STAT 1293 - Quiz 1

Gordon Lu

7/5/2020

Problem 1: Random Numbers (10 points)

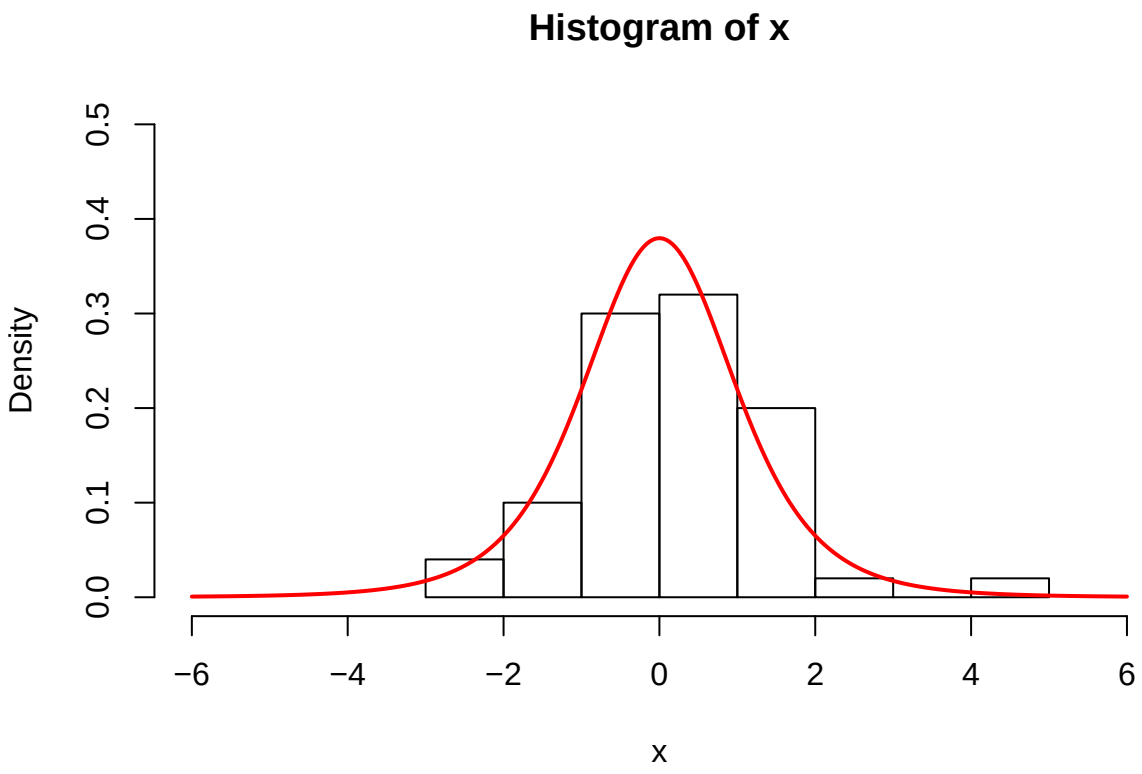
1a) Generate 50 observations from the t_5 distribution. Create a histogram. Overlay the histogram with a standard normal density curve (in red).

Solution:

```
x <- rt(50, 5) #Generate 50 observations from t_5 distribution
hist(x, freq = F, xlim = c(-6, 6), ylim = c(0, 0.5)) #Plot histogram
min(x)
```

```
## [1] -2.593625
```

```
y = seq(-6, 6, 0.01)
lines(y, dt(y, 5), col = 2, lwd = 2) #Overlay density curve
```



1b) Generate 30 observations from the χ^2_{15} distribution. Create a stem-and-leaf plot.

Solution:

```
w <- rchisq(30, 15) #Generate 30 observation form the chi^2_{15} distribution
stem(w) #Create a stem and leaf plot
```

```
##
##   The decimal point is at the |
##
##    6 | 6
##    8 | 0
##   10 | 06234
##   12 | 2955
##   14 | 0583
##   16 | 308
##   18 | 89
##   20 | 383
##   22 | 5027
##   24 | 04
##   26 | 0
```

Problem 2: Objects (10 points)

2a) Create the following matrix using the `matrix` function.

Solution:

```
v <- seq(0, 24, 3)
matA <- matrix(v, 3, 3, byrow = T)
matA
```

```
##      [,1] [,2] [,3]
## [1,]    0    3    6
## [2,]    9   12   15
## [3,]   18   21   24
```

2b) Create the following matrix by combining rows of vectors.

Solution:

```
fruit <- c("Apple", "Banana", "Orange")
m <- c("Allen", "Bill", "Owen")
f <- c("Alice", "Bella", "Olivia")
matB <- rbind(fruit, m, f)
matB
```

```
##       [,1]    [,2]     [,3]
## fruit "Apple" "Banana" "Orange"
## m     "Allen" "Bill"   "Owen"
## f     "Alice" "Bella"  "Olivia"
```

2c) Create a numeric vector `v1` with the following values.

