# Homework Exercises Week 07

Alison Lawyer

2024-11-05

```
# keep this chunk in all your RMarkdown scripts
knitr::opts_chunk$set(echo = TRUE)
knitr::opts_chunk$set(tidy.opts = list(width.cutoff = 60), tidy = TRUE)
```

```
# List required packages
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(openxlsx)
library(ggplot2)
library(lsr)
```

## HOMEWORK EXERCISES

For your homework exercises, you will explore another dataset that is suited for analysis with the chi-squared test statistic. The dataset "diet.xlsx" contains counts of herring gull pairs (a gull species common across the northern hemisphere) with different diets, broken down by the number of chicks hatched for each pair in one breeding season. Counts are the number of pairs in each group. Your goal is to determine whether diet was associated with hatching success in these herring gulls.

### Part 1

Import the data from file "diet.xlsx". Reshape the data to be ready for a chi-squared test.

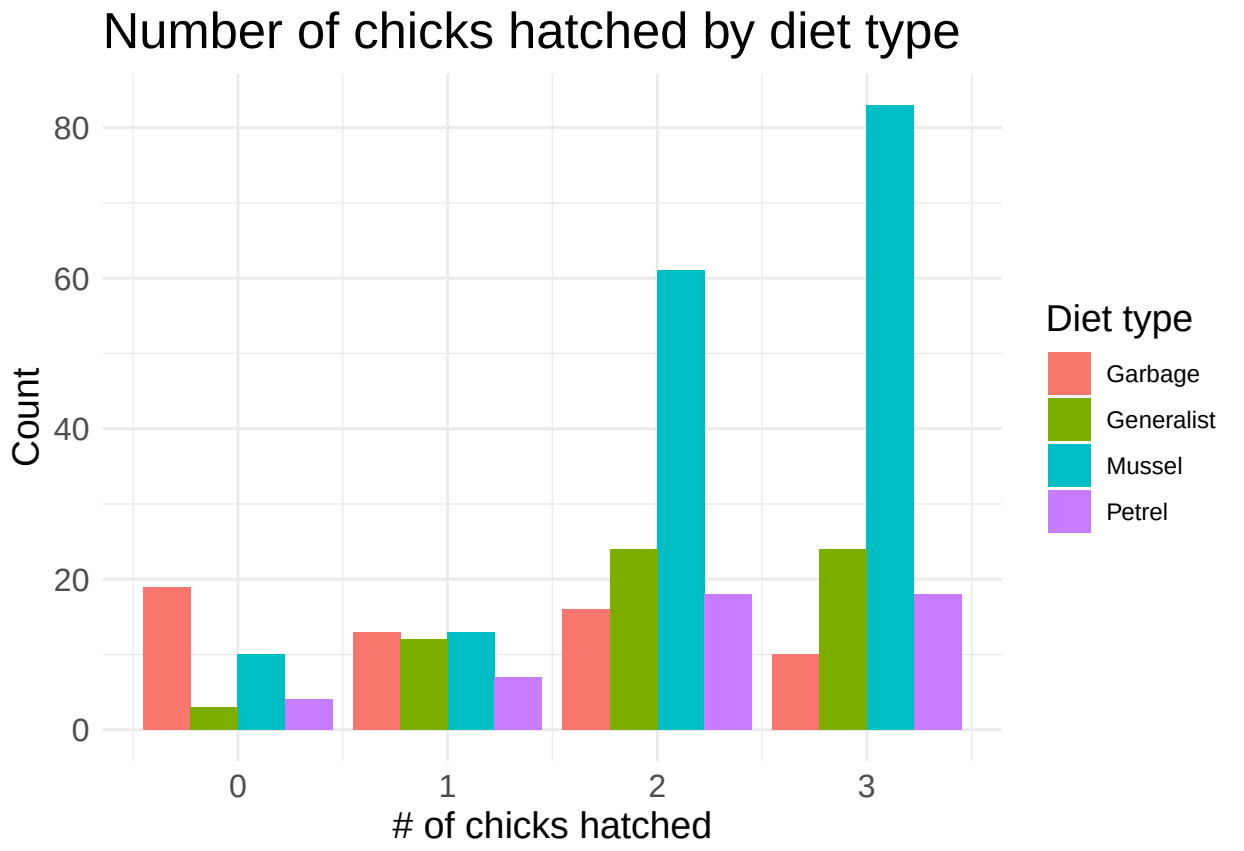
```
diet <- read.xlsx("diet.xlsx")
diet_wide <- diet %>%
    pivot_wider(names_from = Number.of.chicks.hatched, values_from = Count)
```

## Part 2

Plot the data with ggplot. Decide on the best way to visualize based on your objective. Explain why you plotted it the way you did and not another way. Be sure to fine tune your graph with proper labels, font sizes, fill colors and captions.

