

ECON3150/4150 Spring 2015

Lecture 5 and 6, February 2

Siv-Elisabeth Skjelbred

University of Oslo

Last updated: February 2, 2015

Outline

- Mathematics
- Proof that $\hat{\beta}_1$ is unbiased and consistent
- Illustrate unbiasedness and consistency through simulation
- Hypothesis testing and confidence intervals

Basic mathematical tools

$$\sum_{i=1}^n c = nc \text{ for any constant } c$$

$$\sum_{i=1}^n cX_i = c \sum_{i=1}^n X_i \text{ for any constant } c$$

$$\sum_{i=1}^n (X_i + Y_i) = \sum_{i=1}^n X_i + \sum_{i=1}^n Y_i$$

Note also something that is NOT possible:

$$\sum_{i=1}^n X_i Y_i \neq \sum_{i=1}^n X_i \sum_{i=1}^n Y_i$$

Mathematics of OLS components

For a given set of data observations $X_i (i = 1, 2, \dots, n)$ and $y_i (i = 1, 2, \dots, n)$ it can be shown that:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0 \quad (\text{I})$$

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i (X_i - \bar{X}) \quad (\text{II})$$

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) &= \sum_{i=1}^n (X_i - \bar{X})Y_i - n\bar{X}\bar{Y} \\ &= \sum_{i=1}^n (Y_i - \bar{Y})X_i \\ &= \sum_{i=1}^n (X_i - \bar{X})Y_i \end{aligned} \quad (\text{III})$$

Sample covariance

Sample variance of X:

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Sample covariance:

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

Sample correlation and $\hat{\beta}_1$

The sample correlation coefficient is given by:

$$r_{X,Y} = \frac{s_{XY}}{s_X s_Y}$$

In the simple linear regression model:

$$\hat{\beta}_1 = r_{XY} \frac{s_y}{s_x}$$

- s_x and s_y are the sample standard deviations of X and Y.

Mathematics of TSS

$$TSS = SSR + ESS$$

Because:

$$2 \sum_{i=1}^n \hat{u}_i (\hat{Y}_i - \bar{Y}) = 0$$

Proof on blackboard.

Mathematical properties of OLS

The OLS residuals are defined as: $\hat{u}_i = Y_i - \hat{Y}_i, i = 1, 2, \dots, n$, it can then be shown that:

$$\bar{\hat{u}} = \frac{1}{n} \sum_{i=1}^n \hat{u}_i = 0$$

which comes from the property that:

$$\sum_{i=1}^n \hat{u}_i = \sum_{i=1}^n (Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_i) = 0$$

given by the first order condition for OLS.

Mathematical properties of OLS

$$\frac{1}{n} \sum_{i=1}^n (\hat{u}_i - \bar{\bar{u}})(X_i - \bar{X}) = 0$$

Similarly it can be shown that:

$$\frac{1}{n} \sum_{i=1}^n (\hat{u}_i - \bar{\bar{u}})(\hat{Y}_i - \bar{Y}) = 0$$

Alternative specification OLS

The OLS minimization problem:

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$$

can alternatively be written as:

$$S(\alpha, \beta_1) = \sum_{i=1}^n (Y_i - \alpha - \beta_1 (X_i - \bar{X}))^2$$

where the intercept parameter is redefined to: $\alpha = \beta_0 + \beta_1 \bar{X}$

Alternative specification OLS

$$\frac{\partial S(\alpha, \beta_1)}{\partial \alpha} = -2 \sum_{i=1}^n [Y_i - \alpha - \beta_1(X_i - \bar{X})]$$

$$\frac{\partial S(\alpha, \beta_1)}{\partial \beta_1} = -2 \sum_{i=1}^n [Y_i - \alpha - \beta_1(X_i - \bar{X})] * (X_i - \bar{X})$$

$\hat{\alpha}$ and $\hat{\beta}_1$ are the values of α and β_1 for which the FOC is equal to zero.
Solution:

$$\hat{\alpha} = \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$$

and β_1 as before.

Unbiased estimators

Show that:

$$E(\hat{\beta}_1) = \beta_1$$

Unbiased estimators

Show that:

$$E(\hat{\beta}_1) = \beta_1$$

Similarly it can be shown that

$$E(\hat{\beta}_0) = \beta_0$$

Calculate yourself and control with solution to exercise 4.7.

