---
title: Lab Assignment 2
author: Ben Tankus

```r
library(magrittr)
library(vcdExtra)
```

```
## Warning: package 'vcdExtra' was built under R version 4.0.4
```

```
## Loading required package: vcd
```

```
## Warning: package 'vcd' was built under R version 4.0.4
```

```
## Loading required package: grid
```

```
## Loading required package: gnm
```

```
## Warning: package 'gnm' was built under R version 4.0.4
```

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.3
```

```
## -- Attaching packages ------------------------------------------ tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```
## Warning: package 'tidyr' was built under R version 4.0.3
```

```
## Warning: package 'readr' was built under R version 4.0.3
```

```
## Warning: package 'purrr' was built under R version 4.0.3
```

```
## Warning: package 'dplyr' was built under R version 4.0.3
```

```
## Warning: package 'forcats' was built under R version 4.0.3
```

```
## -- Conflicts --------------------------------------------- tidyverse_conflicts() --
## x tidyr::extract()   masks magrittr::extract()
## x dplyr::filter()    masks stats::filter()
## x dplyr::lag()       masks stats::lag()
## x purrr::set_names() masks magrittr::set_names()
## x dplyr::summarise() masks vcdExtra::summarise()
```

Do be sure to include, near the top of your script, a code chunk that loads any non-default packages you use (such as *vcdExtra* or *magrittr*).

Within the question sections you create, you can chunk your code in whatever way seems reasonable. Incorporate any written answers outside the code chunks using Markdown formatting as needed (see "Help" -> "RMarkdown Quick Reference" for text formatting help).

To ensure that your .Rmd script will be fully self-contained (i.e. it will not depend on objects that were defined during the lab, and could be run as-is if you sent it to someone else), you should **clear the workspace** before you begin.

To clear the workspace, click the broom icon in the Environment pane.

Once you've answered the questions in your new .Rmd script and you have verified your code is self-contained, you should Run -> Run All Chunks and generate a new HTML preview of your document, just to check that everything looks like you want it to. Having concluded that your .Rmd script produces an HTML report that includes all the output you want, submit **the Lab_Assignment_2.Rmd file** on Canvas as your Lab Assignment 2.

Feel free to post on the discussion board if you have any questions or encounter any difficulties with this process.

Questions:

NOTE: The first two questions are for credit. The next two are optional/for fun.

Question 1

The following data come from one of Gregor Mendel's famous experiments with pea plants. In this particular experiment, Mendel examined two categorical traits in pea seeds, each with two possible values: seed color – yellow or green – and seed shape – round or angular. According to Mendel's hypothesis, the inheritance pattern of these traits was characterized by "independent assortment" – that is, there was no "genetic linkage" between these traits.

```r
mendel <- as.table(
  matrix(c(315, 108, 101, 32),
         nrow = 2,
         dimnames =
           list(Color= c("Yellow", "Green"),
                Shape = c("Round", "Angular"))))

mosaic(mendel)
```

Shape

Round   Angular
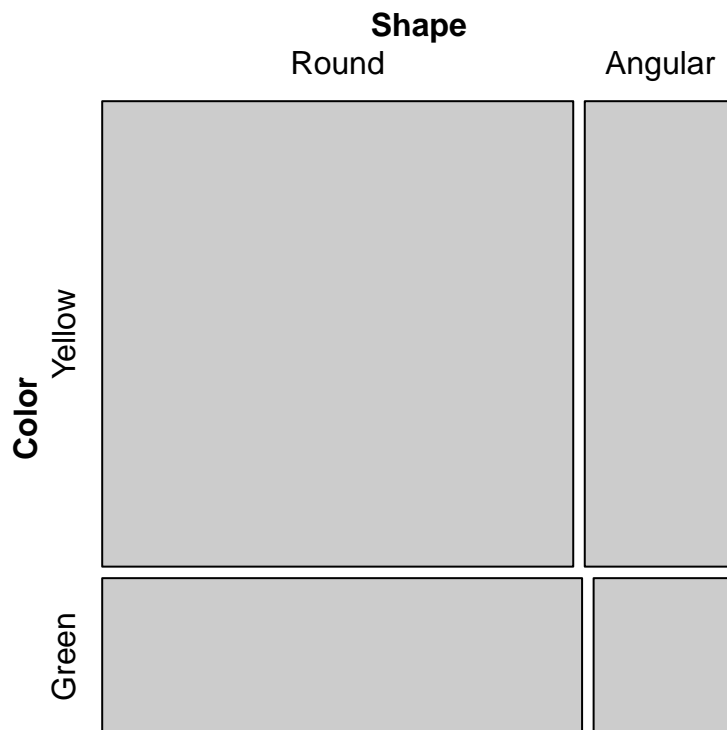

