MTH 324, Lab #11, Fall 2022   Name ____

Instructions: Follow along with the tutorial portion of the lab. Replicate the code examples in R on your own, along with the demonstration. Then use those examples as a model to answer the questions/perform the tasks that follow. Copy and paste the results of your code to answer questions where directed. Submit your response file and the code used (both for the tutorial and part two). Your code file and your lab response file should each include your name inside.

Contingency Table tests

We have several tests we can conduct on a two-way table/contingency table/crosstabs. We'll go through several examples.

Consider the sports and gender table we looked at in lecture.

		Sport Preference			
		Archery	**Boxing**	**Cycling**	
Gender	**Female**	35	15	50	**100**
	Male	10	30	60	**100**
		45	**45**	**110**	**200**

We enter it into R in this form (without the totals column or row).

```
observed_table <- matrix(c(35, 15, 50, 10, 30, 60), nrow = 2, ncol = 3, byrow = T)
rownames(observed_table) <- c('Female', 'Male')
colnames(observed_table) <- c('Archery', 'Boxing', 'Cycling')
observed_table
```

We can conduct a test of independence using the chisq.test() function.

```
chisq.test(observed_table)
```

The output displays as text.

```
        Pearson's Chi-squared test

data:  observed_table
X-squared = 19.798, df = 2, p-value = 5.023e-05
```

As we say in the lecture, the test is statistically significant, so the variables are dependent.

If we save the output of the test as a variable, we can access various components of the test, such as the expected table, and so forth.

We can look at Fisher's Exact Test for this data (even though the sample size is larger than we normally use for this test).

```
library(stats)
fisher.test(observed_table)
```

We obtain a similar outcome for the typical χ^2 test, which is what we would expect.

```
        Fisher's Exact Test for Count Data

data:  observed_table
p-value = 3.897e-05
alternative hypothesis: two.sided
```

With the computational power of R, we can use this test even when it is not strictly necessary.

If our table is bigger than 2×2, we can do a pairwise proportion analysis to see if only some of the levels of the variables are different than expected.

```
library(rstatix)
pairwise_fisher_test(as.matrix(observed_table), p.adjust.method = "fdr")
```

The results are shown below. Note that a "tibble" is just a special kind of dataframe used with packages in the tidyverse.

```
# A tibble: 3 x 6
  group1  group2      n         p    p.adj p.adj.signif
* <chr>   <chr>   <dbl>     <dbl>    <dbl> <chr>
1 Archery Boxing     90 0.0000423 0.000127 ***
2 Archery Cycling   155 0.00032   0.00048  ***
3 Boxing  Cycling   155 0.21      0.21     ns
```

We can conduct a permutation test on a contingency table using a function in the CNPS package. We can specify the number of trials.

```
library(CNPS)

permu_table(observed_table, samplenum = 10000)
```

The output calculates the P-value for us.

```
                 Pearson's Chi-squared test
chi-squared_obs = 19.798
simulated method to calculate :
        p-value =  < 2.2e-16
alternative hypothesis:
        There exists at least one pair of groups that are not independent.
```

We can set options to fix the row or column totals as part of the test.

A useful way to visualize a contingency table is using a mosaic plot.

```
library(mosaic)
mosaicplot(observed_table, main = "Mosaic plot of observed results")
```

To conduct a goodness-of-fit test in R, we need to enter our observations and expectations (in the form of probabilities) as vectors.

```
observed <- c(50, 60, 40, 47, 53)
expected <- c(.2, .2, .2, .2, .2)
chisq.test(x=observed, p=expected)
```

The probabilities must add up to 1. The same chisq.test() function does our test for us.

```
         Chi-squared test for given probabilities

data:  observed
X-squared = 4.36, df = 4, p-value = 0.3595
```

If you are modeling a particular distribution, you can use built-in probability distribution functions to fill in the expected probabilities.

Tasks

Use the examples above to complete the following tasks.

1. The first table on the left displayed the expected percentages of the number of pets owned in a household nationally. A sample of 1000 people were surveyed, and the observed counts are shown in the table on the right. Conduct a goodness-of-fit test on the data. Be sure to clearly state the null and alternative hypotheses.

Number of Pets	Percent
0	18
1	25
2	30
3	18
4+	9

Number of Pets	Frequency
0	210
1	240
2	320
3	140
4+	90

2. De Anza College is interested in the relationship between anxiety level and the need to succeed in school. A random sample of 400 students took a test that measured anxiety level and need to succeed in school. The table shows the results. De Anza College wants to know if anxiety level and need to succeed in school are independent events. Conduct a test of independence. Clearly state the null and alternative hypothesis and describe the results of the test in context. Make a graph of the table. (Experiment with bar graphs and mosaic plots. Which do you prefer?)

Need to Succeed in School	High Anxiety	Med-high Anxiety	Medium Anxiety	Med-low Anxiety	Low Anxiety	Row Total
High Need	35	42	53	15	10	155
Medium Need	18	48	63	33	31	193
Low Need	4	5	11	15	17	52
Column Total	57	95	127	63	58	400

3. Do private practice doctors and hospital doctors have the same distribution of working hours? Suppose that a sample of 100 private practice doctors and 150 hospital doctors are selected at random and asked about the number of hours a week they work. Use Fisher's Exact test to conduct your test. Follow up with pairwise comparisons if you reject the null hypothesis.

	20–30	**30–40**	**40–50**	**50–60**
Private Practice	16	40	38	6
Hospital	8	44	59	39

4. Use the sample data in #3 to conduct a permutation test. Compare the results.

References:

1. Discovering Statistics Using R. Andy Field, Jeremy Miles, Zoe Field. (2012)
2. https://book.stat420.org/applied_statistics.pdf
3. https://scholarworks.montana.edu/xmlui/handle/1/2999
4. https://www.rstudio.com/resources/cheatsheets/
5. https://www.datacamp.com/tutorial/contingency-analysis-r
6. https://www.reneshbedre.com/blog/fisher-exact-test.html
7. https://online.stat.psu.edu/stat504/lesson/4/4.5
8. https://search.r-project.org/CRAN/refmans/CNPS/html/permu_table.html
9. https://www.statology.org/chi-square-goodness-of-fit-test-in-r/