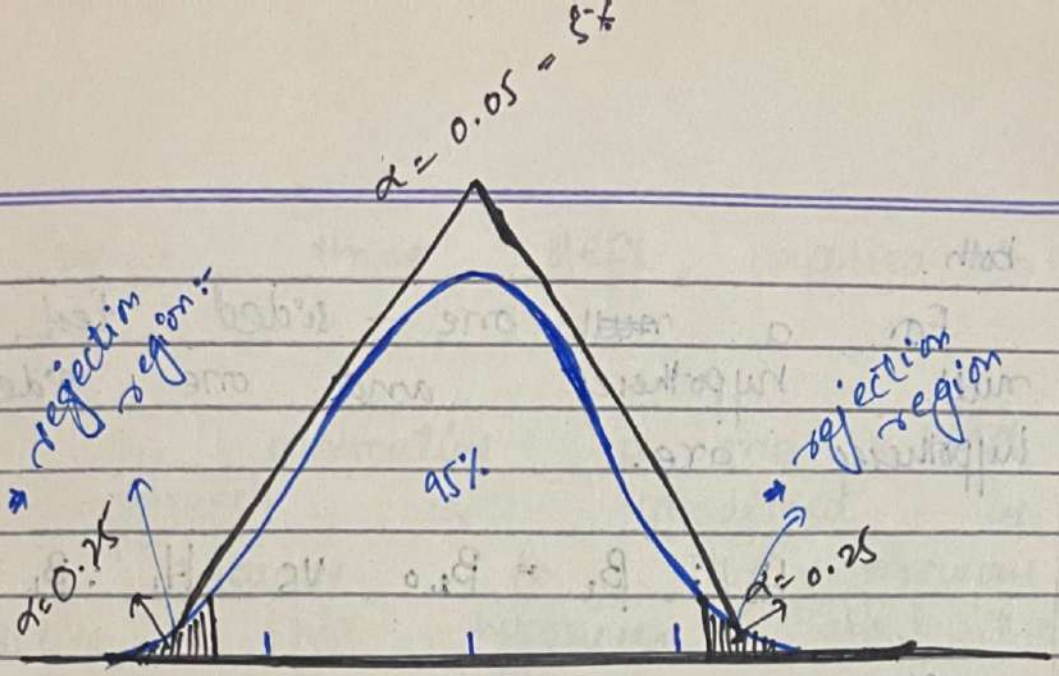
equivalently, if $|t^{act}| > 1.96$.

The standard error and t-Statistic and p-value testing $\beta_1 = 0$ are computed automatically by regression software.

A p-value of less than 5% provides evidence against the null hypotheses in the sense that, under the null hypotheses, the probability of obtaining a value of $\hat{\beta}_1$ at least as far from the null as that actually observed is less than 5%. If so the null hypotheses is rejected at a 5% significance level.

Alternatively, the hypothesis can be ~~rejected~~ tested at the 5% significance level simply by comparing the absolute value of the t-Statistic to 1.96, the critical value for a two-sided test, and rejecting the null hypothesis at a 5% level if $|t^{act}| > 1.96$.

Graphically,




If that sample t-statistic value falls in any of the rejection regions that means we can reject our null hypotheses.

→ One-sided hypotheses Concerning $\beta_1$ & In one-sided hypotheses the null and alternative hypotheses are.

**Null :-** The effect is greater than or equal to zero

**Alternative :-** The effect is less than zero

A one sided hypotheses test is a statistical test in which the critical area of a distribution is one-sided so that it is either greater than or less than a certain value, but not

both.

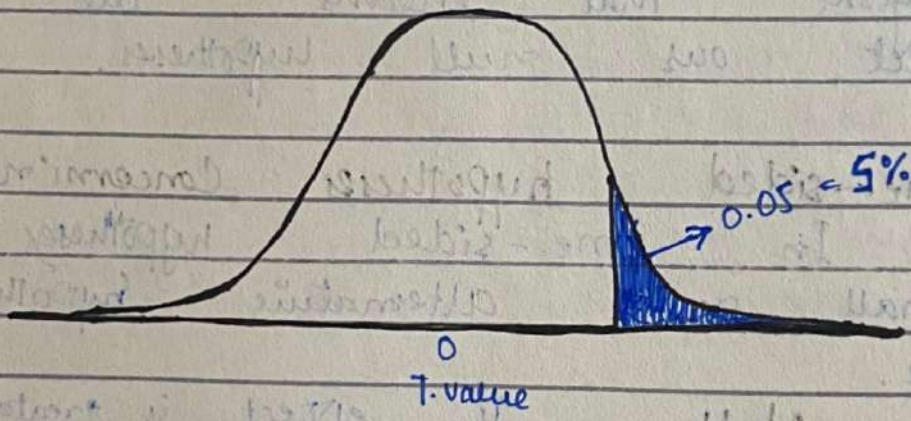
For a ~~null~~ one – sided test, the null hypotheses and one sided alternative hypotheses are.

