

Computer Assignment

ECO 4000 Statistical Analysis for Economics and Finance

23 April, 2022

Maximum Points : 25

Due Date : 4 May, 2022 (Before Class)

What are the root causes of terrorism? Poverty? Repressive political regimes? Religious or ethnic conflicts arising from heterogeneous populations? In this assignment you will take a look at some empirical evidence on cross-country sources of terrorism. Variables in the data set, **terrorism.dta**, are defined in the table below.

Variable	Definition
ftmpop	Number of fatalities from terrorist incidents in the country, 1998-2004, per million population (U.S. State Department)
evmpop	Number of terrorist events in the country, 1998-2004, per million population (U.S. State Department)
gdppc	GDP per capita in the country (World Bank)
lackpf	Index of the lack of political freedoms (Freedom House), 1-7 scale, 7 = extremely limited political freedoms
language	Index of linguistic fractionalization (0 to 1 scale, 0 = no fractionalization)
ethnic	Index of ethnic fractionalization (0 to 1 scale, 0 = no fractionalization)
religion	Index of religious fractionalization (0 to 1 scale, 0 = no fractionalization)
mideast, latinam, easteuropa, africa, eastasia	= 1 if the country is in the indicated region, = 0 otherwise

To do this problem set, you will need to create (generate) some new variables, which are functions of the variables in **terrorism.dta**. You should restrict your analysis to the observations for which there are data on GDP per capita (i.e. for which `gdppc` is non-missing) and for which there are terrorist fatalities (i.e. for which `ftmpop` > 0). [R Hint: Use **filter()** from `dplyr` package]

Part 1. Preliminary Data Analysis (4 points)

- Produce the scatter plot of `ftmpop` vs. `gdppc`. (Keep in mind, it is always Y vs X when you draw a plot. So, `ftmpop` will be on Y-axis and `gdppc` will be on X-axis) (1 point)
- Create the variables `lnftmpop = log(ftmpop)` and `lngdppc = log(gdppc)`. Produce the scatter plot of `lnftmpop` vs. `lngdppc`. [R Hint: use the command **`mutate()`** to create new columns.] (1 point)
- Produce the scatter plot of `lnftmpop` v. `lackpf`. (1 point)
- Using the scatter plots from (a) and (b), would you suggest using the variables (i) `ftmpop` and `gdppc` or (ii) `lnftmpop` and `lngdppc` for modeling using linear regression? (1 point)

Part 2. Estimate the regressions in given in the table and fill in the empty entries including R^2 , \bar{R}^2 , and n . (12 points) (2 points for each regression + 1 point for creating each of the required new variables)

R Hint: to create the square of a variable, for example `lackpf`, use the command **`mutate(lackpf2 = lackpf^2)`**. Also, instead of filling in the empty entries, I would just create a result table using **`stargazer`** package. The code for that could be found on our website. But of course, it will be your choice. But you must show the codes of regressions that you run.

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	<i>lnftmpop</i>	<i>lnftmpop</i>	<i>lnftmpop</i>	<i>lnftmpop</i>	<i>lnftmpop</i>
Regressor:					
<i>lngdppc</i>	()	()	()	()	()
<i>(lngdppc)²</i>	—	—	()	—	—
<i>lackpf</i>	—	()	()	()	()
<i>lackpf²</i>	—	—	()	()	()
<i>ethnic</i>	—	—	—	()	()
<i>religion</i>	—	—	—	()	()
<i>Mideast</i>	—	—	—	—	()
Other regional dummies (<i>latinam</i> , <i>easteurope</i> , <i>africa</i> , <i>eastasia</i>)?	No	No	No	No	Yes
\bar{R}^2					
R^2					
n					

The “other regional dummies” included in regression (5) are *latinam*, *easteurope*, *africa*, and *eastasia* (the omitted case is Western Europe combined with North America).

Part 3. Answer the following questions:

- a. Using regression (1), test the hypothesis that the coefficient on *lngdppc* is zero, against the alternative that it is nonzero, at the 5% significance level. Explain in words what the coefficient means (i.e. interpret the coefficient of interest here). (2 points)
- b. Using regression (3), test the hypothesis that the coefficients on *lngdppc* and *lngdppc*² are both zero, against the alternative that one or the other (or both) coefficient is nonzero, at the 5% significance level. (show the code) [R Hint : use **linearHypothesis()** function for f-test from **car** package. See an example **here**] (2 points)
- c. Explain why the coefficient on *lngdppc* in regressions (1) and (3) differ. (2 points)
- d. Using regression (5), test the null hypothesis (at the 5% significance level) that the coefficients on the “other regional dummies” all are zero, against the alternative hypothesis that at least one is nonzero. (show the code) What is number of restrictions *q* in your test? What is the critical value of your test? (3 points)

Bonus Question

Using regression (4), estimate the effect on *lnftmpop* of changing from *lackpf* = 7 (extremely limited political freedoms) to *lackpf* = 5 (some political freedoms), holding constant the values of the other regressors in regression (4). (5 points)