ECON4150 - Introductory Econometrics

Lecture 3: Review of Statistics & OLS

Monique de Haan
(moniqued@econ.uio.no)

Stock and Watson Chapter 3-4

Lecture outline

- Comparing means from different populations
  - Ideal randomized experiment
- Using the *t*-statistic when *n* is small
- Relationship between two random variables
  - California test score data
  - scatter plot
  - sample covariance
  - sample correlation
- Linear regression with 1 regressor
  - derivation of the OLS estimators
  - measures of fit (R^2 and *SER*)

Comparing means from different populations

- Previous lecture we tested the hypothesis that the mean wage of individuals with a master degree equals 60000
- Suppose we would like to test whether the mean wages of men and women with a master degree differ by an amount d_0

$$H_0 : \mu_{w^M} - \mu_{w^F} = d_0 \qquad H_1 : \mu_{w^M} - \mu_{w^F} \neq d_0$$

- To test the null hypothesis against the two-sided alternative we follow the 4 steps with some adjustments

Step 1: Estimate $(\mu_{w^M} - \mu_{w^F})$ by $\left(\overline{W}_M - \overline{W}_F\right)$

- Because a weighted average of 2 independent normal random variables is itself normally distributed we have ($Cov\left(\overline{W}_M, \overline{W}_F\right) = 0$)

$$\overline{W}_M - \overline{W}_F \sim N\left(\mu_{w^M} - \mu_{w^F} \ , \ \frac{\sigma_{W_M}}{n_M} + \frac{\sigma_{W_F}}{n_F}\right)$$

Comparing means from different populations

Step 2: Estimate σ_{W_M} and σ_{W_F} to obtain $SE\left(\overline{W}_M - \overline{W}_F\right)$

$$SE\left(\overline{W}_M - \overline{W}_F\right) = \sqrt{\frac{s^2_{W_M}}{n_M} + \frac{s^2_{W_F}}{n_F}}$$

Step 3: compute the t-statistic

$$t^{act} = \frac{\left(\overline{W}_M - \overline{W}_F\right) - d_0}{SE\left(\overline{W}_M - \overline{W}_F\right)}$$

Step 4: Reject H_0 at a 5% significance level if

- $|t^{act}| > 1.96$
- or if $p - value < 0.05$

Comparing means from different populations

Suppose we have random samples of 500 men and 500 women with a master degree

and we would like to test that the mean wages are equal:

$$H_0 : \mu_{w^M} - \mu_{w^F} = 0 \qquad H_1 : \mu_{w^M} - \mu_{w^F} \neq 0$$

Step 1: $\overline{W}_M - \overline{W}_F = 64159.45 - 53163.41 = 10996.04$

Step 2: $SE\left(\overline{W}_M - \overline{W}_F\right) = 1240.709$

Step 3: $t^{act} = \frac{\left(\overline{W}_M - \overline{W}_F\right) - 0}{SE\left(\overline{W}_M - \overline{W}_F\right)} = \frac{10996.04}{1240.709} = 8.86$

Step 4: Since we use a 5% significance level, we reject H_0 because $|t^{act}| = 8.86 > 1.96$

