MA2611 Lab 3

Xin Yu Wu

2024-09-20

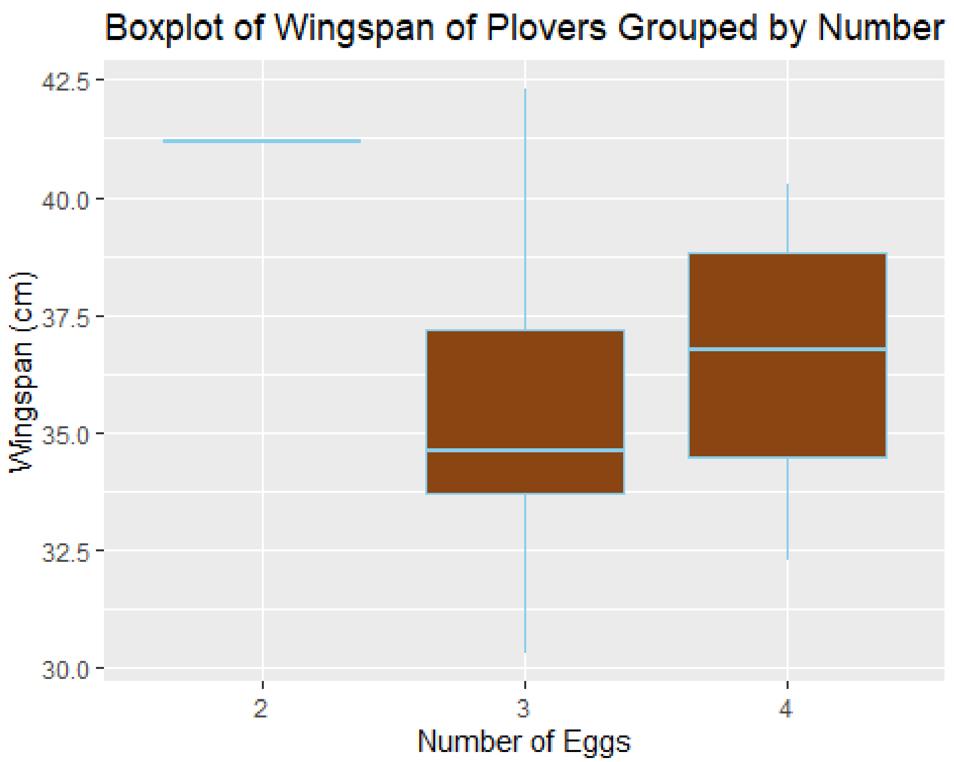
```r
if(!require("ggplot2")){
install.packages("ggplot2")
require ("ggplot2")
}
```

```
## Loading required package: ggplot2
```

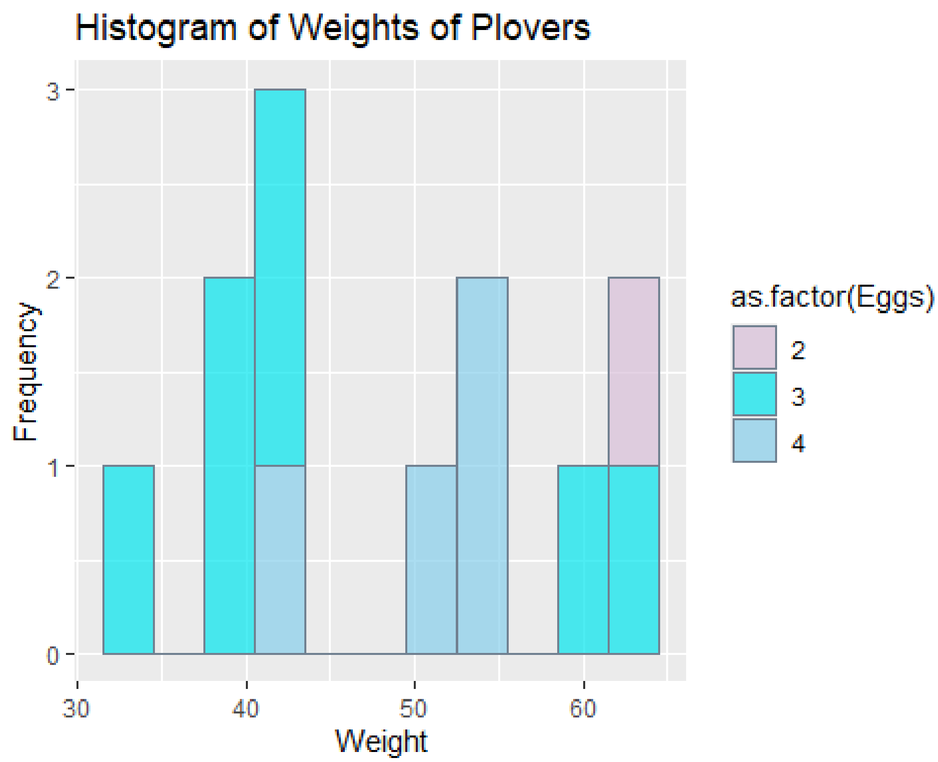
```r
  library(ggplot2)

library(knitr)


plovers <- read.csv("plovers.csv",header=TRUE)
```

