PSY308d.DA1

Daniel Pinedo

April 2, 2019

**Main Analyses:**

A set of researchers from a marketing company conducted a survey to investigate reading habits of Americans (this is a real dataset). Although a study has been published on this already, they have tasked you with investigating it further in order to understand the relationships between certain demographic variables of their sampled readers, inculding sex, employment, martial status, race, and age.

While your advisor thinks there might be some interesting results that could come from this data for an upcoming conference presentation, she only wants a sample write-up of a Results and Discussion section to start based on the following proposed RQs.

Research Questions to Investigate:  
 \*1:\* Is there a relationship betwween sex and employment?  
 \*2:\* Is there a relationship between sex and education?   
 \*3:\* Is there a relationship between martial status and employment?

**Conceptual Addition:** Following your analyses - please incorporate into the Discussion section of your write-up the following conceptual addition:

1.) Your advisor would also like your notes on a potentially interesting caveat for using this information to make generalizations to the general public. Specifically, your advisor wants you to test that the proportions of the sample match those which may be expected in the U.S. population for race, per the proportions below:

White: 80%  
Black or African American: 12%  
Asian or Pacific Islander: 3%  
 Mixed Race: 2%  
 Native American/American Indian: 1%  
 Other: 2%

2.) Following testing, interpret these results and discuss how this may affect your ability to interpret the data and make generalizations. What suggestions could you make moving forward?

*Please report all relevant statistics per APA format and write for a professional audience.*

library(pacman) #Package used to load all packages using p\_load(); will install missing packages

## Warning: package 'pacman' was built under R version 3.5.3

p\_load(vcd, MASS, jmv, gmodels)

dat <- read.csv("https://www.dropbox.com/s/zhhyiegj8gyakuu/Reading.csv?dl=1")  
head(dat) # check to see if labels are needed. In this case they are not. See Chi2Demo.Rmd for how-to.

## Age Sex Race Married Married.status  
## 1 66 Male <NA> No Divorced  
## 2 46 Male Native American/American Indian Yes Married  
## 3 32 Male Mixed race No Never been married  
## 4 27 Male Mixed race Yes Married  
## 5 16 Female Mixed race No Never been married  
## 6 55 Female Asian or Pacific Islander No Divorced  
## Education Employement  
## 1 College graduate Retired  
## 2 High school graduate Employed full-time  
## 3 High school graduate Employed full-time  
## 4 High school graduate Employed full-time  
## 5 High school incomplete Employed part-time  
## 6 Some college, no 4-year degree Have own business/self-employed  
## Incomes  
## 1 $20,000 to under $30,000  
## 2 Less than $10,000  
## 3 Less than $10,000  
## 4 $40,000 to under $50,000  
## 5 $10,000 to under $20,000  
## 6 $40,000 to under $50,000  
## How.many.books.did.you.read.during.last.12months.  
## 1 97  
## 2 97  
## 3 97  
## 4 97  
## 5 97  
## 6 97  
## Read.any.printed.books.during.last.12months.  
## 1 Yes  
## 2 Yes  
## 3 No  
## 4 Yes  
## 5 Yes  
## 6 Yes  
## Read.any.audiobooks.during.last.12months.  
## 1 No  
## 2 Yes  
## 3 Yes  
## 4 No  
## 5 Yes  
## 6 Yes  
## Read.any.e.books.during.last.12months.  
## 1 Yes  
## 2 Yes  
## 3 Yes  
## 4 Yes  
## 5 No  
## 6 Yes  
## Last.book.you.read..youâ..  
## 1 Purchased the book  
## 2 Purchased the book  
## 3 Borrowed the book from a friend or family member  
## 4 Borrowed the book from a library  
## 5 Purchased the book  
## 6 Purchased the book  
## Do.you.happen.to.read.any.daily.news.or.newspapers.  
## 1 No  
## 2 Yes  
## 3 Yes  
## 4 Yes  
## 5 Yes  
## 6 No  
## Do.you.happen.to.read.any.magazines.or.journals.  
## 1 Yes  
## 2 Yes  
## 3 Yes  
## 4 No  
## 5 No  
## 6 No

Take a look at the data set,

# Run descriptives for categorical variables  
glimpse <- descriptives(data = dat,   
 vars = c('Sex', 'Employement', 'Education', 'Married', 'Race'),   
 freq = TRUE)  
glimpse

