PSY 221A Homework 2

Cristopher Garduno Luna

10/13/2017

Cohen Chapter 1

A1

a. Nominal scale
   i. Academic departments - These are descriptive categories that do not fit into some order.
   ii. Gender - These are also descriptive categories that do not fit into some order

b. Ordinal scale
   i. Runner's finishing position in a race - has an order but runners don't necessarily finish in equal intervals
   ii. Class rank in a school - ranked without necessarily having equal distances between students' class rank metric such as GPA

c. Interval scale
   i. SAT scores - quantified discrete scores with equal intervals between them, but no true zero
   ii. Celsius temperature - same as above, the intervals between temperatures are equal, but there is no true zero on this scale

d. Ratio scale
   i. Kelvin temperature - continuous temperature scale with an absolute zero
   ii. Number of children - discrete scale with an absolute zero

e. Continuous variable
   i. Weight - no limitation to precision as long as tools are available
   ii. Height of students in PSY 221A - again, no precision limitation if tools allow, and values can be non-integers

f. Discrete variable
   i. Number of whole chips in a Lays bag - presumably between 1-5, this is obtained by counting the number of whole chips and this will always be an integer value
   ii. Number of red marbles in a jar - Value will be counted and always an integer

A6

a. The types of psychotherapy are the independent variables.
b. The rating of the progress of each subject is the dependent variable.
c. Types of psychotherapies would form a nominal scale.
d. Suppose the psychotherapy is being used to treat patients with depression, the scale used to measure the rating of progress of each subject can be an ordinal scale where the possibilities are 'not happy', 'content', or 'happy'.

A8

a. Statistic - The average income for 100 US citizens describes a subset of the population
b. Parameter - The average income of all US citizens describes the whole population
c. Statistic - The average age among respondents to any survey (assuming that the population is NOT defined to be survey respondents to this magazine)

B1

a. $\sum_{i=2}^{5} X_i = 4 + 6 + 8 + 10 = 28$
b. $\sum_{i=1}^{4} Y_i = 3 + 5 + 7 + 9 = 24$
c. $\sum 5X_i = 5 \sum X_i = 5(2 + 4 + 6 + 8 + 10) = 5(30) = 150$
d. $\sum 3Y_i = 3 \sum Y_i = 3(3 + 5 + 7 + 9 + 11) = 3(35) = 105$
e. $\sum X_i^2 = 2^2 + 4^2 + 6^2 + 8^2 + 10^2 = 220$
f. $(\sum 5X_i)^2 = (150)^2 = 22,550$
g. $\sum Y_i^2 = 3^2 + 5^2 + 7^2 + 9^2 + 11^2 = 285$

C1, including optional section

```
# Source haven from library, if not installed use install.packages("haven")
library(haven)
```

```
## Warning: package 'haven' was built under R version 3.2.5
```

```
# Create object for filepath and read file into dataframe object
filelocation = "~/Desktop/UCSB/fall2017/psych221a/hw/data_hw1.sav"
dataset = as.data.frame(read_sav(filelocation))

# Replace NA data with 999
# This may be unnecessary because we can exclude NA values in R
#dataset$Mathquiz[is.na(dataset$Mathquiz)] = 999

# Generate label string vectors
genderl   = c("Female", "Male")
majorl    = c("Psychology", "Premed", "Biology", "Sociology", "Economics")
reasonl   = c("Program requirement", "Personal interest", "Advisor recommendation")
exp_condl = c("Easy", "Moderate", "Difficult", "Impossible")
coffeel   = c("not a regular coffee drinker", "regularly drinks coffee")

# Generate factors
Gender   = factor(dataset$Gender,   level = c(1:2), genderl)
Major    = factor(dataset$Major,    level = c(1:5), majorl)
Reason   = factor(dataset$Reason,   level = c(1:3), reasonl)
Exp_cond = factor(dataset$Exp_cond, level = c(1:4), exp_condl)
Coffee   = factor(dataset$Coffee,   level = c(0:1), coffeel)
```