```r
chisq.test(mendel, p = c(9/16, 3/16, 3/16, 1/16))
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  mendel
## X-squared = 0.051332, df = 1, p-value = 0.8208
```

How consistent are these data with Mendel's hypothesis of independence? Notice that in this particular case, we're not interested in a *departure* from independence, but rather a *confirmation* of independence. Is a Chi-squared test appropriate here? Why or why not?

> The chi-squared test is aproprate here because it is testing for independence. I would also recommend using a mosaic plot.

> The Mosaic plot suggests independence, and a chisq test provides a pvalue of 0.821 which also suggests independence of the data.

Question 2

We mentioned previously, and it's mentioned in the narrated lecture materials that there's an equivalence between the Chi-squared test and the difference in proportions test in the case of 2x2 tables. Take a look for the Mendel data:

```
chisq.test(mendel, correct = FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  mendel
## X-squared = 0.11634, df = 1, p-value = 0.733
```

```
prop.test(mendel, correct = FALSE)
```

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  mendel
## X-squared = 0.11634, df = 1, p-value = 0.733
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.09506177  0.06662771
## sample estimates:
##    prop 1    prop 2
## 0.7572115 0.7714286
```

The output from the second chunk refers to a `two.sided` test, with an interpretation in terms of a signed difference between two proportions, whereas the `chisq.test()` is a one-sided test, with an interpretation involving deviation from the hypothesis of independent rows and columns. It's not *entirely* obvious why these results should be the same. The following exercise can help you see what's going on.

Use `rnorm()` and `rchisq()` to produce and store the following vectors:

1000 draws from the standard normal distribution, 1000 draws from the Chi-squared distribution with 1 degree of freedom.

(a) Construct separate histograms/density plots of the raw chi-squared values and the *squares* of the standard normal values. What do you observe?

```
hist(rnorm(1000,0,1)^2, main = 'Norm^2')
```

Norm^2
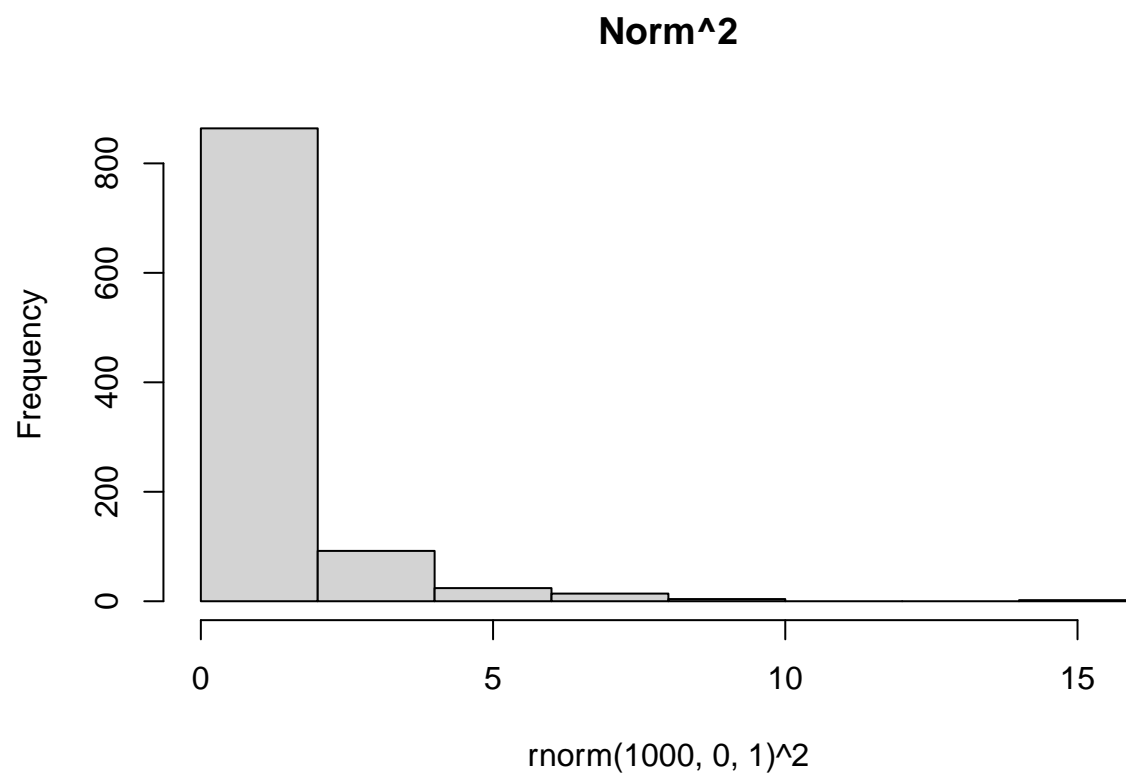

```r
hist(rchisq(1000,1), main = 'ChiSq')
```