Consistency of $\hat{\beta}_1$

Consistency

A variable is consistent if the spread around the true parameter approaches zero as n increases.

The spread is measured by the variance:

$$\begin{aligned} \text{var}(\hat{\beta}_1) &= \text{var} \left(\frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \right) \\ &= \frac{\text{Var}(Y_i)}{ns_X^2} \end{aligned}$$

- Consistent as the larger n the smaller the variance
- The larger the variance of X the lower the variance of $\hat{\beta}_1$
- Lower variance of Y lower variance of $\hat{\beta}_1$

- We have shown that the estimators are unbiased, but are they most efficient?
- The linearity excludes non-linear functions of Y_i
- Since OLS minimizes the spread around the regression line, it has the lowest variance of the linear estimators.

Distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$

Normality assumption

The population error u is independent of the explanatory variables and is normally distributed with zero mean and variance σ^2 : $u \sim \text{Normal}(0, \sigma^2)$

Distribution of estimators

- Under the OLS assumptions including the normality assumption, sampling distribution of the OLS estimators is normal.
- $\hat{\beta}_1 \sim \text{Normal}[\beta_1, \text{Var}(\hat{\beta}_1)]$
- Thus $(\hat{\beta}_1 - \beta_1)/\text{sd}(\hat{\beta}_1) \sim \text{Normal}(0, 1)$
- : This comes from:
 - A random variable which is a linear function of a normally distributed variable is itself normally distributed.
 - If we assume that $u \sim N(0, \sigma^2)$ then Y_i is normally distributed.
 - Since the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ is linear functions of the Y_i 's then the estimators are normally distributed.

Normality assumption

- Whenever y takes on just a few values it cannot have anything close to a normal distribution.
- The exact normality of OLS depends on the normality of the error.
- If the $\hat{\beta}_1$ is not normally distributed the t-statistic does not have t distribution.
- The normal distribution of u is the same as the distribution of Y given X .
- In large samples we can invoke the CLT to conclude that the OLS satisfy asymptotic normality.

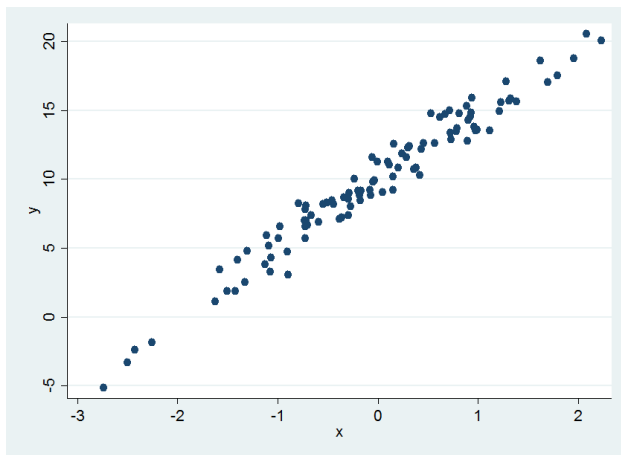
Simulation

A simulation is a fictitious computer representation of reality.

- 1 Choose the sample size n
- 2 Choose the parameter values and functional form of the population regression function.
- 3 Generate n values of x randomly in Stata
- 4 Choose probability distribution of the error term and generate n values of u
- 5 Estimate the model
- 6 Repeat step 1 through 5 multiple times and look at the summary statistics over the repetitions.

Simulation

A random realization of X and u for 100 observations with the true population function: $Y = 10 + 5x + u$.



How does OLS perform in estimating the underlying population function?

Simulation

```
1 . reg y x
```

Source	SS	df	MS
Model	2438.45884	1	2438.45884
Residual	100.53965	98	1.02591479
Total	2538.99849	99	25.6464494

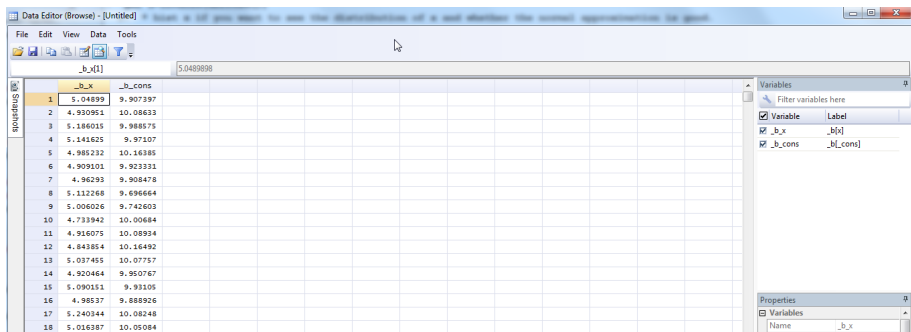
Number of obs = 100
F(1, 98) = 2376.86
Prob > F = 0.0000
R-squared = 0.9604
Adj R-squared = 0.9600
Root MSE = 1.0129

