University of Cape Town

Centre for Actuarial Research (CARe)

DOC4002F – Demographic Data and Statistics

Professor : Vissého Adjiwanou, Ph.D. Academic year: 2018

Tutorial 8: Linear regression: properties of an estimator (part 1)

Exercise 1:

Suppose $y = \alpha + \beta x + \varepsilon$, where the ε's are iid with mean 0 and variance σ^2. Suppose the data are divided evenly into two groups denoted by subscripts a and b, and $\boldsymbol{\beta}$ is estimated by $\beta^* = (y_a - y_b)/(x_a - x_b)$ where ya is the average of all the y observations in group a, etc.

1. Define the algebraic form of y_a, y_b, x_a, x_b
2. Show that β^* is unbiased estimator of β
3. Find the variance of β^*, $Var(\beta^*)$
4. Demonstrate that $Var(\beta^*) = (2\sigma)^2/T(x_a-x_b)^2$ in the specific case where the number of observation in each group is T/2, T being the total number of observations
5. How would you allocate observations into the two groups? Why?

Exercise 2:

The Mean Square Error of the estimator $\hat{\theta}$ is defined by $MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$, $\hat{\theta} - \theta$ is called the sampling error, and $E(\hat{\theta}) - \theta$, is the bias.

1. Define in your own word your understanding of MSE
2. Demonstrate that $MSE(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})])^2 + [E(\hat{\theta}) - \theta]^2$ = Variance + Square bias

Exercise 3:

Suppose we have the sample x_1, x_2, and x_3, drawn randomly from a distribution with mean 4 and varince 9. You have two estimates of the mean: $\mu^* = (x_1 + x_2 + x_3)/3$ and $\mu^{**} = (x_1 + x_2 + x_3)/4$

1. Compute the expected value of the two estimators
2. Compute their variances
3. Which estimator will you choice? Why?

Exercise 4: Estimating regression parameters (manually)

The true relationship between X and Y in the population is given by: $Y_i = 2 + 3X_i + \varepsilon_i$.

Suppose the values of X in the sample of 10 observations are 1,2,…..,10. The values of the disturbances are drawn at random from a normal population with zero mean and unit variance.

$\varepsilon_1 = 0.464$	$\varepsilon_6 = 0.296$
$\varepsilon_2 = 0.137$	$\varepsilon_7 = -0.288$
$\varepsilon_3 = 2.455$	$\varepsilon_8 = 1.298$
$\varepsilon_4 = -0.323$	$\varepsilon_9 = 0.241$
$\varepsilon_5 = -0.068$	$\varepsilon_{10} = -0.957$

1. Determine the 10 observed values of X and Y
2. Put the information on a graph using R (ggplot)
3. Use the least square formulas to estimate (manually) the regression coefficients and their standard errors, and compare the results with the true values
4. Draw the estimated line using R in the same graph as in 2.
5. Estimate the parameters using R

Exercise 5: Estimating regression points estimates and variances (manually)

Consider the data below on prices and quantities of oranges sold in a supermarket on twelve consecutive days.

Table 7-2

Price: ¢/lb	Quantity: lb
100	55
90	70
80	90
70	100
70	90
70	105
70	80
65	110
60	125
60	115
55	130
50	130

Let Xi be the price charged and Yi the quantity sold on sold on the *i*th day. Let us further postulate that the demand function is of the form $Y_i = \alpha + ßX_i + \varepsilon_i$, and such that the basic assumptions of the classical normal regression mode are satisfied.

1. Estimates the parameters of the model (manually)
2. Estimated the sample regression line and show it in graph using R
3. Compute the biased and unbiased estimates of the error term variance
4. Compute the estimated variance of α^* and $ß^*$
5. Derive the R^2. Comment.
6. Compare your results with what you get using R. Relate each estimate from R to what you estimate manually.
7. What do you have from the R regression table that you have not already compute manually? Do you have any idea to what to do it.