Lab 7 KEY

Anwesha Guha & Janette Avelar

2/22/2022

Part I

Suppose two researchers want to know whether there is a significant relationship between the defendant's race and the death sentence when the victim is white. They collected data on 825 cases and recorded information on two variables for each case: race (white vs. non-white) and death sentence (yes vs. no). Please use either Unah and Boger.sav or Unah and Boger.csv data files for this analysis.

Set up and import the files you need. Your approach may look different here and you may not need all the libraries below in the code.

```
library(rio)
library(here)
library(psych)
library(tidyverse)
library(kableExtra)

ecls <- import(here("data", "ecls-k-sub.csv"))
uab <- import(here("data", "Unah and Boger.csv"))
```

1. Write the null hypotheses associated with a chi-square test of independence for this study.

Null: Defendant's death sentence is not associated with their race.

2. Fill in the following 2 x 2 contingency table to report observed frequencies for this dataset.

```
table(uab$race, uab$death.sentence)
```

```
##
##             No Yes
##   Nonwhite 251  33
##   White    508  33
```

Fill in the table.

Table 1: The observed frequencies for the race of defendants and death sentence

	No	Yes
Nonwhite	251	33
White	508	33

3. Run a chi-square test of independence. Summarize your findings for this analysis in APA style write-up.

Run Chi-squared test:

```
fit <- chisq.test(x = uab$race,
                  y = uab$death.sentence)

fit
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  uab$race and uab$death.sentence
## X-squared = 6.9781, df = 1, p-value = 0.008251
```

#Compute Cramer's V:

```
fit$expected
```

```
##           uab$death.sentence
## uab$race       No   Yes
##   Nonwhite 261.28 22.72
##   White    497.72 43.28
```

```
fit$observed
```

```
##           uab$death.sentence
## uab$race    No Yes
##   Nonwhite 251  33
##   White    508  33
```

```
chi <- fit$statistic
chi # 6.978
```

```
## X-squared
##  6.978117
```

```
N <- sum(fit$observed)
N # sample size is 825
```

```
## [1] 825
```

```r
L <- min(nrow(fit$observed), ncol(fit$observed)) # smallest number of either the row or col
L # 2
```

```
## [1] 2
```

```r
# compute Cramer's V
sqrt(chi/N*(L-1)) # 0.0919
```

```
##  X-squared
## 0.09196914
```

```r
# you can also input the numbers manually into the formula instead of using variables and get the same
# this approach might be easier/more straightforward if you know the numbers
sqrt(6.978117/825*(2-1))
```

```
## [1] 0.09196915
```

A chi-square test was conducted to test whether or not there is a significant relationship between a defendant's race and death sentence. The findings indicated that there was a significant relationship between race and death sentence, $X_1^2 = 6.98$, $p = 0.008$, *Cramer's* $V = 0.092$.

4. Compute the proportion (or percentage) of death sentences for White and Non-white individuals in this sample. Interpret the direction of the difference. What would you say about it? Write a brief narrative.

```r
fit$observed
```

```
##           uab$death.sentence
## uab$race    No Yes
##   Nonwhite 251  33
##   White    508  33
```

There are 33/541 or 6.10% of defendants who are white have gotten the death sentences, as opposed to 33/284 or 11.62% of defendants who are non-white have gotten the death sentence. This shows that non-white defendants are about twice as likely as white defendants to be subjected to the death penalty.

Part II

For Part II, you will use a subset of the ECLS-K dataset (https://nces.ed.gov/ecls/). The dataset is available on the website in SPSS format (ecls-k-sub.sav) and CSV format (ecls-k-sub.csv). Please refer to Lab 2 for more information about this dataset.

Suppose that you have a research hypothesis. Those students from households with different poverty levels differ in the proportion of first-time kindergartners. The variables of interest are X2POVTY and X1FIRKDG.

X2POVTY is a categorical variable with three levels in this dataset: 1: Below poverty threshold 2: At or above the poverty threshold, below 200 percent of the poverty threshold 3: At or above 200 percent of the poverty threshold

X1FIRKDG is a categorical variable with two levels in this dataset: 1: Yes 2: No

First, recode values to NA and convert our variables of interest to factors.

```
ecls$X2POVTY[ecls$X2POVTY==-9] <- NA
ecls$X2POVTY <- factor(ecls$X2POVTY)

ecls$X1FIRKDG[ecls$X1FIRKDG<0] <- NA
ecls$X1FIRKDG <- factor(ecls$X1FIRKDG)
```

1. Write the null hypotheses associated with a chi-square test of independence for this study.

Null: Being a first-time kindergartener is not associated with poverty level.