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	4.86004	.0996868	48.75	0.000	4.662214	5.057865
_cons	9.951091	.1013099	98.22	0.000	9.750045	10.15214

The coefficients are close to the true population coefficients.

Simulation

So running one simulation got us close to the estimate, how if we simulate 1000 times? Gives us 1000 estimates for β_0 and β_1



The screenshot shows the SPSS Data Editor window with a dataset containing 1000 rows of simulation results. The first two columns are labeled `_b_x` and `_b_cons`. The first row is highlighted in orange. The right sidebar shows the 'Variables' list with `_b_x` and `_b_cons` selected.

	<code>_b_x</code>	<code>_b_cons</code>
1	5.04899	9.907397
2	4.930951	10.08633
3	5.186015	9.988575
4	5.141625	9.97107
5	4.985232	10.16385
6	4.909101	9.923331
7	4.96293	9.908478
8	5.112268	9.696664
9	5.006026	9.742603
10	4.733942	10.00684
11	4.916075	10.08934
12	4.843854	10.16492
13	5.037455	10.07757
14	4.920464	9.950767
15	5.090151	9.93105
16	4.98537	9.888926
17	5.240344	10.08248
18	5.016387	10.05084

Simulation

```
1 . sum
```

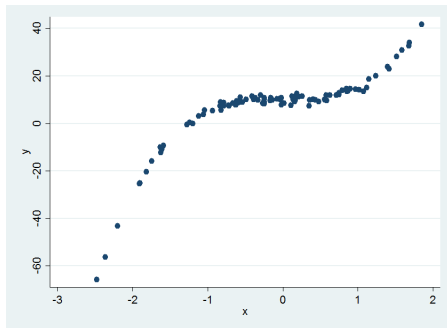
Variable	Obs	Mean	Std. Dev.	Min	Max
_b_x	1000	5.000788	.1048841	4.679994	5.318002
_b_cons	1000	9.997947	.0994027	9.696664	10.33181

The estimated OLS coefficients approximate to the true population coefficient. Thus OLS gives an unbiased estimate for the slope coefficient and the constant term.

Simulation

Specifying wrong functional form:

- The true functional form is $Y = \beta_0 + \beta_1 X^3 + u$
- But we run the regression $Y = \beta_0 + \beta_1 X + u$



2 . sum

Variable	Obs	Mean	Std. Dev.	Min	Max
_b_x	1000	14.80622	3.053321	7.893195	28.17236
_b_cons	1000	10.05226	1.22186	5.38526	14.49691

Simulation

Assumption violation: The expected value of the error term is not zero, but 3? So $u \sim N(3, 1)$

Variable	Obs	Mean	Std. Dev.	Min	Max
_b_x	1000	4.998194	.0975124	4.67178	5.298087
_b_cons	1000	13.0041	.1030487	12.65324	13.33964

- As long as X and u are uncorrelated $\hat{\beta}_1$ is unbiased.
- The constant term and the error term is correlated in this situation so β_0 is biased.

Simulation and variance

10 observations repeated on 1000 samples.

```
2 . sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
_b_x	1000	4.997962	.3770747	3.270008	6.512727
_b_cons	1000	10.01276	.3351356	8.923193	11.20773

100 observations repeated on 1000 samples.

```
1 . sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
_b_x	1000	5.000788	.1048841	4.679994	5.318002
_b_cons	1000	9.997947	.0994027	9.696664	10.33181

Hypothesis testing of the regression coefficients

Distribution of $\hat{\beta}_1$

- Given that $\hat{\beta}_1$ is either normally distributed or approximately normally distributed the t statistic for $\hat{\beta}_1$ is t-distributed.
- Thus: $(\hat{\beta}_1 - \beta_1)/se(\hat{\beta}_1) \sim t_{n-2}$
- As the degrees of freedom in the t-distribution gets large, the t distribution approaches the standard normal distribution.

