

EDUC/PSY 6600: Unit 6 Homework

Categorical Data - Binomial and Chi Squared Tests

Your Name

Spring 2018

Contents

NEEDS TO BE UPDATED!!!	3
PREPARATION	3
Load Packages	3
Other Datasets for Section B's	3
Ihno's Dataset for Section C's	4
Chapter 9. Linear Correlation	5
Section B	5
9B-5 Calculating Pearson's r	5
9B-6 One vs. Two Sided Alternative	6
Section C	7
9C-1. Scatterplots - Eyeball method for estimating correlation	7
9C-2a. Calculating Pearson's r	9
9C-2b. Effect of Excluding Extreme Values	10
9C-3. Reporting APA Style	11
9C-4. Missing Values	12
Chapter 10. Linear Regression	13
Section B	13
10B-6 Swapping x and y	13
10B-9 Predictions and Residuals	14
10B-10 Raw Correlation vs. Partial Correlation	16
Section C	17
10C-1. Linear Regression	17
10C-2. Subgroups Analysis	18
Chapter 19: Binomial Distribution	19
Section B	19
coupleIQs - Paired Measures: Teenage Couple IQs	19
19B 2: Sign Test (1-sided)	19
schizo_mental - Repeated Measures: Mental condition over time	21
19B 3: Sign Test	21
Chapter 20: Chi-Squared Tests	22
Section A	22
Tutorials: Chi-squared Test	22
Goodness of Fit (1-way) - Equally Likely	22
Goodness of Fit (1-way) - Hypothesised Probabilities	23
Test for Independence (2-way) - vs. Association	24
Blind taste test of soft drinks (given counts)	25
20A 3: 1-way Chi-squared Test - Goodness-of-Fit (equally likely)	25
Psychiatric Hospital Admits by Season	27
20A 7: 1-way Chi-squared Test - Goodness-of-Fit (equally likely)	27

20A 8: 1-way Chi-squared Test - Goodness-of-Fit (hypothesised probabilities)	28
Section B	29
react_wealth Reaction to Wealth	29
20B 4: 2-way Chi-squared Test - Independence	29
speed_voice Dichotimize Reaction Time to Voice Calling for Help	30
20B 8: 2-way Chi-squared Test - Independence	30
Section C	32
ihno_clean Ihno's Dataset	32
20C 1a: 1-way Chi-squared Test - Goodness-of-Fit (equally likely)	32
20C 1b: Repeat, separately for each gender	33
20C 3: 2-way Chi-squared Test - Independence	34

NEEDS TO BE UPDATED!!!

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(multcomp)      # Simultaneous Inference in General Parametric Models
library(pander)       # Make tables display nicer
library(educ6600)     # datasets for section B
```

Other Datasets for Section B's

```
data("coupleIQs")
data("schizo_metal")
data("woman_parents")
data("react_wealth")
data("speed_voice")
data("data_12b4")
```

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 9. Linear Correlation

Section B

9B-5 Calculating Pearson's r

TEXTBOOK QUESTION: *A psychiatrist has noticed that the schizophrenics who have been in the hospital the longest score the lowest on a mental orientation test. The data for 10 schizophrenics are listed in the following table. (a) Calculate Pearson's r for the data. (b) Test for statistical significance at the .05 level (two-tailed).*

schizo

	id	yr_hos	ori_test
1	1	5	22
2	2	7	26
3	3	12	16
4	4	5	20
5	5	11	18
6	6	3	30
7	7	7	14
8	8	2	24
9	9	9	15
10	10	6	19

DIRECTIONS: Calculate Pearson's r between `yr_hos` and `ori_test` in the `schizo` dataset. Also, test against the two-sided alternative.

The `cor.test()` function needs at least two arguments:

1. the formula: `~ continuous_var1 + continuous_var2`
2. the dataset: `data = .` we use the period to signify that the dataset is being piped from above

NOTE: The `cor.test()` function computes the Pearson correlation coefficient by default (`method = "pearson"`), but you may also specify the Kendall (`method = "kendall"`) or Spearman (`method = "spearman"`) methods. It also defaults to testing for the two-sided alternative and computing a 95% confidence interval (`conf.level = 0.95`). You will not need to change these options in this assignment.

