

Seminar 1 - OLS

1. Which of the following can cause an OLS estimator to be biased? Explain.
 - (a) Heteroskedasticity
 - (b) Omitting an important variable
 - (c) A sample correlation of .95 between two regressors included in the model
2. In the models below, show that $\hat{\beta}$ is an unbiased estimator of β . Be explicit about the key assumptions.
 - (a) using scalar notation: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$
 - (b) using vector notation: $y_i = \mathbf{x}'_i \boldsymbol{\beta} + \epsilon_i$
 - (c) using matrix notation: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
3. Say that the variance of ϵ does not depend on \mathbf{x} , i.e. $\text{var}(\epsilon|\mathbf{x}) = \sigma^2$. Show that $\text{var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$.
4. Derive $\text{var}(\hat{\beta}_1|\mathbf{x})$ in 2(a), assuming $\text{var}(\epsilon_i|x_i) = \sigma_i^2$
5. Assume $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$. Assume that $E(\epsilon_i|x_{2i}) \neq 0$. Can $\hat{\beta}_1$ still be an unbiased estimator of β_1 ?
6. Consider the regression output below, where the log of hourly wages (lnwage) is regressed on years of schooling (eduy) and age.
 - (a) Interpret the coefficient on -eduy- in the first regression.
 - (b) What is the wage of people with zero years of schooling? And with 10 years of schooling?
 - (c) Calculate the 90% confidence interval of the coefficient on -eduy-
 - (d) What is the interpretation of the R-squared?

The second regression adds -age- as a regressor.

- (e) Calculate the F -statistic for $H_0 : \beta_{age} = 0$ using i) the sum of squares from the regression output, ii) the R-squares. Compare your results to what you get using the t-statistic in the second regression.
- (f) What is the correlation between -age- and -eduy-?

```

. reg lnw eduy

      Source |           SS          df           MS              Number of obs =      31237
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      Model |    633.04524            1    633.04524              F( 1, 31235) = 4498.68
      Residual | 4395.32675    31235    .140718001              Prob > F      = 0.0000
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      Total | 5028.37199    31236    .160980023              R-squared     = 0.1259
                                          Adj R-squared = 0.1259
                                          Root MSE    = .37512

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      lnwage |           Coef.      Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      eduy |    .0528497      .000788      67.07   0.000     .0513053     .0543941
      _cons |    1.602484      .0097426    164.48   0.000     1.583388     1.62158

. reg lnw eduy age

      Source |           SS          df           MS              Number of obs =      31237
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      Model |    1085.5649            2    542.782451              F( 2, 31234) = 4299.80
      Residual | 3942.80709    31234    .126234459              Prob > F      = 0.0000
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      Total | 5028.37199    31236    .160980023              R-squared     = 0.2159
                                          Adj R-squared = 0.2158
                                          Root MSE    = .35529

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      lnwage |           Coef.      Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      eduy |    .0629493      .0007651     82.27   0.000     .0614496     .0644489
      age |    .0145086      .0002423     59.87   0.000     .0140336     .0149836
      _cons |    .9515069      .0142606     66.72   0.000     .9235556     .9794582

. sum lnw eduy age

      Variable |           Obs          Mean        Std. Dev.        Min          Max
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      lnw |    31237      2.240249      .4012232    -.7533873     6.214762
      eduy |    31237     12.06752      2.693686           5           17
      age |    31237     36.4681      8.505208          20           50

```

Recommended exercises from Wooldridge

- W4.1 a-c - interpreting non-linear LHS, delta method.
- W4.6 endogeneity of predetermined variable
- W4.7 endogeneity in regression
- W4.8a-c - average partial effect