Title: Homework 2 (No official title - Working title will be based on the dataset chosen)

The task is to perform some simple (bivariate) tests of inferential statistics. You can use the database you created for homework assignment 1, or any other database containing at least 3 variables and at least 30-50 observations. The idea is to run one or two simple tests, you can decide which type of tests to use (e.g. cross tabs + Chi2 tests, comparisons of means, correlation, regression analysis). You will need to choose which method suits your purposes best. You should start with explicitly stating the question(s) you want to investigate. Then you should state the working and null hypothesis of your question(s). You should choose the right form of statistical test, perform the test and interpret the findings (size of effects, standard errors, statistical significance/ error probability). Please also state what you have to conclude on the basis of the hypothesis/es posed. You should interpret these findings not only in terms of statistical significance, but substantive size. Finally, you should conclude by arguing how one could possibly criticize your finding and/ or how you would refine the investigation method (more data, other tests, other variables…).

There is no precise number of tests I suggest. Depending on your data, it might make sense to perform different tests (e.g. one correlation, one regression; or several different regressions). In most cases more than one or two tests does not make sense, because it takes too much time to explain the results. This implies that you should also think about which tests you want to present.

Please also check carefully the assumptions of each test. For regression analysis, for instance, you don't have to go through all the assumptions of regression analysis (so-called regression diagnostics), but you should show us that you aware of some of the problems and perhaps illustrate them using your dataset. Knowing the limits of a method is very important. For instance, if you choose a test, but you feel or know there might be a more suitable one (which we didn't cover in class/ the labs) that's not a problem. Use the test you know, but flag out that there might be better alternatives available (which and why are these better).

Those with advanced knowledge of statistics should feel free to use more sophisticated models or different tests. Be aware that it will then be required to understand these advanced techniques sufficiently well. Don't try to show off. It is better to use simple tests you understand well, than very sophisticated tests you don't really know how to operate. Everyone can click buttons in a computer software. It is knowing which button to click and what to make of the results which matters. For all other students with less statistical training simple bi-variate tests are enough. If you do them very well, you will receive full grade.

Basically, the task is to write a paper with any hypotheses based on the dataset provided (see the attachement, but you can use other dataset if you want from e.g. World Bank, EuroStat, UN, etc). Paper should have at least 3 variables and 28-30 variables (e.g., GDPpc PPP, Literacy rate, etc for 28-30 or more countries)(provided dataset has this all). It is important to create at least 2 graphs/charts/scatter boxes/diagrams with detailed explanation using the inferential statistics terms. Such software like STATA/SPSS/EXCEL should be used in creation of the charts and graphs (any of this).

Please find attached the requirements for my quantitative homework and a sample of previously accomplished homework, which of course should not be 100% followed in the new work.