Comparing means from different populations

Difference in mean wages between men and women with a master degree

This is how to do the test in Stata:

```
. ttest wage, by(female)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     500    64159.45    847.7946    18957.26    62493.76    65825.13
       1 |     500    53163.41    905.8709    20255.89    51383.62     54943.2
---------+--------------------------------------------------------------------
combined |   1,000    58661.43    643.9819     20364.5    57397.72    59925.14
---------+--------------------------------------------------------------------
    diff |            10996.04    1240.709                 8561.34    13430.73
------------------------------------------------------------------------------
    diff = mean( 0) - mean( 1)                                    t =   8.8627
Ho: diff = 0                                     degrees of freedom =      998

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Confidence interval for the difference in population means

- The method for constructing a confidence interval for 1 population mean can be easily extended to the difference between 2 population means
- A hypothesized value of the difference in means d_0 will be rejected if $|t| > 1.96$
- and will be in the confidence set if $|t| \leq 1.96$
- Thus the 95% confidence interval for $(\mu_{W_M} - \mu_{W_F})$ are the values of d_0 within ± 1.96 standard errors of $\left(\overline{W}_M - \overline{W}_F\right)$

95% confidence interval for $\mu_{W_M} - \mu_{W_F}$

$$\left(\overline{W}_M - \overline{W}_F\right) \pm 1.96 \cdot SE\left(\overline{W}_M - \overline{W}_F\right)$$

$$10996.04 \pm 1.96 \cdot 1240.709$$

$$\{8561.34\ ,\ 13430.73\}$$

Comparing means from different populations
Example: An ideal randomized experiment

In this course we will focus on estimating causal effects:

the expected effect on Y of a change in X

A causal effect can be measured by an **ideal randomized experiment**:

- Subjects are selected by simple random sampling from the population of interest
- Subjects are randomly assigned to a treatment or control group
- Treatment group receives treatment of interest ($X = 1$), control group receives no treatment ($X = 0$).
- The mean causal effect is the difference between the mean outcome when treated and the mean outcome when untreated

$$\textit{Mean causal effect} = \mu_{X=1} - \mu_{X=0}$$

Comparing means from different populations
Example: An ideal randomized experiment

If we want to know whether the treatment is effective we can test:

$$H_0 : \mu_{X=1} - \mu_{X=0} = 0 \qquad H_1 : \mu_{X=1} - \mu_{X=0} \neq 0$$

Step 1: Estimate $(\mu_{X=1} - \mu_{X=0})$ by computing the difference in mean outcomes of individuals in the treatment and control group:

$$\overline{Y}_{Treated} - \overline{Y}_{Control}$$

Step 2: Compute $SE\left(\overline{Y}_{Treated} - \overline{Y}_{Control}\right)$

Step 3: Compute $t^{act} = \frac{\left(\overline{Y}_{Treated} - \overline{Y}_{Control}\right) - 0}{SE\left(\overline{Y}_{Treated} - \overline{Y}_{Control}\right)}$

Step 4: Reject the null hypothesis of no treatment effect at a 5% significance level if $|t^{act}| > 1.96$

Using the t-statistic when n is small

- The test on the previous slide is based on the sample size n being large
- Especially in actual randomized experiments n can be small
- If the hypothesis test concerns 1 population mean, the t-statistic

$$t^{act} = \frac{\overline{Y} - \mu_{Y,0}}{SE\left(\overline{Y}\right)}$$

  - is *not* normally distributed for small n!
  - has the student-t distribution in the special case that the population distribution of Y is normal.

- If the hypothesis test concerns the difference in 2 population means, the t-statistic

$$t^{act} = \frac{\left(\overline{Y}_M - \overline{Y}_F\right) - d_0}{SE\left(\overline{Y}_M - \overline{Y}_F\right)}$$

  - is *not* normally distributed for small n!
  - does not have a student-t distribution even if the population distributions are normal!

Relationship between two random variables

- In general, questions in econometrics involve a relationship between 2 (or more) random variables:
  - What is the relation between education and earnings?
  - What is the relation between interest rates and economic growth?
  - What is the relation between the beer tax and traffic fatalities?
  - What is the relation between class size and student test scores?
- In this and coming lectures we will focus on the last of these questions.

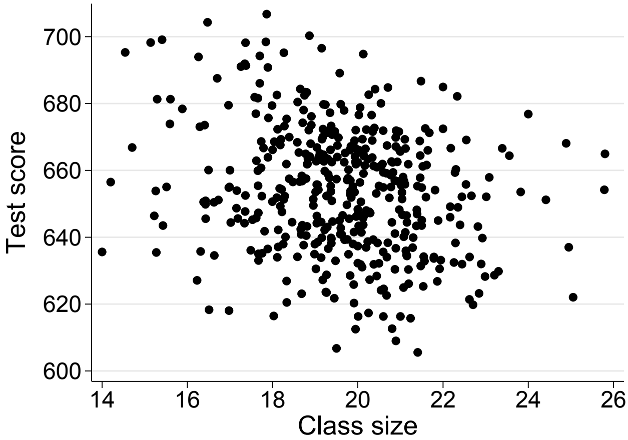
California test score data

- We will use a data set that contains data on test performance, school characteristics and student demographic backgrounds.
- The data are from 420 districts in California.
- Data were obtained from the California Department of Education
- Main variables of interest:
    - *TestScore* is the district average of the reading and math scores of 5th grade students
    - *ClassSize* is defined as the number of students divided by the number of full-time equivalent teachers in the district.

The relation between class size and test scores

- To examine the relation between class size and test scores we can make a scatter plot

A scatter plot is a plot of n observations on X_i and Y_i in which each observation is represented by the point (X_i, Y_i)



Sample covariance

- The covariance is a measure of the extend to which two random variables X and Y move together,

$$Cov(X,Y) = \sigma_{XY} \quad = \quad E\left[(X-\mu_X)\cdot(Y-\mu_Y)\right]$$

- The population covariance is unobserved but can be estimated by the **sample covariance** s_{XY}

$$s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)$$

- If $(X_i, Y_i)$are i.i.d and have finite fourth moments $E\left(X^4\right) < \infty$ & $E\left(Y^4\right) < \infty$

$$s_{XY} \xrightarrow{p} \sigma_{XY}$$

- The sample covariance between class size and test scores s_{CT} =-8.16

Sample correlation

- What does it mean for the sample covariance between test scores and class size to equal -8.16?
- The units of the covariance are the units of test scores multiplies by the units of class size
- The **sample correlation** r_{XY} measures the strength of the linear association between X and Y that is unit-free and lies between -1 and 1

$$r_{XY} = \frac{s_{XY}}{s_X s_Y}$$

- The sample correlation between class size and test scores $r_{CT} =$-0.23

Sample covariance and correlation in Stata

To compute the sample covariance in Stata:

```
. corr test_score class_size, covariance
(obs=420)

             | test_s~e class_~e
-------------+------------------
  test_score |   363.03
  class_size | -8.15932  3.57895
```

To compute the sample correlation in Stata:

```
. corr test_score class_size
(obs=420)

             | test_s~e class_~e
-------------+------------------
  test_score |   1.0000
  class_size |  -0.2264   1.0000
```

Linear regression with one regressor

Linear regression with one regressor

Suppose we would like to answer the following question:

What is the effect on district test scores if we would increase district average class size by 1 student?

We would like to know

$$\beta_{ClassSize} = \frac{\triangle Test\ score}{\triangle Class\ size}$$

$\beta_{ClassSize}$ is the definition of the slope of a straight line relating test scores and class size

$$Test\ score = \beta_0 + \beta_{ClassSize} \times Class\ size$$

where β_0 is the intercept of the straight line.

Linear regression with one regressor

- The average test score in district *i* does not only depend on the average class size
- It also depends on factors such as
  - Quality of the teachers
  - Student background
  - quality of text books
  - .....
- The equation describing the linear relation between *Test score* and *Class size* is better written as

$$Test\ score_i = \beta_0 + \beta_{ClassSize} \times Class\ size_i + u_i$$

where u_i lumps together all other district characteristics that affect average test scores.

Terminology for the Linear Regression Model with One Regressor

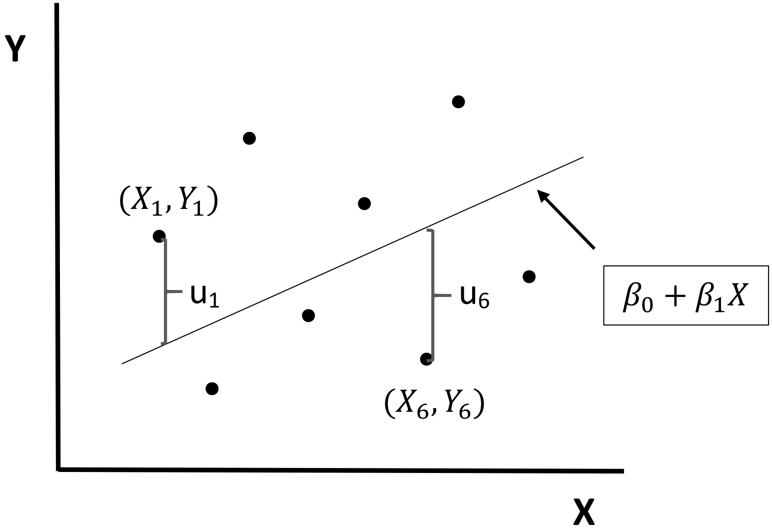
The linear regression model with one regressor is denoted by

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

where

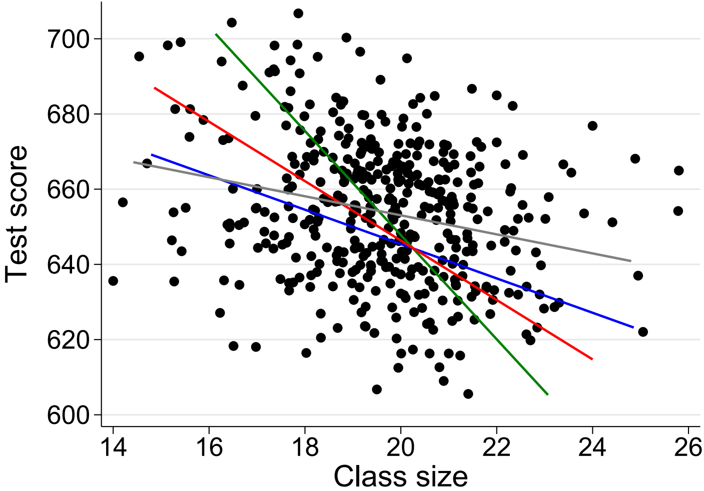
- Y_i is the dependent variable
- X_i is the independent variable or regressor
- $\beta_0 + \beta_1 X_i$ is the population regression line
- β_0 is the intercept of the population regression line (expected value of Y when $X = 0$)
- β_1 is the slope of the population regression line
- u_i is the error term (all other factors determining Y_i)

Linear regression with one regressor



Linear regression with one regressor

- In general we don't know β_0 and β_1 and we have to estimate them using a random sample of data.
- How to find the line that fits the data best?



The Ordinary Least Squares Estimator (OLS)

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 the squared mistakes made in predicting Y given X

