

# Psy/Educ 6600: Unit 6 Homework

## Categorical Data - Binomial and Chi Squared Tests

*Your Name*

*Spring 2018*

## Contents

<b>PREPARATION</b>	<b>2</b>
Load Packages . . . . .	2
Other Datasets for Section B's . . . . .	2
Ihno's Dataset for Section C's . . . . .	3
 <b>Chapter 19: Binomial Distribution</b>	 <b>4</b>
coupleIQs - Paired Measures: Teenage Couple IQs . . . . .	4
11B 6: Matched t-Test (1-sided) . . . . .	4
19B 2: Sign Test (1-sided) . . . . .	5
schizo_mental - Repeated Measures: Mental condition over time . . . . .	6
19B 3: Sign Test . . . . .	6
 <b>Chapter 20: Chi-Squared Tests</b>	 <b>7</b>
Tutorials: Chi-squared Test . . . . .	7
Goodness of Fit (1-way) - Equally Likely . . . . .	7
Goodness of Fit (1-way) - Hypothesised Probabilities . . . . .	8
Test for Independence (2-way) - vs. Association . . . . .	9
Blind taste test of soft drinks (given counts) . . . . .	10
20A 3: 1-way Chi-squared Test - Goodness-of-Fit (equally likely) . . . . .	10
Psychiatric Hospital Admits by Season . . . . .	11
20A 7: 1-way Chi-squared Test - Goodness-of-Fit (equally likely) . . . . .	11
20A 8: 1-way Chi-squared Test - Goodness-of-Fit (hypothesised probabilities) . . . . .	12
react_wealth Reaction to Wealth . . . . .	13
20B 4: 2-way Chi-squared Test - Independence . . . . .	13
speed_voice Dichotomize Reaction Time to Voice Calling for Help . . . . .	14
20B 8: 2-way Chi-squared Test - Independence . . . . .	14
ihno_clean Ihno's Dataset . . . . .	16
20C 1a: 1-way Chi-squared Test - Goodness-of-Fit (equally likely) . . . . .	16
20C 1b: Repeat, separately for each gender . . . . .	17
20C 3: 2-way Chi-squared Test - Independence . . . . .	18

# PREPARATION

## Load Packages

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

```
library(tidyverse)    # Loads several very helpful 'tidy' packages
library(readxl)       # Read in Excel datasets
library(furniture)    # Nice tables (by our own Tyson Barrett)
library(afex)         # Analysis of Factorial Experiments
library(emmeans)      # Estimated marginal means (Least-squares means)
library(lsmeans)      # Least-Squares Means
library(multcomp)     # Simultaneous Inference in General Parametric Models
library(pander)       # Make tables display nicer
```

## Other Datasets for Section B's

```
coupleIQs <- data.frame(couple = 1:10,
                        boy      = c(110, 100, 120, 90, 108, 115, 122, 110, 127, 118),
                        girl     = c(105, 108, 110, 95, 105, 125, 118, 116, 118, 126))

schizo_metal <- data.frame(patient = 1:9,
                          year_2   = c(5, 7, 4, 2, 5, 3, 5, 6, 4),
                          year_3   = c(3, 6, 5, 2, 4, 4, 6, 5, 3))

woman_parents <- data.frame(home_broken   = c(7, 3),
                          home_complete = c(8, 12),
                          row.names = c("self_divorced", "self_married")) %>%
  as.matrix() %>%
  as.table()

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))
```

## Ihno's Dataset for Section C's

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)
```

## Chapter 19: Binomial Distribution

### coupleIQs - Paired Measures: Teenage Couple IQs

#### 11B 6: Matched t-Test (1-sided)

**TEXTBOOK QUESTION:** *Do teenage boys tend to date teenage girls who have a lower IQ than they do? To try to answer this question, 10 teenage couples (i.e., who are dating regularly) are randomly selected, and each member of each couple is given an IQ test. The results are given in the following table (each column represents a different couple):*

```
coupleIQs %>% pander::pander()
```

couple	boy	girl
1	110	105
2	100	108
3	120	110
4	90	95
5	108	105
6	115	125
7	122	118
8	110	116
9	127	118
10	118	126

**TEXTBOOK QUESTION:** *Perform a one-tailed matched  $t$  test ( $\alpha = .05$ ) to determine whether the boys have higher IQs than their girlfriends. What can you conclude?*

**NOTE:** The homework skeleton shows the two-tailed test by mistake.