Solution:

```
v1 <- seq(0.1, 0.5, 0.1)
v1
```

```
## [1] 0.1 0.2 0.3 0.4 0.5
```

2d) Convert `v1` to a factor, call it `v2`. Label the levels as `a, b, c, d, e`.

Solution:

```
v2 <- factor(v1, levels = seq(0.1, 0.5, 0.1))
levels(v2) <- letters[1:5]
v2
```

```
## [1] a b c d e
## Levels: a b c d e
```

2e) Create a logical vector by comparing `v1` with `0.225`.

Solution:

```
bool_vec <- v1 > 0.225
bool_vec
```

```
## [1] FALSE FALSE  TRUE  TRUE  TRUE
```

Problem 3: Statistics and Graphs (Quantitative) (20 points)

3a) Calculate the five-number summary, mean, and the standard deviation of travel time in New York state.

Solution:

```
nytravel <- read.table("C:/Users/gordo/Desktop/nytravel.txt", header = T)
summary(nytravel$Minutes) #five-number summary
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    5.00   15.00   22.50   31.25   41.25   85.00
```

```
mean(nytravel$Minutes) #calculate mean
```

```
## [1] 31.25
```

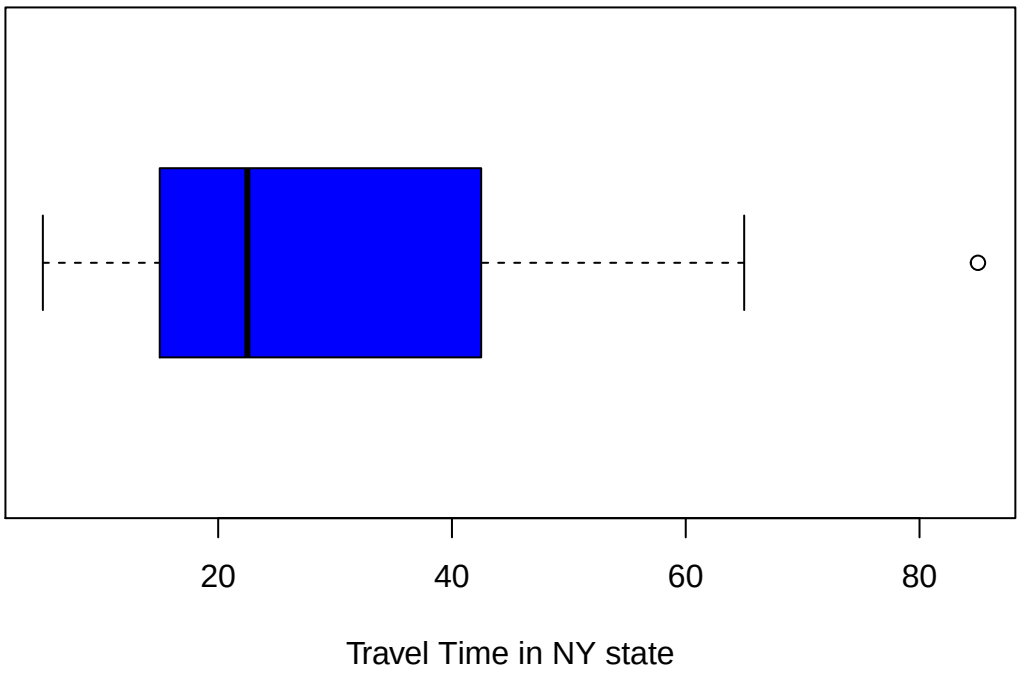
```
sd(nytravel$Minutes) #calculate sd
```

```
## [1] 21.87735
```

3b) Create a horizontal blue boxplot of `Minutes`.

Solution:

```
boxplot(nytravel$Minutes, horizontal = TRUE, col = 4, xlab = "Travel Time in NY state")
```



3c) Calculate the five-number summary, of the commute time for 15 workers in North Carolina.

Solution:

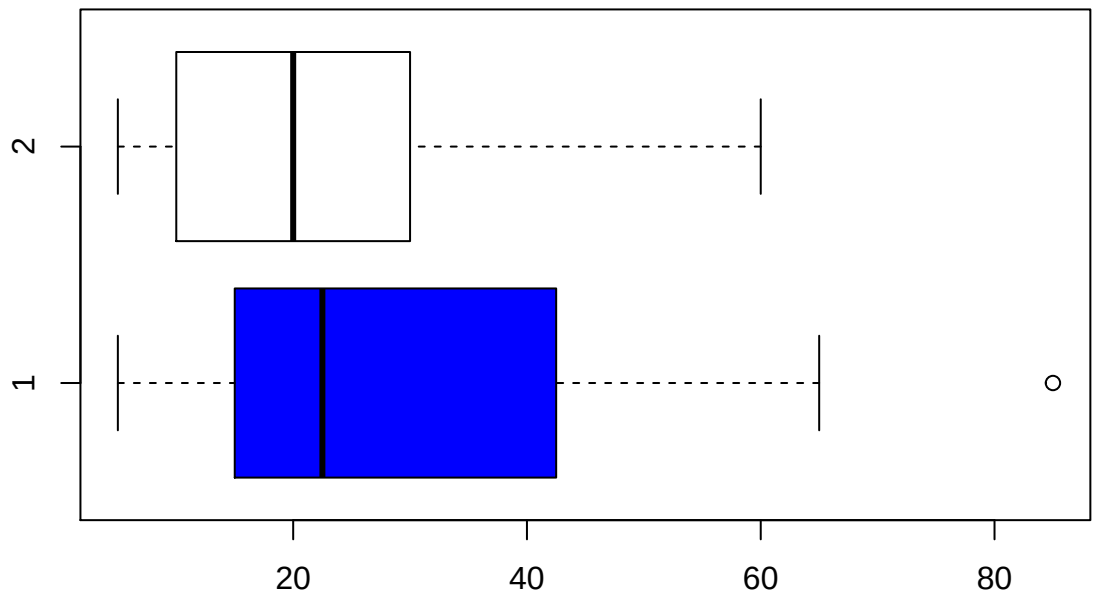
```
nctravel <- read.table("C:/Users/gordo/Desktop/nctravel.txt", header = T)
summary(nctravel$Minutes)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    5.00   10.00   20.00   22.47   30.00   60.00
```

3d) Create a horizontal side-by-side boxplot and compare the commute time of workers in NY and NC.

Solution:

```
boxplot(nytravel$Minutes, nctravel$Minutes, horizontal = T, col = c("blue", 0))
```



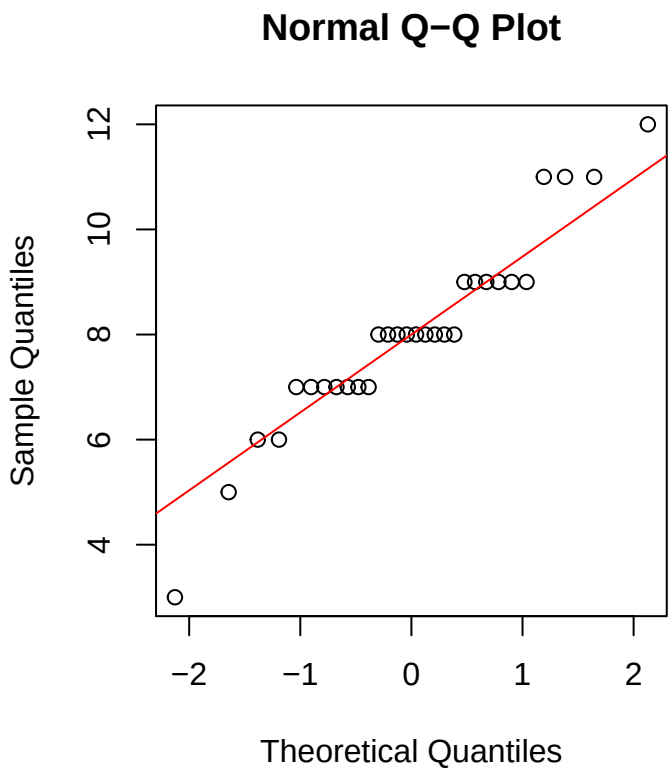
In general, it appears that those who commute to NY typically have a longer travel time to work than those from NC.

Problem 4: Probability Distributions (10 points)

4a) Generate 30 observations from `Bin(20, 0.4)`. Create a `Q-Q` plot and add a red reference line.

Solution:

```
par(pty = "s")
bin_v <- rbinom(30, 20, 0.4)
qqnorm(bin_v)
qqline(bin_v, col = 2)
```



4b) Generate 100 observations from `N(20, 5)`. Calculate the empirical 20%, 40%, 60%, and 80% percentiles.

Solution:

```
rand_norm <- rnorm(100, 20, 5)
percentile_vec <- seq(0.2, 0.8, 0.2)
quantile(rand_norm, percentile_vec)
```

```
##      20%      40%      60%      80%
## 17.22713 19.04456 21.19458 23.63746
```

4c) Calculate the following probabilities:

$P(T_8 > 3)$ where T_8 means a `t` random variable with 8 degrees of freedom.

Solution:

```
1 - pt(3, 8) #P(T_{8} < 3 would just be pt(3,8)
```

```
## [1] 0.008535841
```

$P(F_{3,5} < 1)$ where $F_{3,5}$ is an `F` random variable with 3 and 5 degrees of freedom.

Solution:

```
pf(1,3,5) #calculate probability F_{3,5} < 1
```

```
## [1] 0.5351452
```