Problem L.10

```r
#a)
ggplot(data = plovers, aes(x = as.factor(Eggs), y = Wingspan, fill =
as.factor(Eggs))) +
  geom_boxplot(fill = "chocolate4", color = "skyblue") +
  labs(title = "Boxplot of Wingspan of Plovers Grouped by Number of Eggs",
       x = "Number of Eggs",
       y = "Wingspan (cm)")
```




```r
#b)
ggplot(data = plovers, aes(x = Weight, fill = as.factor(Eggs)) ) +
  geom_histogram(binwidth = 3, color = "slategray", alpha = 0.7) +
  scale_fill_manual(values = c("thistle", "turquoise2", "skyblue")) +
labs(title="Histogram of Weights of Plovers",
     x="Weight",
     y="Frequency")
```

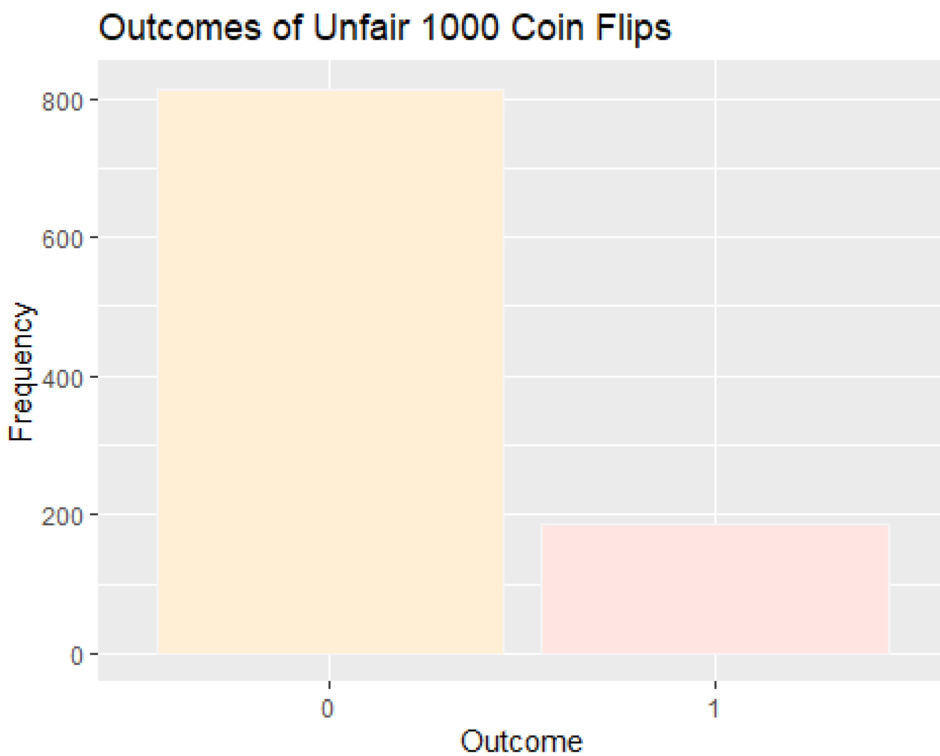


#c) Based on the data from the box plot and histogram from parts a and b, it appears that both wingspan and weight vary with the number of eggs in the plover nest. Moreover, there is a more definitive conclusion that could be reached with a larger data set. The box plot shows significant differences in wingspan, as evidenced by the variation in means across the plots. Additionally, the histograms reveal that the weights of plovers also differ according to the egg count, indicated by the varying skewness of the distributions.

