Chisq_AK

Caleb Rother, Jacob C. Cooper

2025-08-26

```r
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   3.5.2     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- 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
```

Question 9.1

```r
# download La Selva data
laselva <- read_csv("https://zenodo.org/records/4979259/files/LaSelvaBirdTrendsDatatable.csv")
```

```
## Rows: 202 Columns: 20
## -- Column specification ---------------------------------------------------
## Delimiter: ","
## chr  (8): ScientificName, CommonName, QualChg, Prob>ChiSq, Diet_AB, Diet_BS,...
## dbl (10): MaxN, MeanAbund, NYears, TrendEstimate, StdError, L-RChiSquare, Me...
## lgl  (2): Extirpated?, Invading?
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# isolate the big groundbirds
big_groundbirds <- laselva[1:6,]
```

```r
#Test 1
bigyearsall <- chisq.test(big_groundbirds$NYears)
bigyearsall
```

```
##
##  Chi-squared test for given probabilities
##
## data:  big_groundbirds$NYears
## X-squared = 0.70769, df = 5, p-value = 0.9825
```

```r
bigyearsp <- bigyearsall$p.value
print(round(bigyearsp, 2))
```

```
## [1] 0.98
```

```
#Test 2
bigmaxall <- chisq.test(big_groundbirds$MaxN)
bigmaxall
```

```
##
##  Chi-squared test for given probabilities
##
## data:  big_groundbirds$MaxN
## X-squared = 76.447, df = 5, p-value = 4.64e-15
```

```
print(round(bigmaxall$p.value, 4))
```

```
## [1] 0
```

Part 1

We do not reject the null for different numbers of years being encountered ($\chi^2_{(5)} = 0.71, p = 0.98$).

Part 2

We do reject the null for high counts ($\chi^2_{(5)} = 76.45, p < 0.0001$).

Question 9.2

Null: There is no difference in traffic depending on time on the east sidewalk. Null: There is no difference in traffic depending on time on the west sidewalk. Null: There is no difference in traffic between east and west sidewalks.

```
bicycles <- read_csv("https://zenodo.org/records/2648564/files/Fremont_Bridge_Hourly_Bicycle_Counts_by_
```

```
## Rows: 56904 Columns: 3
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## chr (1): Date
## dbl (2): Fremont Bridge East Sidewalk, Fremont Bridge West Sidewalk
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
bicycles <- bicycles |>
  # convert to data frame
  as.data.frame() |>
  # convert to date format
  mutate(Date = as.POSIXct(Date, format = "%m/%d/%Y %I:%M:%S %p")) |>
  #subset to March 30-31, 2019
  filter(year(Date) == 2019) |>
  filter(month(Date) == 3) |>
  filter(day(Date) == 30 | day(Date) == 31)
march_31 <- bicycles |>
  filter(day(Date) == 31) |>
  filter(hour(Date) > 7 & hour(Date) < 18)
```

```
#Eastbound
eastall <- chisq.test(march_31$`Fremont Bridge East Sidewalk`)
eastall
```

```
##
##  Chi-squared test for given probabilities
##
## data:  march_31$`Fremont Bridge East Sidewalk`
## X-squared = 98.542, df = 9, p-value < 2.2e-16
```

Part 1

Eastbound traffic is not equally likely at all times of day ($\chi^2_{(9)} = 98.54, p < 0.0001$). We reject H_0.

Part 2

```
#Westbound
westall <- chisq.test(march_31$`Fremont Bridge West Sidewalk`)
westall
```

```
##
##  Chi-squared test for given probabilities
##
## data:  march_31$`Fremont Bridge West Sidewalk`
## X-squared = 106.81, df = 9, p-value < 2.2e-16
```

```
print(round(westall$p.value, 4))
```

```
## [1] 0
```

Westbound traffic is not equally likely at all times of day ($\chi^2_{(9)} = 106.81, p < 0.0001$). We reject H_0.

Part 3: Method 1 (Table)

```
#Compared
table_dir <- rbind(East = march_31$`Fremont Bridge East Sidewalk`,
                   West = march_31$`Fremont Bridge West Sidewalk`)
eastwestall <- chisq.test(table_dir, correct=FALSE)
eastwestall
```

```
##
##  Pearson's Chi-squared test
##
## data:  table_dir
## X-squared = 3.2449, df = 9, p-value = 0.9538
```

Traffic at a particular time of day is equally likely between traffic directions ($\chi^2_{(9)} = 3.24$, $p = 0.95$). We fail to reject H_0.

Part 2: Method 2 (putting each separately)

```
chisq.test(march_31$`Fremont Bridge East Sidewalk`, march_31$`Fremont Bridge West Sidewalk`)
```

```
## Warning in chisq.test(march_31$`Fremont Bridge East Sidewalk`,
## march_31$`Fremont Bridge West Sidewalk`): Chi-squared approximation may be
## incorrect
```

```
##
##  Pearson's Chi-squared test
##
```

```
## data:  march_31$`Fremont Bridge East Sidewalk` and march_31$`Fremont Bridge West Sidewalk`
## X-squared = 90, df = 81, p-value = 0.2313
```

Traffic at a particular time of day is equally likely between traffic directions ($\chi^2_{(81)} = 90$, $p = 0.23$). We fail to reject H_0.

Question 9.3

Method 1: Table

```
march_30 <- bicycles  |>
  filter(day(Date) == 30, hour(Date) > 7 & hour(Date) < 18) |>
  select(-`Fremont Bridge West Sidewalk`) |>
  mutate(Date = hour(Date)) |>
  rename("March 30" = `Fremont Bridge East Sidewalk`)

march_31_east <- march_31 |>
  select(-`Fremont Bridge West Sidewalk`) |>
  mutate(Date = hour(Date)) |>
  rename("March 31" = `Fremont Bridge East Sidewalk`)

date_compare <- full_join(march_31_east, march_30, by = "Date") |>
  rename("Hour" = Date) |>
  arrange(Hour)

# Create table for chi-sq
date_table <- rbind(`March 31` = date_compare$`March 31`,
                    `March 30` = date_compare$`March 30`)

chisq_date <- chisq.test(date_table)
chisq_date
```

```
##
##  Pearson's Chi-squared test
##
## data:  date_table
## X-squared = 3.9253, df = 9, p-value = 0.9163
```

We fail to reject the null. There is no difference between days with respect to traffic ($\chi^2_{(9)} = 3.93$, $p = 0.92$).

Method 2: Listing values

```
chisq.test(date_compare$`March 31`, date_compare$`March 30`)
```

```
## Warning in chisq.test(date_compare$`March 31`, date_compare$`March 30`):
## Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  date_compare$`March 31` and date_compare$`March 30`
## X-squared = 80, df = 72, p-value = 0.2424
```

We fail to reject the null. There is no difference between days with respect to traffic ($\chi^2_{(72)} = 80$, $p = 0.24$).