ECU22032 Statistics Assignment 2022/23

Michael Wycherley Trinity College Dublin

Due before noon on Thursday 6th April 2023

All solutions should be **typed** (you can copy and paste tables and regression results from excel into word).
Do not forget to include the cover sheet.

Note: It is good practice when testing a hypothesis to state both the null and alternative hypotheses, as well as the decision rule.

We are going to use the 1998 ISSP cross country survey to test the idea of a "Protestant work ethic", the data is available on Blackboard.

1. **Average hours worked [30 marks]**

   (a) Estimate mean hours worked for each of the two religious groups and their variances

   (b) Use the sample to test whether Protestants in general work more than Catholics, assuming the population variances are different

   (c) What is the power of the test if, in general, the average working hours of Protestants is the same as that of Catholics?

   (d) Test whether the populations variances are different. If there is evidence that they are the same then repeat part b making the appropriate assumption about the variances.

   (e) Does this suggest the existence of a Protestant work ethic? Explain why or why not.

   (f) How might we better address the question of whether a Protestant work ethic exists?

2. **Regression [40 marks]**.

   Estimate the regression below

$$
\begin{aligned}
AverageHoursWorked_i &= \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \varepsilon_i \\
X_{1,i} &= \text{takes the value 1 if Catholic and 0 if Protestant} \\
X_{2,i} &= \text{increasing in the self reported strength of their faith} \\
X_{3,i} &= \text{increasing in the amount of education}
\end{aligned}
$$

   (a) How do you interpret the coefficient estimates and the R^2?

   (b) Are the coefficient estimates statistically significant?

   (c) What does the F-test tell us?

(d) Do these results agree with those in question one and if not why might this be?

It's possible that the effect of religion depends on the level of educaiton or the strength of faith, so estimate a new regression as follows

$$AverageHoursWorked_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \beta_4 X_{1,i} X_{2,i} + \beta_5 X_{1,i} X_{3,i} + \varepsilon_i$$

(e) Compare this regression to the previous one.

(f) What, if anything, does this tell us about the possibility of a Protestant work ethic? What data and methods might help us to analyse the topic better?

3. **Principles of Regression [30 marks]**

Consider the following multiple regression model:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \varepsilon_i.$$

(a) Explain how the ordinary least squares estimator for $\beta = (\beta_0, \beta_1, \beta_2)$ is determined and how the expressions for the coefficient estimators $b = (b_0, b_1, b_2)$ are derived[1].

(b) Which assumptions are needed to make b an unbiased estimator for β and why? An unbiased estimator is one where the expected value (not the actual value obtained from the sample) of $b = \beta$.

(c) Explain how one can test the hypothesis that $\beta_1 = 1$.Which additional assumptions are needed and why?

(d) Suppose that $x_{i,1} = -2x_{i,2}$. What will happen if you try to estimate the above model?

(e) Again assuming $x_{i,1} = -2x_{i,2}$, how will estimating $y_i = \beta_0^* + \beta_1^* x_{i,2} + \varepsilon_i^*$ differ from estimating $y_i = \beta_0 + \beta_1 x_{i,1} + \varepsilon_i$ and what does this imply for regressions if variables are measured in different units (e.g. metres or km)?

(f) Now suppose that $x_{i,1} = -2x_{i,2} + u_i$ where $x_{i,2}$ and u_i are uncorrelated. How will estimating $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \varepsilon_i$ differ from estimating $y_i = \beta_0^* + \beta_1^* x_{i,1} + \beta_2^* u_i + \varepsilon_i^*$?

[1] You are not being asked to derive the estimators, but to explain how they are obtained.