Problem L.11

```r
#a)
coin.1000 <-sample(x=c(0,1),size=1000,replace=TRUE,prob=c(0.8,0.2))
coin.data <-data.frame(Values=table(coin.1000),Names=as.factor(c("0","1")))

ggplot(coin.data,aes(x=Names,y=Values.Freq)) +
geom_col(fill=c("papayawhip","mistyrose"),
colour="whitesmoke") +
labs(title="Outcomes of Unfair 1000 Coin Flips",x="Outcome",y="Frequency")
```

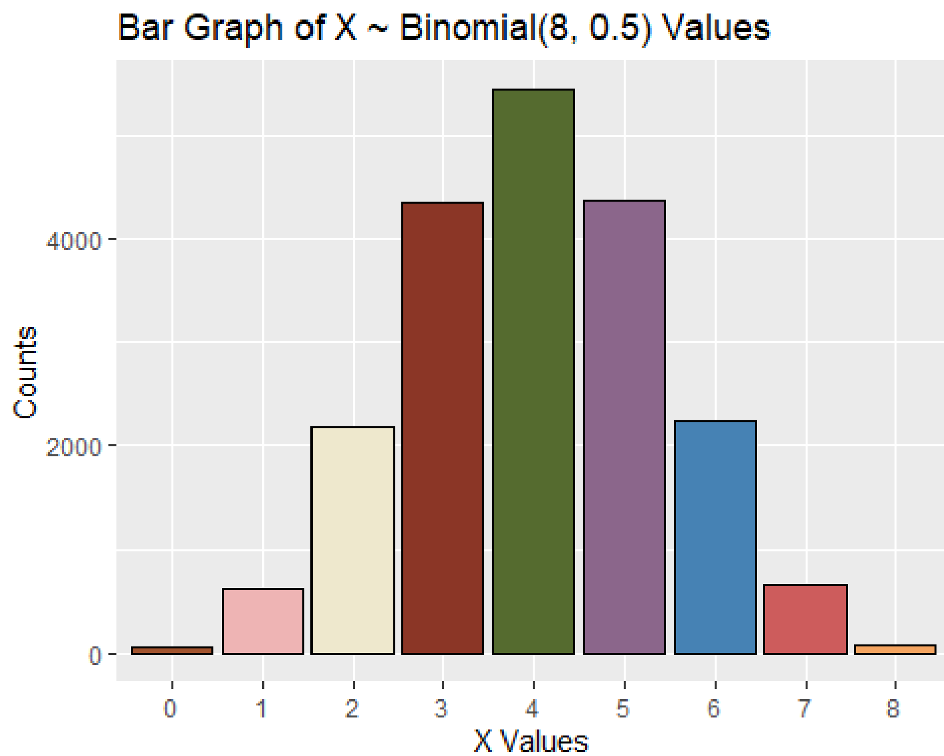


Problem L.11 and L.12

```
#c)
X_bin <- rbinom(n = 20000, size = 8, prob = 0.5)
X.data <-data.frame(Values=table(X_bin),Names=as.factor(c("0","1","2","3",
"4","5","6","7","8")))
X.data

##   Values.X_bin Values.Freq Names
## 1            0          60     0
## 2            1         632     1
## 3            2        2178     2
## 4            3        4344     3
## 5            4        5438     4
## 6            5        4363     5
## 7            6        2246     6
## 8            7         668     7
## 9            8          71     8

ggplot(X.data,aes(x=Names,y=Values.Freq)) +
geom_col(fill=c("sienna","rosybrown2","cornsilk2","tomato4","darkolivegreen",
"plum4","steelblue","indianred","sandybrown"),
colour="black") +
labs(title="Bar Graph of X ~ Binomial(8, 0.5) Values",x="X
Values",y="Counts")
```




```
``` r
#a)

#Exact probability using sampled data:

mean(X_bin == 5)

## [1] 0.21815

mean(X_bin >= 4 & X_bin < 7)

## [1] 0.60235

#Exact probability using binom distribution:

sum(X_bin == 5)/20000

## [1] 0.21815

#The probability P is 4 <= X < 7
prob4 <- sum(X_bin == 4)/20000
prob5 <- sum(X_bin == 5)/20000
prob6 <- sum(X_bin == 6)/20000
probCombine <- prob4 + prob5 + prob6

print(probCombine)
```
```

```
## [1] 0.60235
```

#a) The estimated probability using sampled data is: P(X=5) = 0.21685 P(4 <= X < 7) = 0.6079. The estimated probability using binomial distribution is: P(X=5) = 0.21685 P(4 <= X < 7) = 0.6079. In this case, the estimated values are the same as the exact probabilities down to the 4th decimal place.

```
#b)
#sample
expected_val = mean(X_bin)
standard_dev = sd(X_bin)
print(expected_val)

## [1] 4.01535

print(standard_dev)

## [1] 1.417679

#exact using formula
Exact_expected_val = 8 * 0.5
Exact_standard_deviation <- sqrt(8 * 0.5 * (1 - 0.5))
print(Exact_expected_val)

## [1] 4

print(Exact_standard_deviation)

## [1] 1.414214
```

#b) The expected value for X using the randomly sampled data is 4.0042. The standard deviation for X using the randomly sampled data is 1.410738. The exact expected value for X is 4.0. The exact standard deviation for X is 1.414214. The estimated and exact values are very close, but not exact, with a thousandth decimal difference.