

Exercise Sheet 2 – B203 Econometrics
Costas Meghir

1. Consider two different ways of fitting a line to data:
 - a. **Least squares**, which minimises the residual sum of squares – we did this in the lectures. Verify that you have understood how the coefficients are estimated. Use the data below to estimate the slope and intercept coefficients using least squares.
 - b. **The Wald estimator**. Split the data into two sub-samples. The first includes all observations with values of the explanatory variable X below the median; the second includes the observations with X above the median. Take the average of X and Y in each sub-sample. Join the two averages. The slope of this line is an estimator of the slope coefficient. The intercept of this line is an estimator of the intercept (a).
 - c. Show graphically how you would compute the Wald Estimator using the data from the previous exercise sheet (provided below)

Experiment	Price			Quantity Sold
1.	10			23.39
2.	20			20.32
3.	8			24.74
4.	22			22.2
5.	28			14.81
6.	30			12.63
7.	33			13.03
8.	38			12.17

- d. The formula for the Wald estimator for the slope is $\tilde{b} = \frac{\bar{Y}_1 - \bar{Y}_2}{\bar{X}_1 - \bar{X}_2}$, where the bar refers to average and the 1 or 2 subscript refers to the below and above median X sub-sample respectively as described above. Calculate the slope and compare it to the slope you obtained using the Least Squares formula. Now calculate the estimator of the intercept.
 - e. Show that the slope coefficient obtained using the Wald estimator and the one obtained using least squares are both unbiased.
 - f. Why do you think they differ? On what basis do you think we will choose between the two estimators?
2. Show mathematically the relationship between the correlation of Y and X and the least squares estimator of the slope coefficient in the regression of Y on X .
3. You are provided with a STATA data set called “pssobh.dta”. This includes one sample from the US (PSID, data==1) one from Germany (SOEP data==2) and UK (BHPS, data==3). Estimate the impact of age on log wage (lnw) for each country and for males and females separately. Make a table showing the percentage growth of wages with age in each country by gender. Which country has the largest wage growth over the lifecycle? How do wage growths differ for men and women.? Plot the estimated wage-age profile for each country on the same graph for men; Repeat that for women. Why do you think that wages grow with age?
4. Use any of these regressions to verify, using STATA, that the R^2 is the square of the correlation coefficient between actual values of Y and the fitted values of Y .