```
# Pearson's r: yr_hos & ori_test
```

9B-6 One vs. Two Sided Alternative

TEXTBOOK QUESTION: *If a test is reliable, each participant will tend to get the same score each time he or she takes the test. Therefore, the correlation between two administrations of the test (test-retest reliability) should be high. The reliability of the verbal GRE score was tested using five participants, as shown below. (a) Calculate Pearson's r for the test-retest reliability of the verbal GRE score. (b) Test the significance of this correlation with $\alpha = .05$ (one-tailed). Would this correlation be significant with a twotailed test?*

GRE

	id	verbalGRE_1	verbalGRE_2
1	1	540	570
2	2	510	520
3	3	580	600
4	4	550	530
5	5	520	520

DIRECTIONS: Calculate Pearson's r between `verbalGRE_1` and `verbalGRE_2` in the GRE dataset TWICE. The first time test for the **one-sided** alternative and the second time for the **two-sided** alternative.

The `cor.test()` function defaults to the `alternative = "two.sided"`. If you would like a one-sided alternative, you must choose which side you would like to test: `alternative = "greater"` or `alternative = "less"`

```
# Pearson's r: verbalGRE_1 & verbalGRE_2 --> ONE tail
```

```
# Pearson's r: verbalGRE_1 & verbalGRE_2 --> TWO tails
```

Section C

9C-1. Scatterplots - Eyeball method for estimating correlation

TEXTBOOK QUESTION: (A) Create a scatter plot of *phobia* versus *statquiz*. From looking at the plot, do you think the Pearson's r will be positive or negative? Large, medium, or small? (B) Create a scatter plot of baseline anxiety versus postquiz anxiety. From looking at the plot, do you think the Pearson's r will be positive or negative? Large, medium, or small?

DIRECTIONS: Create two scatter plots: the first with **phobia** on the horizontal axis (x) and **statquiz** on the vertical axis (y) and the second with **anx_base** on the x-axis and **anx_post** on the y-axis. Then answer the rest of the question in the printed homework packet.

NOTE: You may use the `geom_count()` function instead of the `geom_point()` function due to the high number of points that are 'over plotted' or on top of each other, since the two measures are quite coarsely captured.

```
# Scatterplot: phobia vs. statquiz
```

```
# Scatterplot: an $\alpha$ _base vs. an $\alpha$ _post
```


9C-2a. Calculating Pearson's r

TEXTBOOK QUESTION: Compute the Pearson's r between *phobia* and *statquiz* for all students; also, find the Pearson's r between baseline and postquiz anxiety.

DIRECTIONS: Compute Pearson's r : first for *phobia* and *statquiz*, followed by *anx_base* and *anx_post* using the `cor.test()` function.

```
# Pearson's r: phobia & statquiz
```

```
# Pearson's r: anx_base & anx_post
```

9C-2b. Effect of Excluding Extreme Values

TEXTBOOK QUESTION: Use *Select Cases* to delete any student whose baseline anxiety is over 29, and repeat part (B) of the first exercise. Also, rerun the correlation of baseline and postquiz anxiety. What happened to the Pearson's r ? Use the change in the scatter plot to explain the change in the correlation coefficient.

DIRECTIONS: Create a scatterplot for `anx_base` and `anx_post`, AFTER first using a `dplyr::filter()` function in a prepatroy step to restrict to the subsample of students with baseline anxiety of 29 and below.

```
# Scatterplot: anx_base vs. anx_post <-- restricting to baseline anxiety of 29 and lower
```

DIRECTIONS: Compute Pearson's r : for `anx_base` and `anx_post`, AFTER first using a `dplyr::filter()` function in a prepatroy step to restrict to the subsample of students with baseline anxiety of 29 and below.