T-statistic

In general the t-statistics has the form:

$$t = \frac{\text{estimator} - \text{hypothesised value}}{\text{standard error of the estimator}}$$

- Since the standard error is always positive the t-statistic has the same sign as the difference between the estimator and the hypothesized value.
- For a given standard error the larger value of the estimator the larger value of the t-statistic.
- If the null hypothesis is that the true parameter is zero, a large estimator provides evidence against the null.
- T values sufficiently far from the hypothesized value result in rejection of the null.

Repetition hypothesis testing

- 1 Formulate the null and alternative hypothesis
- 2 Compute the standard error of the variable of interest
- 3 Compute the t-statistics
- 4 Find the rejection rule
- 5 Make your conclusion

Testing hypotheses about β_1

- 1 Formulate the null and alternative hypothesis

$$H_0 : \beta_1 = \beta_{1,0} \text{ vs. } H_1 : \beta_1 \neq \beta_{1,0} \text{ or: } H_1 : \beta_1 < \beta_{1,0} \text{ or: } H_1 : \beta_1 > \beta_{1,0}$$

- 2 Compute the standard error:

$$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}_{\beta_1}^2}$$

- 3 Compute the t-statistics:

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

Step 4: Determine rejection rule

- The rejection rule depends on your desired significance level, i.e the probability of rejecting H_0 when it is in fact true.
- A 5% significance level means that we mistakenly reject H_0 5% of the time.
- Given the significance level we can find the critical value of t .
- The critical value increases as significance level falls, thus a null hypothesis that is rejected at a 5% level is automatically rejected at the 10% level as well.

Rejection rules t-statistic

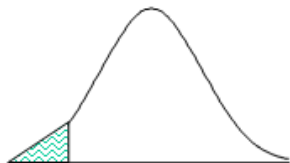
Compare the critical value to the t-statistic actually calculated:

- One sided: $H_1 : \beta_1 > 0 : t^{act} > t^c$
- One sided: $H_1 : \beta_1 < 0 : t^{act} < -t^c$
- Two sided: $H_1 : \beta_1 \neq 0 : |t^{act}| > t^c$

One-sided vs two-sided test



Positive one-tailed test

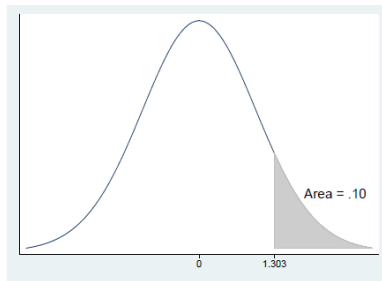
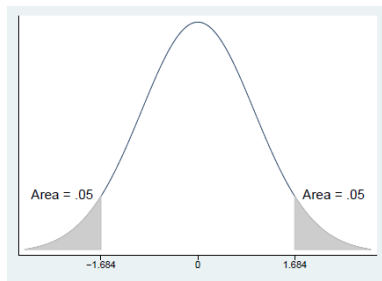


Negative one-tailed test



Two-tailed test

One-sided vs two-sided test



An illustration of the difference between two-sided and one sided test with 40 degrees of freedom.

Step 5: Make conclusion

- When H_0 is rejected at the 5% level we usually say that X is statistically significant, or statistically different from zero, at the 5% level.
- If H_0 is not rejected we say that X is statistically insignificant at the 5% level.
- If we fail to reject H_0 we never say that we accept H_0 because there are many other values for β_1 which cannot be rejected and they cannot all be true.

Computing p-values for t-tests

- An alternative to comparing the t-statistic to the critical t-statistic is to compute the p-value for the t-statistic.
- With p-value you do not have to commit to the significance level ahead of time.
- The p-value is more informative as it gives you the smallest significance level at which the null hypothesis would have been rejected.
- A null that is rejected at a 5% level thus must have a p-value smaller than 5%.

Computing p-values for t-tests

- The p-value (in SLRM) is calculated by computing the probability that a t random variable with $(n-2)$ degrees of freedom is larger than t^{act} in absolute value.
- Thus the p-value is the significance level of the test when we use the value of the test statistic as the critical value for the test.