$H_0 : \beta_1 = \beta_{1,0}$ vs $H_1 : \beta_1 < \beta_{1,0}$

or

$H_0 : \beta_1 = \beta_{1,0}$ vs $H_1 : \beta_1 > \beta_{1,0}$

(one sided alternative)




⇒ Testing hypotheses about the intercept $\beta_0$ &

The null hypotheses concerning the intercept and the two-sided alternative are.

$H_0 : \beta_0 = \beta_{0,0}$ $H_1 : \beta_0 \neq \beta_{0,0}$

The general approach to testing this null hypotheses consists of

the same three steps applied to $\beta_1$.

If the alternative is one-sided, this approach is modified as was discussed in the previous subsection for hypotheses about the slope.

⇒ Hypotheses Test are useful if you have a specific null hypotheses in mind. Being able to accept or reject this null hypotheses based on the statistical hypotheses / evidence.

Yet there are many times that no sigle hypotheses about a regression coefficient is dominant, and instead one would like to know a range of values of coefficients that are consistant with the data. This calls for construction a confidence interval.

⇒ Confidence Intervals for a regression Coefficients :

Because any statistical estimate of the slope $\beta_1$ necessarily has sampling uncertainity, we cannot determine the true value of.

$\beta_1$ exactly from the sample of data. It is possible, however, to use the OLS estimator and its standard error to construct a confidence interval for a slope $\beta_1$ or for the intercept $\beta_0$

→ Confidence interval for $\beta_1$ :

Confidence interval means that, for example as we take 100 samples of size 420, then 95 samples will contains true population parametric value and 5 samples will not contain (Note :- if $\hat{\beta}_1$ is normally distributed).

Recall that a 95% confidence is, equivalently:

↳ A set of points that cannot be rejected at a 5% significance level.

↳ A set-valued function of a data that contains the true parameter value 95% of the time in repeated samples.

Because the t-statistic for $\beta_1$ is $N(0,1)$ in large samples,

construction of a 95% confidence for $\beta_1$ is just like the case in the sample mean.

95% confidence interval for $\beta_1$ =

$$= \{ \hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1) \}$$

Confidence interval example :-

Test score and STR

Estimated regression line :

$$TS = 698.9 - 2.28 \times STR$$

$SE(\hat{\beta}_0) = 10.4$ $SE(\hat{\beta}_1) = 0.52$

95% confidence interval for $\hat{\beta}_1$ :

$$\{ \hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1) \} = \{ -2.28 \pm 1.96 \times 0.52 \}$$

$$= \{ -3.30, -1.26 \}.$$

So, in the above example the $\beta_1 \neq 0$ because it does contain zero so we reject the null hypothesis.

The following two statements are equivalent.

is The 95% confidence interval

does not include zero;
(ii) Null hypothesis $\beta_1 = 0$ is rejected at the 5% level.

Now:- implication of this example:
Our hypothetical superintendent is contemplating reducing the STR by 2. Because the 95% confidence interval for $\beta_1$ is $[-3.30, -1.26]$, the effect of reducing the STR by 2 could be as great as $-3.30 \times -2 = 6.60$ or as little as $-1.26 \times -2 = 2.52$.
Thus decreasing the student-teacher ratio by 2 is predicted to increase test scores by between 2.52 and 6.60 points, with a 95% confidence level.

⇒ Regression When X is a Binary Variable:
Regression analysis can also be used when the regressor is binary - that is, when it takes on only two values, 0 or 1.

For example:-

- X = 1 if small class size, = 0 if not.
- X = 1 if female, = 0 if male.

↳ X = 1 if Rural, = 0 if Urban.

Binary regressors are sometimes called "Dummy" variables.

→ Interpretation of the regression Coefficients:

The mechanics of regression with a binary regressor are the same as if it was/is continuous. The interpretation of $\beta_1$, however, is different, and it turns out that regression with a binary variable is equivalent to performing a difference of means analysis.

Simple regression model = $Y_i = \beta_0 + \beta_1 X_i + u_i$, where X is binary ($X_i = 0$ or 1).

(i) When $X_i = 0$

$Y_i = \beta_0 + u_i$

that is, $E(Y_i | X_i = 0) = \beta_0$

(ii) When $X_i = 1$, $Y_i = \beta_0 + \beta_1 * u_i$

↳ The mean of $Y_i$ is $\beta_0 + \beta_1$

↳ That is, $E(Y_i | X_i = 1) = \beta_0 + \beta_1$.

Here we now that $X_i$ is not continuous; so, it is not useful to

think of $\beta_1$ as a slope; indeed, because $X_i$ can take on only "two" values two values, there is no "line", so it makes no sense to talk about the slope. Thus we will not refer to $\beta_1$ as a slope in equation $Y_i = \beta_0 + \beta_1 X_i + u_i$ (when $X_i$ is binary); instead we will simply refer to $\beta_1$ as the coefficient multiplying $X_i$.