```
# Pearson's r: anx_base & anx_post <-- restricting to baseline anxiety of 29 and lower
```

9C-3. Reporting APA Style

TEXTBOOK QUESTION: (a) Compute Pearson's r 's among the three measures of anxiety. Write up the results in APA style. (b) Compute the average of the three measures of anxiety, and then compute the correlation between each measure of anxiety and the average, ~~so that the output contains a single column of correlations (do this by creating and appropriately modifying a syntax file).~~

DIRECTIONS: First, compute a new variable called `anx_mean` that is the average of all three of the anxiety measures using the `furniture::rowmeans()` function. Then use the `furniture::tableC()` function to create a correlation matrix for all FOUR anxiety measres.

```
# Pearson's r: anx_mean, anx_base, anx_pre, & anx_post
```

9C-4. Missing Values

TEXTBOOK QUESTION: (a) Compute Pearson's r for the following list of variables: *mathquiz*, *statquiz*, and *phobia*. (b) Repeat part a after selecting *exclude cases listwise*. Which correlation was changed? Explain why.

Directions: Compute the correlation matrix between *mathquiz*, *statquiz*, and *phobia* using the `furniture::tableC()` function two times; first with all defaults and again with listwise deletion.

Note: The `furniture::tableC()` function defaults to `na.rm = FALSE` which displays NA for any correlation between a pair of variables where even one subject is missing one value. To use listwise deletion, specify the option `na.rm = TRUE`.

```
# Pearson's r: (default: na.rm = FALSE)
```

```
# Pearson's r: "complete.obs" (list-wise deletion)
```

Chapter 10. Linear Regression

Section B

10B-6 Swapping x and y

TEXTBOOK QUESTION: *A cognitive psychologist is interested in the relationship between spatial ability (e.g., ability to rotate objects mentally) and mathematical ability, so she measures 12 participants on both variables. The data appear in the following table. (a) Find the regression equation for predicting the math score from the spatial ability score. (b) Find the regression equation for predicting the spatial ability score from the math score. (c) According to your answer to part a, what math score is predicted from a spatial ability score of 20? (d) According to your answer to part b, what spatial ability score is predicted from a math score of 20?*

test_scores

	id	spatial	math
1	1	13	19
2	2	32	25
3	3	41	31
4	4	26	18
5	5	28	37
6	6	12	16
7	7	19	14
8	8	33	28
9	9	24	20
10	10	46	39
11	11	22	21
12	12	17	15

DIRECTIONS: Use the `lm()` function to fit a linear model or linear regression model TWICE for `math` and `spacial` in the `test_scores` dataset, specifying which is x and which is y.

The `lm()` function needs at least two arguments:

1. the formula: `continuous_y ~ continuous_x`
2. the dataset: `data = .` we use the period to signify that the dataset is being piped from above

NOTE: To view more complete information, add a `summary()` step using a pipe AFTER the `lm()` step

```
# Linear model: y = math & x = spatial
```

```
# Linear model: y = spatial & x = math
```

10B-9 Predictions and Residuals

TEXTBOOK QUESTION: *If you calculate the correlation between shoe size and reading level in a group of elementary school children, the correlation will turn out to be quite large, provided that you have a large range of ages in your sample. The fact that each variable is correlated with age means that they will be somewhat correlated with each other. The following table illustrates this point. Shoe size is measured in inches, for this example, reading level is by grade (4.0 is average for the fourth grade), and age is measured in years. (a) Find the regression equation for predicting shoe size from age. (b) Find the regression equation for predicting reading level from age. (c) Use the equations from parts a and b to make shoe size and reading level predictions for each child. Subtract each prediction from its actual value to find the residual.*

`child_vars`

	child	shoe	read	age
1	1	5.2	1.7	5
2	2	4.7	1.5	6
3	3	7.0	2.7	7
4	4	5.8	3.1	8
5	5	7.2	3.9	9
6	6	6.9	4.5	10
7	7	7.7	5.1	11
8	8	8.0	7.4	12

DIRECTIONS: Use the `lm()` function to fit a linear model or linear regression model TWICE for `shoe` and `read` each predicted in turn by `age` in the `child_vars` dataset, specifying which is x and which is y.

Linear model: $y = \text{shoe}$ & $x = \text{age}$