2. Fill in the following 3 x 2 contingency table to report observed frequencies for this dataset.

```
table(ecls$X2POVTY, ecls$X1FIRKDG)
```

```
##
##       1   2
##   1 257  24
##   2 237  11
##   3 646  16
```

This table isn't easy to read. Let's relabel the factors to something that makes mroe sense and is easier to interpret.

Remember, from above:

X2POVTY is a categorical variable with three levels in this dataset: 1: Below poverty threshold 2: At or above the poverty threshold, below 200 percent of the poverty threshold 3: At or above 200 percent of the poverty threshold

X1FIRKDG is a categorical variable with two levels in this dataset: 1: Yes 2: No

```
# for X1FIRKDG
levels(ecls$X1FIRKDG) <- c("Yes", "No")
str(ecls$X1FIRKDG)
```

```
##  Factor w/ 2 levels "Yes","No": 1 1 NA 1 NA 1 1 1 1 1 ...
```

```
# for X2POVTY
levels(ecls$X2POVTY) <- c("< poverty threshold", ">= poverty threshold, but <200%", ">= 200% poverty thr
str(ecls$X2POVTY)
```

```
##  Factor w/ 3 levels "< poverty threshold",..: 2 3 1 2 NA 3 2 3 3 2 ...
```

Better...

```
table(ecls$X2POVTY, ecls$X1FIRKDG)
```

```
##
##                                  Yes  No
##   < poverty threshold             257  24
##   >= poverty threshold, but <200% 237  11
##   >= 200% poverty threshold       646  16
```

Fill in the table.

Table 2: The observed frequencies for the poverty levels for first-time kindergarteners

	Yes	No
< poverty threshold	257	24
>= poverty threshold, but <200%	237	11
>= 200% poverty threshold	646	16

3. Run a chi-square test of independence. Summarize your findings for this analysis in APA style write-up.

Run Chi-squared test:

```
fit <- chisq.test(x = ecls$X2POVTY,
                  y = ecls$X1FIRKDG)

fit
```

```
##
##  Pearson's Chi-squared test
##
## data:  ecls$X2POVTY and ecls$X1FIRKDG
## X-squared = 18.068, df = 2, p-value = 0.0001193
```

#Compute Cramer's V:

```
fit$expected
```

```
##                                 ecls$X1FIRKDG
## ecls$X2POVTY                          Yes       No
##   < poverty threshold             268.9673 12.03275
##   >= poverty threshold, but <200% 237.3804 10.61965
##   >= 200% poverty threshold       633.6524 28.34761
```

```
fit$observed
```

```
##                                 ecls$X1FIRKDG
## ecls$X2POVTY                     Yes  No
##   < poverty threshold             257  24
##   >= poverty threshold, but <200% 237  11
##   >= 200% poverty threshold       646  16
```

```r
chi <- fit$statistic
chi # 18.06778
```

```
## X-squared
##  18.06778
```

```r
N <- sum(fit$observed)
N # sample size is 1191
```

```
## [1] 1191
```

```r
L <- min(nrow(fit$observed), ncol(fit$observed))
L #2
```

```
## [1] 2
```

```r
#compute Cramer's V
sqrt(chi/N*(L-1)) # 0.1231676
```

```
## X-squared
## 0.1231676
```

```r
# you can also input the numbers manually into the formula instead of using variables and get the same
# this approach might be easier/more straightforward if you know the numbers
sqrt(18.06778/1191*(2-1))
```

```
## [1] 0.1231676
```

A chi-square test was conducted to test whether or not there is a significant relationship between different poverty levels of student households and first-time kindergartner status. The findings indicated that there was a significant relationship between poverty level and first-time kindergartner status, $X_1^2 = 18.07$, $p = 0.001$, *Cramer's* $V = 0.12$.

4. Compute the proportion (or percentage) of first-time kindergartners for each poverty group. Interpret the direction of differences. What would you say about it? Write a brief narrative.

```r
fit$observed
```

```
##                                   ecls$X1FIRKDG
## ecls$X2POVTY                       Yes  No
##   < poverty threshold              257  24
##   >= poverty threshold, but <200% 237  11
##   >= 200% poverty threshold        646  16
```

There are 257/281 or 91.46% of students below the poverty level who are first-time kindergartners, as opposed to 237/248 or 95.56% for those at or above poverty but below 200% of poverty threshold and 646/662 or 97.58% for students 200% above the poverty threshold. As students move from one poverty level sub-group to another (increasing in income), the proportion of first-time kindergartners also increases