Statistics for economics
Computer test 2

School of Economics, University College Dublin

Spring 2017

Download `coffeeyield.RData` from the blackboard website and load it in R. The data is taken from the 2002 Nature paper "The Value of Bees to the Coffee Harvest" by Roubik, and contains data on coffee yields (in kg/ha^{-1}). It includes the following variables

- `country`, country name

- `world`, factor variable indicating whether the country belongs to the old or new world

- `world.d`, binary indicator whether the country belongs to the old (0) or new world (1)

- `yield61to80`, coffee yield between 1961-1980

- `yield81to01`, coffee yield between 1981-2001

Use the data to answer the questions in the next section. Each question is worth one point, and if you answer all nine questions you get a bonus point, for a total of 10 points. Write your answers down in your favourite word processor and include the figures you produce. When you're done you need to send the document with your answers to stijn.vanweezel@ucd.ie

Questions

1. Create a histogram for the coffee yield between 1981-2001, setting the breaks equal to 10. Describe the distribution.

2. Between 1981-2001, what is the average and standard deviation of the coffee yield in the new world and in the old world?

3. Use a scatterplot to show the coffee yield between 1961-1980 versus 1981-2001. Is there a relation between the two periods?

4. Add points for the new world countries to the scatterplot.[1] What would you conclude on the basis of the figure?

   [1] Can use `pch=19` or `col="blue"` or both.

5. By world group[2], create a boxplot for the coffee yield between 1981-2001. Which conclusions do you draw based on the figure?

   [2] So for old and new world countries

6. Are there countries that see a reduction in yield between the two periods? If so, how many?

7. Use a paired t-test to examine whether the average yield in old world countries is the same for the two periods.

8. Use a paired t-test to examine whether the average yield in new world countries is higher between 1981-2001 compared to 1961-1980. What does the uncertainty interval tell?

9. Estimate a regression model where you regress the average yield between 1981-2001 on the yield between 1961-1981 and the world indicator. Analyse the estimated effects.