Cohen Chapter 2

A2

```
# Input data
Recall = c(25, 23, 26, 24, 19, 25, 24, 28, 26, 21, 24, 24,
           29, 23, 19, 24, 23, 24, 25, 23, 24, 25, 26, 28, 25)

# Create and display frequency distribution
freq_recall = table(Recall)
freq_recall
```

```
## Recall
## 19 21 23 24 25 26 28 29
##  2  1  4  7  5  3  2  1
```

```
# Generate data frame
data = data.frame(rev(as.vector(freq_recall)))
rownames(data) = rev(names(freq_recall))
colnames(data) = c("Freq")
```

Next we need to calculate the relative frequency (rf), cumulative frequency (cf), cumulative relative frequency (crf), and cumulative percentage frequency (cpf)

```
# Generate vectors for rf, cf, crf, and cpf
rf  = rev(as.vector(freq_recall))/sum(as.vector(freq_recall))
cf  = rev(as.vector(cumsum(freq_recall)))
crf = cf/sum(as.vector(freq_recall))
cpf = crf*100

data$RelFreq    = rf
data$CumFreq    = cf
data$CumRelFreq = crf
data$CumPerFreq = cpf
data
```

```
##    Freq RelFreq CumFreq CumRelFreq CumPerFreq
## 29    1    0.04      25       1.00        100
## 28    2    0.08      24       0.96         96
## 26    3    0.12      22       0.88         88
## 25    5    0.20      19       0.76         76
## 24    7    0.28      14       0.56         56
## 23    4    0.16       7       0.28         28
## 21    1    0.04       3       0.12         12
## 19    2    0.08       2       0.08          8
```

a. As we can see from the table above, the proportion of participants that recalled exactly 24 words is **0.28**.

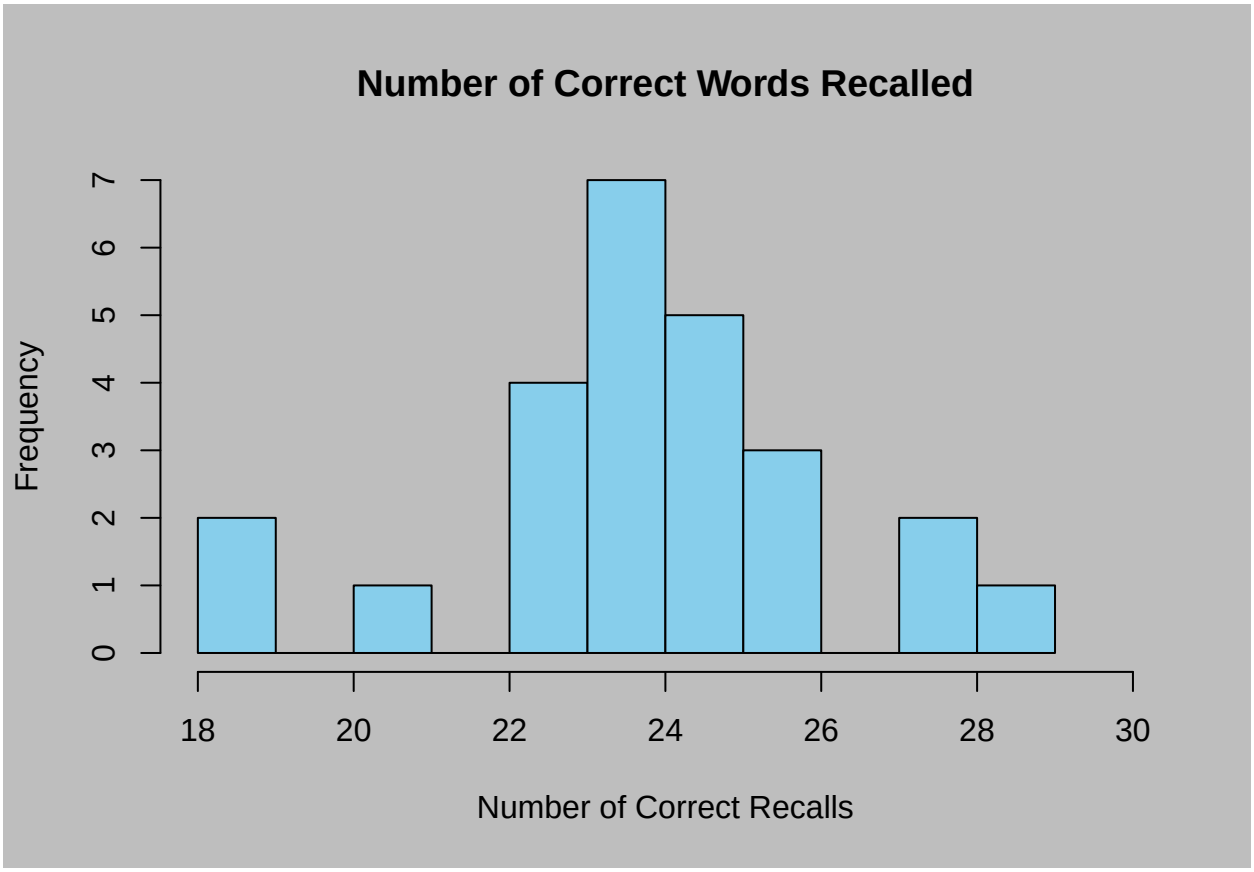
b. No more than **7** participants recalled 23 words. This represents a proportion of **0.28** of the total.

