

620-371: Linear Models

Practice Class 5

31st March, 2009

In this practice class, we will replicate and extend the results of the practice class from last week, using R. From last week: we model an individual's income at age 30 against the number of years of formal education (with a linear model). The following data is collected:

Years of formal education (x)	Income (\$k) (y)
8	8
12	15
14	16
16	20
16	25
20	40

1. Construct X and \mathbf{y} objects in R. Also set the number of samples, n , and the number of parameters, p .
2. Estimate the parameters β using the formula for the least squares estimator.
3. Create a scatterplot of the data and add the regression line to it.
4. Create a data frame with the data in two components, `income` and `education`.
5. Use the `lm` function to estimate the parameters again. Check that you have the same result as in question 2.
6. Produce diagnostic plots for the model.
Do the remaining questions in two ways: using matrix manipulation as described in the lecture slides, and using your `lm` model. Check that you get the same results!
7. Estimate the income of a person who has had 18 years of formal education.
8. Calculate the residuals of the model.
9. Estimate the variance of the error, σ^2 .
10. Find 95% confidence intervals for both parameters.
11. Find a 95% confidence interval for the average income of a person who has had 18 years of formal education.