Week3_Lab_Assignment

Summer Heschong

2025-02-04

#Setup

```
library(here)
```

```
## here() starts at /Users/summerheschong/Documents/GitHub/stats_spring25
```

```
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 1.9.4     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(moments)
```

#(1) Discrete Probability Distributions

##a. Say you flip a fair, unweighted coin 20 times. What is the probability of obtaining 5 or fewer heads?

```
#probability of 5 or fewer successes
pbinom(q = 3, size = 20, prob = 0.50)
```

```
## [1] 0.001288414
```

The probability of obtaining 5 or fewer heads is 0.12%

##b. Now, you switch to using an unfair, weighted coin whose probability of landing on heads is 80%. If you flip the new, weighted coin 15 times, what is the probability of obtaining 10 or more heads?

```
#calculate probability of 9 or fewer successes and subtract from 1 to get
#probability of 10 or more successes
1 - pbinom(q = 9, size = 15, prob = 0.80)
```

```
## [1] 0.9389486
```

The probability of obtaining 10 or more heads is 94%

##c. What is the probability that only 6 whale strikes will occur this coming year? What is the probability that 6 or fewer whale strikes will occur this coming year?

```
#probability of only 6 whale strikes
dpois(x = 6, lambda = 12)
```

```
## [1] 0.02548128
```

```
#probability of 6 or fewer whale strikes
ppois(q = 6, lambda = 12)
```

```
## [1] 0.04582231
```

The probability of only 6 whale strikes occurring in the coming year is 2.5% The probability of 6 or fewer whale strikes is 4.6%

#(2) Assessing Normality

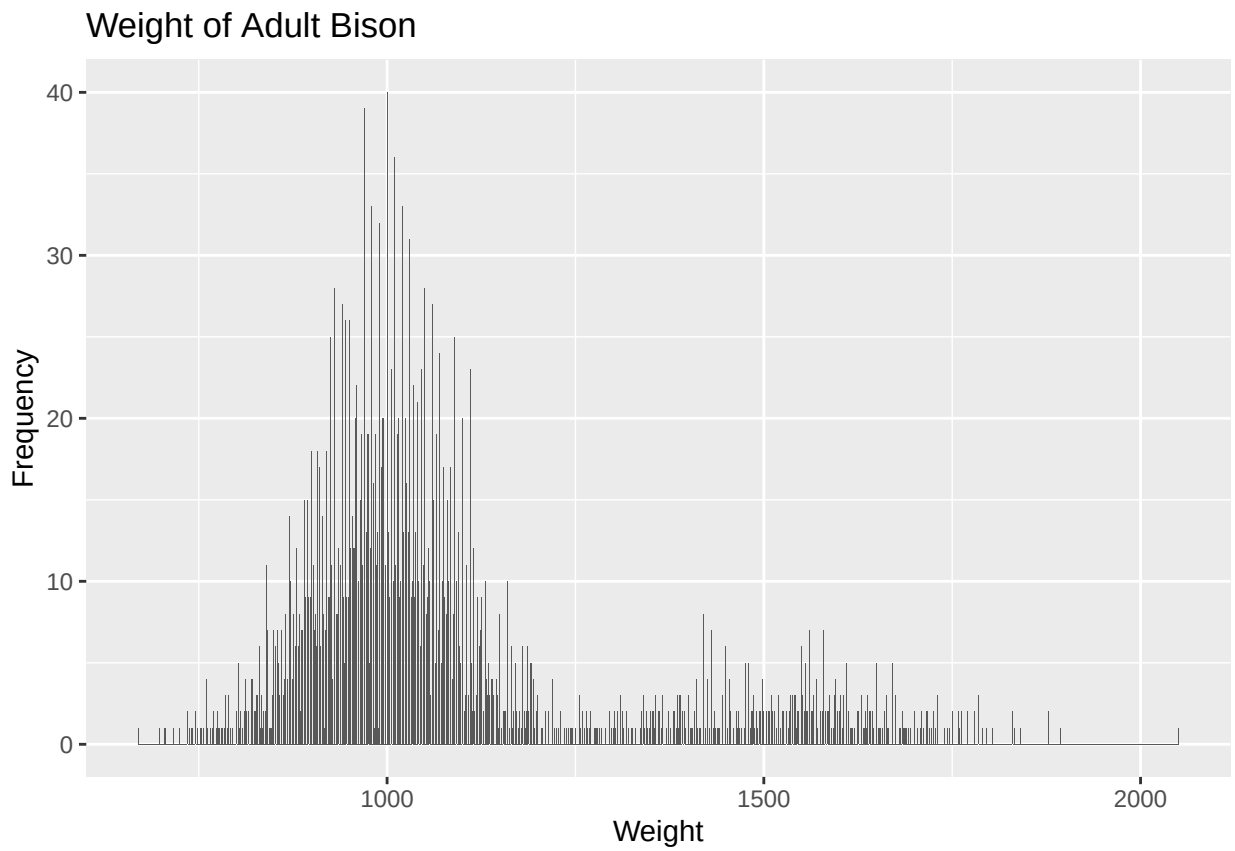
##a. Are weights of adult bison normally distributed?

```
#load in data frame
bison_data_raw <- read_csv(here('Data/Raw/knz_bison.csv'))
```

