Midterm

Alizeh Azhar

10/22/2021

Holiday Lights (18 points)

a. How many bulbs should the manager expect to inspect before they find a burnt out bulb?

```
0.06*20
```

```
## [1] 1.2
```

b. A string of lights has twenty bulbs. What is the probability that less than seven are burnt out?

```
pbinom(7, size = 20, prob = 0.06)
```

```
## [1] 0.999989
```

c. What is the probability that exactly three of the twenty bulbs on a string of lights are burnt out?

```
dbinom(3, size = 20, prob = 0.06)
```

```
## [1] 0.08600667
```

d. What is the probability that the manager finds their first burnt out bulb on the tenth inspection?

```
dgeom(10, prob = 0.06)
```

```
## [1] 0.03231691
```

e. What is the probability that the manager inspects at least twelve bulbs before they find one that is burnt out

```
1- pbinom(12, size = 20, prob = 0.06)
```

```
## [1] 6.780576e-12
```

f. How many bulbs out of twenty on a string of lights should the manager expect to be burnt out?

```
20*0.06
```

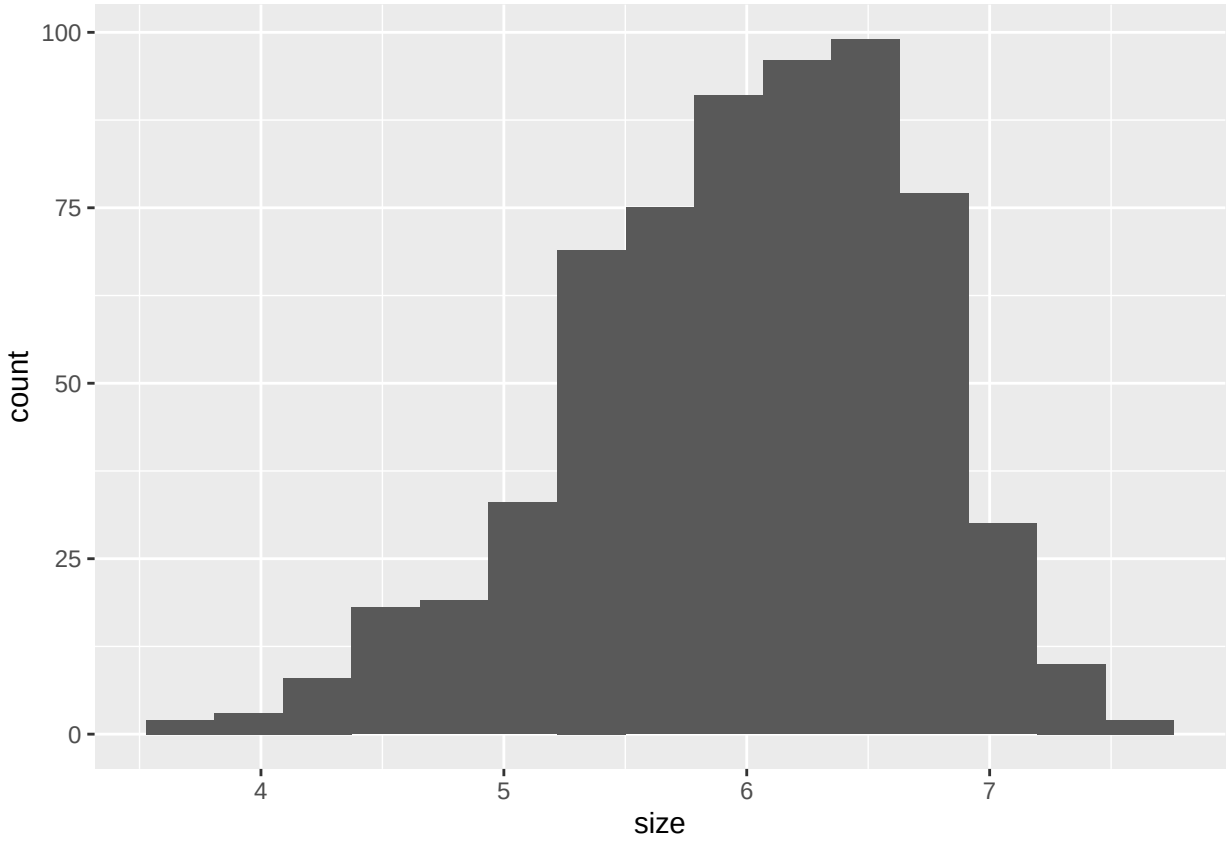
```
## [1] 1.2
```

Tree Sizes (7 pts)

```
trees <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/geepack/sitka89.csv")
```

a. Using `ggplot()`, create a histogram of size, the size of the trees at the farm as measured in $log(height \times diameter^2)$. Your histogram should have 15 bins (3 pts).

```
library(ggplot2)
ggplot(data = trees, aes(size))+ geom_histogram(bins = 15)
```



b. Comment on the distribution of the tree sizes (4 pts).