c. The percentile rank of a participant that scored 25 is **76**, while for a participant that scored 27 it is **88**.

d. The first quartile is where crf = 0.25, and the closest score to this is **23**. The third quartile is where the crf = 0.75, and the closest score to this is **25**.

e. See the histogram below.

```r
par(bg = "gray")
hist(Recall, breaks = c(18:29), xlim = c(18,30), col = "skyblue",
     xlab = "Number of Correct Recalls", main = "Number of Correct Words Recalled")
```



A5

```r
# Input data
Scores = c(10, 3, 8, 7, 1, 6, 5, 9, 8, 4, 2, 7, 7, 10, 9, 6, 8, 3, 8, 5)

# Generate frequency table and display it
freq_scores = table(Scores)
freq_scores
```

```
## Scores
##  1  2  3  4  5  6  7  8  9 10
##  1  1  2  1  2  2  3  4  2  2
```

```r
# Generate data frame
data = data.frame(rev(as.vector(freq_scores)))
rownames(data) = rev(names(freq_scores))
colnames(data) = c("Freq")

# Calculate rf, cf, crf, cpf
rf  = rev(as.vector(freq_scores))/sum(as.vector(freq_scores))
cf  = rev(as.vector(cumsum(freq_scores)))
```

```
crf = cf/sum(as.vector(freq_scores))
cpf = crf*100

data$RelFreq    = rf
data$CumFreq    = cf
data$CumRelFreq = crf
data$CumPerFreq = cpf
data
```

```
##    Freq RelFreq CumFreq CumRelFreq CumPerFreq
## 10    2    0.10      20       1.00        100
## 9     2    0.10      18       0.90         90
## 8     4    0.20      16       0.80         80
## 7     3    0.15      12       0.60         60
## 6     2    0.10       9       0.45         45
## 5     2    0.10       7       0.35         35
## 4     1    0.05       5       0.25         25
## 3     2    0.10       4       0.20         20
## 2     1    0.05       2       0.10         10
## 1     1    0.05       1       0.05          5
```

a. **2** students obtained a perfect score, which represents a proportion of **0.10** of the total.