##   
## DESCRIPTIVES  
##   
## Descriptives   
## ------------------------------------------------------------------   
## Sex Employement Education Married Race   
## ------------------------------------------------------------------   
## N 2442 2442 2438 2442 2398   
## Missing 0 0 4 0 44   
## Mean   
## Median   
## Minimum   
## Maximum   
## ------------------------------------------------------------------   
##   
##   
## FREQUENCIES  
##   
## Frequencies of Sex   
## --------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## --------------------------------------------------   
## Female 1329 54.4 54.4   
## Male 1113 45.6 100.0   
## --------------------------------------------------   
##   
##   
## Frequencies of Employement   
## ---------------------------------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## ---------------------------------------------------------------------------   
## Disabled 46 1.9 1.9   
## Employed full-time 1089 44.6 46.5   
## Employed part-time 322 13.2 59.7   
## Have own business/self-employed 47 1.9 61.6   
## Not employed for pay 402 16.5 78.1   
## Other 13 0.5 78.6   
## Retired 501 20.5 99.1   
## Student 22 0.9 100.0   
## ---------------------------------------------------------------------------   
##   
##   
## Frequencies of Education   
## ----------------------------------------------------------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## ----------------------------------------------------------------------------------------------------   
## College graduate 568 23.3 23.3   
## High school graduate 522 21.4 44.7   
## High school incomplete 205 8.4 53.1   
## None 30 1.2 54.3   
## Post-graduate training/professional school after college 485 19.9 74.2   
## Some college, no 4-year degree 579 23.7 98.0   
## Technical, trade or vocational school AFTER high school 49 2.0 100.0   
## ----------------------------------------------------------------------------------------------------   
##   
##   
## Frequencies of Married   
## --------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## --------------------------------------------------   
## No 1109 45.4 45.4   
## Yes 1333 54.6 100.0   
## --------------------------------------------------   
##   
##   
## Frequencies of Race   
## ---------------------------------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## ---------------------------------------------------------------------------   
## Asian or Pacific Islander 62 2.6 2.6   
## Black or African-American 282 11.8 14.3   
## Mixed race 55 2.3 16.6   
## Native American/American Indian 24 1.0 17.6   
## Other 48 2.0 19.6   
## White 1927 80.4 100.0   
## ---------------------------------------------------------------------------

# First thing to note is the Row for "Missing" DATA --> Different N's and the line that indicates how many are missing shows there are missing cases in our sample of items to be tested  
# There are 44 missing total all from the "Race" category {no further analysis of missing values needed - see /Regression/Regression\_DIagnostics for how-to}  
#Option: (1) delete list-wise

Remove missing cases and view data again with **observed frequencies**

# Option: Listwise deletion of missing data. New dataset is named "dat.no.NA"  
dat.no.NA <- na.omit(dat)  
  
# check descriptives again  
# no missing cases  
glimpse.no.NA <- descriptives(data = dat.no.NA,   
 vars = c('Sex', 'Employement', 'Education', 'Married', 'Race'),   
 freq = TRUE)  
glimpse.no.NA

##   
## DESCRIPTIVES  
##   
## Descriptives   
## ------------------------------------------------------------------   
## Sex Employement Education Married Race   
## ------------------------------------------------------------------   
## N 2134 2134 2134 2134 2134   
## Missing 0 0 0 0 0   
## Mean   
## Median   
## Minimum   
## Maximum   
## ------------------------------------------------------------------   
##   
##   
## FREQUENCIES  
##   
## Frequencies of Sex   
## --------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## --------------------------------------------------   
## Female 1151 53.9 53.9   
## Male 983 46.1 100.0   
## --------------------------------------------------   
##   
##   
## Frequencies of Employement   
## ---------------------------------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## ---------------------------------------------------------------------------   
## Disabled 43 2.0 2.0   
## Employed full-time 963 45.1 47.1   
## Employed part-time 296 13.9 61.0   
## Have own business/self-employed 42 2.0 63.0   
## Not employed for pay 361 16.9 79.9   
## Other 7 0.3 80.2   
## Retired 401 18.8 99.0   
## Student 21 1.0 100.0   
## ---------------------------------------------------------------------------   
##   
##   
## Frequencies of Education   
## ----------------------------------------------------------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## ----------------------------------------------------------------------------------------------------   
## College graduate 490 23.0 23.0   
## High school graduate 465 21.8 44.8   
## High school incomplete 185 8.7 53.4   
## None 25 1.2 54.6   
## Post-graduate training/professional school after college 406 19.0 73.6   
## Some college, no 4-year degree 520 24.4 98.0   
## Technical, trade or vocational school AFTER high school 43 2.0 100.0   
## ----------------------------------------------------------------------------------------------------   
##   
##   
## Frequencies of Married   
## --------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## --------------------------------------------------   
## No 1007 47.2 47.2   
## Yes 1127 52.8 100.0   
## --------------------------------------------------   
##   
##   
## Frequencies of Race   
## ---------------------------------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## ---------------------------------------------------------------------------   
## Asian or Pacific Islander 54 2.5 2.5   
## Black or African-American 259 12.1 14.7   
## Mixed race 49 2.3 17.0   
## Native American/American Indian 21 1.0 17.9   
## Other 43 2.0 20.0   
## White 1708 80.0 100.0   
## ---------------------------------------------------------------------------

