Erasmus Econometrics - Test Assignment 1

Harro Cyranka

7/15/2017

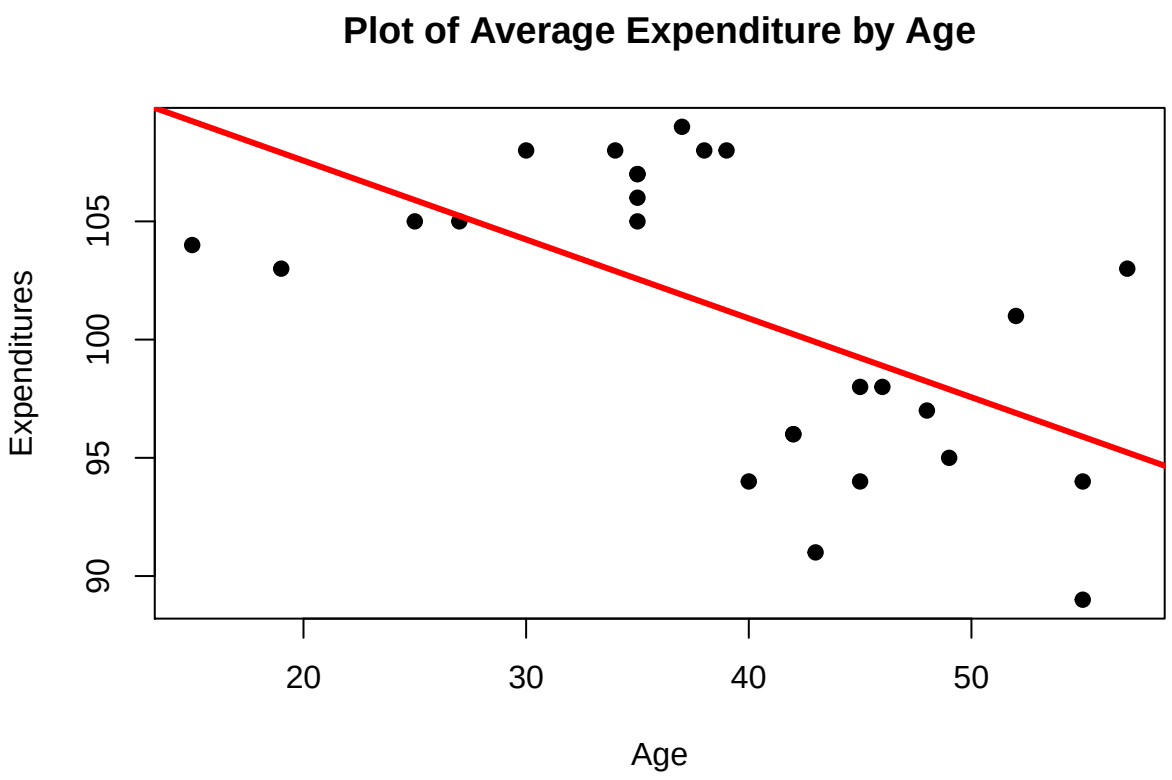
Introduction

In this assignment, I try to explore a dataset whose observations do not satisfy all the standard assumptions of linear regression. Specifically, the dataset violates assumption 6 that the coefficients α and β are the same for all observations.

Estimating OLS Coefficients, Age Standard Error and Age T-Value

```
## [1] "Estimate of the Intercept: 114.24"
## [1] "Estimate of the Age Coefficient: -0.33"
## [1] "Standard Error of Age Coefficient: 0.0954"
## [1] "T-statistic of Age Coefficient: -3.4979"
## [1] "P-Value Associated with Age Coefficient: 0.0019"
```

Creating a Scatter Diagram to Visualize the Model



Comments on the scatterplot

The scatterplot provides a graphical representation of the first model. There is a negative association between age and average expenditures. The value of the coefficient suggests that a marginal increase in age is accompanied by a reduction of U$0.33 in average expenditures. The coefficient is statistically significant, allowing us to reject the null hypothesis that there is no linear association between the two variables. However, the scatterplot also reveals two separate groups of observations and that the model is probably not capturing the true relationship between age and average expenditure.

Creating Separate Models

Model A: Estimating OLS Coefficients, Age Standard Error and Age T-Value for Customers Under 40

```
## [1] "Estimate of the Intercept: 100.23"
## [1] "Estimate of the Age Coefficient: 0.2"
## [1] "Standard Error of Age Coefficient: 0.0444"
## [1] "T Statistic of Age Coefficient: 4.4605"
## [1] "P-Value Associated with Age Coefficient: 0.001"
```

Model B: Estimating OLS Coefficients, Age Standard Error and Age T-Value for Customers Over 40

```
## [1] "Estimate of the Intercept: 88.87"
## [1] "Estimate of the Age Coefficient: 0.15"
## [1] "Standard Error of Age Coefficient: 0.1974"
## [1] "T Statistic of Age Coefficient: 0.7421"
## [1] "P-Value Associated with Age Coefficient: 0.4736"
```

Comments on new results

There are a handful of differences that are worth exploring:

i) The coefficient for age, which was initially negative in the first model, became positive in the two most recent models. For customers under 40, a marginal increase in age is accompanied by an expected increase of U$0.2 in average daily expenditures. For customers over 40, a marginal increase in age is associated with an expected increase of U$0.15 in average daily expenditures.

ii) Whereas we are able to reject the null hypothesis in the first model and in the model for customers under 40, it is not possible to reject it for customers over 40, for the p-value associated with the coefficient b under the assumption that the null hypothesis is true is equal to 0.4736.