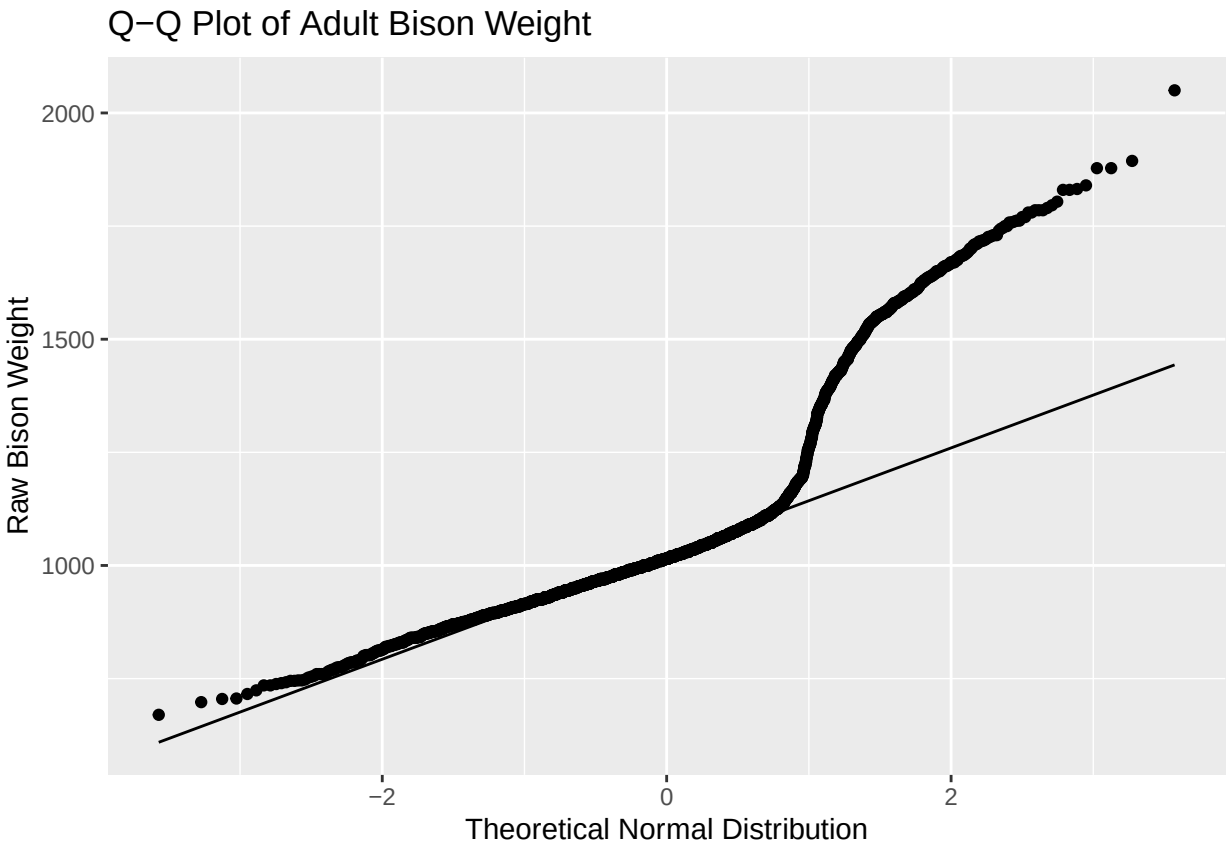
```
## Rows: 8325 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): data_code, animal_code, animal_sex
## dbl (5): rec_year, rec_month, rec_day, animal_weight, animal_yob
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#create column displaying bison age, and filter for adult bison (age >3)
bison_data <- bison_data_raw %>%
  mutate(animal_age = rec_year - animal_yob) %>%
  filter(animal_age > 3)

#create histogram of adult bison weight
#(number of bins = (2)2820^1/3 = 1880, bin width = 1380/1880 = 0.7340426)
fig1 <- ggplot(data = bison_data, aes(x = animal_weight)) +
  geom_histogram(bins = 1880, binwidth = 0.73) +
  labs(title = 'Weight of Adult Bison',
       x = 'Weight',
       y = 'Frequency')
print(fig1)
```




```
#Create QQ plot of adult bison weight
fig2 <- ggplot(data = bison_data, aes(sample = animal_weight)) +
  geom_qq() +
  geom_qq_line() +
  labs(title = "Q-Q Plot of Adult Bison Weight",
       x = 'Theoretical Normal Distribution',
       y = 'Raw Bison Weight')
print(fig2)
```