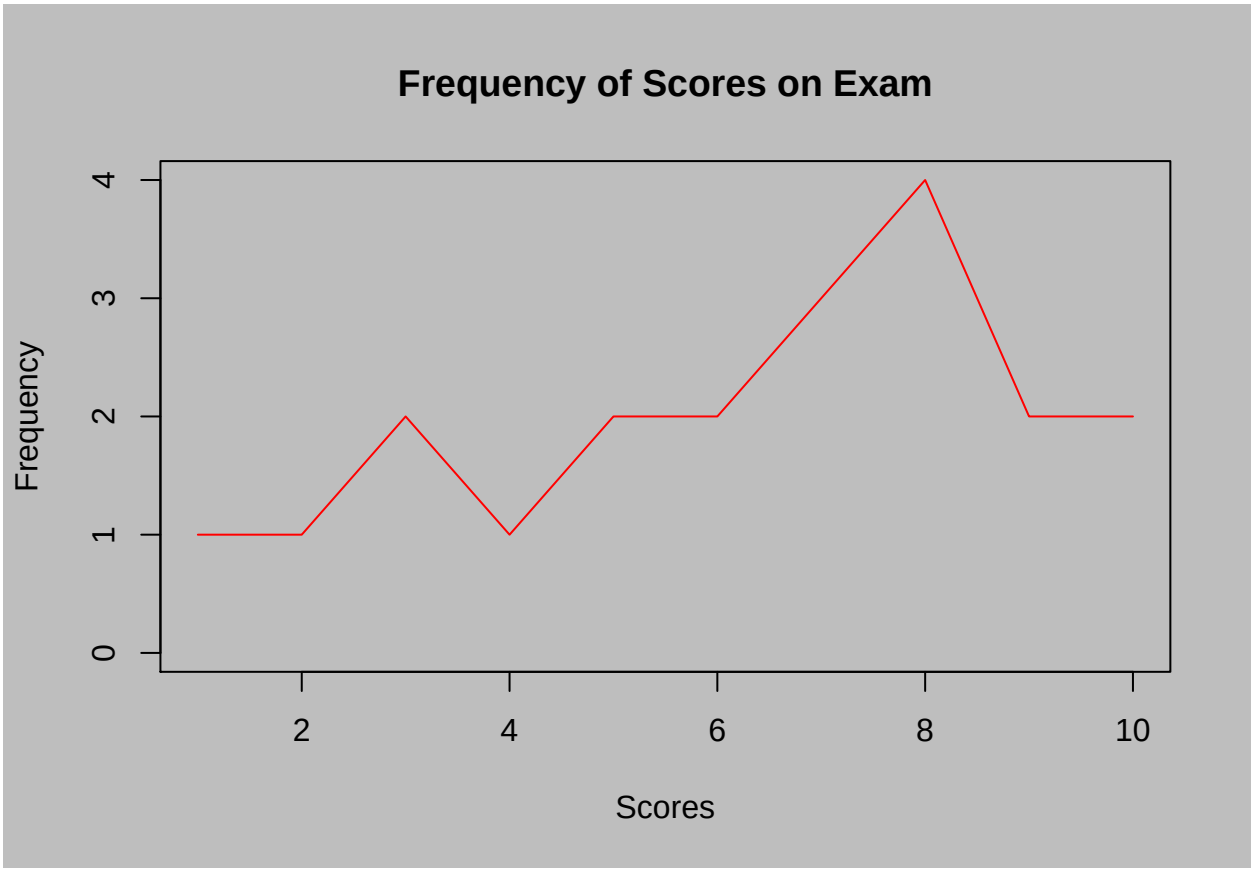
b. A score of **6** is closest to the 50th percentile.

c. The percentile rank of a student who scored a 5 is **35**, and for one who scored a 9 it is **90**.

d. The proportion of students that scored a 9 or more is **0.20**.

e. See frequency polygon below.

```
par(bg = "gray")
plot(as.vector(freq_scores), type = "l", ylim = c(0, 4), xlab = "Scores",
     ylab = "Frequency", main = "Frequency of Scores on Exam", col = "red")
```



B2a

```
# Input data
scores = c(83, 76, 80, 81, 74, 68, 92, 64, 95, 96, 55, 70, 78, 86, 85,
           94, 76, 77, 82, 85, 81, 71, 72, 99, 63, 75, 76, 83, 92, 79,
           82, 69, 91, 84, 87, 90, 80, 65, 84, 87, 97, 61, 73, 75, 77,
           86, 89, 92, 79, 80, 85, 87, 82, 94, 90, 89, 85, 84, 86, 56)

# Convert to grouped frequency table and display
freq_scores = table(cut(scores, 10))
data = as.data.frame(rev(freq_scores))
colnames(data) = "Frequency"
data
```