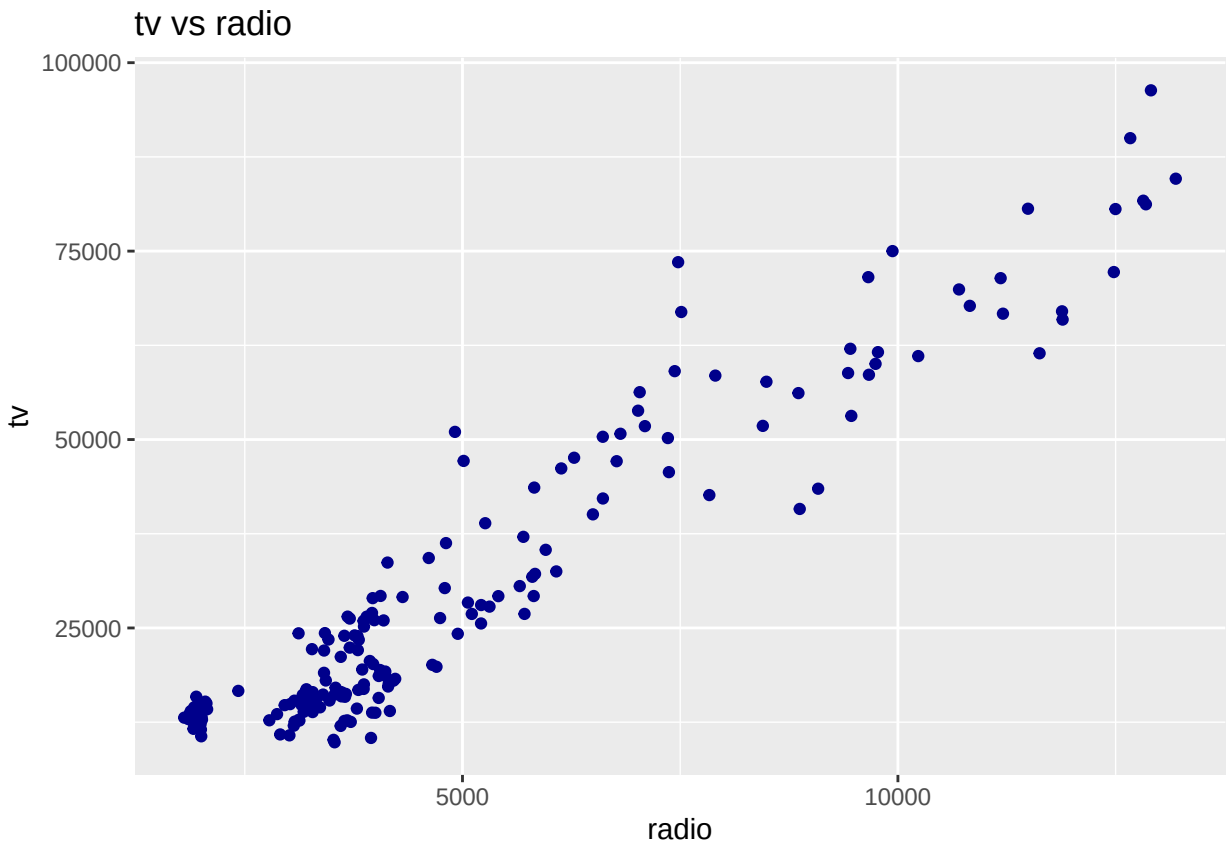
The distribution of the tree sizes is left skewed because most of the data is in the right side of the histogram. It is negatively skewed

Advertising (8 points)

```
advertising <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/AER/DutchAdvert.csv")
```

a. Using `ggplot()`, create a scatterplot of tv vs. radio advertising dollars. Change the color of the points to a color of your choice (3 pts).

```
library(ggplot2)
ggplot(data = advertising, aes(radio, tv))+ geom_point(color="darkblue")+ ggtitle("tv vs radio")
```



b. Comment on the relationship (4 pts).

