IS5 in R: Comparing Counts (Chapter 19)

Margaret Chien and Nicholas Horton (nhorton@amherst.edu)

July 17, 2018

Introduction and background

This document is intended to help describe how to undertake analyses introduced as examples in the Fifth Edition of *Intro Stats* (2018) by De Veaux, Velleman, and Bock. More information about the book can be found at http://wps.aw.com/aw_deveaux_stats_series. This file as well as the associated R Markdown reproducible analysis source file used to create it can be found at http://nhorton.people.amherst.edu/is5.

This work leverages initiatives undertaken by Project MOSAIC (http://www.mosaic-web.org), an NSF-funded effort to improve the teaching of statistics, calculus, science and computing in the undergraduate curriculum. In particular, we utilize the `mosaic` package, which was written to simplify the use of R for introductory statistics courses. A short summary of the R needed to teach introductory statistics can be found in the mosaic package vignettes (http://cran.r-project.org/web/packages/mosaic). A paper describing the mosaic approach was published in the *R Journal*: https://journal.r-project.org/archive/2017/RJ-2017-024.

Chapter 19: Comparing Counts

```
library(mosaic)
library(readr)
library(janitor)
Zodiac <- read_csv("http://nhorton.people.amherst.edu/is5/data/Zodiac.csv")
```

```
## Parsed with column specification:
## cols(
##   Month = col_character(),
##   Births = col_integer(),
##   Expected = col_double(),
##   Residual = col_double()
## )
```

By default, `read_csv()` prints the variable names. These messages can be suppressed using the `message=FALSE` code chunk option to save space and improve readability.

```
Zodiac %>%
  select(Month, Births)
```

```
## # A tibble: 12 x 2
##    Month       Births
##    <chr>        <int>
##  1 Pisces          29
##  2 Aquarius        24
##  3 Aries           23
##  4 Cancer          23
##  5 Capricorn       22
##  6 Scorpio         21
##  7 Taurus          20
##  8 Leo             20
##  9 Saggitarius     19
```

```
## 10 Virgo           19
## 11 Libra           18
## 12 Gemini          18
```

Section 19.1: Goodness-of-Fit Tests

```
# page 611
BaseBallBirths <- read_csv("http://nhorton.people.amherst.edu/is5/data/Ballplayer_births.csv") %>%
  clean_names()
```

```
## Parsed with column specification:
## cols(
##   Month = col_integer(),
##   `Ballplayer Count` = col_integer()
## )
```

Here we use the `clean_names()` function from the `janitor` package to sanitize the names of the columns (which would otherwise contain special characters or whitespace).

Section 19.2: Chi-Square Test of Homogeneity

Section 19.3: Examining the Residuals

Section 19.4: Chi-Square Test of Independence