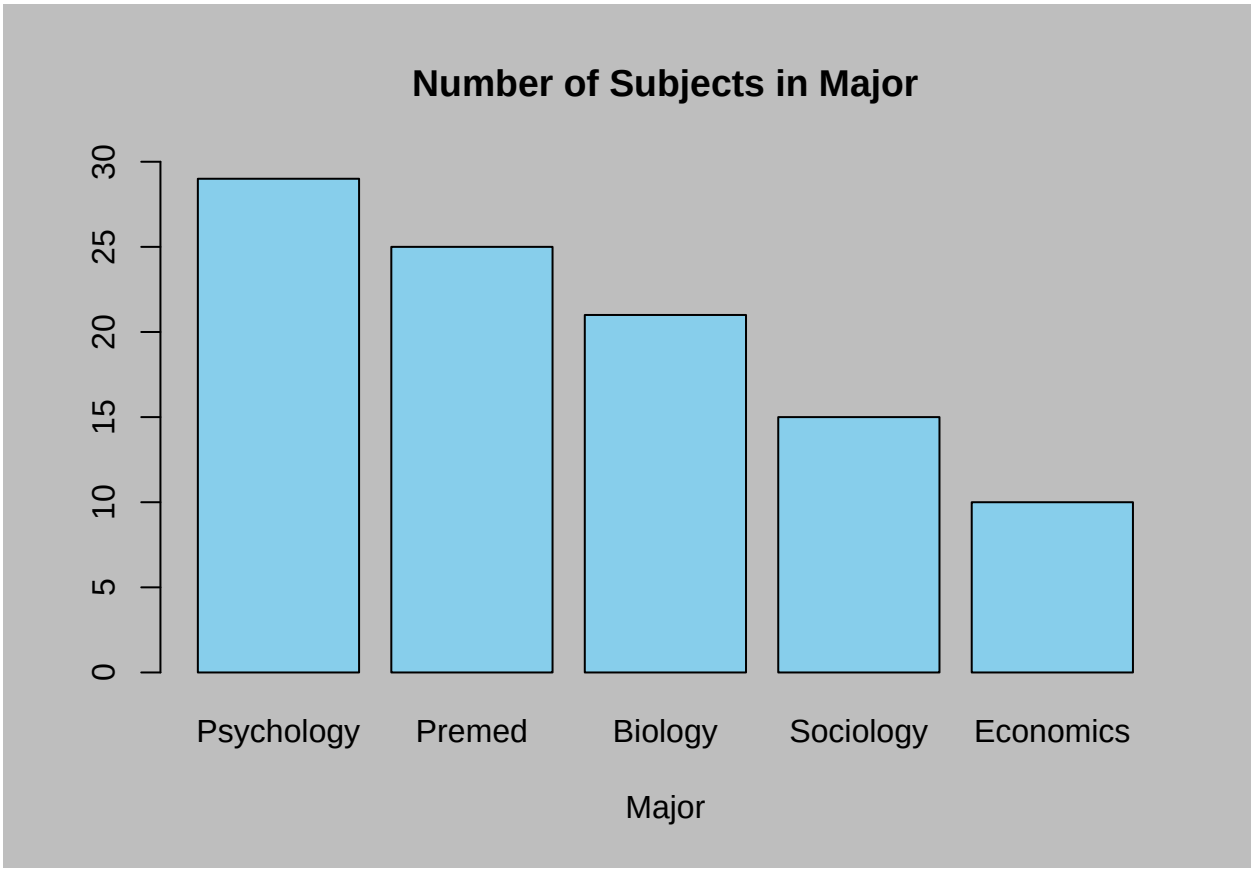
```
##             Frequency
## (94.6,99]           4
## (90.2,94.6]         6
## (85.8,90.2]        10
## (81.4,85.8]        12
## (77,81.4]           8
## (72.6,77]           9
## (68.2,72.6]         4
## (63.8,68.2]         3
## (59.4,63.8]         2
## (55,59.4]           2
```

C1

```
# Generate frequncy table for Major, then show table
freq_maj = table(Major)
freq_maj
```

```
## Major
## Psychology     Premed    Biology  Sociology  Economics
##         29         25         21         15         10
```

```
# Generate bar chart for major frequencies
par(bg = "gray")
barplot(freq_maj, ylim = c(0,30), col = "skyblue", xlab = "Major",
        main = "Number of Subjects in Major")
```

