# **BQA WEEK 2**

## **EXERCISE 3-6:**

## **BIVARIATE PARAMETRIC METHODS**

## **ANSWERS**

**+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++**

### 3. CORRELATION COEFFICIENTS

#### 3.1 CORRELATION MATRIX

|  |  |
| --- | --- |
| **One variable in the list appears to be less strongly correlated with the others. Which is it?** | ***Ppltrst – the variable which asks about the statement “most people can be trusted”. This may be because the question is in a rather different form than the others*** |
| **Two groups of variables exhibit a particularly strong relationship with each other. What are those groups of variables, and why do you think they are particularly strongly correlated?** | ***One pair is the European parliament and the UN – both supranational organizations. The other is parliament, politicians and political parties – again, similar institutions.*** |

#### 3.2 FINDING THE SLOPE OF A SCATTERPLOT USING OLS

|  |  |
| --- | --- |
| **Look at the line of best fit, and write down its approximate slope, (the increase in the Y variable associated with a one-unit increase in the X variable)** | ***Do this by inspection of the graph – check that you know it should be about 12*** |
| **Now calculate the slope formally using OLS regression , along with a p-value (the syntax is regress LS GETBY)** | ***12.5469 p=0.000*** |

|  |  |
| --- | --- |
| **Now do the same for the relationship between IMPRICH and IMPCREATE: write down your estimate of the slope by looking at the scatterplot:** | ***By inspection – negative and around 0.5*** |
| **And calculate it formally using OLS regression, along with a p-value** | ***-.6169371 (not significant, p=0.066)*** |

|  |  |
| --- | --- |
| **Question: how are these findings similar, and how are they different, to what we found last week at an aggregate country level?** | ***When looking at country means, there is a clear negative relationship between responses on the importance of being rich and the other variables. This is not the case at the individual level. Think about why!*** |

#### 3.3 RELATIONSHIP BETWEEN CORRELATION COEFFICIENT AND EFFECT SIZE

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Unstandardized variables**  **(GovSat and progressive)** | **Standardized**  **(stdGovSat and stdprogressive)** | **Standardized, taking complete care of missing values** |
| **Correlation coefficient (use corr)** | ***0.1693*** | ***0.1693*** | ***0.1693*** |
| **Slope (use regress)** | ***1.05745*** | ***.1645816*** | ***.1693382*** |

### 4. T-TESTS

#### 4.1 ONE SAMPLE T-TESTS

***Extension (not necessary unless you have some spare time): Perform a t-test which investigates whether the value of LifeSat is equal to the mean of that variable plus 1.96 times the standard error of the mean***

#### 4.2 TWO-SAMPLE T-TEST

|  |
| --- |
| ***There are significant differences in life satisfaction by gender in Spain, but not in the UK.*** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Countries in which people with degrees score significantly *lower* on the progressive values scale than people with lower levels of education** | **Countries in which people with degrees score on average *lower* on the progressive values scale than people with lower levels of education, but the difference is insignificant** | **Countries in which people with degrees score on average *higher* on the progressive values scale than people with lower levels of education, but the difference is insignificant** | **Countries in which people with degrees score significantly *higher* on the progressive values scale than people with lower levels of education** |
|  |  | ***Bulgaria, Czech Rep, Ukraine*** | ***Belgium, Cyprus, Germany, Denmark, Estonia, Spain, Finland, France, GB, Ireland, Israel, Lithuania, Netherlands, Norway, Poland, Portugal, Russia, Sweden, Slovenia, Slovakia*** |

### 5. ANOVA

* Now, run the ANOVA by typing **oneway LifeSat nadults2, tabulate**
* Follow it up with **pwmean LifeSat, over(nadults2)**

### 6. OLS

Also run an OLS regression, using the syntax **regress LifeSat i.nadults2**

Enter both sets of results into the table below. They should be identical (or equivalent).

You will have to use a post-estimation test to work out the final cell in the table for OLS, and remember that the sample means *per se* are not reported for OLS – rather, you can calculate them from the constant term and the other coefficients.

|  |  |  |
| --- | --- | --- |
|  | **ANOVA** | **OLS** |
| **F-test on whole distribution** | ***247.18*** | ***247.18*** |
| **p-value on F-test** | ***0.0000*** | ***0.0000*** |
| **Sample mean over those living in single-adult households** | ***54.943375*** | ***54.94337*** |
| **Sample mean over those living in two-adult households** | ***61.273273*** | ***54.94337 + 6.329899 = 61.273273*** |
| **Difference between those living in 1 and 2 adult households** | ***6.329899*** | ***6.329899*** |
| **CI or p-value of that difference** | ***[5.97572,6.684077]*** | ***P=0.000***  ***[5.97572-6.684077]*** |
| **Difference between those living in 4- and 5- adult households** | ***.5077584*** | ***.5077584*** |
| **CI or p-value of that difference** | ***[-.6115592,1.627076]*** | ***P=0.3739*** |