Psy/Educ 6600: Categorical Data Analysis

Chapter 20: Chi Squared TEsts

Your Name

Spring 2020

Contents

List of Tables

List of Figures

PREPARATION

Packages

Make sure the packages are **installed** *(Package tab)*

```
library(readxl)
library(magrittr)     # Forward pipes in R
library(tidyverse)    # Loads several very helpful 'tidy' packages
library(furniture)    # Nice tables (by our own Tyson Barrett)
library(effectsize)   # effect size calculation
```

SECTION A

Datasets

```
table_soda <- c(X = 27,
                Y = 15,
                Z = 24) %>%
  as.table()

table_season <- c(spring = 30,
                  summer = 40,
                  fall   = 20,
                  winter = 10) %>%
  as.table()

table_dx <- c(Schizophrenic = 60,
              Depressed     = 30,
              Bipolar       = 10) %>%
  as.table()
```

`table_soda` Blind taste test of soft drinks (given counts)

20A 3: 1-way Chi-squared Test - Goodness-of-Fit (equally likely)

TEXTBOOK QUESTION: *A soft drink manufacturer is conducting a blind taste test to compare its best-selling product (`X`) with two leading competitors (`Y` and `Z`). Each subject tastes all three and selects the one that tastes best to him or her. (a) What is the appropriate null hypothesis for this study? (b) If 27 subjects prefer product `X`, 15 prefer product `Y`, and 24 prefer product `Z`, can you reject the null hypothesis at the .05 level?*

DIRECTIONS: Use the `chisq.test()` function to perform a Goodnes-of-Fit or one-way Chi-Squared test to see if the observed counts (`table_soda`) are significantly different from being equally distributed among the three soft drinks. Save the fitted model as `chisq_soda`.

> **NOTE:** You do not need to declare any options inside the `chisq.test()` function, as the default is to use equally likely probabilities.

```
# Run the 1-way chi-square test for equally likely
```

DIRECTIONS: Folow the tutorial to create a table comparing the observed and expected counts.

> **HINT** You may *copy-and-paste* the code from the chunked named `tutorial_chiSq_GoF_EL_counts`, but remember to change the name of the model (appears before the $-sign in two places).

```
# Request the observed and expected counts
```

DIRECTIONS: Place the model's name (`chisq_soda`) in the following chunck, so that when run it will display the full output of the Chi-squared test.

```
# Diplay the full output
```

`table_season` Psychiatric Hospital Admits by Season

20A 7: 1-way Chi-squared Test - Goodness-of-Fit (equally likely)

TEXTBOOK QUESTION: *It has been suggested that admissions to psychiatric hospitals may vary by season. One hypothetical hospital admitted 100 patients last year: 30 in the spring; 40 in the summer; 20 in the fall; and 10 in the winter. Use the chi-square test to evaluate the hypothesis that mental illness emergencies are evenly distributed throughout the year.*

DIRECTIONS: Use the `chisq.test()` function to perform a Goodnes-of-Fit or one-way Chi-Squared test to see if the observed counts (`table_season`) are significantly different from being equally distributed among the four seasons. Save the fitted model as `chisq_season`.

> **NOTE:** You do not need to declare any options inside the `chisq.test()` function, as the default is to use equally likely probabilities.

```
# Run the 1-way chi-square test for equally likely
```

DIRECTIONS: Folow the tutorial to create a table comparing the observed and expected counts.

> **HINT** You may *copy-and-paste* the code from the chunked named `tutorial_chiSq_GoF_EL_counts`, but remember to change the name of the model (appears before the $-sign in two places).

```
# Request the observed and expected counts
```

DIRECTIONS: Place the model's name (`chisq_season`) in the following chunck, so that when run it will display the full output of the Chi-squared test.

```
# Diplay the full output
```

`table_dx` Psychiatric Hospital Admits by Diagnosis

20A 8: 1-way Chi-squared Test - Goodness-of-Fit (hypothesised probabilities)

TEXTBOOK QUESTION: *Of the 100 psychiatric patients referred to in the previous exercise, 60 were diagnosed as schizophrenic, 30 were severely depressed, and 10 had a bipolar disorder. Assuming that the national percentages for psychiatric admissions are 55% schizophrenic, 39% depressive, and 6% bipolar, use the chi-square test to evaluate the null hypothesis that this particular hospital is receiving a random selection of psychiatric patients from the national population.*

DIRECTIONS: Use the `chisq.test()` function to perform a Goodnes-of-Fit or one-way Chi-Squared test to see if the observed counts (`table_dx`) are significantly different from being equally distributed among the three diagnoses. Save the fitted model as `chisq_dx`.