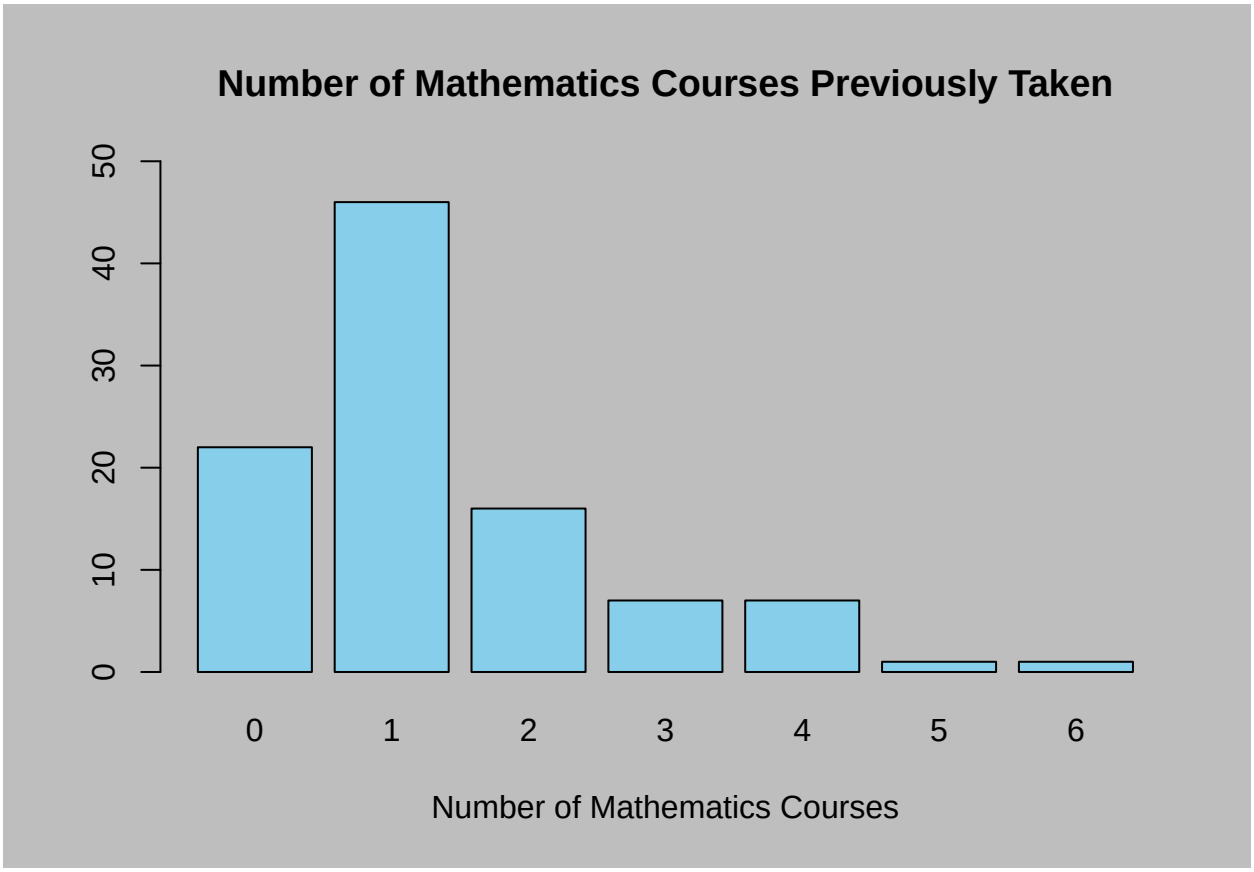


C2

```
# Generate frequencies and plot for 'Prevmath'
freq_preMath = table(dataset$Prevmath)
names(dimnames(freq_preMath)) = "Previous Math Courses Taken: 0 to 6"
freq_preMath
```

```
## Previous Math Courses Taken: 0 to 6
##  0  1  2  3  4  5  6
## 22 46 16  7  7  1  1
```

```
par(bg = "gray")
barplot(freq_preMath, ylim = c(0, 50), col = "skyblue",
        xlab = "Number of Mathematics Courses",
        main = "Number of Mathematics Courses Previously Taken")
```