- Let b_0 and b_1 be estimators of β_0 and β_1
- The predicted value of Y_i given X_i using these estimators is $b_0 + b_1 X_i$
- The prediction mistake is

$$Y_i - (b_0 + b_1 X_i) = Y_i - b_0 - b_1 X_i$$

- The estimators of the slope and intercept that minimize

$$\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

are called the ordinary least squares (OLS) estimators of β_0 and β_1

$\overline{Y}$ is the ordinary least squares estimator of μ_Y

- Suppose there is no X only Y

$$Y_i = \mu_Y + u_i$$

- Let m be an estimator of μ_Y
- The least squares estimator minimizes

$$\sum_{i=1}^{n} (Y_i - m)^2$$

- Taking the derivative w.r.t m and setting it to zero gives

$$\begin{aligned}
\frac{\partial}{\partial m} \sum_{i=1}^{n} (Y_i - m)^2 \quad = \quad & -2 \sum_{i=1}^{n} (Y_i - m) && = 0 \\
& -2 \sum_{i=1}^{n} Y_i + 2 \cdot n \cdot m && = 0 \\
& \frac{1}{n} \sum_{i=1}^{n} Y_i - m && = 0
\end{aligned}$$

- Solving for m gives

$$m = \frac{1}{n} \sum_{i=1}^{n} Y_i = \overline{Y}$$

The Simple Linear Regression Model

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

- OLS minimizes sum of squared prediction mistakes:

$$\sum_{i=1}^{n} \widehat{u}_i^2 = \sum_{i=1}^{n} \left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i \right)^2$$

- Step 1:

$$\frac{\partial}{\partial \widehat{\beta}_0} \sum_{i=1}^{n} \left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i \right)^2 = 0$$

- Step 2:

$$\frac{\partial}{\partial \widehat{\beta}_1} \sum_{i=1}^{n} \left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i \right)^2 = 0$$

Step 1: OLS estimator of β_0

$$
\begin{aligned}
\frac{\partial}{\partial \widehat{\beta}_0} \sum_{i=1}^n u_i^2 = \quad & -2\sum_{i=1}^n \left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i\right) & = 0 \\
= \quad & \frac{1}{n}\left(\sum_{i=1}^n Y_i - \sum_{i=1}^n \widehat{\beta}_0 - \sum_{i=1}^n \widehat{\beta}_1 X_i\right) & = 0 \\
= \quad & \frac{1}{n}\sum_{i=1}^n Y_i - \frac{1}{n} n\widehat{\beta}_0 - \widehat{\beta}_1 \frac{1}{n}\sum_{i=1}^n X_i & = 0 \\
= \quad & \overline{Y_i} - \widehat{\beta}_0 - \widehat{\beta}_1 \overline{X_i} & = 0
\end{aligned}
$$

- This gives

$$
\boxed{\widehat{\beta_0} = \overline{Y} - \widehat{\beta}_1 \overline{X}}
$$

Step 2: OLS estimator of β_1

$$\frac{\partial}{\partial \widehat{\beta}_1} \sum_{i=1}^n u_i^2 = -2 \cdot \sum_{i=1}^n -X_i \left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i\right) = 0$$