```
coupleIQs %>%  
  tidyr::gather(key = sex,  
                value = score,  
                boy, girl) %>%  
  t.test(score ~ sex,  
        data = .,  
        alternative = "greater",  
        paired = TRUE)
```

Paired t-test

```
data:  score by sex  
t = -0.25048, df = 9, p-value = 0.5961  
alternative hypothesis: true difference in means is greater than 0  
95 percent confidence interval:  
 -4.990976      Inf  
sample estimates:  
mean of the differences  
      -0.6
```

## 19B 2: Sign Test (1-sided)

**TEXTBOOK QUESTION:** Perform the sign test on the data from Exercise 11B6 using the same alpha level and number of tails. Did you reach the same conclusion with the sign test as with the matched t test? If not, explain the discrepancy.

```
coupleIQs_wide <- coupleIQs %>%
  dplyr::mutate(sign = case_when(boy > girl ~ "positive",
                                boy < girl ~ "negative"))

coupleIQs_wide %>% pander::pander()
```

couple	boy	girl	sign
1	110	105	positive
2	100	108	negative
3	120	110	positive
4	90	95	negative
5	108	105	positive
6	115	125	negative
7	122	118	positive
8	110	116	negative
9	127	118	positive
10	118	126	negative

```
coupleIQs_table <- coupleIQs_wide %>%
  dplyr::select(sign) %>%
  table()

coupleIQs_table %>% pander::pander()
```

negative	positive
5	5

```
coupleIQs_table %>%
  binom.test(alternative = "greater")
```

Exact binomial test

```
data: .
number of successes = 5, number of trials = 10, p-value = 0.623
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.2224411 1.0000000
sample estimates:
probability of success
               0.5
```

## schizo\_mental - Repeated Measures: Mental condition over time

### 19B 3: Sign Test

```
schizo_mental_wide <- schizo_mental %>%  
  dplyr::mutate(sign = case_when(year_2 > year_3 ~ "decline",  
                                year_2 < year_3 ~ "gain"))  
  
schizo_mental_wide %>% pander::pander()
```

patient	year_2	year_3	sign
1	5	3	decline
2	7	6	decline
3	4	5	gain
4	2	2	NA
5	5	4	decline
6	3	4	gain
7	5	6	gain
8	6	5	decline
9	4	3	decline

```
schizo_mental_table <- schizo_mental_wide %>%  
  dplyr::select(sign) %>%  
  table()  
  
schizo_mental_table %>% pander::pander()
```

decline	gain
5	3

```
schizo_mental_table %>%  
  binom.test()      # default: alternative = "two.sided"
```

Exact binomial test

```
data: .  
number of successes = 5, number of trials = 8, p-value = 0.7266  
alternative hypothesis: true probability of success is not equal to 0.5  
95 percent confidence interval:  
 0.2448632 0.9147666  
sample estimates:  
probability of success  
      0.625
```

## Chapter 20: Chi-Squared Tests

### Tutorials: Chi-squared Test

#### Goodness of Fit (1-way) - Equally Likely

**TEXTBOOK Example:** Often, especially in an experimental context, the expected frequencies are based on more abstract theoretical considerations. For instance, imagine that a developmental psychologist is studying color preference in toddlers. Each child is told that he or she can take one toy out of four that are offered. All four toys are identical except for color: red, blue, yellow, or green. Forty children are run in the experiment, and their color preferences are as follows: red, 13; blue, 9; yellow, 15; and green, 3. These are the obtained frequencies. The expected frequencies depend on the null hypothesis. If the null hypothesis is that toddlers in general have no preference for color, we would expect the choices of colors to be equally divided among the entire population of toddlers. Hence, the expected frequencies would be 10 for each color.

---

Use the `chisq.test()` function to perform a Goodness-of-Fit or one-way Chi-Squared test to see if the observed counts are significantly different from being equally distributed.

**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
chisq_toy_color <- c(red = 13, blue = 9,
                    yellow = 15, green = 3) %>%
  chisq.test() # defaults to Equally likely
```

---

The following code chunk shows how to create and display a table of the observed and expected counts for any 1-way Chi-squared test.