##b. Calculate and report the skewness and kurtosis values for the weight of adult bison

```
#calculate skewness
skewness(bison_data$animal_weight)
```

```
## [1] 1.531318
```

```
#calculate kurtosis
kurtosis(bison_data$animal_weight)
```

```
## [1] 4.764673
```

The skewness for adult bison weight is 1.531318, and the kurtosis is 4.764673.

##c. Why would you or would you not assume these data are normally distributed?

Answer: I would not assume that these data are normally distributed. First, from the information provided we know that bison are sexually dimorphic and males weigh more than females which could contribute to bimodal data. Second, in the figures above, the data does not appear normal. And, finally, the absolute value of the skewness is >1 which indicates high skewness, and the kurtosis is >3 making it leptokurtic.

#Continuous Probability Distributions

##a. What are the mean and standard deviation values for weights of adult, female bison?

```r
#filter for female bison
female_bison_data <- filter(bison_data, animal_sex == 'F')

#mean weight
mean(female_bison_data$animal_weight)
```

```
## [1] 992.5751
```

```r
#standard deviation weight
sd(female_bison_data$animal_weight)
```

```
## [1] 90.00992
```

The mean adult, female bison weight is 992.5751, and the standard deviation is 90.00992.

##b. What is the probability of capturing an adult, female bison weighing less than 900 lbs? Weighing between 900 and 1100 lbs? What is the value below which 95% of adult female bison are expected to weigh?

```r
#Prob of female less than 900lbs
pnorm(q = 900, mean = 993, sd = 90)
```

```
## [1] 0.150724
```

```r
#Prob of female between 900 and 1100lbs?
#prob of female less than 1100lbs?
pnorm(q = 1100, mean = 993, sd = 90)
```

```
## [1] 0.8827583
```

```r
#subtract smaller prob from larger
0.8827583 - 0.150724
```

```
## [1] 0.7320343
```

```r
#value below which 95% are expected to weigh
qnorm(p = .95, mean = 993, sd = 90)
```

```
## [1] 1141.037
```

The probability of capturing an adult, female bison weighing less than 900lbs is 15%. The probability of capturing an adult, female bison weighing between 900 and 1100 lbs is 73%. The value below which 95% of adult, female bison are expected to weigh is 1141 lbs.