For the two sided test:

$$\begin{aligned}\text{p-value} &= Pr_{H_0}(|t| > |t^{act}|) = 2P(t > t^{act}) \\ &= Pr(|Z| > |t^{act}|) = 2\phi(-|t^{act}|) \text{ in large samples}\end{aligned}$$

Computing p-values for t-tests

- One sided test: $H_1 : \beta_1 > 0$
- If $\hat{\beta}_1 < 0$ we know that the p-value is greater than 0.5 and there is no need to calculate it.
- If $\hat{\beta}_1 > 0$ then $t > 0$ and the p-value is half of the two-sided p-value.
- Since the t-distribution is symmetric around zero the reversed applied to the one sided test that $\beta_1 < 0$

$$p - value = Pr_{H_0}(t < t^{act}) = Pr_{H_0}(t > |t^{act}|)$$

Interpreting p-values

- The p-value is the probability of observing a t statistic as extreme as we did if the null hypothesis is true.
- Small p-values are evidence against the null, large p-values provide little evidence against the null.
- If for example the p-value = 0.5 then we would observe a value of the t statistic as extreme as we did in 50% of all random samples when the null hypothesis is true.
- If α denotes the significance level then H_0 is rejected if $\text{p-value} < \alpha$.

Interpreting p-values

p-value

Correct interpretation: Assuming that the null is true you would obtain the observed difference or more in $p\%$ of studies due to random sampling error.

Wrong interpretation: P-value is the probability of making a mistake by rejecting a true null hypothesis.

- The p-value is calculated based on the assumption that the null is true for the population, thus it cannot tell you the probability that the null is true or false.
- A low p-value indicates that your data are unlikely assuming a true null, but it cannot evaluate whether it is more likely that the low p-value comes from the null being true but having an unlikely sample or the null being false.

An example

The MEAP93 data contains observations on 408 school districts on average teacher salary in thousands of dollars (*sal*) and the percentage of students passing the MEAP math. A regression gives the following output:

$$\widehat{math10} = \underset{(3.22)}{8.28} + \underset{(0.10)}{0.498}sal$$

- The constant term is 8.28 with a standard error of 3.22
- The slope parameter is 0.498 with a standard error of 0.1

An example

$$\hat{math10} = \underset{(3.22)}{8.28} + \underset{(0.10)}{0.498}sal$$

$$t = \frac{0.498 - 0}{0.10} = 4.98$$

$$p - value = 2\phi(-4.98) < 0.00001$$

- The 5% critical value is given by: $t_{406}^c = 1.96$
- We can reject the null that salary does not affect the percentage of students passing the math10.

An example

The stata output for the same regression shows the same conclusion.

```
1 . reg math10 sal
```

Source	SS	df	MS
Model	2562.57022	1	2562.57022
Residual	42254.6103	406	104.075395
Total	44817.1805	407	110.115923

Number of obs = 408
F(1, 406) = 24.62
Prob > F = 0.0000
R-squared = 0.0572
Adj R-squared = 0.0549
Root MSE = 10.202

math10	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sal	.4980309	.1003674	4.96	0.000	.3007264 .6953355
_cons	8.282175	3.228869	2.57	0.011	1.934787 14.62956

Economic versus statistical significance

- The statistical significance of a variable is determined entirely by the size of the computed t-statistic.
- A coefficient can be statistically significant either because the coefficient is large, or because the standard error is small.
- With large samples parameters can be estimated very precisely which usually results in statistical significance.
- The economic significance is related to the size (and sign) of $\hat{\beta}_1$.
- Thus you should also discuss whether the coefficient is economically important (i.e. the magnitude of the coefficient)

Confidence intervals

Confidence interval

Confidence interval

A confidence interval is a rule used to construct a random interval so that a certain percentage of all data sets, determined by the confidence level, yields an interval that contains the population value.

Confidence level

The percentage of samples in which we want our confidence interval to contain the population value.

Confidence interval

Two equivalent definitions of confidence interval:

- A 95% CI is the set of values that cannot be rejected using a two-sided hypothesis test with a 5% significance level.
- An interval that has a 95% probability of containing the true value of β_1 . In 95% of possible samples that can be drawn the confidence interval will contain the true value of β_1 .

