Midterm

Alizeh Azhar

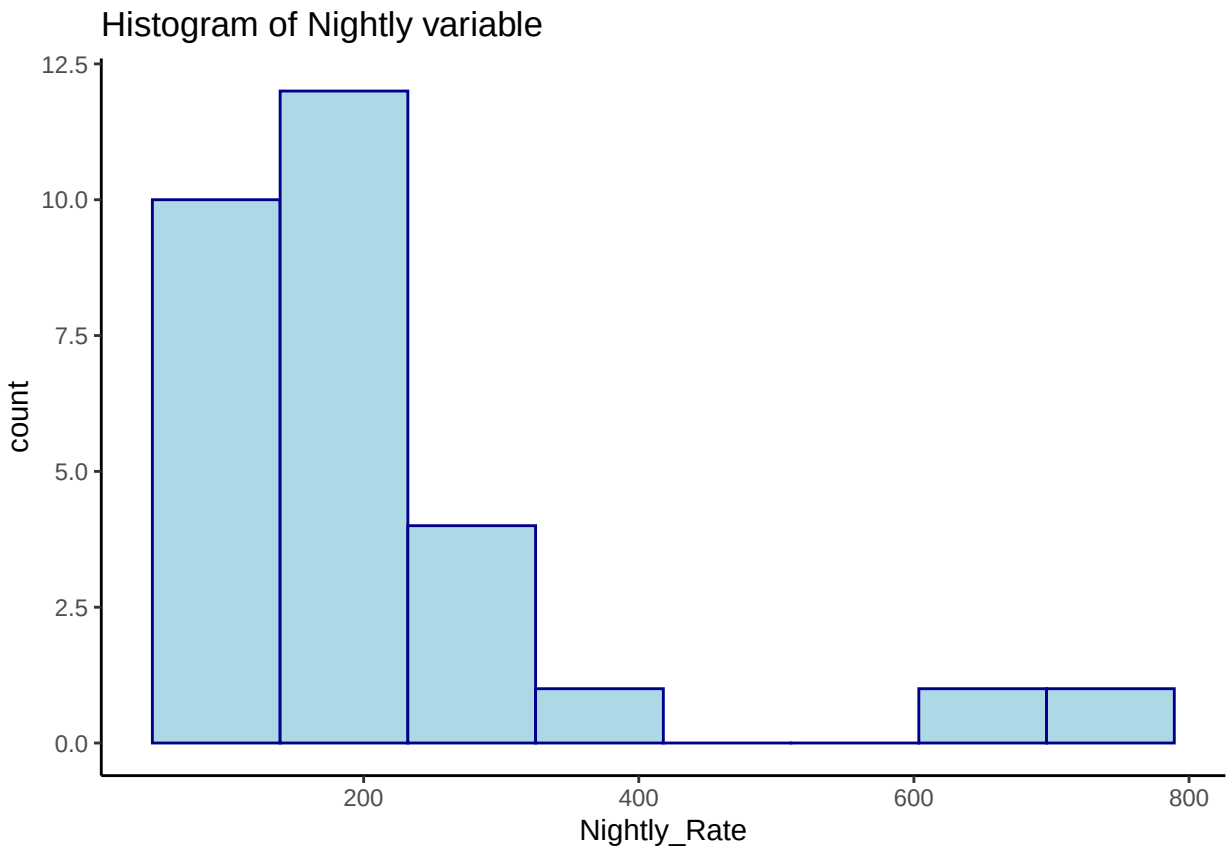
10/23/2021

Please complete this midterm using RMarkdown. You may use your notes, the textbook, homework solutions, etc. when working on the problems–the only thing you may not do is discuss the midterm with your classmates or others. If you have any questions, please email me.

Hotel Prices (7 pts)

(a) Using `ggplot()`, create a histogram of `Nightly_Rate`. Your histogram should have 8 bins (3 pts).

```
library(ggplot2)
ggplot(data = lodging_prices, aes(Nightly_Rate))+
  geom_histogram(color="darkblue",fill="lightblue",bins = 8)+
  ggtitle("Histogram of Nightly variable")+
  theme_classic()
```