ChiSq

rchisq(1000, 1)
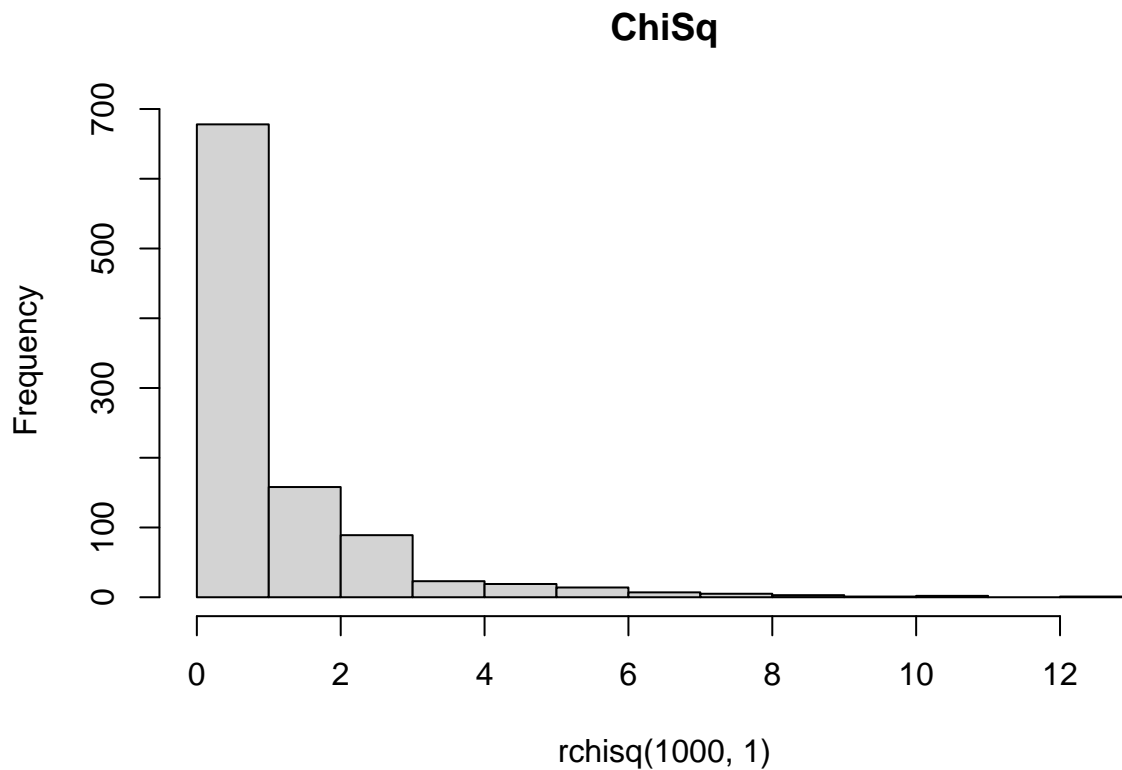
It looks like the squares of the normal distribution are very similar to the chisquared distribution with one degree of freedom.

(b) Recall (or obtain from `qnorm(0.975)` that the *2-sided 95%* critical values for the standard normal distribution are ± 1.96. Use `qchisq()` to compare the *square* of these 2-sided 95% critical values from the standard normal to the *one-sided* 95% critical value for the Chi-squared(1). What does this suggest about tests based on these two reference distributions?

```
qnorm(0.975)^2
```

```
## [1] 3.841459
```

```
qchisq(0.95,1)
```

```
## [1] 3.841459
```

The critical value of the square of the 2 sided qnorm is the same as a one-sided chisquared critical value at 1 degree of freedom. This suggests that the chisquared test is based on a one-sided distribution as default.

6