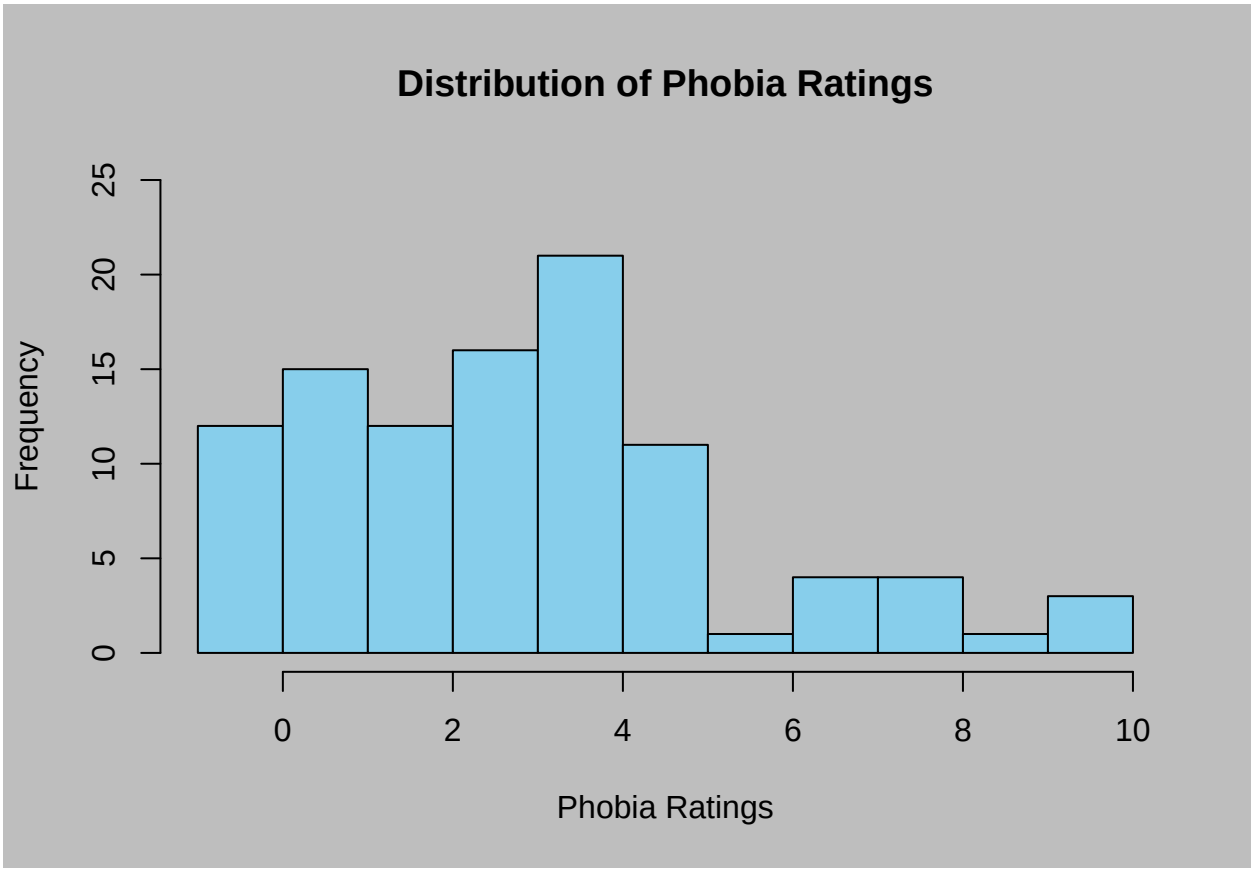


Phobia is a variable that exists on a continuum, so we use a histogram to represent the data.

```
# Generate frequnecies and plot for 'Phobia'
freq_phobia = table(dataset$Phobia)
names(dimnames(freq_phobia)) = "Phobia Measure: 0 = Not at all, 10 = Extremely"
freq_phobia
```

```
## Phobia Measure: 0 = Not at all, 10 = Extremely
##  0  1  2  3  4  5  6  7  8  9 10
## 12 15 12 16 21 11  1  4  4  1  3
```

```
par(bg = "gray")
hist(dataset$Phobia, breaks = c(-1:10), ylim = c(0, 25), col = "skyblue",
     xlab = "Phobia Ratings", main = "Distribution of Phobia Ratings")
```