$\beta_1$ is not a slope so What is it? The best way to interpret $\beta_0$ and $\beta_1$ in a regression with a binary regressor is to consider, one at a time, the two possible cases, $X_i = 0$ and $X_i = 1$

So we know that

(i) when $X_i = 0$.
then $Y_i = \beta_0$

(ii) And when $X_i = 1$
then $Y_i = \beta_0 + \beta_1$

So, the difference between the conditional expectation of $Y_i$ when $X_i = 1$ and when $X_i = 0$ or

$B_{1c} = (E(Y_i|X_i=1) - (E(Y_i|X_i=0)$

So $(\beta_0 + \beta_1) - \beta_0 = \beta_1$ is the difference b/w these two means.

→ Heteroskedasticity and Homoskedasticity.

Our only assumption about the distribution of $u_i$ conditional on $X_i$ is that it has mean of zero. If furthermore, the variance of this conditional distribution does not depend on $X_i$, then the error term is said to be homoskedastic, otherwise the error term is heteroskedastic

→ If $Var(u|X=x)$ is constant - that is, if the variance of the conditional distribution of $u$ given $X$ does not depend on $X$ — then $u$ is said to be homoskedastic, otherwise, $u$ is heteroskedastic.

Equal group variances = Homoskedasticity.
Unequal group variances = Heteroskedasticity.

When :

$$E(u|X_i) = 0$$

↳ The variance of u does not depend on X.

⇒ We now have two formulas for standard error for $\hat{\beta}_1$.

↳ Homoskedasticity-only standard errors.

↳ These are valid only for if errors are homoskedastic.

↳ The usual standard errors - to differentiate the two, it is conventional to call these heteroskedascity-robust standard error, because they are valid wheather or not the errors are heteroskedastic.

↳ The main advantage of the homoskedascity-only standard error is that the formula is simpler. But the disadvantage is that the formula is only correct if the errors are homoskedastic.

↳ If the error are either homoskedastic or heteroskedastic and you use heteroskedastic-robust standard

errors, you are OK.

↳ If the error is heteroskedastic and you use the homoskedasticity-only formula for standard error, your standard error will be wrong.

↳ The two formulas coincide (when n is large) in the special case of homoskedasticity.

↳ So, you should always heteroskedasticity-robust standard errors.

⇒ The extended least squares Assumptions.

(i) $E(u|x_i) = 0$
(ii) $(X_i, Y_i)$, $i = 1, \ldots, n$, are i.i.d.
(iii) Large outliers are unlikely.
(iv) u is homoskedastic
(v) u is distributed $N(0, \sigma^2)$

⇒ The Gauss-Markov Theorem &
If the three OLS assumptions hold and if errors are homoskedastic, then the OLS estimator is Best (most efficient) Linear conditionally Unbaised Estimator (BLUE)

# Chapter # 06 &

## Linear regression with multiple Regressors &

This Chapter explains how to estimate the coefficients of the multiple linear regression model.

As we know that school districts with low STR tends to have higher test scores in California data set, perhaps students from districts with small classes have other advantages that help them perform well on standardized tests. Could this have produced misleading results, and, if so what can be done?

Omitted factor, such as student characteristics, can, in fact, make the OLS estimator of the effect of class size on test scores misleading or more precisely baised. This Chapter explains the "omitted variable bias", and introduces multiple regression, a method that can eliminate omitted variable bias

→ Definition of Omitted Variable Bias:

→ Omitted variable bias:

If the regressor (the STR) is correlated with a variable that has been omitted from the analysis (The el-pct) and that determines in part, the dependent variable (Test score), then the OLS estimator will have omitted variable bias.

Omitted variable bias occures when two conditions are true:-

(i) When the omitted variable is correlated with the included regressor, and

(ii) When the omitted variable is a determinant of the dependent variable.

Both conditions must hold for the omission of Z to result in omitted variable bias.

In the test score example:-

1) English language ability (Wheather the student has english a a 2nd language) plausibly effects the standardized test scores; so

(el. pct) or Z.

U is determinant of Y

8) Immigrant communities tend to be less affluent and thus have smaller school budgets - and higher STR, so Z is correlated with X.

So, this results into omitted variable bias.

→ Omitted variable bias and the first least squares assumption.

Omitted variable bias means that the 1st test squares assumption - that $E(u_i|X_i) = 0$ is incorrect. To see why, recall that the error term $u_i$ in the linear regression model with a single regressor represents all factors, other than $X_i$, that are determinants of $Y_i$. If one of these other factors is correlated with $X_i$ this means that the error term (which contains this factor) is correlated with $X_i$. Because $u_i$ and $X_i$ are correlated the conditional mean of $u_i$ given $X_i$ is nonzero. This correlation therefore violates the first least

squares assumption, and the consequence is serious. The OLS estimator is biased. This bias does not vanish even in very large samples, and the OLS estimator is inconsistant.

⇒ The Formula for Omitted Variable Bias

The discussion about the omitted variable bias can be summerized mathematically by a formula for this bias.

let the correlation b/w X and $u_i$ be $corr(X_i, u_i) = \rho_{Xu}$.

