W203 Lab 3

Mohammad Jawad Habib

April 3, 2016

Part 1: Multiple Choice

Q1 : b (Natural Experiment) Q2 : b, c Q3 : b Q4 : c Q5 : d, needs research Q6 : a, confirm wording of question Q7 : b Q8 : b

Part 2: Test Selection

Q9 : e, Chi-square test Q10: d, ANOVA Q11: a, t-test Q12: b, Pearson correlation Q13: d, Chi-square test

Part 3: Data Analysis and Short Answer

```
# setwd("W203 Week 12/Lab 3")
load("GSS.Rdata")
```

14. Task 1: Chi-Square Test on Marital Status and Political Orientation

A. Null and Alternative Hypothesis

H0: marital status and political orientation are independent H1: marital status and poliical orientation are not independent (that is, knowning about marital status can help predict the political orientation).

We can assume that the various marital statuses are independent of each other i.e. one person is not married and widowed at the same time. We can assume independence for political categorization as well.

We will also see below that none of the expected frequencies are below 5.

```
mp <- data.frame(GSS$marital, GSS$politics)

# remove the value coded as "NA" from "marital"
mp$GSS.marital[mp$GSS.marital == "NA"] <- NA

# remove "NA" as a factor from marital
mp$GSS.marital <- factor(mp$GSS.marital,
                         c("married", "widowed", "divorced", "separated", "never married"))

# remove all rows that have NA for "marital" or "politics"
mp <- mp[complete.cases(mp),]

# run the Chi Square test
mpcs <- chisq.test(table(mp))

# see that no expected frequencies are below 5
mpcs$expected
```

```
##                  GSS.politics
## GSS.marital         Liberal   Tend Lib  Moderate  Tend Cons Conservative
##   married       102.391123 102.924411 281.04230 132.255201   150.386963
##   widowed        19.839112  19.942441  54.45423  25.625520    29.138696
##   divorced       27.162275  27.303745  74.55479  35.084605    39.894591
##   separated       5.192788   5.219834  14.25312   6.707351     7.626907
##   never married  37.414702  37.609570 102.69556  48.327323    54.952843
```

B. Test Statistics and p-value

We get a test statistic and p-value as follows:

```r
# test statistic
mpcs$statistic
```

```
## X-squared
##   44.2255
```

```r
# p-value
mpcs$p.value
```

```
## [1] 0.0001822704
```

Given the p-value above we can reject the null hypothesis. We can say that marital status does seem to be related to political orientation.

C. Effect Size Calculation

We will use Cramer's V for the effect size.

```r
cv <- sqrt(mpcs$statistic / (length(mp$GSS.marital)*min(nrow(mpcs$observed) - 1, ncol(mpcs$observed) - 
names(cv) <- "Cramer's V"
cv
```

```
## Cramer's V
## 0.08756363
```

D. Interpretation

Our Chi-Square test reveals that "politics" is significantly related to "marital" status:

$$\chi^2(16) = 44.225, p < 0.01$$

. The contingency table is shown below.

```r
mpcs$observed
```

```
##                  GSS.politics
## GSS.marital     Liberal Tend Lib Moderate Tend Cons Conservative
##   married            93       92      271       140          173
##   widowed            15       16       57        24           37
##   divorced           22       36       79        38           29
##   separated           7        3       22         6            1
##   never married      55       46       98        40           42
```

We can also see from the standardized residuals which ones are significant outside of +/- 1.96 ($p < 0.05$).

```r
mpcs$stdres > 1.96 | mpcs$stdres < -1.96
```

```
##                  GSS.politics
## GSS.marital     Liberal Tend Lib Moderate Tend Cons Conservative
##   married         FALSE    FALSE    FALSE     FALSE         TRUE
##   widowed         FALSE    FALSE    FALSE     FALSE        FALSE
##   divorced        FALSE    FALSE    FALSE     FALSE         TRUE
##   separated       FALSE    FALSE     TRUE     FALSE         TRUE
##   never married    TRUE    FALSE    FALSE     FALSE         TRUE
```

Being "married" is significantly related with "Conservative" political view. "widowed" is not significantly related with any political view. "divorced" is significantly related with "Conservative". "separated" is significantly related with "Moderate" and "Conservative". "never married" is significantly related with "Liberal" and "Conservative".

15. Task 2: Pearson Correlation on Age when Married and Hours of TV watched

A. Null and Alternative Hypothesis

H0: there is no relationship between agewed and tvhours ($r = 0$) H1: there is a positive or negative relationship between agewed and tvhours

B. Test Statistics and p-value

GSS website is not very clear on how missing values in `agewed` and `tvhours` are coded. For `agewed` let's assume that 0 and 99 are dummy values. for `tvhours` we can assume that anything greater than 24 is a dummy value because there are only 24 hours in the day.

Note: our sample size is large so we can assume normality of our sampling distribution. This assumption is required for establishing whether correlation coefficient is significant.

```r
at <- data.frame(GSS$agewed, GSS$tvhours)
at <- at[!(at$GSS.agewed %in% c(0, 99)) & at$GSS.tvhours <= 24,]

library(Hmisc)

atpc <- rcorr(as.matrix(at))
atpc
```

```
##             GSS.agewed GSS.tvhours
## GSS.agewed        1.00       -0.03
```

```
## GSS.tvhours      -0.03        1.00
##
## n= 1194
##
##
## P
##             GSS.agewed GSS.tvhours
## GSS.agewed             0.3009
## GSS.tvhours 0.3009
```

```r
cor.test(at$GSS.agewed, at$GSS.tvhours)
```

```
##
##  Pearson's product-moment correlation
##
## data:  at$GSS.agewed and at$GSS.tvhours
## t = -1.0349, df = 1192, p-value = 0.3009
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.08654554  0.02681630
## sample estimates:
##        cor
## -0.02996096
```

From the above, we get a test statistic (r) of -0.03 and a p-value of 0.3009. And we get a 95% confidence interval that passes through zero (-0.087, 0.027)

C. Interpretation

The results indicate that agewed is not correlated with tvhours. We also get

$$r^2 = 9e - 04$$

which means that only 0.09% of variance is shared between agewed and tvhours.