Problem Set #4

EH6127 - Quantitative Methods

Steven V. Miller

This homework makes use of data available in `{stevedata}` and implies the use of `{tidyverse}` to answer the questions. `{tidyverse}` is not necessary to answer these questions though it will assuredly make the process easier. Load these two libraries to get started answering these questions.

```
library(tidyverse)
library(stevedata)
```

Comparative Public Health: The Political Economy of Human Misery and Well-Being

This homework will refer to the `GHR04` data set that is available in `{stevedata}`. This data set is capable of (almost perfectly) recreating the analyses done by Ghobarah et al. (2004).[1] You can find out more information about the data by visiting this part of the package's website, or with the following command.

```
?GHR04
```

Here's a little preview of these data.

```
GHR04
```

```
## # A tibble: 182 x 15
##    country       iso3c pubhlthexppgdp totexphlth  hale log_gdppc  gini log_educ
##    <chr>         <chr>          <dbl>      <dbl> <dbl>     <dbl> <dbl>    <dbl>
##  1 StLucia       LCA             2.62       5.44  62       NA    0.41     1.68
##  2 StKitts-Nevis KNA             3.11       6.22  59.6     NA    0.474    1.86
##  3 C-Kinshasa    COD             0.0318     3.81  34.4      6.57 0.441    1.21
##  4 Georgia       GEO             0.377      5.09  58.2      7.7  0.319    2.34
##  5 China         CHN             0.663      4.84  62.1      8.19 0.389    1.88
##  6 Sierra Leone  SLE             0.469      3.14  29.5      6.27 0.609    0.945
##  7 India         IND             0.672      4.71  52        7.45 0.356    1.61
##  8 Vietnam       VNM             0.967      4.50  58.9      7.48 0.346    2.12
##  9 Indonesia     IDN             0.622      4.36  57.4      7.95 0.349    1.70
## 10 Pakistan      PAK             0.919      4.19  48.1      7.6  0.299    1.54
```

[1] Ghobarah, Hazem Adam, Paul Huth, and Bruce Russett. 2004. "Comparative Public Health: The Political Economy of Human Misery and Well-Being" *International Studies Quarterly* 48: 73-94. The authors make these data available but not a script that is more explicit about what exactly they are doing. Thus, the replication here is basically total but not perfect/identical. Students are responsible for reading this article in order to make sense of what is being asked in this problem set.

```
## # i 172 more rows
## # i 7 more variables: log_vanhanen <dbl>, rivalry <dbl>, polity <dbl>,
## #   prvhlthexpgdp <dbl>, urban_growth <dbl>, cwdeaths <dbl>, contig_cw <dbl>
```

Answer these questions. This particular homework may seem brutal because it will demand that you read the Ghobarah et al. (2004) article (see footnote citation) and look *carefully* at the codebook for this data set.

1. (2 POINTS) Using the data set provided to you, and the `lm()` function in base R, reproduce Table I (for which public health expenditures as a percent of GDP is the dependent variable). A successful answer will involve both code and console output.
2. (2 POINTS) Using the data set provided to you, and the `lm()` function in base R, reproduce Table II (for which total expenditures on health is the dependent variable). A successful answer will involve both code and console output.
3. (2 POINTS) Using the data set provided to you, and the `lm()` function in base R, reproduce Table III (for which health-adjusted life expectancy is the dependent variable). A successful answer will involve both code and console output.
4. (2 POINTS) Answer one of the following prompts:
    a. For the regression model on health-adjusted life expectancy, re-run the model (but omit Rwanda from the analysis). Show me the code you used to do this and highlight any differences in sign/significance you see comparing this model to the model above.
    b. In the regression model you estimated for Question 3, the intercept is "statistically significant." What is that value for the intercept actually communicating? Does it make sense?