Devide by -2 and substitute for $\widehat{\beta}_0$:

$$= \sum_{i=1}^n X_i \left(Y_i - \left(\overline{Y} - \widehat{\beta}_1 \overline{X}\right) - \widehat{\beta}_1 X_i\right) = 0$$

rewrite

$$\sum_{i=1}^n X_i \left(\left(Y_i - \overline{Y}\right) - \left(\widehat{\beta}_1 X_i - \widehat{\beta}_1 \overline{X}\right)\right)$$

rewrite

$$= \sum_{i=1}^n X_i \left(Y_i - \overline{Y}\right) - \widehat{\beta}_1 \sum_{i=1}^n X_i \left(X_i - \overline{X}\right) = 0$$

Algebra trick

$$= \sum_{i=1}^n \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) - \widehat{\beta}_1 \sum_{i=1}^n \left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right) = 0$$

Step 2: OLS estimator of β_1

Algebra trick:

$$\begin{aligned}
\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right) &= \sum_{i=1}^{n} X_iY_i - \sum_{i=1}^{n} X_i\overline{Y} - \sum_{i=1}^{n} \overline{X}Y_i + \sum_{i=1}^{n} \overline{XY} \\
&= \sum_{i=1}^{n} X_iY_i - \sum_{i=1}^{n} X_i\overline{Y} - n\overline{X}\left(\tfrac{1}{n}\sum_{i=1}^{n} Y_i\right) + n\overline{XY} \\
&= \sum_{i=1}^{n} X_iY_i - \sum_{i=1}^{n} X_i\overline{Y} - n\overline{XY} + \overline{nXY} \\
&= \sum_{i=1}^{n} X_iY_i - \sum_{i=1}^{n} X_i\overline{Y} \\
&= \sum_{i=1}^{n} X_i\left(Y_i-\overline{Y}\right)
\end{aligned}$$

By a similar reasoning:

$$\sum_{i=1}^{n} X_i\left(X_i-\overline{X}\right) = \sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(X_i-\overline{X}\right).$$

Step 2: OLS estimator of β_1

$$\frac{\partial}{\partial \widehat{\beta}_1} \sum_{i=1}^n u_i^2 = \sum_{i=1}^n \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) - \widehat{\beta}_1 \sum_{i=1}^n \left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right) = 0$$

Solving for $\widehat{\beta}_1$ gives the OLS estimator:

$$\boxed{\widehat{\beta_1} = \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^n (X_i - X)(X_i - X)} = \frac{\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right)} = \frac{s_{xy}}{s_x^2}}$$

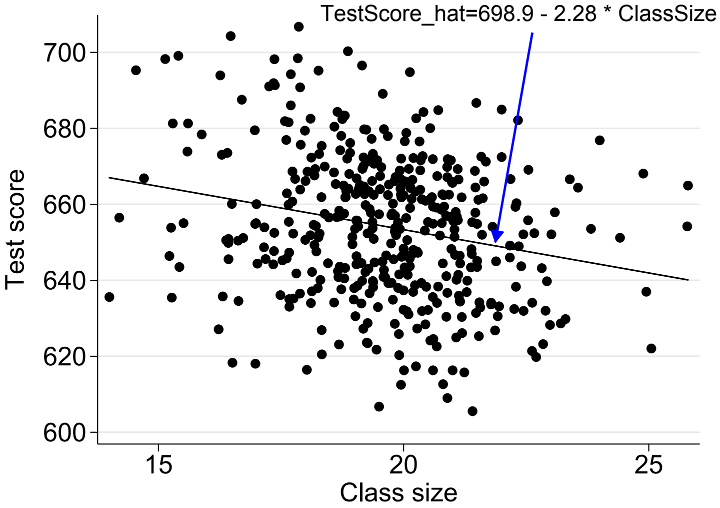
The OLS predicted values $\widehat{Y}_i$ and residuals $\widehat{u}_i$ are:

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i$$

$$\widehat{u}_i = Y_i - \widehat{Y}_i$$

The Simple Linear Regression Model

Example: Class size and test scores



The Simple Linear Regression Model

Example: Class size and test scores

$$TestScore_i = \beta_0 + \beta_1 ClassSize_i + u_i$$