**HINT** You may *copy-and-paste* this code for the rest of the assignment, but make sure to change the name of the model (`chisq_toy_color` appears twice before the `$`-sign).

```
# Request the observed and expected counts
rbind(Observed = chisq_toy_color$observed,
      Expected = chisq_toy_color$expected) %>%
  pander::pander()
```

---

	red	blue	yellow	green
Observed	13	9	15	3
Expected	10	10	10	10

---

---

To display the full output, type and run the name the model is save as.

```
# Display the full output
chisq_toy_color
```

Chi-squared test for given probabilities

```
data: .
X-squared = 8.4, df = 3, p-value = 0.03843
```

## Goodness of Fit (1-way) - Hypothesised Probabilities

**TEXTBOOK Example:** Imagine that the population of a city is made up of three ethnic groups, which I will label A, B, and C. The obtained frequencies were 28, 18, and 2. You could test the null hypothesis that sample is representative of a population proportions which is half group A and a third group B.

The `chisq.test()` function may also be used to perform a Goodness-of-Fit or one-way Chi-Squared test to see if the observed counts are significantly different from those expected from a set of hypothesised probabilities.

**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(p1, p2, ..., pk)` within the `chisq.test()` function. The  $p_i$ 's may be typed as decimals or fractions, but make sure they add up to exactly 1!

```
# Run the 1-way chi-square test for hypothesised probabilities
chisq_ethnic <- c(A = 28,
                 B = 18,
                 C = 2) %>%
  chisq.test(p = c(1/2, 1/3, 1/6))      # declare the probabilities
```

Use the same code chunk to display a table of the observed and expected counts for any 1-way Chi-squared test.

**HINT** You may *copy-and-paste* this code for the rest of the assignment, but make sure to change the name of the model (`chisq_ethnic` appears twice before the `$`-sign).

```
# Request the observed and expected counts
rbind(Observed = chisq_ethnic$observed,
      Expected = chisq_ethnic$expected) %>%
  pander::pander()
```

	A	B	C
Observed	28	18	2
Expected	24	16	8

To display the full output, type and run the name the model is save as.

```
# Display the full output
chisq_ethnic
```

Chi-squared test for given probabilities

```
data: .
X-squared = 5.4167, df = 2, p-value = 0.06665
```



## Test for Independence (2-way) - vs. Association

**TEXTBOOK Example:** Suppose that the researcher has interviewed 30 women who have been married: 10 whose parents were divorced and 20 whose parents were married. Half of the 30 women in this hypothetical study have gone through their own divorce; the other half are still married for the first time. To know whether the divorce of a person's parents makes the person more likely to divorce, we need to see the breakdown in each category- that is, how many currently divorced women come from "broken" homes and how many do not, and similarly for those still married. These frequency data are generally presented in a contingency (or cross-classification) table:

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

	home_broken	home_complete	Sum
self_divorced	7	8	15
self_married	3	12	15
Sum	10	20	30

The `chisq.test()` function may also be used to perform a two-way Chi-Squared test for independence. In this case, the observed counts are compared to those expected if there is no association between the two factors.

```
# Run the 2-way chi-square test for independence
chisq_divorces <- woman_parents %>%
  chisq.test(correct = FALSE) #IF 2x2, add correct = FALSE
```

To display the counts expected if the variables are independent, start with the model name and add `$expected` at the end. Then pipe on both the `addmargins()` and `pander::pander()` functions to print the counts.

```
# Request the expected counts based on "no association"
chisq_divorces$expected %>%
  pander::pander()
```

	home_broken	home_complete
self_divorced	5	10
self_married	5	10

To display the full output, type and run the name the model is save as.

```
# Display the full output
chisq_divorces
```

Pearson's Chi-squared test

```
data: .
X-squared = 2.4, df = 1, p-value = 0.1213
```

## 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 ( $c(X = 27, Y = 15, Z = 24)$ ) 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 chunk, so that when run it will display the full output of the Chi-squared test.

```
# Diplay the full output
```

## 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 (`c(spring = 30, summer = 40, fall = 20, winter = 10)`) 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 chunk, so that when run it will display the full output of the Chi-squared test.

```
# Diplay the full output
```

## 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 (`c(Schizophrenic = 60, Depressed = 30, Bipolar = 10)`) are significantly different from being equally distributed among the three diagnoses. Save the fitted model as `chisq_rx`.

**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 chunk, so that when run it will display the full output of the Chi-squared test.

```
# Diplay the full output
```

## 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` is 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 chunk, so that when run it will display the full output of the Chi-squared test.

```
# Display the full output
```

## 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 chunk, so that when run it will display the full output of the Chi-squared test.

```
# Display the full output
```

**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

## 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 tabulate 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.

```
# 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_ihno_major`) in the following chunk, 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 those expected if 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-tabulate 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.