Suppose that the 2nd and 3rd OLS assumption hold, but the 1st does not because $\rho_{Xu}$ is nonzero. Then the OLS estimator has the limit.

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \underbrace{\rho_{Xu} \frac{\sigma_u}{\sigma_X}}_{\text{bias}}$$

This formula summerizes several of the ideas discussed about omitted variable bias:

1) Omitted variable bias is a problem whether the sample

size $\beta_1$ is large or small, Because $\hat{\beta}_1$ does not converge in probability to the true value $\beta_1$, $\hat{\beta}_1$ is biased and inconsistent. So, the term $\rho_{Xu} \frac{\sigma_u}{\sigma_X}$ in formula is the bias in $\hat{\beta}_1$ that persists even in large sample.

2) Whether this bias is large or small in practice depends on the correlation $\rho_{Xu}$ b/w the regressor and the error term. The larger $|\rho_{Xu}|$ is, the larger the bias.

3) The direction of bias in $\hat{\beta}_1$ depends on whether X and u are positively or negatively correlated. For example the percentage of students learning English has a negative effect on district test score, so the percentage of english learners enter the error term with a negative sign. Thus the student teacher ratio (X) would be negatively correlated with the error term (u) so $\rho_{Xu} < 0$ and the coefficients on the student-teacher ratio $\hat{\beta}_1$ would be biased towards a negative number.

⇒ The multiple regression model:

Consider the case of two regressors.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i,$$
$$i = 1, \ldots, n$$

$Y$ = Dependent variable.

$X_1, X_2$ = two independent variables.

$\beta_0$ = population intercept.

$\beta_1$ = $\frac{\Delta Y}{\Delta X_1}$, holding $X_2$ constant.

$\beta_2$ = $\Delta Y / \Delta X_2$, holding $X_1$ constant.

$u_i$ = the regression error term.

⇒ The OLS Estimator in multiple Regressions:

This section describes how the coefficients of the multiple regression model can be estimated using OLS.

The OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are the values of $b_0, b_1, \ldots, b_k$ that minimizes the sum of squared predictions mistakes $\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{1i})^2$.

The OLS predicted values $\hat{Y}_i$ and residuals are:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki}, \quad i = 1, \ldots, n,$$

and

$$\hat{u}_i = Y_i - \hat{Y}_i, \quad i = 1, \ldots, n.$$

⇒ Measures of Fit in multiple Regression

Three commonly used summary statistics in multiple regression are the standard error of the regression, the regression $R^2$, and the adjusted $R^2$ (also known as $\bar{R}^2$). All three statistics measure how well the OLS estimate of the multiple regression line describes, or "fits" the data.

→ The standard error of regression (SER)

The SER estimates the standard deviation of the error term $u_i$. Thus the SER is a measure of spread of the distribution of Y around the regression line. In multiple regression, the SER is

$$SER = \frac{SSR}{n-k-1}$$

The only difference b/w SER in single and multiple regression is that here the divisor is $n-k-1$ rather than $n-2$.

If there is a single regressor. $k=1$ then both the formulas becomes the same. When $n$ is large, the effect in both formulas become negligible.

$$SER = \frac{SSR}{n-2}, \qquad SER = \frac{SSR}{n-k-1}$$

when n is large it becomes negligible.

→ $R^2$ & The regression $R^2$ is the fraction of the sample variance of $Y_i$ explained by the regressors. Equivalently $R^2$ is

$$R^2 = \frac{ESS}{TSS}, \text{ or } R^2 = 1 - \frac{RSS}{TSS}$$

In multiple regression when $R^2$ increases, whenever a regressor is added unless the estimated coefficient on the added regressor is exactly zero, so in general the RSS will decrease when a new variable is added, But this means that $R^2$ generally increases when a new regressor is added.

→ The Adjusted $R^2$" ($\bar{R}^2$) &

Because the $R^2$ increases when a new variable is added, an increase in $R^2$ does not mean that adding a variable actually improves the fit of the model. In this sense, the $R^2$ gives an inflated estimate of how well the regression fits the data. One way to correct for this is to deflate or reduce the $R^2$ by some factor, and this is what the adjusted $R^2$, or $\bar{R}^2$ does.

The adjusted $R^2$ is the modified version of the $R^2$ that does not necessarily increases when a regressor is added. The $\bar{R}^2$ is

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\left(\frac{SSR}{TSS}\right).$$

→(ii) $(n-1)/(n-k-1)$ is always greater than 1, so $R^2$ is always less than $\bar{R}^2$.
→(iii) Adding new regressors will decrease SSR, which increase $\bar{R}^2$. Whether $\bar{R}^2$ ↑ or ↓ depends on which of these effect is strong.
(iv) The $\bar{R}^2$ can be negative, when SSR↓ in small amount and $\frac{n-1}{n-k-1}$ increases in large amount.

→ The Least Squares Assumptions in Multiple Regression:

The multiple regression model is written as.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

(1) The conditional distribution of u given the X's has mean zero, that is, $E(u|X_1 = x_1, \ldots, X_k = x_k) = 0$.

