Assignment 2 POLI SCI

Faizan Khalid Mohsin

February 25, 2018

## R Markdown

#### Q (a) Recodethevariablemeasuringthevotechoiceinthepresidentialelection(V162034a) bykeepingonlythecategoriesforthetwomajorcandidates—HillaryClintonandDonaldTrump—andbysettingtheothercategoriesasmissingvalues. PrinttheRcodeyou haveused.

require(dplyr)

## Loading required package: dplyr

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

data1 = read.csv("anes2016.csv")  
  
# str(data1)  
# names(data1)  
# dim(data1)  
# data1$V162034a[1]  
# head(data1)  
# class(data1)  
# donald = data1$V162034a[1]  
# ifelse(c("a", "b", "abd", "a", "a", "cdd") %in% c("a", "b"), c("a", "b", "abd", "a", "a", "cdd"), NA)   
  
newcan\_data1 = mutate(data1, new\_candidates =   
 as.factor(ifelse(as.character(data1$V162034a) %in% c("1. Hillary Clinton", "2. Donald Trump"), as.character(data1$V162034a), NA)))  
  
# str(newcan\_data1)

#### Q (b) Using a contingency table, examine the association between the variable V162148, which measures whether respondents support government efforts to reduce income inequalities, and the vote choice variable created in 1(a). Treat the vote choice as the dependent variable. Report the conditional distributions and compute a significance test that is appropriate in this context. Report your results and the R code you have used.

require(knitr)

## Loading required package: knitr

# Produce the contingency table.   
  
contingency\_table = table(newcan\_data1$new\_candidates, newcan\_data1$V162148)  
  
kable(contingency\_table, caption = "Table 1. Respondents support of government efforts to reduce income inequalities per candidate")

Table 1. Respondents support of government efforts to reduce income inequalities per candidate

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1. Favor | 2. Oppose | 3. Neither favor nor oppose |
| 1. Hillary Clinton | 760 | 195 | 299 |
| 2. Donald Trump | 257 | 610 | 291 |

# Find conditional probabilities:   
  
#Conditional Distribution for Donald Trump  
contingency\_table[1,]/sum(contingency\_table)

## 1. Favor 2. Oppose   
## 0.31509121 0.08084577   
## 3. Neither favor nor oppose   
## 0.12396352

#Conditional Distribution for Hillary Clinton  
contingency\_table[2,]/sum(contingency\_table)

## 1. Favor 2. Oppose   
## 0.1065506 0.2529022   
## 3. Neither favor nor oppose   
## 0.1206468

# Do the Chi-square test  
  
chisq.test(contingency\_table)

##   
## Pearson's Chi-squared test  
##   
## data: contingency\_table  
## X-squared = 459.74, df = 2, p-value < 2.2e-16

summary(contingency\_table) # also does the job.

## Number of cases in table: 2412   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 459.7, df = 2, p-value = 1.475e-100

#### Q 1 (c) In a few sentences, interpret the findings obtained in 1(b). Briefly explain the nature ofthestatisticsyouhaveused,anddiscussyourresultsinsubstantiveterms.

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

#### Q 1 (d) The variable V161342 is the gender of the respondent. Compute a 95% confidence interval around the difference between the proportion of women who voted for Donald Trump and the proportion of men who voted for him.

#### Q 1 (e) In one or two sentences, interpret the result obtained in 1(d).

### 2. These second problem requires downloading a section of the Quality of Government Dataset (qog\_exercise.csv) available on Blackboard. The codebook describing variables is also included. (6points)

#### Q 2 (a) Create a binary variable that equals 1 if a country is a democracy, and 0 otherwise, using the regime variable named *ht\_regtype1*. Reportyourcode.

raw\_data2 = read.csv("qog\_exercise.csv")  
str(raw\_data2)

## 'data.frame': 194 obs. of 5 variables:  
## $ ccode : int 4 8 12 20 24 28 31 32 36 40 ...  
## $ cname : Factor w/ 194 levels "Afghanistan",..: 1 2 3 4 5 6 11 7 9 10 ...  
## $ ht\_regtype1 : int 4 100 4 NA 4 100 4 100 100 100 ...  
## $ voh\_gti : num 8.74 0.13 5.3 NA 1.35 ...  
## $ wdi\_expmilgdp: num 1.14 1.49 4.46 NA 3.3 ...

data2 = raw\_data2 %>% mutate(Democratic = ifelse(ht\_regtype1 == 100, 1, 0)) # what about NA values. Here we assume if and only if ht\_regtype1 is 100, it is a Democratic Country.  
   
str(data2)

## 'data.frame': 194 obs. of 6 variables:  
## $ ccode : int 4 8 12 20 24 28 31 32 36 40 ...  
## $ cname : Factor w/ 194 levels "Afghanistan",..: 1 2 3 4 5 6 11 7 9 10 ...  
## $ ht\_regtype1 : int 4 100 4 NA 4 100 4 100 100 100 ...  
## $ voh\_gti : num 8.74 0.13 5.3 NA 1.35 ...  
## $ wdi\_expmilgdp: num 1.14 1.49 4.46 NA 3.3 ...  
## $ Democratic : num 0 1 0 NA 0 1 0 1 1 1 ...