Answer: I chose to plot the data this way because it allows us to easily visualize the number of chicks that were hatched based on diet type. An alternative way would be to swap the x-axis and the grouping/fill variable so that diet type is on the x-axis and number of chicks hatched would be how the groups/bars are divided. That would give us a similar visualization but I found it was easier to digest the data with the number of chicks hatched on the x-axis and the grouping being diet type.

```
ggplot(diet, aes(x = Number.of.chicks.hatched, y = Count, fill = Diet)) +
    geom_col(position = "dodge") + labs(title = "Number of chicks hatched by diet type",
    x = "# of chicks hatched") + guides(fill = guide_legend(title = "Diet type")) +
    theme_minimal() + theme(axis.text = element_text(size = 12),
    axis.title = element_text(size = 14), legend.title = element_text(size = 14),
    title = element_text(size = 16))
```




```
# ggplot(diet, aes(x = Diet, y = Count, group =
# Number.of.chicks.hatched, fill =
# Number.of.chicks.hatched)) + geom_col(position = 'dodge')
# + labs(title = 'Number of chicks hatched by diet type', x
# = '# of chicks hatched') + guides(fill =
# guide_legend(title = 'Diet type')) + theme_minimal() +
# theme(axis.text = element_text(size = 12), axis.title =
```

```
# element_text(size = 14), legend.title = element_text(size
# = 14), title = element_text(size = 16))
```

## Part 3

Based on what you know, are the assumptions of a chi-squared test met? Why or why not?

Answer: Yes, the assumptions are met: the diet types are nominal variables, over 80% of the groups have values over 5, and we can assume each observation is independent of the others.

## Part 4

Conduct the test. Print the test statistic, p-value, and degrees of freedom. Interpret your findings.

Answer: The p-value of this test is incredibly small, therefore we can comfortably reject the null hypothesis. This result seems to agree with the visual representation of our data above, indicating that diet does indeed have an effect on the success of hatching chicks.

```
diet_chi_test <- chisq.test(x = diet_wide[, 2:5], correct = FALSE)
diet_chi_test
```

```
##
##  Pearson's Chi-squared test
##
## data:  diet_wide[, 2:5]
## X-squared = 53.723, df = 9, p-value = 2.131e-08
```

## Part 5

Print the standardized residuals for each observed count. What particular groups, if any, were responsible for the result of your hypothesis test above?

Answer: The group with the garbage diet appears to have significant impact on the results of the above test, as several of the values are close to or greater than +/- 1.96. The group with the mussel diet has a moderate significance (with values close to or slightly larger than +/- 1.96), but this group does not have as big of an impact as the garbage diet group. The other groups seem insignificant.

```
diet_chi_test$residuals
```

```
##               0          1          2          3
## [1,] -1.8757552 -1.9915911  0.2178122  1.9139836
## [2,] -0.4675414  0.2732438  0.3192542 -0.2160590
## [3,]  5.1138926  1.8661767 -1.0140853 -2.7661436
## [4,] -1.4489701  1.2159604  0.3426365 -0.2754821
```

## Part 6

Calculate the effect size. Interpret the result.

Answer: This is a low-to-moderate effect size. This could reflect the significant p-value, but also takes into account the majority of our residuals being insignificant.

```r
cramersV(diet_wide[, 2:5], correct = FALSE)
```

```
## [1] 0.2312056
```

## Part 7

Summarize your findings in a concise statement that includes references to the chi-squared statistic, degrees of freedom, p-value, residuals (if applicable), and effect size.

Answer: We used the chi-squared test to test the hypothesis that there was a relationship between successful chicks hatching and diet type and we found a statistically significant association (X^2 (9)=53.723, p<0.05). Based on standardized residuals, fewer chicks were hatched with the garbage diet, while more chicks were hatched with the mussel diet.