(2) $(X_{1i}, \ldots, X_{ki}, Y_i)$, $i = 1, \ldots, n$, are i.i.d.

(3) Large outliers are unlikely.

(4) There is no perfect multicolinearity.

⇒ Perfect multicolinearity:
Perfect multicolinearity is when one of the regressors is an exact linear function of the other regressor.

For example:

⇒ In the previous regression, $\beta_1$ is the effect on Testscore of a unit change in STR, and also holding STR constant, so

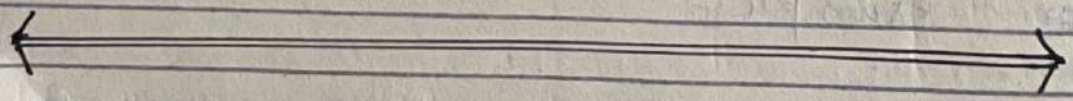
it makes no sense, so if you have perfect multicolinearity your statistical software will let you know — either by crashing or giving an error message or by "dropping" one of the variables arbitrarily.

↳ The solution to perfect multicolinearity is to modify your list of regressors so that you no longer have perfect multicolinearity.

→ Imperfect multicolinearity:

Imperfect multicolinearity occurs when the two or more regressors are very highly correlated but not perfectly correlated.

Imperfect multicolinearity does not pose any problem for the theory of the OLS estimators. So, the imperfect multicolinearity is not necessarily an error



# Chapter # 07 "Hypothesis test and confidence intervals in Multiple Regression."

As disussed in Chapter 6, multiple regression analysis provides a way to mitigate the problem of omitted variable bias by including the additional regressors, thereby controlling the effects of those additional regressors. The coefficients of the multiple regression model can be estimated by OLS. Like all estimators, the OLS estimator has sampling uncertainity because its value differs from one sample to the next.

This Chapter presents method for quantifying the sampling uncertainity of the OLS estimator through the use of standard errors, statistical hypothesis tests, and confidence intervals. One new possibility that arise in multiple regression is a hypothesis that simultaneously involves two or more coefficients of a regression. The General approach to testing such "joint" hypothesis involves a new test statistic, the F-statistic

⇒ Hypothesis Tests and Confidence intervals for a single Coefficient

This section describes how to test hypotheses, and how to construct confidence intervals for a single coefficient in a multiple regression model.

4) → Hypothesis tests for a single Coefficient

Suppose that you want to test the hypothesis that a change in the student teacher ratio has no effect on test score, holding constant the PCTEL in the district. This corresponds to hypothesizing that the true coefficient $\beta_2$ on the STR is zero ($\beta_1 = 0$) in the population regression of test scores on STR and PctEL.

So here the null hypothesis is $H_0 = \beta_1 = 0$, and the alternative hypothesis $H_1 = \beta_1 \neq 0$ (two sided alternative).

The procedure for testing this

null hypothesis when there is a single regression is done in three steps

(i) Compute the standard error of $\hat{\beta}_1$, $SE(\hat{\beta}_1)$

(ii) Compute the t-statistics,

$$t = \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)}$$

(iii) Compute the p-value,

$$p\text{-value} = 2\Phi(-|t^{act}|),$$

Where $t^{act}$ is the value of the t-statistics actually computed. Reject the hypothesis at the 5% significance level if the p-value is less then 0.05 or, equivalently, if $|t^{act}| > 1.96$.

(iii) Confidence intervals for a single Coefficients

The method for constructing a confidence interval in the multiple regression model is also the same as in the single regressor model.

The 95% confidence interval is

(i) 95% confidence interval for $\beta_1$ =

$$\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)$$

(ii) A 90% confidence interval is obtained by replacing 1.96 in above equation with 1.64

The method for conducting a hypothesis test and the method for conducting a confidence interval rely on large samples.

⇒ An angry taxpayer asserts that the population values of both the coefficient on STR ($\beta_1$) and coefficient on spending per pupil ($\beta_2$) are zero; that is, he hypothesizes that both $\beta_1 = 0$ and $\beta_2 = 0$. So,

→ the tax payer's hypothesis is a joint hypothesis, and to test it we need a new tool the F. Statistic.

## ⇒ Tests of Joint Hypotheses :

This section describes how to formulate joint hypothesis on multiple regression coefficient and how to test them using an F - Statistic.

Testing hypothesis on two or more Coefficients :

Joint Null Hypothesis :

Consider the regression

$$\widehat{TS} = 649.6 - 0.29\,STR + 3.87\,Expn - 0.656\,PctEL$$
$$(15.5) \quad (0.48) \quad (1.59) \quad (0.032)$$

One angry taxpayer hypothesizes that neither the student - teacher ratio nor expenditure per pupil have an effect on test score, once we control for the percentage of English learners.