Confidence interval of $\hat{\beta}_1$

- The t-statistic will reject the hypothesized value , $\beta_{0,1}$, (at a 5% level) whenever it is outside the range:

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

Example

$$\hat{math10} = 8.28 + 0.498salary$$

$(3.22) \qquad (0.10)$

Confidence interval for β_{salary}

$$0.498 - 1.96 * 0.10 = 0.302$$

$$0.498 + 1.96 * 0.10 = 0.694$$

```
1 . reg math10 sal
```

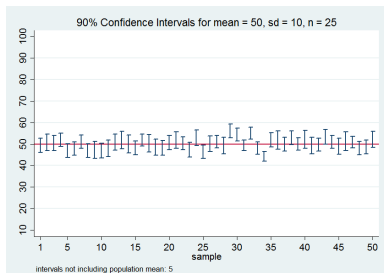
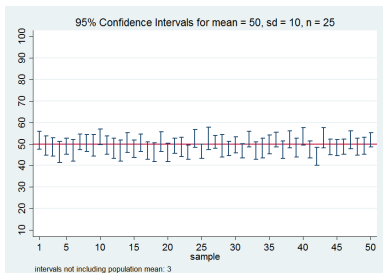
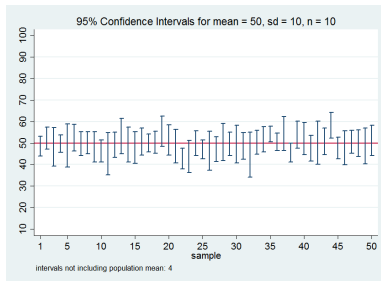
Source	SS	df	MS
Model	2562.57022	1	2562.57022
Residual	42254.6103	406	104.075395
Total	44817.1805	407	110.115923

Number of obs = 408
F(1, 406) = 24.62
Prob > F = 0.0000
R-squared = 0.0572
Adj R-squared = 0.0549
Root MSE = 10.202

math10	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sal	.4980309	.1003674	4.96	0.000	.3007264	.6953355
_cons	8.282175	3.228869	2.57	0.011	1.934787	14.62956

Illustration CI intervals

Simulated confidence intervals for 50 samples each with a mean of 50 and standard deviation of 10.



Confidence interval

- The confidence interval can be used to construct a confidence interval for the predicted effect of a general change in X .
- The 95% confidence interval for $\beta_1 \Delta X$

$$[(\hat{\beta}_1 - 1.96\hat{\beta}_1)\Delta X, (\hat{\beta}_1 + 1.96\hat{\beta}_1)\Delta X]$$

When X is binary

Regression when X is a binary variable

- A lot of information relevant for econometric analysis is qualitative.
- This information can be summarized with one or multiple binary variables.
- In econometrics binary variables are typically called dummy variables.
- In defining a dummy variable we must decide which event is assigned the value one and which is assigned the value 0.
- The name typically indicates the event with value one.
 - Female (1=female, 0=male)
 - Higher_educ (1=college or more, 0=less than college)
 - Public_transport (1=use public transport to work, 0=do not use public transport)
 - Drug (1=received the drug, 0= received placebo)

Regression when X is a binary variable

The population regression model with the binary variable D_i ($D=1$ if female, $D=0$ if male) is:

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

when i is a male ($D=0$) we get:

$$Y_i = \beta_0 + u_i \rightarrow E(Y_i | D = 0) = \beta_0$$

while if i is a female ($D=1$) we get:

$$Y_i = \beta_0 + \beta_1 + u_i \rightarrow E(Y_i | D = 1) = \beta_0 + \beta_1$$

Thus $\beta_1 = E(Y_i | Female) - E(Y_i | male)$

Dummy variables

- The group with an indicator of 1 is the base group, the group against which comparisons are made.
- It does not matter how we choose the base group, but it is important to keep track of who is the base group.
- If two population means are the same then β_1 is zero.

Example

Data from additional E4.1

- Data from on average hourly earnings from a sample of full-time workers.
- Female = 1 the person is female, female = 0 the person is male.

```
1 . reg ahe female
```

Source	SS	df	MS	Number of obs = 7711		
Model	13091.0876	1	13091.0876	F(1, 7709) = 129.46		
Residual	779560.368	7709	101.12341	Prob > F = 0.0000		
Total	792651.456	7710	102.80823	R-squared = 0.0165		
				Adj R-squared = 0.0164		
				Root MSE = 10.056		

ahe	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
female	-2.629912	.2311422	-11.38	0.000	-3.083013	-2.17681
_cons	20.11387	.1520326	132.30	0.000	19.81584	20.41189

- The computed t statistic is higher than the 5% critical value (1.96)
- The p-value is lower than 0.05

β_1 and is statistically significant at a 5% level and since it is negative it indicates that women earn less than men.

