



Quantitative Methods for Economics

Tutorial 7

Katherine Eyal



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 2.5 South Africa License](https://creativecommons.org/licenses/by-nc-sa/2.5/za/).



TUTORIAL 7

20 September 2010

ECO3021S

Part A: Problems

1. You are given the following age and price data for 10 randomly selected Toyota Tazzes between 1 and 6 years old. Here, *age* is in years, and *price* is in thousands of Rands.

<i>age</i>	6	6	6	2	2	5	4	5	1	4
<i>price</i>	205	195	210	340	299	230	270	243	340	240

(**Note:** You may NOT use Stata for this question. You may use Excel, but NOT the Data Analysis tools. You may use a calculator.)

- (a) Estimate the relationship between *age* and *price* using OLS; i.e., obtain the intercept and slope estimates in the equation

$$\widehat{price} = \widehat{\beta}_0 + \widehat{\beta}_1 age.$$

Comment on the direction of the relationship. Does the intercept have a useful interpretation here? Explain. How much lower is *price* predicted to be if *age* increased by two years?

- (b) Compute the fitted values and residuals for each observation, and verify that the residuals (approximately) sum to zero.
 - (c) What is the predicted value of *price* when *age* = 3?
 - (d) How much of the variation in *price* for these ten cars is explained by *age*? Explain.
2. Consider the savings function

$$\begin{aligned} sav &= \beta_0 + \beta_1 inc + u \\ u &= \sqrt{inc} \cdot e \end{aligned}$$

where *e* is a random variable with $E(e) = 0$ and $\text{Var}(e) = \sigma_e^2$. Assume that *e* is independent of *inc*.

- (a) Show that $E(u|inc) = 0$, so that the key zero conditional mean assumption (Assumption SLR.4 in Wooldridge) is satisfied. [*Hint:* If *e* is independent of *inc*, then $E(e|inc) = E(e)$.]

- (b) Show that $\text{Var}(u|inc) = \sigma_e^2 inc$, so that the homoskedasticity assumption (Assumption SLR.5 in Wooldridge) is violated. In particular, the variance of *sav* increases with *inc*. [Hint: $\text{Var}(e|inc) = \text{Var}(e)$, if e and inc are independent.]
- (c) Provide a discussion that supports the assumption that the variance of savings increases with family income.
3. Suppose an econometric analysis finds that in Ghana average annual per capita grain consumption follows the relationship

$$\log(\textit{grain}) = 170 + 0.8 \log(\textit{annual per capita income}) - 0.5 \log(\textit{price of grain})$$

where *grain* is measured in kilograms per person, *annual per capita income* in Kenyan shillings, and *price of grain* in Kenyan shillings per kilogram.

- (a) What is the per capita grain consumption when per capita income equals 1,000 Kenyan shillings and the price of grain is 10 Kenyan shillings per kilogram?
- (b) If income rises from 1,000 to 1,010 Kenyan shillings (and the price remains unchanged at 10 Kenyan shillings per kilogram), compute what happens to grain production. How does this calculation compare with the estimated income elasticity of grain consumption?
- (c) If the price of grain rises from 10 to 20 Kenyan shillings per kilogram (and income remains unchanged at 1,000 Kenyan shillings), compute what happens to grain production. How does this calculation compare with the estimated price elasticity of grain consumption?

Part B: Computer Exercises

1. For many years, housing economists believed that households spent a constant fraction of income on housing, as in

$$\textit{housing expenditure} = \beta(\textit{income}) + u.$$

The data set HOUSING.DTA contains housing expenditures (*housing*) and total expenditures (*total*) for a sample of 19th century Belgian workers collected by Edouard Ducpetiaux.

- (a) Find the averages of total expenditure and housing expenditure in this sample.
- (b) Estimate β , using total expenditure for total income.
- (c) If income rises by 100 (it averages around 900 in this sample), what increase in estimated expected housing expenditure results?