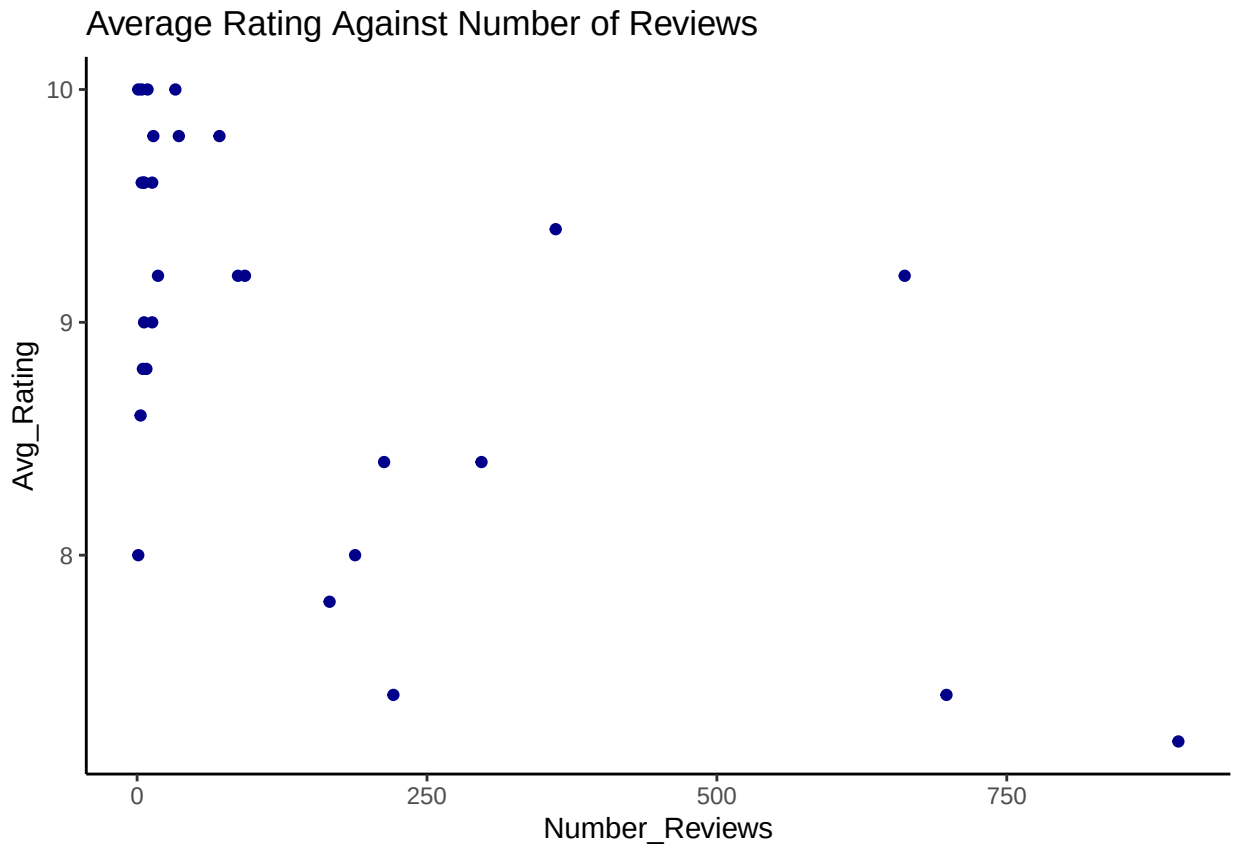


(b) Comment on the distribution of nightly rates (4 pts)

The distribution of nightly rates is skewed to the right because most of the data are on the left side histogram but few larger values on the right.

(c) Using `ggplot()`, create a scatterplot of `Avg_Rating` against `Number_Reviews`. Change the color of the points to a color of your choice (3 pts).

```
library(ggplot2)
ggplot(data = lodging_prices, aes(y=Avg_Rating, x=Number_Reviews))+
  geom_point(color="darkblue")+ ggtitle("Average Rating Against Number of Reviews")+
  theme_classic()
```



(d) Comment on the relationship (4 pts).
The data show a downhill pattern as you move from left to right. This indicates that there ia negative relationship between average rating and number of reviews.

(e) Calculate the correlation between `Number_Reviews` and `Avg_Rating` advertising dollars (1 pt).

```
cor(lodging_prices$Avg_Rating,lodging_prices$Number_Reviews)
```

```
## [1] -0.5838885
```

Moose Tours (18 points)

(a) How many tours should the manager expect to run before the tour is unsuccessful (that is, they do not see a moose)?

```
36*(1-0.97)
```

```
## [1] 1.08
```

(b) Pemi Valley Moose Tours plans on carrying out 36 tours in the month of October (one day per week, twice each Saturday and Saturday, no tours on Mondays). What is the probability that more than 4 of their tours are unsuccessful?

```
pbinom(4, size = 36, prob = 0.03, lower.tail = F)
```

```
## [1] 0.004217609
```

(c) What is the probability that all of their tours are successful?

```
dbinom(x=36, size = 36, prob = 0.97)
```

```
## [1] 0.3340277
```

(d) What is the probability that there are exactly 3 tours before the first unsuccessful tour?

```
sum(dbinom(1:3, size = 36, prob = 0.97))
```

```
## [1] 3.632457e-47
```

(e) What is the probability that there are at least 8 succesful tours before the first unsuccessful tour?

```
pbinom(7, size = 36, prob = 0.97, lower.tail = F)
```

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

(f) How many tours of the 36 in October should Pemi Valley Moose Tours expect to be successful?

```
36*0.97
```

```
## [1] 34.92
```

Fall Foliage Activities (15 points)

(a) What is the probability that a randomly chosen tourist visits Artist's Bluff (2 pts)?

```
58/192
```

```
## [1] 0.3020833
```

