Lab 15 - Multivariate Regression & Interpretation

Your name here

November 30, 2017

Complete the following exercises below and include all code used to find the answers. Knit together the PDF document and commit both the Lab 15 RMD file and the PDF document to Git. Push the changes to GitHub so both documents are visible in your public GitHub repository.

```
library(ggplot2)
library(tidyverse)
library(readxl)

gtd_filter <- c("country_txt", "nkill", "iyear", "imonth", "nwound", "success")
gtd_full <- read_csv("GTD FULL DB.csv")
IMF_Real_GDP_Growth <- read_excel("IMF Real GDP Growth.xls")


#Set the ranges for the data using filter
plotTerror <- filter(gtd_full, country_txt == c("Nigeria", "Niger", "Chad", "Cameroon"))
plotTerror <- select(plotTerror,gtd_filter)

#Fix the GDP growth data
temp <- gather(IMF_Real_GDP_Growth, key = Year, value = RGDP, ...=- X__1)

#Rename columns to setup keys for join
names(temp)[names(temp) == 'X__1'] <- 'Countries'
names(plotTerror)[names(plotTerror) == 'country_txt'] <- 'Countries'
names(plotTerror)[names(plotTerror) == 'iyear'] <- "Year"

#Join the two datasets

joinT <- merge(plotTerror, temp, by = c("Countries", "Year"))


#Create new catagory - totalcas
joinT[is.na(joinT)] <- 0
joinT <- mutate(joinT, totalcas = (nkill + nwound))
test <- joinT %>%
  select(-nwound) %>%
  group_by(Countries, Year) %>%
  mutate(TOTALCAS = sum(totalcas)) %>%
  select(-nkill, -totalcas) %>%
  distinct() %>%
  ungroup() %>%
  mutate(Year = as.character(Year)) %>%
  right_join(temp, by = c("Countries", "Year"))

test[is.na(test)] <- 0
test <- select(test, -RGDP.x)
test <- arrange(test, Countries, Year)
test$Year <- as.numeric(as.character(test$Year))
```

```
nigeriaTest <- filter(test, Countries == "Nigeria")
nigerTest <- filter(test, Countries == "Niger")
chadTest <- filter(test, Countries == "Chad")
cameroonTest <- filter(test, Countries == "Cameroon")
```

1. Select a second explanatory variable from your dataset that you think has implications for the theoretical association of your focal relationship.

```
testmod <-lm(RGDP.y ~ TOTALCAS, factor(Year), data = test)
summary(testmod)
```

```
##
## Call:
## lm(formula = RGDP.y ~ TOTALCAS, data = test, subset = factor(Year))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6447 -0.4398  0.7602  1.7654  1.8602
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.739792   0.142106  26.317  < 2e-16 ***
## TOTALCAS    0.006168   0.001468   4.201 3.44e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.447 on 326 degrees of freedom
## Multiple R-squared:  0.05135,    Adjusted R-squared:  0.04844
## F-statistic: 17.64 on 1 and 326 DF,  p-value: 3.44e-05
```

a. Describe the theoretical reasoning for selecting this variable.

Following previous guesses about casualty counts, this model attempts to find an association between the RGDP and TOTALCAS by taking the years into account.

b. What type of relationship do you think this variable has with your focal variables? Given that, what do you expect to happen to your focal relationship when it is added to the model?

I was hoping that adding Year into the variables would help solve the issue of how complex GDP is.

c. Is it a continuous or categorical variable? What implications does this have for a multivariate regression equation?

It's a continous variable, which is probably what is causing me issues. Using time leads to spurious results.

d. Conduct a multivariate linear regression with this additional explanatory variable and save the model object. Print out the full results by calling `summary()` on your model object.

e. Describe the results of the multivariate analysis, highlighting:

- the apparent association between the control variable and the focal response variable
- how the focal association changed when you incorporatedthe control variable
- the implications of these results for your focal association

f. How well does this model fit the data? Is it an improvement over the bivariate model? Why or why not?

The model doesnt' fit very well, however it fits much better than the bivariate did.

2. Select any additional variables you want to incorporate into your final model. For each additional variable added to the model answer the following questions:

I have no other variabels I intend to use.