- (d) What economic argument would you make against housing absorbing a constant share of income?
- (e) What are some determinants of housing that are captured by u ?
2. Use the data in CHARITY.DTA to answer the following questions. We are interested in the relationship between charitable contributions made by a person ($gift$) and the number of mailings per year sent by a charitable organisation to that person ($mailsyar$).
- (a) What is the average gift in the sample of 4,268 people (in Dutch guilders)? What percentage of people gave no gift?
- (b) What is the average mailings per year? What are the minimum and maximum values?
- (c) Estimate the model
- $$gift = \beta_0 + \beta_1 mailsyar + u$$
- by OLS and interpret your results.
- (d) Interpret the slope coefficient. If each mailing costs one guilder, is the charity expected to make a net gain on each mailing? Does this mean the charity makes a net gain on every mailing? Explain.
- (e) What is the smallest predicted charitable contribution in the sample? Using this simple regression analysis, can you ever predict zero for $gift$?

TUTORIAL 7 SOLUTIONS

20 September 2010

ECO3021S

Part A: Problems

1. You are given the following age and price data for 10 randomly selected Toyota Tazzes between 1 and 6 years old. Here, *age* is in years, and *price* is in thousands of Rands.

<i>age</i>	6	6	6	2	2	5	4	5	1	4
<i>price</i>	205	195	210	340	299	230	270	243	340	240

(**Note:** You may NOT use Stata for this question. You may use Excel, but NOT the Data Analysis tools. You may use a calculator.)

- (a) Estimate the relationship between *age* and *price* using OLS; i.e., obtain the intercept and slope estimates in the equation

$$\widehat{price} = \widehat{\beta}_0 + \widehat{\beta}_1 age.$$

Comment on the direction of the relationship. Does the intercept have a useful interpretation here? Explain. How much lower is *price* predicted to be if *age* increased by two years?

- (b) Compute the fitted values and residuals for each observation, and verify that the residuals (approximately) sum to zero.
- (c) What is the predicted value of *price* when *age* = 3?
- (d) How much of the variation in *price* for these ten cars is explained by *age*? Explain.

SOLUTION:

- (a) Estimate intercept and slope using OLS: $\widehat{\beta}_1 = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) / \sum_{i=1}^n (x_i - \bar{x})^2$ and $\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$, to get $\widehat{\beta}_1 = -27.90$ and $\widehat{\beta}_0 = 371.6$.

There is a negative relationship between the age of the Toyota Tazz and its selling price: the value of a car decreases as it gets older, because of wear and tear. The intercept reflects the price of a new Toyota Tazz, i.e. when *age* = 0. However, we must be cautious when interpreting the intercept because our data do not include any observations where *age* = 0.

We can express the predicted change in price as a function of the change in the age of the car: $\Delta \widehat{price} = -27.90 (\Delta age)$. Thus, if age increases by two years, $\Delta age = 2$, then *price* is predicted to decrease by about 55.8.

- (b) Fitted values: $\widehat{price}_i = 371.6 - 27.90 \text{ age}_i$. The residuals are the difference between the observed and the fitted values.

<i>Fitted values</i>	<i>Residuals</i>
204.2	0.8
204.2	-9.2
204.2	5.8
315.8	24.2
315.8	-16.8
232.1	-2.1
260.0	10.0
231.1	10.9
343.7	-3.7
260.0	-20.0

You can verify that the residuals sum to -0.1 which is close to 0.

- (c) When the age of the car is 3, the predicted price of the car is 287.9.
- (d) The fraction of the sample variation in price that is explained by age is given by the *R*-squared or coefficient of determination. $R\text{-squared} = SSE/SST = 1 - SSR/SST$.
 $SSR = 1623.71$ and $SST = 25681.6$. $R\text{-squared} = 0.937 \Rightarrow$ Almost 94% of the variation in the price of the car is explained by its age.

2. Consider the savings function

$$\begin{aligned} sav &= \beta_0 + \beta_1 inc + u \\ u &= \sqrt{inc} \cdot e \end{aligned}$$

where e is a random variable with $E(e) = 0$ and $\text{Var}(e) = \sigma_e^2$. Assume that e is independent of inc .

