LAB 3

PART I.

1. b) Natural experiment
2. a) Pretest-postest experimental design
3. b) Parametric tests typically have higher statistical power than non-parametric test when parametric assumptions are met.
4. c) Dependent-sample t-test
5. d) 42
6. c) It depends on your theory and how you came to observe this difference in means.
7. b) External validity
8. b) Difference of means, divided by the pooled standard deviation of the variable.

PART II.

9. e) Chi-square test
10. d) ANOVA
11. a) t-test
12. b) Pearson correlation
13. d) Chi-square test

PART III.

14.

a)

Null hypothesis: There is no relationship between marital status and political orientation (the portions of political orientation are the same across different marital statuses).

Alternative hypothesis: There is some relationship between marital status and political orientation (the portions of political orientation are not the same across different marital statuses).

b)

All expected values are greater than 5, so we do not need to do a Fisher's exact test. My chi-squared value is 44.2255 and my p-value is 0.0001822704.

c)

I calculated a Crammer's V of 0.08756363 (a weak effect).

d)

I reject my null hypothesis that there is no relationship between marital status and political orientation ($\chi^2 = 44.2255$, $p < .001$) but understand that this effect is so practically weak (*Crammer's V = 0.08756363*) that it is likely uninteresting.

15.

a)

Null hypothesis: There is no relationship between age when married and hours of TV watched.

Alternative hypothesis: There is some relationship between age when married and hours of TV watched.

b)

I find that the age when married is significantly related to the hours of TV watched ($r = 0.11$, $p < .001$).

c)

I reject my null hypothesis that there is no relationship between age when married and hours of TV watched. The age when married was significantly and positively related to the hours of TV watched ($r = 0.11$, $p < .001$). However, I understand that this relationship has small practical significance.

16.

a)

The mean of my new variable (married) is 0.2857143, indicating that 28.57% of my sample of 23 years old are married.

b)

Null hypothesis: There is no association between being married (*married=1*) and number of children (*childs*) for respondents who are 23 years old.

Alternative hypothesis: There is an association between being married (*married=1*) and number of children (*childs*) for respondents who are 23 years old.

c)

My test statistic, W, is 19 and my p-value is 0.0002656.

d)

I converted the z-score into an effect size estimate, r. My r value was -0.6891632, indicating a large effect.

e)

I reject my null hypothesis that there is no association between being married and number of children for 23 year olds, $W = 19$, $p < .001$. This is a large effect, $r = -.689$. I recommend that future future studies look at my data and test a directional hypothesis on newly collected data.

17.

a)

Null hypothesis: There is no association between religious affiliation (relig) and age when married (agewed).

Alternative hypothesis: There is an association between religious affiliation (relig) and age when when married (agewed).

b)

My F-value is 8.197 and my p-value is 1.61e-06.

c)

Yes, there are statistically significant differences between pairs of groups. We perform a pairwise t test with a Bonferroni correction and learn that there are statistically significant differences between Catholics and Protestants, Jews and Protestants, and Others and Protestants.

d)

I reject my null hypothesis that there is no association between religious affiliation and age when married, $F = 9.197$, $p < .001$. This statistically significant association between religious affiliation and age when married is practically weak, $\omega^2 = 0.0235$.