> **NOTE:** You **DO** need to declare the probabilities, as the default is to use equally likely probabilities. You may do this by including `p = c(.55, .39, .06)` within the `chisq.test()` function.

```
# Run the 1-way chi-square test for hypothesized probabilities
```

DIRECTIONS: Folow the tutorial to create a table comparing the observed and expected counts.

> **HINT** You may *copy-and-paste* the code from the chunked named `tutorial_chiSq_GoF_EL_counts`, but remember to change the name of the model (appears before the $-sign in two places).

```
# Request the observed and expected counts
```

DIRECTIONS: Place the model's name (`chisq_rx`) in the following chunck, so that when run it will display the full output of the Chi-squared test.

```
# Diplay the full output
```

Section B

Datasets

```
react_wealth <- data.frame(poor    = c(16, 8,  6),
                           middle  = c(10, 6, 14),
                           wealthy = c( 7, 5, 18),
                           row.names = c("ignores", "talks", "helps")) %>%
  as.matrix() %>%
  as.table()


speed_voice <- data.frame(child = c(5, 2),
                          woman = c(3, 4),
                          man   = c(1, 6),
                          row.names = c("fast", "slow")) %>%
  as.matrix() %>%
  as.table()


data_12b4 <- data.frame(child = c(10, 12, 15, 11,  5,  7,  2),
                        woman = c(17, 13, 16, 12,  7,  8,  3),
                        man   = c(20, 25, 14, 17, 12, 18,  7))
```

`react_wealth` Reaction to Wealth

20B 4: 2-way Chi-squared Test - Independence

TEXTBOOK QUESTION: *A social psychologist is studying whether people are more likely to help a poor person or a rich person who they find lying on the floor. The three conditions all involve an elderly woman who falls down in a shopping mall (when only one person at a time is nearby). The independent variable concerns the apparent wealth of the woman; she is dressed to appear either poor, wealthy, or middle class. The reaction of each bystander is classified in one of three ways: ignoring her, asking if she is all right, and helping her to her feet. The data appear in the contingency table below. (a) Test the null hypothesis at the .01 level. Is there evidence for an association between the apparent wealth of the victim and the amount of help provided by a bystander? ~~(b) Calculate Cramer's phi for these data. What can you say about the strength of the relationship between the two variables?~~*

```
# Display the observed counts
react_wealth %>%
  addmargins() %>%
  pander::pander()
```

	poor	middle	wealthy	Sum
ignores	16	10	7	33
talks	8	6	5	19
helps	6	14	18	38
Sum	30	30	30	90

DIRECTIONS: Use the `chisq.test()` function to perform a two-way Chi-Squared test for independence to see if the observed counts provide evidence of an association between the level of wealth and reaction. Save the fitted model as `chisq_react_wealth`.

> **NOTE:** You do not need to declare any options inside the `chisq.test()` function, as the default is test for independence when given a table. The `correct = FALSE` id needed only for $2x2$ tables.

```
# Run the 2-way chi-square test for independence
```

DIRECTIONS: Display the counts expected if reaction is independent of wealth by starting with the model name `chisq_react_wealth` and adding `$expected` at the end.

```
# Request the expected counts based on "no association"
```

DIRECTIONS: Place the model's name (`chisq_react_wealth`) in the following chunck, so that when run it will display the full output of the Chi-squared test.

```
# Diplay the full output
```

DIRECTIONS: Place the data table (`react_wealth`) into the function `cramers_v()` from the `effectsize` package to get the effect size, which is called "phi" ϕ or "Cramer's V.