- (a) Show that $E(u|inc) = 0$, so that the key zero conditional mean assumption (Assumption SLR.4 in Wooldridge) is satisfied. [*Hint*: If e is independent of inc , then $E(e|inc) = E(e)$.]
- (b) Show that $\text{Var}(u|inc) = \sigma_e^2 inc$, so that the homoskedasticity assumption (Assumption SLR.5 in Wooldridge) is violated. In particular, the variance of *sav* increases with *inc*. [*Hint*: $\text{Var}(e|inc) = \text{Var}(e)$, if e and inc are independent.]

- (c) Provide a discussion that supports the assumption that the variance of savings increases with family income.

SOLUTION:

- (a) When we condition on inc in computing an expectation, \sqrt{inc} becomes a constant. So $E(u|inc) = E(\sqrt{inc} \cdot e|inc) = \sqrt{inc} \cdot E(e|inc) = \sqrt{inc} \cdot 0$ because $E(e|inc) = E(e) = 0$.
- (b) Again, when we condition on inc in computing a variance, \sqrt{inc} becomes a constant. So $Var(u|inc) = Var(\sqrt{inc} \cdot e|inc) = (\sqrt{inc})^2 Var(e|inc) = \sigma_\varepsilon^2 inc$ because $Var(e|inc) = \sigma_\varepsilon^2$.
- (c) Families with low incomes do not have much discretion about spending; typically, a low-income family must spend on food, clothing, housing, and other necessities. Higher income people have more discretion, and some might choose more consumption while others more saving. This discretion suggests wider variability in saving among higher income families.
3. Suppose an econometric analysis finds that in Ghana average annual per capita grain consumption follows the relationship

$$\log(\textit{grain}) = 170 + 0.8 \log(\textit{annual per capita income}) - 0.5 \log(\textit{price of grain})$$

where \textit{grain} is measured in kilograms per person, $\textit{annual per capita income}$ in Kenyan shillings, and $\textit{price of grain}$ in Kenyan shillings per kilogram.

- (a) What is the per capita grain consumption when per capita income equals 1,000 Kenyan shillings and the price of grain is 10 Kenyan shillings per kilogram?
- (b) If income rises from 1,000 to 1,010 Kenyan shillings (and the price remains unchanged at 10 Kenyan shillings per kilogram), compute what happens to grain production. How does this calculation compare with the estimated income elasticity of grain consumption?
- (c) If the price of grain rises from 10 to 20 Kenyan shillings per kilogram (and income remains unchanged at 1,000 Kenyan shillings), compute what happens to grain production. How does this calculation compare with the estimated price elasticity of grain consumption?

SOLUTION:

- (a) $\log(\textit{grain}) = 170 + 0.8 \log(1000) - 0.5 \log(10)$
 Thus, annual grain consumption = $\textit{grain} = \exp(170) \cdot 1000^{0.8} \cdot 10^{-0.5} = 5.37108 \times 10^{75}$ kilograms per person

- (b) Assuming production (supply) equals consumption (demand)

New annual grain production = $\exp(170) \cdot 1010^{0.8} \cdot 10^{-0.5} = 5.414 \times 10^{75}$ kilograms per person

So when income increases by 1% (holding the price constant), grain consumption per capita (production per capita) increases by 0.79% $[(5.414 - 5.371)/5.371 \times 100]$.

The coefficient on $\log(\textit{annual per capita income})$ is the estimated income elasticity of grain consumption since we have a log-log model. Our calculation is very close to the estimated elasticity.

Note that the estimated income elasticity of grain consumption is slightly inelastic; this may be because it is a staple food and so demand changes are slightly less responsive to changes in income.

- (c) New grain production = $\exp(170) \cdot 1000^{0.8} \cdot 20^{-0.5} = 3.797 \times 10^{75}$ kilograms per person

So, when the price increases by 100% (holding income constant), annual production/consumption decreases by 41%. $[(3.797 - 5.371)/5.371 \times 100]$

This is similar to the price elasticity of grain consumption estimated in our model of 0.5, which implies that for a 100% increase in the price of grain, consumption decreases by 50%.