```
. regress test_score class_size, robust

Linear regression                               Number of obs     =        420
                                                  F(1, 418)         =      19.26
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.0512
                                                  Root MSE          =     18.581
```

test_score	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
class_size	-2.279808	.5194892	-4.39	0.000	-3.300945	-1.258671
_cons	698.933	10.36436	67.44	0.000	678.5602	719.3057

- $\widehat{\beta}_1 = -2.27$ A reduction in class size by 1 student is associated with an increase in test scores by 2.27 points
- $\widehat{\beta}_0 = 698.93$ The expected test score when class size is zero equals 698.93 (what does it mean for class size to be zero)?

$\overline{Y}$ is the ordinary least squares estimator of μ_Y

Example: test scores

The sample mean of district average test scores $\overline{TestScore} = 654.16$

```
. mean test_score

Mean estimation                   Number of obs   =        420
```

	Mean	Std. Err.	[95% Conf.	Interval]
test_score	**654.1565**	**.9297082**	**652.3291**	**655.984**

As shown on slide 24 we can also obtain the sample mean by OLS

```
. regress test_score
```

Source	SS	df	MS
Model	**0**	**0**	**.**
Residual	**152109.594**	**419**	**363.030056**
Total	**152109.594**	**419**	**363.030056**

Number of obs	=	**420**
F(0, 419)	=	**0.00**
Prob > F	=	**.**
R-squared	=	**0.0000**
Adj R-squared	=	**0.0000**
Root MSE	=	**19.053**

test_score	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_cons	**654.1565**	**.9297082**	**703.61**	**0.000**	**652.3291**	**655.984**

Measures of fit

How well does the estimated regression line describe the data?

- Does the regressor X account for much or for little variation in Y ?
- Are the observations in the scatter plot clustered closely around the regression line?

Two measures of how well the OLS line fits the data.

The R^2 measures the fraction of the variation in Y_i explained/predicted by X_i

The standard error of the regression SER measures how far Y_i typically is from its predicted value

The R^2

R^2 is the fraction of the sample variance of Y_i explained/predicted by X_i

We can write

$$Y_i = \widehat{Y}_i + \widehat{u}_i$$

which implies that the R^2 is the ratio of the sample variance of $\widehat{Y}_i$ and the sample variance of Y_i

$$R^2 = \frac{\textit{Explained sum of squares}}{\textit{Total sum of squares}} = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}\left(\widehat{Y}_i - \overline{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}$$

The R^2 ranges from 0 to 1

- If $R^2 = 0$, X_i explains no none of the variation in Y_i
- If $R^2 = 1$, X_i explains all of the variation in Y_i ($Y_i = \widehat{Y}_i$)
- in practice $0 < R^2 < 1$

The R^2

The total sum of squares *TSS* can be divided in the explained sum of squares *ESS* and the residual sum of squares *SSR*:

$$
\begin{aligned}
TSS &= ESS + SSR \\
\textstyle\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2 &= \textstyle\sum_{i=1}^{n} \left(\widehat{Y}_i - \overline{Y}\right)^2 + \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2 \\
\textstyle\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2 &= \textstyle\sum_{i=1}^{n} \left(\widehat{Y}_i - \overline{Y}\right)^2 + \sum_{i=1}^{n} \widehat{u}_i^2
\end{aligned}
$$

This implies that the R^2 can also be written as

$$
R^2 = \frac{ESS}{TSS} = \frac{TSS - SSR}{TSS} = 1 - \frac{SSR}{TSS} = \frac{\sum_{i=1}^{n} \widehat{u}_i^2}{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2}
$$

The R^2

Example: Class size and test scores

```
. regress test_score class_size, robust

Linear regression                               Number of obs     =        420
                                                  F(1, 418)          =      19.26
                                                  Prob > F           =     0.0000
                                                  R-squared          =     0.0512
                                                  Root MSE           =     18.581
```

test_score	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
class_size	-2.279808	.5194892	-4.39	0.000	-3.300945	-1.258671
_cons	698.933	10.36436	67.44	0.000	678.5602	719.3057

$R^2 = 0.0512$

Note: the R^2 is uninformative about whether an increase in class size *causes* a reduction in test scores!

The standard error of the regression

- Another measures of fit is the SER.

The standard error of the regression (SER) is an estimator of the standard deviation of the regression error u_i

$$SER = s_{\hat{u}} = \sqrt{s^2_{\hat{u}}} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2}$$

It measures the spread of the observations around the regression line in the units of the dependent variable

- The divisor *n-2* is used because 2 degrees of freedom were lost in estimating the two regression coefficients β_0 and β_1.

The standard error of the regression

Example: Class size and test scores

```
. regress test_score class_size, robust

Linear regression                               Number of obs     =        420
                                                  F(1, 418)         =      19.26
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.0512
                                                  Root MSE          =     18.581
```

test_score	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
class_size	-2.279808	.5194892	-4.39	0.000	-3.300945	-1.258671
_cons	698.933	10.36436	67.44	0.000	678.5602	719.3057

In Stata the SER is denoted as Root MSE.

$SER = 18.6$