```
# Linear model:  $y = \text{read}$  &  $x = \text{age}$ 
```

DIRECTIONS: Starting with the `child_vars` dataset, create four new variables, each with a separate `dplyr::mutate()` function step. Pipe it all together and save it as new dataset with the `child_new <-` assignment operator to use in the next step.

1. `shoe_pred` Use the appropriate regression equation
2. `shoe_resid` Subtract: `shoe` (the original) minus `shoe_resid` (the residual)
3. `read_pred` Use the appropriate regression equation
4. `read_resid` Subtract: `read` (the original) minus `read_resid` (the residual)

```
# create new variables --> save as: child_new
```

Note: Remove the hashtag symbol at the first of the code line below to show your new variables.

```
#child_new
```

10B-10 Raw Correlation vs. Partial Correlation

TEXTBOOK QUESTION: (a) Calculate Pearson's r for shoe size and reading level using the data from Exercise 9. (b) Calculate Pearson's r for the two sets of residuals you found in part c of Exercise 9. (c) Compare your answer in part b with your answer to part a. The correlation in part b is the partial correlation between shoe size and reading level after the confounding effect of age has been removed from each variable (see Chapter 17 for a much easier way to obtain partial correlations).

DIRECTIONS: Calculate Pearson's r between shoe and read in the child_new dataset.

```
# Pearson's r: shoe & read
```

DIRECTIONS: Calculate Pearson's r between shoe_resid and read_resid in the child_new dataset.

```
# Pearson's r: shoe_resid & read_resid
```


Section C

10C-1. Linear Regression

TEXTBOOK QUESTION: *Perform a linear regression to predict statquiz from phobia, and write out the raw-score regression formula. Do the slope and Y intercept differ significantly from zero? Explain how you know. What stats quiz score would be predicted for a student with a phobia rating of 9? Approximately what phobia rating would a student need to have in order for her predicted statquiz score to be 7.2?*

Directions: Use the `lm()` function to fit a linear model or linear regression model predicting `statquiz` from `phobia`.

```
# Linear model: y = statquiz & x = phobia
```

10C-2. Subgroups Analysis

TEXTBOOK QUESTION: (a) Perform a linear regression to predict pre-quiz anxiety from phobia, and write out the raw-score regression formula. (b) Repeat part a separately for men and women. For each gender, what prequiz anxiety rating would be predicted for someone reporting a phobia rating of 8? For which gender should you really not be making predictions at all? Explain.

Directions: Use the `lm()` function to fit a linear model or linear regression model predicting `anx_pre` from `phobia`. Then repeat the same model TWICE more: first among just men and then for just women.

Note: Use the `dplyr::filter()` function to subset the sample BEFORE fitting the model. Also, be aware of which type of variable you are using: `genderF == "Male"` or `gender == 2` works, but `gender == male` does NOT.

```
# Linear model: y = anx_pre & x = phobia <-- full sample
```

```
# Linear model: y = anx_pre & x = phobia <-- subset of men
```

```
# Linear model: y = anx_pre & x = phobia <-- subset of women
```

Chapter 19: Binomial Distribution

Section B

coupleIQs - Paired Measures: Teenage Couple IQs

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 %>% pandr::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 %>% pandr::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

Section A

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) %>%
  pandoc::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
chisq_soda <- c(X = 27,
               Y = 15,
               Z = 24) %>%
  chisq.test()
```

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
rbind(Observed = chisq_soda$observed,
      Expected = chisq_soda$expected)
```

```
      X  Y  Z
Observed 27 15 24
Expected 22 22 22
```

```
# Request the observed and expected counts
rbind(Observed = chisq_soda$observed,
      Expected = chisq_soda$expected) %>%
  pandrer::pander(caption = "Counts for Soda Prefereces, Observed and Expected")
```

Table 9: Counts for Soda Prefereces, Observed and Expected

	X	Y	Z
Observed	27	15	24
Expected	22	22	22

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
chisq_soda
```

Chi-squared test for given probabilities

data: .

X-squared = 3.5455, df = 2, p-value = 0.1699

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.

```
# Display the full output
```

Section B

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 Dichotomize 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

Section C

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.