Proportions and percentages as dependent variables

- The proportional change is the change in a variable relative to its initial value, mathematically, the change divided by the initial value.
- The percentage change is the proportionate change in a variable, multiplied by 100.
- The percentage point change is the difference between two percentages.

Proportions and percentages as dependent variables

In a dataset on CEO's where y is annual salary in thousands of dollars and X is the average return on equity (roe) the following OLS regression line can be obtained:

$$\text{salary} = \beta_0 + \beta_1 \text{roe} + u$$

- ROE is defined in terms of net income as a percentage of common equity, thus if $\text{roe}=10$, the average return on equity is 10%.
- The slope parameter β_1 measures the change in annual salary, in thousands of dollars, when return on equity increase by one percentage point.

Homoskedasticity

The dummy variable example can shed light on what is meant by homoskedasticity:

- The definition of homoskedasticity requires the error term to be independent of X , i.e it must not depend on female in our example.
- For women the error term (u_i) is the deviation of the i^{th} woman's earning from the population mean earnings for women.
- For men the error term (u_i) is the deviation of the i^{th} man's earning from the population mean earnings for men.
- Thus the variance of earnings must be the same for men as it is for women.

Homoskedasticity

Is the assumption realistic?

- Highly paid women are more rare than highly paid men suggesting that the distribution of earnings among women is tighter than among men.
- It is plausible that the variance of the error term for women is less than the one for men.
- Stata makes it easy to control for heteroskedasticity and nothing is lost by using the heteroskedasticity robust standard errors, thus always using the robust ones is the best thing.

Homoskedasticity

Homoskedasticity assumption:

```
1 . reg ahe female
```

Source	SS	df	MS	Number of obs = 7711		
Model	13091.0876	1	13091.0876	F(1, 7709) = 129.46		
Residual	779560.368	7709	101.12341	Prob > F = 0.0000		
				R-squared = 0.0165		
				Adj R-squared = 0.0164		
				Root MSE = 10.056		
Total	792651.456	7710	102.80823			

ahe	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
female	-2.629912	.2311422	-11.38	0.000	-3.083013	-2.17681
_cons	20.11387	.1520326	132.30	0.000	19.81584	20.41189

Heteroskedasticity robust:

```
1 . reg ahe female, robust
```

Linear regression

Number of obs = 7711
F(1, 7709) = 134.80
Prob > F = 0.0000
R-squared = 0.0165
Root MSE = 10.056

ahe	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
female	-2.629912	.2265122	-11.61	0.000	-3.073937	-2.185886
_cons	20.11387	.1614226	124.60	0.000	19.79744	20.4303

Implication of heteroskedasticity

- If the regression errors are homoskedastic and normally distributed and if the homoskedasticity-only t-statistics is used, then critical values should be taken from the Student t distribution.
- In econometric applications the errors are rarely homoskedastic and normally distributed, but as long as n is large and we compute heteroskedasticity robust standard errors we can compute t-statistics and hence p-values and confidence intervals as normal.

Prediction

```
1 . reg ahe age
```

Source	SS	df	MS
Model	23005.7375	1	23005.7375
Residual	769645.718	7709	99.8372964
Total	792651.456	7710	102.80823

Number of obs = 7711
F(1, 7709) = 230.43
Prob > F = 0.0000
R-squared = 0.0290
Adj R-squared = 0.0289
Root MSE = 9.9919

ahe	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
age	.6049863	.0398542	15.18	0.000	.5268613 .6831113
_cons	1.082275	1.184255	0.91	0.361	-1.239187 3.403737

The regression result gives:

$$\hat{Y} = 1.08 + 0.60age$$

Predictions:

- A 26 year old worker is predicted to have an average hourly wage of: \$ 16.68 (1.08+0.6*26).
- For each year of education you are predicted to earn \$ 0.6 more.

Note of caution:

- The test statistic: The t-value and hence the p-value and confidence interval is only as good as the underlying assumptions used to construct it.
- If any of the underlying assumptions are violated the test statistic is not reliable.
- Most often the violated assumption is the zero conditional mean assumption, X is often correlated with the error term.
- More about this in the next lecture when we talk about omitted variable bias.