The relationship is linear. If tv increases, radio also increases.

c. Calculate the correlation between tv and radio advertising dollars (1 pt).

```
cor(advertising$tv, advertising$radio)
```

```
## [1] 0.9491555
```

Bookstore Coupons (15 points)

a. What is the probability that exactly five shoppers come in to use the coupon?

```
dpois(5, lambda = 3.5)
```

```
## [1] 0.1321686
```

b. What is the probability that no more than two customers come in to use the coupon?

```
ppois(2, lambda = 3.5)
```

```
## [1] 0.3208472
```

c. The coupon is valid for the entire month of December (31 days). What is the distribution of the average number of customers using the coupon?

Average number of customers $\sim N(\mu = 3.5, \sigma = \sqrt{3.5})$

d. What is the probability that the average number of coupon users is between two and four?

```
pnorm(4, mean = 3.5, sd=sqrt(3.5))- pnorm(2, mean = 3.5, sd=sqrt(3.5))
```

```
## [1] 0.3940269
```

e. What is the probability that the average number of coupon users is at least three?

```
1- pnorm(3,  mean = 3.5, sd=sqrt(3.5))
```

```
## [1] 0.605366
```

Bookstore Sales (15 points)

(a) What is the probability that the daily sales are more than $521?

```
1- pnorm(521, mean = 478, sd= 49)
```

```
## [1] 0.1900937
```

b. The coupon the bookstore owner is planning to offer is a 10% off coupon–that is, customers will pay 90% of the total price. If the daily sales are represented by B, write an expression that represents the daily sales if all customers use the coupon.

$$B \sim N(\mu = 0.9 \times 478, \sigma = 0.9 \times 49)$$

c. How much should the bookstore owner expect the daily sales to be with the coupon? What is the standard deviation of sales with the coupon?

```
average <- 0.9*478; average
```

```
## [1] 430.2
```

```
st_deviation <- 0.9*49; st_deviation
```

```
## [1] 44.1
```

The expected daily sales is 430.2 and the standard deviation of the codes is 44.1

d. What is the probability that the daily sales are at least $400?

```
1- pnorm(400, mean = average, sd= st_deviation)
```

```
## [1] 0.7532672
```

e. What is the probability that the daily sales are at most $386?

```
pnorm(386, mean = average, sd= st_deviation)
```

```
## [1] 0.1581072
```

Coffee Specials (15 points)

(a) What is the probability that a randomly chosen customer bought a latte (2 pts)?

```
144/288
```

```
## [1] 0.5
```

(b) What is the probability that a randomly chosen customer bought gingerbread (2 pts)?

```
20/288
```

```
## [1] 0.06944444
```

(c) What is the probability that a randomly chosen customer bought a mocha and panettone (2 pts)?

```
16/288
```

```
## [1] 0.05555556
```

(d) What is the probability that a randomly chosen customer bought a cappuccino, given that they bought a sugar cookie (2 pts)?

```
25/150
```

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

(e) What is the probability that a randomly chosen customer bought panettone, given that they bought a cappuccino (2 pts)?

```
8/48
```

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

(f) Are lattes and gingerbread independent? Give a mathematical justification (3 pts).

```
45/288
```

```
## [1] 0.15625
```

```
144/288*90/288
```

```
## [1] 0.15625
```

$P(lattes \cap gingerbread) = 0.15625$ and $P(lattes) \cdot P(gingerbread) = 0.15625$
Since the two probabilities are equal, lattes and gingerbread are independent.

Stationary Sale (8 points)

(a) Find the value of $P(\$25.96)$ that will make this distribution a valid distribution (2 pts).

```
1-(0.16+0.18+0.26+0.26)
```

```
## [1] 0.14
```

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

```
a <- c(0,7.99,13.98,19.97,25.96)
b <- c(0.18,0.16,0.26,0.26,0.14)

exp <- sum(a*b);exp
```

```
## [1] 13.7398
```

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

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

```
## [1] 70.28407
```

(d) What is the probability a customer buys at least one box of stationary (1 pt)?

```
1-0.18
```

```
## [1] 0.82
```

Customer Plans (14 points)

(a) Build a 90% confidence interval for the true proportion of shoppers who are planning on doing at least some of their holiday shopping in the town's shopping district (9 pts)

```
prop.test(x=191, n=324, conf.level = 0.9)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  191 out of 324, null probability 0.5
## X-squared = 10.028, df = 1, p-value = 0.001542
## alternative hypothesis: true p is not equal to 0.5
## 90 percent confidence interval:
##  0.5424375 0.6350418
## sample estimates:
##         p
## 0.5895062
```

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

```
se <- sqrt((191/324*(1-192/324))/324)
z_star <- qnorm(0.95, mean = 0, sd=1)
MoE <- z_star*se
MoE
```

```
## [1] 0.04478302
```

(c) If the manager increases the confidence level, will the interval become wider or narrower (2 points)?

The interval become narrower