Replication assignment

Johannes Karreth

RPOS 517

Purpose and objectives

Replicating (or more precisely, reproducing) other scholars' work is a key element of the scientific process. Replications can help scholars assess the validity of previous studies. They can also help identify important qualifications of previous findings or expose limitations of extant work. To engage with quantitative social scientific studies, you will replicate (reproduce) a study of your choice or from a list of suggestions using the methods you are learning in our course. This assignment will also give you some insight on how to conduct your own data analysis. You will need to obtain data, describe them, and analyze them in a manner suited for publication. At a minimum, this assignment will give you the tools to complete all necessary steps to use quantitative methods to answer a research question. In an ideal case, a replication will give you the foundation for an opportunity to improve upon existing work and write your own publishable research paper. For more benefits of replication assignments, read, for example, Gary King's article *Replication, Replication* at `http://gking.harvard.edu/files/replication.pdf` and a blog post at `http://wp.me/p315fp-jD`.

Steps of this assignment

By **noon on February 16**, you need to identify a scholarly article from a political science journal that uses quantitative methods (from bivariate associations to multiple linear regression) and for which replication data is publicly available. After you sign up for the article via the sign-up sheet on Blackboard, you will complete the following steps and turn in your final replication project by the beginning of class on week 12:

1. Retrieve the replication data for the article
2. Write an outline of your replication plan (template provided)
3. Write a replication script
4. Conduct the replication analysis of the main model in the article
5. Complete a replication memo, summarizing your findings (template provided)

Finding an article

You need to identify a published article that uses methods that are covered in our seminar. This includes bivariate associations such as t-tests, cross tabs, and bivariate and multiple linear regression. The article should interest you so that you can motivate yourself working on it, perhaps beyond this seminar.

The article also needs to have replication data available. You need to check whether this is the case before you commit to an article. If the data used in the article is not available directly as replication data or via publicly and easily accessible data source, you cannot use the article for this assignment.

This is the most important step of this assignment, so please be sure to reach out to me if you are unsure about any of the aforementioned criteria. On **February 13**, I will provide a list with potential replication articles that you may choose from. Each article can be chosen by only one student. Most articles on that list address political science questions. I encourage you to search for articles in your specific field of interest before February 13.

Signing up for an article

You will sign up for an article using a spreadsheet linked on Blackboard. This spreadsheet contains 19 potential replication articles that you can sign up for (first come, first serve!). Enter your full name in the *Claimed by* column. If you found another article that you would like to replicate, just enter the article info on the bottom of the spreadsheet and put your name next to it. The due date to claim an article is by **noon on February 16**.

Retrieve replication data

By **5pm on Wednesday (February 18)**, you need to send me the replication data for your article in comma-separated values (CSV) format. You learn the tools to create a comma-separated values spreadsheet on Day 2. As a reminder, the sequence you need to follow is:

1. Read your data into R. Depending on the format of the replication data you retrieved, you may have to use commands such as `read.dta()`, `read.csv()`, etc.
2. Once the dataset is in R, use the `write.csv()` command to create a CSV file. Be sure to set the `row.names` option to `FALSE`: `row.names = FALSE`.

An example for this sequence is below. This example assumes that you have a Stata dataset, "`articledata.dta`", saved in your working directory.

```
library(foreign)
replication.dat <- read.dta(file = "articledata.dta")
write.csv(x = replication.dat, file = "myreplicationdata.csv", row.names = FALSE)
```

If you receive the error: `Error in read.dta(file = "articledata.dta") : not a Stata version 5-12 .dta file`, you can use an alternative command to read a Stata dataset into R. For that, you first need to install the package "`readstata13`" the usual way. Then:

```
library(readstata13)
replication.dat <- read.dta13(path = "articledata_Stata13.dta")
write.csv(x = replication.dat, file = "myreplicationdata.csv", row.names = FALSE)
```

Outline of your replication plan

By **noon on Monday, February 23**, you need to send me the outline of your replication plan as a PDF file. This outline needs to use the template linked on the course website. It needs to contain all elements listed in the template.

Replication script

Between Days 4 and 8, you will be learning all the techniques necessary to replicate "your" article. After our meeting on Day 8, you need to write up a complete replication script for your assignment. This file is an R-script; it contains only R code and your comments. Write this document in a way that you will be able to understand it yourself two years from now: comment frequently and explain the steps you are performing. This script needs to use the template linked on the course website. It is due on Blackboard by **noon on Monday, March 30**.

Replication analysis

You should conduct your replication analysis between March 30 and April 13. This analysis executes your replication script and fixes potential errors and glitches in your original script.

Replication memo

Between April 13 and April 20, complete your replication memo, including a table and/or figure that presents the key results of the replication. A template is provided on the course website. This memo builds on the replication plan you submitted on February 18. The memo is due on Blackboard by **noon on Monday, April 20**.

After completion

You will receive feedback from me during and after completion of your memo.