So, we can write this null hypothesis mathematically as

$H_0: \beta_1 = 0$ and $\beta_2 = 0$ Vs $H_1: \beta_1 \neq 0$ and/or $\beta_2 \neq 0$

Here the null hypothesis imposes two restrictions on multiple regression model : $\beta_1 = 0$ and $\beta_2 = 0$

In General the joint hypothesis is a hypothesis that imposes two or more restrictions on the regression coefficients

→ If any one (or more than one) of the equalities under the null hypothesis $H_0$ is false, then the joint null hypothesis itself is false. Thus, the alternative hypothesis is that at least one of the equalities in the null hypothesis $H_0$ does not hold.

→ Why can't I just test the individual coefficients one at a time?

Although it seems it is possible to test a joint hypothesis by using the usual t-statistics to test the restriction one at a time, but the calculations shows that it is unreliable. When the regressors are corelated,

the situations is even more complicated, so it becomes deficult to test the null hypothesis on "one at a time" using the t-statistics.

→ We can test the joint hypothesis that is more powerful, especially when the regressors are highly correlated. That approach is based on the F-statistics.

⇒ The F-statistic

The F-statistic is used to test joint hypothesis about regression coefficients.

A simple F-statistic formula that is easy to understand (it is only valid if the errors are homoskedastic, but it might help intuition).

→ The homoskedasticity-only F-statistic

When the errors are homoskedastic, there is a simple formula for computing the "homoskedasticity-only" F-statistic:

→ Run the two regressions, one under the null hypothesis (the

"restricted" regression) and one under the alternative hypothesis (the "unrestricted" regression)

The Homoskedasticity-only F-statistic is given by the formula.

$$F = \frac{(SSR_{restricted} - SSR_{unrestricted})/q}{SSR_{unrestricted}/(n - K_{unrestricted} - 1)}$$

q = number of restrictions under the null hypothesis and
K = number of regressors in the unrestricted regression.

An alternative equivalent formula for the homoskedasticity-only F-statistic is based on the $R^2$ of the two regressions.

$$F = \frac{(R^2_{unrestricted} - R^2_{restricted})/q}{(1 - R^2_{unrestricted})(n - K_{unrestricted} - 1)}$$

⇒ Examples
(i) Restricted regression:

$\widehat{TS} = 644.7 - 0.671\ PctEL$, $R^2_{restricted} = 0.449$

→(ii). Unrestricted regression :

$$\widehat{TS} = 649.6 - 0.29\,STR + 3.87\,Expn - 0.656\,PctEL$$

$R^2_{un} = 0.4366$, $k_{un} = 3$, $q_r = 2$, $n = 420$.

So

$$F = \frac{(0.4366 - 0.4149)/2}{(1 - 0.4366)(420 - 3 - 1)}$$

$$F = 8.01$$

So it is higher then the critical value so that null hypothesis is rejected that $\beta_1 = 0$ and $\beta_2 = 0$. Because 8.01 exceeds the 1% critical value of 4.61, the hypothesis the hypothesis is rejected at 1% level using the homoskedasticity-only test.

⇒ Testing Single Restrictions involving Multiple coefficients :

Sometimes economic theory suggests a single restriction that involves two or more regression coefficients. For example, theory might suggests a null hypothesis of the form $\beta_1 = \beta_2$; that is the effect of

the first and the 2nd regressor are the same. In this case the task is to test this null hypothesis against the alternative that the two coefficients differ.

$$H_0 : \beta_1 = \beta_2 \quad \text{vs} \quad H_1 : \beta_1 \neq \beta_2$$

This null hypothesis has a single restriction, so $q = 1$, but the restriction involves multiple coefficients ($\beta_1$ and $\beta_2$)

→ Confidence sets for multiple coefficients

The method is conceptually similar to the method in single coefficient using the t-statistic, except that the confidence set for multiple coefficients is based on F-statistics.

→ Model specification for multiple regression

The job of determining which variables to include in multiple regression - that is the problem

of choosing a regression specification can be quite challenging; and no single rule applies in all situations

It is important to rely on your expert knowledge of the empirical problem and to focus on obtaining an unbaised estimate of the causal effect of interest; donot rely solely on purely statistical measures of fit as $R^2$ or $\bar{R}^2$.

→ Omitted Variable Bias in Multiple Regressions

Omitted variable bias is the bias in the OLS estimator that arise when one or more included regressors are correlated with an omitted variable. For omitted variable bias to arise, two things must be true:

(i) At least one of the included regressors must be correlated with the omitted variable.

(ii) The omitted variable bias must be a determinant of the dependent variable Y.

→ Interpreting the $R^2$ and the Adjusted $R^2$ in practice:

The $R^2$ and $\bar{R}^2$ tell you whether the regressors are good at predicting, or "explaining", the values of the dependent variable in the sample of data on hand. If the $R^2$ (or $\bar{R}^2$) is nearly 1, then the regressors produce good predictions of the dependent variable in that sample, in the sense that the variance of the OLS residual is small compared to the variance of the dependent variable. If the $R^2$ or ($\bar{R}^2$) is nearly 0, the opposite is true.