`speed_voice` Dichotimize Reaction Time to Voice Calling for Help

20B 8: 2-way Chi-squared Test - Independence

TEXTBOOK QUESTION: *In Exercise 12B4, the dependent variable was the amount of time a subject listened to taperecorded cries for help from the next room before getting up to do something. If some subjects never respond within the time allotted for the experiment, the validity of using parametric statistical techniques could be questioned. As an alternative, subjects could be classified as fast or slow responders (and possibly, nonresponders). The data from Exercise 12B4 were used to classify subjects as fast responders (less than 12 seconds to respond) or slow responders (12 seconds or more). The resulting contingency table is shown in the following table:*

```
# Display the observed counts
speed_voice %>%
  addmargins() %>%
  pander::pander()
```

	child	woman	man	Sum
fast	5	3	1	9
slow	2	4	6	12
Sum	7	7	7	21

TEXTBOOK QUESTION: *(a) Test the null hypothesis ($\alpha = .05$) that speed of response is independent of type of voice heard.*

DIRECTIONS: Use the `chisq.test()` function to perform a two-way Chi-Squared test for independence to see if the observed counts provide evidence of an association between the level of wealth and reaction. Save the fitted model as `chisq_speed_voice`.

> **NOTE:** You do not need to declare any options inside the `chisq.test()` function, as the default is test for independence when given a table. The `correct = FALSE` id needed only for $2x2$ tables.

```
# Run the 2-way chi-square test for independence
```

DIRECTIONS: Display the counts expected if reaction is independent of wealth by starting with the model name `chisq_speed_voice` and adding `$expected` at the end.

```
# Request the expected counts based on "no association"
```

DIRECTIONS: Place the model's name (`chisq_speed_voice`) in the following chunck, so that when run it will display the full output of the Chi-squared test.

```
# Diplay the full output
```

DIRECTIONS: Place the data table (`speed_voice`) into the function `cramers_v()` from the `effectsize` package to get the effect size, which is called "phi" ϕ or "Cramer's V.

TEXTBOOK QUESTION: *(b) How does your conclusion in part a compare with the conclusion you drew in Exercise 12B4? Categorizing the dependent variable throws away information; how do you think that loss of information affects power?*

```
data_12b4 %>%
  pander::pander()
```

child	woman	man
10	17	20
12	13	25
15	16	14
11	12	17
5	7	12
7	8	18
2	3	7

```
data_12b4 %>%
  tidyr::gather(key = voice,
                value = seconds,
                child, woman, man) %>%
  dplyr::mutate(voice = factor(voice,
                               levels = c("child", "woman", "man"))) %>%
  dplyr::mutate(id = row_number()) %>%
  afex::aov_4(seconds ~ voice + (1|id),
              data = .) %>%
  summary()
```

```
Anova Table (Type 3 tests)

Response: seconds
      num Df den Df    MSE      F     ges  Pr(>F)
voice      2     18 26.476 3.7464 0.29392 0.04362 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Section C

Import Data, Define Factors, and Compute New Variables

Import Data, Define Factors, and Compute New Variables

- Make sure the **dataset** is saved in the same *folder* as this file
- Make sure the that *folder* is the **working directory**

  NOTE: I added the second line to convert all the variables names to lower case. I still kept the `F` as a capital letter at the end of the five factor variables.