APPENDIX.

```
# Alex Smith
# W203 - Lab 3

# set the working directory
setwd("~/Documents/MIDS/Spring14/W203/Lab3")

# load the dataset & examine the first few rows
load("GSS.Rdata")
head(GSS)

# Task 1 (q.14): perform a chi-square test of independence for
# marital status and political orientation

# determine if there are any weird values that we should turn to NAs
summary(GSS$marital)
summary(GSS$politics)

levels(GSS$marital)
levels(GSS$politics)

# remove the NA in marital
GSS$marital[GSS$marital=="NA"]= NA
GSS$marital <- factor(GSS$marital)

# load the gmodels package
library("gmodels")

# perform chi-square test
CrossTable(GSS$marital, GSS$politics, chisq = TRUE, expected = TRUE, sresid = TRUE, format =
 "SPSS")

# determine the effect size by using a cramer's v function
marital_politics = chisq.test(GSS$marital, GSS$politics)
marital_politics

# source of crammers's v function: W203 Async material
crammers_v = function(chisquare)
{
 crammersv = sqrt(chisquare$statistic / (sum(chisquare$observed) * (min(dim(chisquare$observed))-1)))
 print.noquote("Crammer's V: ")
 return(as.numeric(crammersv))
}

crammers_v(marital_politics)

# Task 2 (q.15) conduct a Pearson's correlation analysis to examine the association between
# age when married (agewed) and hours of tv watched (tvhours)

# load the Hmisc package so we can test p-values
library("Hmisc")
```

```
# conduct the Pearson's correlation analysis, remembering to turn the data into a matrix format
rcorr(as.matrix(GSS[,c("agewed","tvhours")]))


# Task 3 (q.16), Conduct a Wilcox rank-sum test to determine whether
# your new "married" variable is associated with the number of children (childs)
# for respondents who are 23 years old.

# create a dummy variable, married, to note if a person is currently married
GSS$married <- ifelse(GSS$marital == "married",1,0)

# mean of married variable for those aged 23 (i.e. portion of sample that is married and 23)
mean(subset(GSS, age == 23, select=married)$married, na.rm = TRUE)

# wilcox rank sum test
married_23 <- wilcox.test(childs ~ married, data = subset(GSS, age == 23), paired = FALSE, na.action =
 na.exclude)
married_23

# calculate effect size
# source of r function from textbook, p.656
rFromWilcox <- function(wilcoxModel, N)
{
 z <- qnorm(wilcoxModel$p.value/2)
 r <- z / sqrt(N)
 cat(wilcoxModel$data.name, "Effect Size, r = ", r)
}

GSS$twenty3 <- ifelse(GSS$age == 23, 1, 0)
TwentyThreeSampleSize = sum(GSS$twenty3)

rFromWilcox(married_23,TwentyThreeSampleSize)

# Task 4 (q.17), Conduct an analysis of variance to determine if
# there is an association between religious affiliation (relig) and
# age when married (agewed).

# check variables for anything weird
table(GSS$agewed)
summary(GSS$agewed)

table(GSS$relig)
summary(GSS$relig)

# recode values of 0 and 99 in agewed variable to "NA"
GSS$agewed[GSS$agewed==0] <- NA
GSS$agewed[GSS$agewed==99] <- NA

# recode values of DK and NA in relig variable to "NA"
GSS$relig[GSS$relig=="DK"] <- NA
GSS$relig[GSS$relig=="NA"] <- NA
GSS$relig <- factor(GSS$relig)
```

```
# we use Levene's test to test homogeneity of variance
library("car")
leveneTest(GSS$agewed,GSS$relig)

# because Levene's test is not significant, we continue with our ANOVA
# load packages for ANOVA
library("compute.es")
library("multcomp")
library("pastecs")

# create ANOVA model
ReligAgeWed <- aov(agewed ~ relig, data = GSS, na.action=na.omit)
summary(ReligAgeWed)

# test individual pairs using a Bonferroni correction
pairwise.t.test(GSS$agewed, GSS$relig, p.adjust.method = "bonferroni")

# calculate the Omega squared effect size
# (SSm - (dfm)MSr)/(SSt + MSr)
# formula is modified version from textbook, p.455 and
# site: http://stats.stackexchange.com/questions/2962/omega-squared-for-measure-of-effect-in-r

omega_squared <- function(anova_model)
{
 sum_stats <- summary(anova_model)[[1]]
 SSm <- sum_stats[["Sum Sq"]][1]
 SSr <- sum_stats[["Sum Sq"]][2]
 SSt <- SSm + SSr
 MSr <- sum_stats[["Mean Sq"]][2]
 DFm <- sum_stats[["Df"]][1]
 omega_sq = (SSm - DFm * MSr)/(SSt + MSr)
 print.noquote("Omega squared: ")
 return(as.numeric(omega_sq))
}

omega_squared(ReligAgeWed)
```