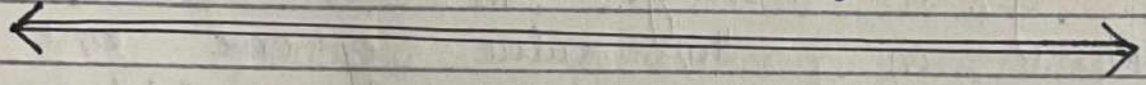
The $R^2$ and $\bar{R}^2$ do not tell you whether:

(1) An increase in the $R^2$ or $\bar{R}^2$ does not necessary mean that an added variable is statistically significant.

(2) A high $R^2$ or $\bar{R}^2$ does not mean that the regressors are true cause of dependent variable.

(3) A high $R^2$ or $\bar{R}^2$ does not mean that there is no omitted variable bias.

(4) A high $\bar{R}^2$ or $R^2$ does not necessarily mean that you have the most appropriate set of regressors, nor does a low $R^2$ or $\bar{R}^2$ necessarily mean that you have an inappropriate set of regressors.



chapter # 08

# "Nonlinear Regression Functions"

In chapter 4 through 7, the population regression function was assumed to be linear. In other words, the slope of the population regression function was constant, so the effect of Y of a unit change in X does not itself depend on the value of X. But what if the effect on Y of a change in X does depend on the value of one or more of the independent variables? If so, the population regression function is non linear.

A non linear function is a function whose line is not constant. A non linear regression function with a slope that is not constant.

→ A General Strategy for modeling non linear regression functions

→ Test scores and District incomes

The relationship blw income and test scores is not a straight line. Rather it is a nolineal.

A non linear function is a function with a slope that is not constant.

One way to approximate such a curve mathematically is to model the relationship as a quadratic function.

That is, we could model test scores as a function of income and the square of income.

A quadratic population regression model relating test scores and income is written mathematically as

$$\text{Test Score}_i = \beta_0 + \beta_1 \text{income}_i + \beta_2 \text{income}_i^2 + u_i$$

It is a quadratic function of the independent variable

The quadratic function captures in the scatterplot: It is steep for low values of district income but flattens out when district income is high. In short, the quadratic regression function function seems to fit the data better than the linear one.

This quadratic regression model becomes linear when $\beta_2 = 0$ so it becomes linear

when (null hypothesis $\beta_x = 0$)

→ Non Linear Regression Population Regression functions. General ideas

If a relation between Y and X is nonlinear.

» The effect on Y of a change in X depends on the value of X — that is, the marginal effect of X is not constant.

» The Effect on Y of a change in X in nonLinear Specifications

The Expected effect on Y of a change in $X_1$ in the non linear regression model:-

The expected change in Y, $\Delta Y$, associated with the change in $X_1$, $\Delta X_1$, holding $X_2, \ldots, X_k$ constant, is the difference b/w the value of the population regression function before and after changing $X_1$, holding $X_2, \ldots, X_k$ constant that is expected change in Y is the difference.

$$\Delta Y = f(X_1 + \Delta X_1, X_2, \ldots, X_K) - f(X_1, X_2, \ldots, X_K)$$

The estimator of this unknown population difference is the difference b/w the predicted values for these two cases. Let $\hat{f}(X_1, X_2, X_3 \ldots, X_K)$ be the predicted values of Y based on the estimator $\hat{f}$ of the population regression function the predicted change in Y is

$$\Delta \hat{Y} = \hat{f}(X_1 + \Delta X_1, X_2, \ldots X_K) - \hat{f}(X_1, X_2, \ldots, X_K)$$

→ **Nonlinear Function of a single Independent variable**

This section provides two methods for modeling a nonlinear regression function that involves only one independent variable

We have two ways to capture nonlinearity in variables
Both are complementary approaches.

1) **Polynomials**

Polynomials is an expression of more than two algebraic terms especially the same several terms that contain different

powers of the same variables

A population regression function is approximated by a quadratic, cubic, or higher-degree polynomial

Approximate the population regression function by a polynomial:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_r X_i^r + u_i$$

This is just the linear multiple regression model - except that the regressors are powers of $X_i$!

Estimation, hypothesis testing, etc. proceeds as in the multiple regression model using OLS.

When $r = 2$ ⇒ quadratic regression model.
When $r = 3$ ⇒ Cubic regression model.

→ Testing null hypothesis that the population regression function is linear:

According to the null hypothesis $H_0$ that the regression is linear and the alternative ($H_1$) that it is polynomial of

degree $r$ corresponds to.

$H_0 : \beta_2 = 0, \beta_3 = 0, \ldots, \beta_r = 0$ vs $H_1$: at least on $\beta_j \neq 0$, $j = 2, \ldots, r$.

→ Which degree polynomial should I use? This is summerized in following steps

1) Pick a maximum value of $r$ and estimate the polynomial regression for that $r$.

2) Use the t-statistic to test the hypothesis that the coefficient on $X^r$ [$\beta_r = 0$] is zero. If you reject this hypothesis, then $X^r$ belongs in the regression, so use the polynomial of degree $r$.