Assumptions - 1. Adequate expected cell counts - 5 or more in 2 x 2 or 5 or more in 80% of cells for larger table - Otherwise, Fisher’s test - 2. Independence of Observations - otherwise McNemar’s test of dependent proportions

## Chi-square Test of Independence

H1: Is Sex dependent upon Employement? Is there a relationship between Sex and Employement? H2: Is Sex dependent upon Education? Is there a relationship between Sex and Education? H3: Is Married dependent upon Employement? Is there a relationship between Married and Employement?

Cramer’s V - small = .1; medium = .3, large = .5; indicates effect size of discrepancy between observed and expected scores

# Chi-square = Sum[(Observed - Expected)^2/Expected]  
# Expected = [(# of row entries for cel)/(# total entries)] \* (# of column entries for cel)  
# Expected indicates expected values for each category if there is no relationship between two categorical variables  
# df = (# rows - 1) \* (# columns - 1)  
# report APA, magnitude of effect (Cramer's V), direction of effect example (more or less than expected in each category - include Contingency Table)  
  
H1 <- jmv::contTables(dat = dat.no.NA,  
 rows = 'Employement',  
 cols = 'Sex',  
 exp = TRUE,  
 phiCra = TRUE)  
  
H1

##   
## CONTINGENCY TABLES  
##   
## Contingency Tables   
## -----------------------------------------------------------------------------   
## Employement Female Male Total   
## -----------------------------------------------------------------------------   
## Disabled Observed 24 19 43   
## Expected 23.19 19.81   
##   
## Employed full-time Observed 443 520 963   
## Expected 519.41 443.59   
##   
## Employed part-time Observed 173 123 296   
## Expected 159.65 136.35   
##   
## Have own business/self-employed Observed 18 24 42   
## Expected 22.65 19.35   
##   
## Not employed for pay Observed 239 122 361   
## Expected 194.71 166.29   
##   
## Other Observed 6 1 7   
## Expected 3.78 3.22   
##   
## Retired Observed 232 169 401   
## Expected 216.28 184.72   
##   
## Student Observed 16 5 21   
## Expected 11.33 9.67   
##   
## Total Observed 1151 983 2134   
## Expected 1151.00 983.00   
## -----------------------------------------------------------------------------   
##   
##   
## <U+03C7>² Tests   
## -------------------------------   
## Value df p   
## -------------------------------   
## <U+03C7>² 60.3 7 < .001   
## N 2134   
## -------------------------------   
##   
##   
## Nominal   
## ----------------------------   
## Value   
## ----------------------------   
## Phi-coefficient NaN   
## Cramer's V 0.168   
## ----------------------------

H2 <- jmv::contTables(dat = dat.no.NA,  
 rows = 'Education',  
 cols = 'Sex',  
 exp = TRUE,  
 phiCra = TRUE)  
  
H2

##   
## CONTINGENCY TABLES  
##   
## Contingency Tables   
## ----------------------------------------------------------------------------------------------------   
## Education Female Male Total   
## ----------------------------------------------------------------------------------------------------   
## College graduate Observed 271 219 490   
## Expected 264.3 225.7   
##   
## High school graduate Observed 246 219 465   
## Expected 250.8 214.2   
##   
## High school incomplete Observed 97 88 185   
## Expected 99.8 85.2   
##   
## None Observed 11 14 25   
## Expected 13.5 11.5   
##   
## Post-graduate training/professional school after college Observed 218 188 406   
## Expected 219.0 187.0   
##   
## Some college, no 4-year degree Observed 286 234 520   
## Expected 280.5 239.5   
##   
## Technical, trade or vocational school AFTER high school Observed 22 21 43   
## Expected 23.2 19.8   
##   
## Total Observed 1151 983 2134   
## Expected 1151.0 983.0   
## ----------------------------------------------------------------------------------------------------   
##   
##   
## <U+03C7>² Tests   
## ------------------------------   
## Value df p   
## ------------------------------   
## <U+03C7>² 2.11 6 0.909   
## N 2134   
## ------------------------------   
##   
##   
## Nominal   
## -----------------------------   
## Value   
## -----------------------------   
## Phi-coefficient NaN   
## Cramer's V 0.0315   
## -----------------------------