#data2$ht\_regtype1

#### Q 2 (b) Using the binary variable created in 2(a), use a significance test to assess whether the mean of the global terrorism index (named voh\_gti) is lower in democracies compared to non-democratic regimes. The global terrorism index ranges from 0 to 10, where higher values indicate that a country is more heavily impacted by terrorism.

# Could use ANOVA or Two Sample t-test  
# Method 1: use t-test first:  
  
# Helper function to create the two samples vectors on which we will perform the two sample t-test.   
  
create\_sample\_vectors = function(data, var\_of\_interest, var\_to\_filter, value1, value2){  
   
 #data: a data frame  
 #var\_of\_interest: a character string with the name of the variable of which we # will compare the means in the t.test  
 #var\_to\_filter: a character string with the name of the variable by which to   
 # group by  
 #value1: the number 1 number which indicates all the democratic countries  
 #value2: the number 0 which indicates all the nondemocratic coutries.   
   
 dem\_terror = subset(data, var\_to\_filter == value1)[, var\_of\_interest]  
 nondem\_terror = subset(data, var\_to\_filter == value2)[, var\_of\_interest]   
   
return(c(dem\_terror, nondem\_terror))  
   
}  
  
#demterror = create\_sample\_vectors(data2, voh\_gti, Democratic, 1, 0)[1]  
#nondemterror = create\_sample\_vectors(data2, voh\_gti, Democratic, 1, 0)[2]  
  
  
dem\_terror = subset(data2, Democratic == 1)[, "voh\_gti"]  
nondem\_terror = subset(data2, Democratic == 0)[,"voh\_gti"]  
  
t.test(dem\_terror, nondem\_terror)

##   
## Welch Two Sample t-test  
##   
## data: dem\_terror and nondem\_terror  
## t = -2.4271, df = 150.43, p-value = 0.0164  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.6207203 -0.1661066  
## sample estimates:  
## mean of x mean of y   
## 1.643462 2.536875

# #### Using dplyr package to do the same thing as above.   
# dem\_terror\_index = data2 %>% filter(Democratic==1) %>% select(voh\_gti)   
# nondem\_terror\_index = data2 %>% filter(Democratic==0) %>% select(voh\_gti)   
# t.test(dem\_terror\_index, nondem\_terror\_index)  
  
  
# # Method 2: Use Anova  
#   
# # First test the assumptions of ANOVA  
#   
# bartlett.test(voh\_gti ~ Democratic, data = data2) # Homogeneity of Variances  
# # Not homegenous.  
#   
# qqnorm(dem\_terror, ylab = "Sample Quantiles for Democratic Country Terror Index") ; qqline(nondem\_terror, col = 2) # Multivariate normality test  
# # Not normally distributed  
#   
# #Should not use anova test. Should use nonparametric equivalent which the T-test below does automatically.   
# result\_terror = aov(voh\_gti ~ Democratic, data = data2)  
# summary(result\_terror)

Next, perform a similar test for the mean level of military expenditures as a percentage of the gross domestic product (the variable named wdi\_expmilgdp). Make sure to report all your steps and the R code you have used.

### This wdi\_expmilgdp is the military expenditure as a percentage of the gdp. ### Hence should not be between the values of 0 and 1. Or is this between 0 and ### 100. So if a country has the value of 3.4 it means their military expenditure ### as a percentatge of the gdp is 3.4%  
  
dem\_military\_exp = data2 %>% filter(Democratic==1) %>% select(wdi\_expmilgdp)  
nondem\_military\_exp = data2 %>% filter(Democratic==0) %>% select(wdi\_expmilgdp)  
  
t.test(dem\_military\_exp, nondem\_military\_exp)

##   
## Welch Two Sample t-test  
##   
## data: dem\_military\_exp and nondem\_military\_exp  
## t = -3.4297, df = 97.354, p-value = 0.0008876  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.4501096 -0.3870335  
## sample estimates:  
## mean of x mean of y   
## 1.478267 2.396839

dem\_terror = subset(data2, Democratic == 1)[, "wdi\_expmilgdp"]  
nondem\_terror = subset(data2, Democratic == 0)[,"wdi\_expmilgdp"]  
  
t.test(dem\_terror, nondem\_terror)

##   
## Welch Two Sample t-test  
##   
## data: dem\_terror and nondem\_terror  
## t = -3.4297, df = 97.354, p-value = 0.0008876  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.4501096 -0.3870335  
## sample estimates:  
## mean of x mean of y   
## 1.478267 2.396839

#### Q 2 (c) In a few sentences,interpret the outcome of both tests conducted in 2(b).

From the tests above we see that non-democratic countries have higher terrorism and also spend more on their military expenditure as a percentage of their gdp compared to democratic countries that have statistically significant less terrorism and also statistically significantly spend less on military expenditure as a percentage of their gdp.