3) If you donot reject $\beta_r = 0$ in step 2, eliminate $X^r$ from the regression and estimate a polynomial regression of degree $r-1$. Test whether the coefficient on $X^{r-1}$ is zero. If you reject, use the polynomial of degree $r-1$

4) If you donot reject $\beta_{r-1}$,

≠0 in step 3, continue this procedure until the coefficient on the highest power is your polynomial is statistically significant.

⇒ Interpretation of coefficients in polynomial regression models

The coefficients in polynomial regressions donot have a simple regression interpretation. The best way to interpret polynomial regressions is to plot the estimated regression function and calculate the estimated effect on Y associated with a change in X for one or more values of X.

⇒ Logarithms:

Usually log of a variable may make non normal data normal, non linear data linear, heteroskedastic data homoskedastic.

Logarithms convert changes in variables into percentage changes

The logarithms function has the following useful properties

$$\ln(1/x) = -\ln(x)$$
$$\ln(ax) = \ln a + \ln x$$
$$\ln(x/a) = \ln x - \ln a$$
$$\ln(x^a) = a \ln x$$

→ **Logarithms and percentages**

The link between logarithms and percentages relies on a key fact:

When $\Delta x$ is small, the difference between the logarithms of $x + \Delta x$ and the logarithms of $x$ is approximately $\Delta x/x$, the percentage change in $x$ is divided by 100. That is,

$$\ln(x + \Delta x) - \ln(x) \cong \frac{\Delta x}{x} \quad \left(\text{when } \frac{\Delta x}{x} \text{ is small}\right)$$

**The Three logarithmic regression models**

There are three different cases in which logarithms might be used.

Logarithms can be used to transform the dependent variable $Y$, an independent variable $X$ or both (but the variable being transformed must be positive). The following

Summarizes these three cases and the interpretation of the regression coefficient $\beta_1$. In each case $\beta_1$ can be estimated using OLS after taking the logarithm of the dependent and/or independent variable.

| Regression Specification | Interpretation of $\beta_1$ |
|---|---|
| 1) $Y_i = \beta_0 + \beta_1 \ln(X) + u_i$ (linear log model) | A 1% change in X is associated with a change of of a $0.01\beta_1$ change in Y. |
| 2) $\ln(Y_i) = \beta_0 + \beta_1 X_1 + u_i$ (log linear model) | A change in X by one unit ($\Delta X = 1$) is associated with $100\beta_1$% change in Y |
| 3) $\ln(Y_i) = \beta_0 + \beta_1 \ln(X) + u_i$ (log-log model) | A 1% change in X is associated with a $\beta_1$% change in Y, so $\beta_1$ is the elasticity of Y with respect to X. |

⇒ Interactions between Independent Variables:
An interaction occurs when an

independent variable has a different effect on the dependent variable (outcome), depending on the values of another independent variable.

We consider three cases

(i) when both independent variables are binary.

(ii) when one is binary and the other is continuous.

(iii) And when both are continuous.

→ Interaction between two binary variables:
The population regression of $Y_i$ on these binary variables is

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i$$

(binary variables)

→ There could be an interaction between the two binary variables.
It is easy to modify the specification so that it does by introducing another regressor, the product of the two binary variables, $D_{1i} \times D_{2i}$. The resulting regression is

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i$$

↳ This this population regression model is called binary variable interaction regression model.

A method for interpreting coefficients in regressions with binary variables first compute the expected values of Y for each possible case described by the set of binary variables. Next compare these expected values. Each coefficient can be expressed either as an expected value or as the difference b/w two or more expected values.

→ Interactions between a continuous and a binary variables
Through the use of interaction term $X_i \times D_i$, the population regression line relating $Y_i$ and the continuous variable $X_i$ can have a slope that depends on the binary variables $D_i$.
There are three possibilities

(i) Different intercept, same slope

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i$$

ii) Different intercept and slope.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + u_i$$

iii) Same intercept and Different slope.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i \times D_i) + u_i$$

(The coefficients of all three can be estimated by OLS)

3)→ Interaction between Two continuous Variables :

This interaction is modeled by augmenting the linear regression model with an interaction term that is product of $X_{1i}$ and $X_{2i}$.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i$$

⇒ Interactions in Multiple regressions :

The interaction term between two independent variables $X_1$ and $X_2$ is there product $X_{1i} \times X_{2i}$. Including this interaction terms allows the effect on Y ~~with~~ of a change

in $X_1$ to depend on the values of $X_2$ and, conversely, allows the effect of a change in $X_2$ to depend on the values of $X_1$

The coefficient on $X_1$ and $X_2$ is the effect of a one unit increase in $X_1$ and $X_2$, above and beyond the sum of the individual effects of a unit increase in $X_1$ alone and a unit increase in $X_2$ along. This is true whether $X_1$ and/or $X_2$ are continuous or binary.

⇒ Non linear Effects on Test scores of a student - Teacher ratio &

Briefly explained in Book with the help of a table