H3 <- jmv::contTables(dat = dat.no.NA,  
 rows = 'Employement',  
 cols = 'Married',  
 exp = TRUE,  
 phiCra = TRUE)  
  
H3

##   
## CONTINGENCY TABLES  
##   
## Contingency Tables   
## ------------------------------------------------------------------------------   
## Employement No Yes Total   
## ------------------------------------------------------------------------------   
## Disabled Observed 26 17 43   
## Expected 20.29 22.71   
##   
## Employed full-time Observed 370 593 963   
## Expected 454.42 508.58   
##   
## Employed part-time Observed 176 120 296   
## Expected 139.68 156.32   
##   
## Have own business/self-employed Observed 10 32 42   
## Expected 19.82 22.18   
##   
## Not employed for pay Observed 243 118 361   
## Expected 170.35 190.65   
##   
## Other Observed 3 4 7   
## Expected 3.30 3.70   
##   
## Retired Observed 162 239 401   
## Expected 189.23 211.77   
##   
## Student Observed 17 4 21   
## Expected 9.91 11.09   
##   
## Total Observed 1007 1127 2134   
## Expected 1007.00 1127.00   
## ------------------------------------------------------------------------------   
##   
##   
## <U+03C7>² Tests   
## -------------------------------   
## Value df p   
## -------------------------------   
## <U+03C7>² 136 7 < .001   
## N 2134   
## -------------------------------   
##   
##   
## Nominal   
## ----------------------------   
## Value   
## ----------------------------   
## Phi-coefficient NaN   
## Cramer's V 0.252   
## ----------------------------

## Conceptual Question: Goodness-of-Fit

H0: Proportions fit expected values for Race Ha: Proportions do not fit expected values for Race

*Order Matters* Asian or Pacific Islander: 3% Black or African American: 12% Mixed Race: 2% Native American/American Indian: 1% Other: 2% White: 80%

# H0 = equal proportions in each category; Ha = unequal proportions in each category  
# Chi-square = Sum[(Observed - Expected)^2/Expected]  
# df = # of categories - 1  
  
# use print(levels(dat.no.NA$Race)) to check for order of list in order to apply ratios accurately  
# in this case it is in alphabetical order  
  
# First checking unweighted expected values  
# H0: equal proportions in each category  
# Ha: unequal proportions in each category  
goodness <- jmv::propTestN(data = dat.no.NA,  
 var = 'Race',  
 expected = TRUE,   
 ratio = c(1, 1, 1, 1, 1, 1))  
  
goodness

##   
## PROPORTION TEST (N OUTCOMES)  
##   
## Proportions   
## ----------------------------------------------------------------------   
## Level Count Proportion   
## ----------------------------------------------------------------------   
## Asian or Pacific Islander Observed 54 0.02530   
## Expected 356 0.167   
##   
## Black or African-American Observed 259 0.12137   
## Expected 356 0.167   
##   
## Mixed race Observed 49 0.02296   
## Expected 356 0.167   
##   
## Native American/American Indian Observed 21 0.00984   
## Expected 356 0.167   
##   
## Other Observed 43 0.02015   
## Expected 356 0.167   
##   
## White Observed 1708 0.80037   
## Expected 356 0.167   
## ----------------------------------------------------------------------   
##   
##   
## <U+03C7>² Goodness of Fit   
## ------------------------   
## <U+03C7>² df p   
## ------------------------   
## 6278 5 < .001   
## ------------------------

# Ha holds  
  
# check with weights added from conceptual question  
goodness.weighted <- jmv::propTestN(data = dat.no.NA,  
 var = 'Race',  
 expected = TRUE,   
 ratio = c(.03, .12, .02, .01, .02, .80))  
  
goodness.weighted

##   
## PROPORTION TEST (N OUTCOMES)  
##   
## Proportions   
## ----------------------------------------------------------------------   
## Level Count Proportion   
## ----------------------------------------------------------------------   
## Asian or Pacific Islander Observed 54 0.02530   
## Expected 64 0.0300   
##   
## Black or African-American Observed 259 0.12137   
## Expected 256 0.1200   
##   
## Mixed race Observed 49 0.02296   
## Expected 43 0.0200   
##   
## Native American/American Indian Observed 21 0.00984   
## Expected 21 0.0100   
##   
## Other Observed 43 0.02015   
## Expected 43 0.0200   
##   
## White Observed 1708 0.80037   
## Expected 1707 0.8000   
## ----------------------------------------------------------------------   
##   
##   
## <U+03C7>² Goodness of Fit   
## -----------------------   
## <U+03C7>² df p   
## -----------------------   
## 2.55 5 0.770   
## -----------------------

# H0 holds