```
ihno_clean <- read_excel("Ihno_dataset.xls") %>%
  dplyr::rename_all(tolower) %>%
  dplyr::mutate(genderF = factor(gender,
                                 levels = c(1, 2),
                                 labels = c("Female",
                                            "Male"))) %>%
  dplyr::mutate(majorF = factor(major,
                                levels = c(1, 2, 3, 4,5),
                                labels = c("Psychology",
                                           "Premed",
                                           "Biology",
                                           "Sociology",
                                           "Economics"))) %>%
  dplyr::mutate(reasonF = factor(reason,
                                 levels = c(1, 2, 3),
                                 labels = c("Program requirement",
                                            "Personal interest",
                                            "Advisor recommendation"))) %>%
  dplyr::mutate(exp_condF = factor(exp_cond,
                                   levels = c(1, 2, 3, 4),
                                   labels = c("Easy",
                                              "Moderate",
                                              "Difficult",
                                              "Impossible"))) %>%
  dplyr::mutate(coffeeF = factor(coffee,
                                 levels = c(0, 1),
                                 labels = c("Not a regular coffee drinker",
                                            "Regularly drinks coffee")))  %>%
  dplyr::mutate(hr_base_bps = hr_base / 60)
```

`ihno_clean` Ihno's Dataset

20C 1a: 1-way Chi-squared Test - Goodness-of-Fit (equally likely)

TEXTBOOK QUESTION: *(a) Perform a one-way chi square test to determine whether you can reject the null hypothesis that, at Ihno's university, there are the same number of students majoring in each of the five areas represented in Ihno's class, if you assume that Ihno's students represent a random sample with respect to major area.*

DIRECTIONS: Use the `chisq.test()` function to perform a Goodnes-of-Fit or one-way Chi-Squared test to see if the observed counts are significantly different from being equally distributed among the five majors. Save the fitted model as `chisq_ihno_major`.

> **HINT:** Since you are working from a full dataset, you will need to pipe a `dplyr::select(majorF)` step onto the `ihno_clean` dataset to first select out just the `majorF` variable and then pipe on the `table()` function to tablulate the observed counts for each major. Then and only then, you may add the `chisq.test()` function.
>
> **NOTE:** You do not need to declare any options inside the `chisq.test()` function, as the default is to use equally likely probabilities.

DIRECTIONS: Folow the tutorial to create a table comparing the observed and expected counts.

> **HINT** You may *copy-and-paste* the code from the chunked named `tutorial_chiSq_GoF_EL_counts`, but remember to change the name of the model (appears before the $-sign in two places).

```
# Request the observed and expected counts
```

DIRECTIONS: Place the model's name (`chisq_ihno_major`) in the following chunck, so that when run it will display the full output of the Chi-squared test.

```
# Diplay the full output
```

20C 1b: Repeat, separately for each gender

TEXTBOOK QUESTION: *(b) Perform the test in part a separately for both the males and the females in Ihno's class.*

DIRECTIONS: Perform the same test you did in part a, but separately for each level of the gender variable.

> **HINT** You may *copy-and-paste* the code from the chunked named `20c1a_chiSq_GoF_EL_test`, but do NOT same the model as anything.
>
> **NOTE:** You will need to add a `dplyr::filter(genderF == "Male")` step before the selecting of major.

> **HINT** You may *copy-and-paste* the code chunk directly above, changing only `"Male"` to `"Female"`.

20C 3: 2-way Chi-squared Test - Independence

TEXTBOOK QUESTION: *Conduct a two-way chi-square analysis of Ihno's data to test the null hypothesis that the proportion of females is the same for each of the five represented majors in the entire university population.* ~~*Request a statistic to describe the strength of the relationship between gender and major.*~~

> **NOTE:** The `furniture` package includes a very helpful function called `tableX()` which creates a nice cross-tabulation given the names of two variables.

```
ihno_clean %>%
  furniture::tableX(genderF, majorF)
```

```
        majorF
genderF  Psychology Premed Biology Sociology Economics Total
  Female 19         11     11      12        4         57
  Male   10         14     10      3         6         43
  Total  29         25     21      15        10        100
```

DIRECTIONS: Use the `chisq.test()` function to perform a two-way Chi-Squared test for independence to see if the observed counts are significantly different from thoes expected is there is no association between gender and major.

> **HINT:** Since you are working from a full dataset, you will need to pipe a `dplyr::select(genderF, majorF)` step onto the `ihno_clean` dataset to first select out just the `genderF` and `majorF` variables. Then pipe on the `table()` function to cross-tablulate the observed counts. Then and only then, you may add the `chisq.test()` function.

> **NOTE:** If you do not save the model to a name, the full output will be displayed.

DIRECTIONS: Follow the data table with the function `cramers_v()` from the `effectsize` package to get the effect size, which is called "phi" ϕ_C or "Cramer's V.