(b) What is the probability that a randomly chosen tourist drove on the Kancamagus Highway (2 pts)?

```
160/192
```

```
## [1] 0.8333333
```

(c) What is the probability that a randomly chosen tourist drove on the highway and did not do any of the hikes in the state park (2 pts)?

```
10/192
```

```
## [1] 0.05208333
```

(d) What is the probability that a randomly chosen tourist drove on the highway, given that they hiked Flume Gorge (2 pts)?

```
80/98
```

```
## [1] 0.8163265
```

(e) What is the probability that a randomly chosen customer did not drive the highway, given that they hiked the Basin (2 pts)?

```
4/24
```

```
## [1] 0.1666667
```

(f) Are driving on the highway and hiking Flume Gorge independent? Give a mathematical justification (3 pts).

```
80/192
```

```
## [1] 0.4166667
```

```
98/192*160/192
```

```
## [1] 0.4253472
```

$P(drove \cap FlumeGorge) = 0.4$ and $P(drove) \cdot P(FlumeGorge) = 0.4$ Since the two probabilities are equal, driving on the highway and hiking Flume Gorge are independent.

Pancakes for Breakfast (15 points)

(a) What is the probability that exactly 100 customers come in during one hour?

```
dpois(100, lambda = 87.5)
```

```
## [1] 0.01698378
```

(b) What is the probability that no more than 105 customers during one hour?

```
ppois(105, lambda = 87.5)
```

```
## [1] 0.9699913
```

(c) The owner takes a random sample of 42 hours during the month of October (because of the fall foliage, October is always a busy time). What is the distribution of the customers arriving each hour?

$$\sim B\left(n = 3675, p = \frac{1}{42}\right)$$

(d) What is the probability that the average number of customers is between 82 and 86?

```
ppois(86, lambda = 87.5)- ppois(82, lambda = 87.5)
```

```
## [1] 0.163516
```

(e) What is the probability that the average number of customers is at least 90?

```
ppois(90, lambda = 87.5, lower.tail = F)
```

```
## [1] 0.3681916
```

More Pancakes (15 points)

(a) What is the probability that a single bill is more than $65?

```
1- pnorm(65, mean = 81, sd= 32)
```

```
## [1] 0.6914625
```

(b) Service at Polly's is very good, and every customer tips exactly 20%. Thus, the actual cost of the meal will be 20% higher than the bill. If a bill is represented by B, write an expression that represents the actual cost of the meal.

$$B \sim N(\mu = 1.2 \times 81, \sigma = 1.2 \times 32)$$

Implying that:

$$B \sim N(\mu = 97.2, \sigma = 38.4)$$

(c) How much should customers expect their meal to cost with tips? What is the standard deviation of cost with tips?

```
average <- 1.2*81; average
```

```
## [1] 97.2
```

```
st_deviation <- 1.2*32; st_deviation
```

```
## [1] 38.4
```

The expected daily sales is 97.2 and the standard deviation of the codes is 38.4

(d) What is the probability that the cost of a meal is between $55 and $115?

```
 pnorm(115, mean = average, sd= st_deviation)- pnorm(55, mean = average, sd= st_deviation)
```

```
## [1] 0.5426188
```

(e) What is the probability that a meal costs at most $40?

```
 pnorm(40, mean = average, sd= st_deviation)
```

```
## [1] 0.06816691
```

T-Shirt Sales (8 points)

(a) Find the value of $P(\$30.96)$ that will make this distribution a valid distribution (2 pts).

```
1-(0.16+0.12+0.22+0.22+0.08)
```

```
## [1] 0.2
```

(b) What is $E(S)$ (2 pts)?

```
a <- c(0,17.98,30.96,43.94,56.92,69.90)
b <- c(0.16,0.12,0.2,0.22,0.22,0.08)
exp <- sum(a*b);exp
```

```
## [1] 36.1308
```

(c) What is $Var(S)$ (3 pts)?

```
variance <- sum((a-exp)^2*b)
variance
```

```
## [1] 453.4783
```

(d) What is the probability a customer buys at least two shirts (1 pt)?

```
1-(0.16+0.12)
```

```
## [1] 0.72
```

Mount Washington (14 points)

(a) Build a 99% confidence interval for the true proportion of days with hurricane-force wind gusts at the top of Mount Washington (9 pts).

```
prop.test(x=13, n=40, conf.level = 0.99)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  13 out of 40, null probability 0.5
## X-squared = 4.225, df = 1, p-value = 0.03983
## alternative hypothesis: true p is not equal to 0.5
## 99 percent confidence interval:
##  0.1623200 0.5404691
## sample estimates:
##     p
## 0.325
```

(b) What is the margin of error of the confidence interval (3 pts)?

```
se <- sqrt((13/40*(1-13/40))/40)
z_star <- qnorm(0.995, mean = 0, sd=1)
MoE <- z_star*se
MoE
```

```
## [1] 0.1907571
```

(c) If the researchers choose to report a 90% confidence instead of 99%, will the interval become wider or narrower (2 points)?
The interval become narrower