Note: If we have a model where $\log y$ is the dependent variable, predicting y by simply exponentiating the predicted value for $\log(y) : \hat{y} = \exp(\widehat{\log y})$ does not work! This method will systematically *underestimate* the expected value of y . This is why our calculations above resulted in elasticities that were smaller than the estimated elasticities. A consistent, but not unbiased, prediction of y when the error term is normal is: $\hat{y} = \exp(\hat{\sigma}^2/2) \exp(\widehat{\log y})$, where $\hat{\sigma}$ is the standard error of the regression. We do not have the standard error of the regression for this question so we cannot calculate consistent predictions of y . Note that there are no unbiased predictions of y when $\log y$ is the dependent variable. It is also possible to predict y when the error term is not normally distributed. We will discuss this in more detail in class when we cover Chapter 6 (page 211 of Wooldridge).

Part B: Computer Exercises

1. For many years, housing economists believed that households spent a constant fraction of income on housing, as in

$$\textit{housing expenditure} = \beta (\textit{income}) + u.$$

The data set HOUSING.DTA contains housing expenditures (*housing*) and total expenditures (*total*) for a sample of 19th century Belgian workers collected by Edouard Ducpetiaux.

- (a) Find the averages of total expenditure and housing expenditure in this sample.
- (b) Estimate β , using total expenditure for total income.
- (c) If income rises by 100 (it averages around 900 in this sample), what increase in estimated expected housing expenditure results?
- (d) What economic argument would you make against housing absorbing a constant share of income?
- (e) What are some determinants of housing that are captured by u ?

SOLUTION:

- (a) The average of total expenditure is 902.82 and for housing expenditure is 72.54. (The Stata command is `sum`.)
 - (b) The estimated slope using total expenditure for total income is 0.075. Stata command: `regress housing total, noconstant`.
 - (c) If income rises by 100, the estimated expected housing expenditure increase is 7.5.
 - (d) Households with low incomes do not have much discretion about spending; typically, a low-income household must spend a certain proportion on the necessities such as housing. Higher income households have more discretion about their spending. For instance, wealthier households may choose to live in very large houses in more expensive locations, which cost a greater proportion of their income.
 - (e) Family size, whether any members of the household work from home, location, etc.
2. Use the data in CHARITY.DTA to answer the following questions. We are interested in the relationship between charitable contributions made by a person (*gift*) and the number of mailings per year sent by a charitable organisation to that person (*mailsy*).
- (a) What is the average gift in the sample of 4,268 people (in Dutch guilders)? What percentage of people gave no gift?
 - (b) What is the average mailings per year? What are the minimum and maximum values?

- (c) Estimate the model

$$gift = \beta_0 + \beta_1 mailsyear + u$$

by OLS and interpret your results.

- (d) Interpret the slope coefficient. If each mailing costs one guilder, is the charity expected to make a net gain on each mailing? Does this mean the charity makes a net gain on every mailing? Explain.
- (e) What is the smallest predicted charitable contribution in the sample? Using this simple regression analysis, can you ever predict zero for *gift*?

SOLUTION:

- (a) The average gift is about 7.44 Dutch guilders. Out of 4,268 respondents, 2,561 did not give a gift, or about 60 percent.
- (b) The average mailings per year is about 2.05. The minimum value is .25 (which presumably means that someone has been on the mailing list for at least four years) and the maximum value is 3.5.
- (c) The estimated equation is

$$\begin{aligned}\widehat{gift} &= 2.01 + 2.65 mailsyear \\ n &= 4,268, \quad R^2 = .0138\end{aligned}$$

- (d) The slope coefficient from part (c) means that each mailing per year is associated with – perhaps even “causes” – an estimated 2.65 additional guilders, on average. Therefore, if each mailing costs one guilder, the expected profit from each mailing is estimated to be 1.65 guilders. This is only the average, however. Some mailings generate no contributions, or a contribution less than the mailing cost; other mailings generated much more than the mailing cost.
- (e) Because the smallest *mailsyear* in the sample is .25, the smallest predicted value of *gift* is $2.01 + 2.65(.25) \approx 2.67$. Even if we look at the overall population, where some people have received no mailings, the smallest predicted value is about two. So, with this estimated equation, we never predict zero charitable gifts.