**Statistical Thinking**

**Problem Set 2**

**CORRELATION COEFFICIENTS**

Your task is to create a correlation matrix of all the variables in the data set relating to conspiracy theories. These are the variables beginning **Glob\_conspiracy**

Now, you are ready to produce your correlation matrix

**pwcorr Glob\_c\***  will produce correlation matrix of variables starting “Glob\_c”

**pwcorr Glob\_c\*, sig** will also give p-values for every cell

|  |  |
| --- | --- |
| **One variable in the list appears to be less strongly correlated with the others. Which is it?** |  |
| **Pick one correlation you find interesting, and use a scatterplot (command: scatter) with name labels (ml(short)) to discuss the cases you think are responsible for this association.** |  |

**ONE SAMPLE T-TESTS**

The syntax for conducting a one-sample t-test in Stata is

**ttest [variable] = [the value you are testing]**

For example,

**ttest earthsun = 75**

provides a test under which the null hypothesis is that the mean of earthsun is equal to 75 (i.e.: the regional average for the percentage of respondents who believe that “the earth goes around the sun” is actually 75% of the population, rather than the recorded mean of our sample, which is just 49%).

The top of the output gives summary statistics; below that, the null hypothesis is set out, and below that, three alternative hypotheses: that the actual mean is greater than, equal to, or less than 75.

Below these three hypotheses are the weight of the distribution in question.

Now test the following hypotheses, and briefly report your findings:

1. That across world regions, on average, a fifth (20%) of the population believe that “the truth about vaccines is being deliberately concealed” (**Glob\_conspiracy\_\_3**)

2. Across world regions, the average number of social networking sites that people make use of is exactly 4 (**n\_social\_media**)

**TWO-SAMPLE T-TEST**

T-tests are more interesting when they are used to compare samples.

For example, if we are interested in the hypothesis that people are more likely to believe conspiracy theories in societies without a history of authoritarian rule, we could generate a variable for whether countries spent most of the last century as liberal democracies or otherwise.

**gen democratic\_legacy = 0**

**replace democratic\_legacy = 1 if country=="Australia" | country=="Denmark" | country=="France" | country=="Sweden" | country=="United States" | country=="United Kingdom" | country=="India" | country=="Japan" | country=="Germany"**

Now we can investigate this hypothesis formally by typing:

**ttest conspiracies, by(democratic\_legacy)**

The “by” option tells Stata that we are performing this test between the different categories of the variable “democratic\_legacy”. Check that you understand all the sections of this command, then look at the output. This time, the differences are expressed in terms of differences between the groups. The difference is statistically significant Pr(T > t) = 0.00

Open and run the do file, “Problem\_Set\_2\_dofile.do” that is in the Problem Set directory on Moodle (same folder as this document).

It will generate and code a variable called “developing,” which is a dummy variable, coded to 0 if a region has GDP per capita at PPP *above* $30,000, and 1 if that region has GDP per capita at PPP *below* that level. (Because this variable is coded at the regional level, many middle-income countries contain a mix of both “developed” and “developing” regions).

Add an “if” statement to perform this test for low development regions (variable: **developing==1**) countries and then for developed regions (**developing==0**). What do you find?

|  |
| --- |
|  |