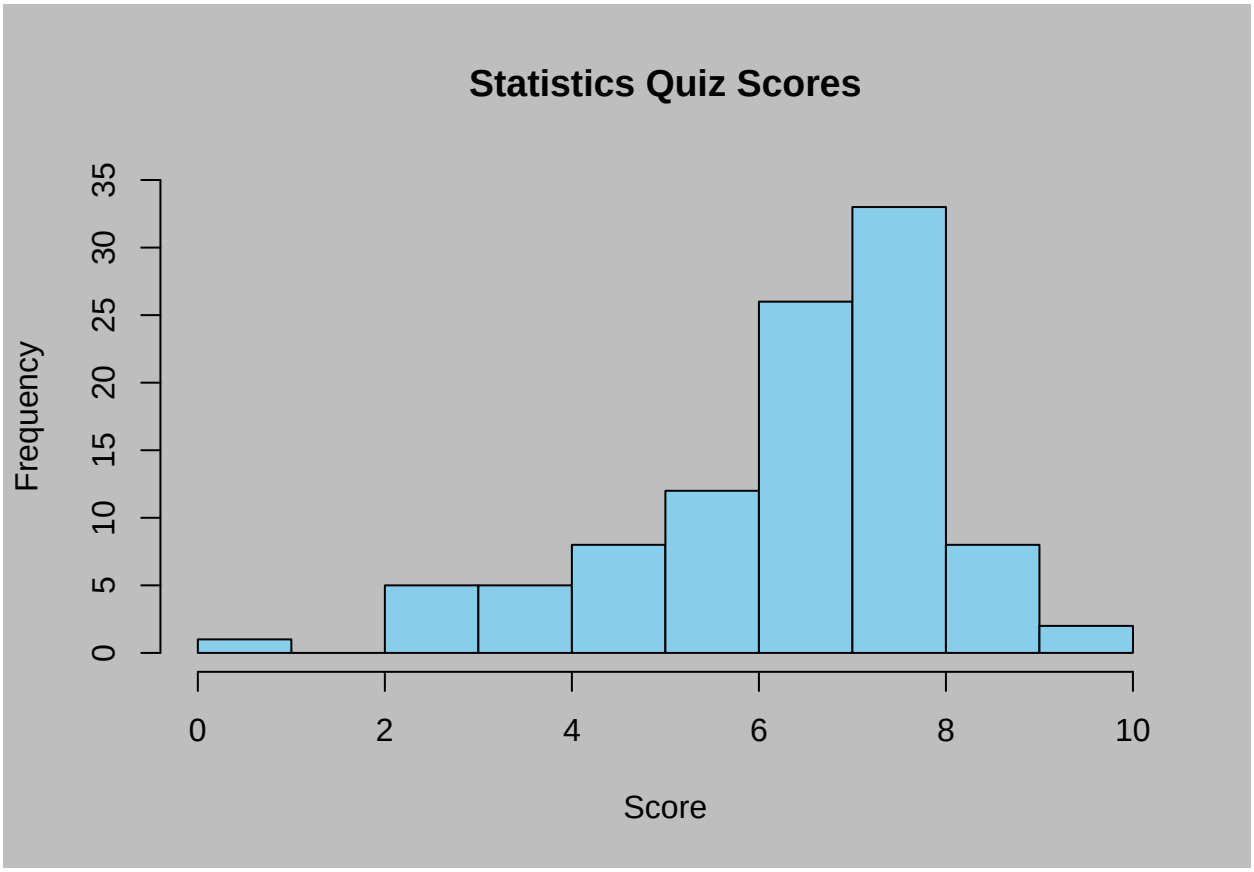


C3

```
# Frequency distribution and hisogram for 'statquiz'
freq_statquiz = table(dataset$Statquiz)
names(dimnames(freq_statquiz)) = "Scores on Statistics Quiz"
freq_statquiz
```

```
## Scores on Statistics Quiz
##  1  3  4  5  6  7  8  9 10
##  1  5  5  8 12 26 33  8  2
```

```
par(bg = "gray")
hist(dataset$Statquiz, breaks = c(0:10), ylim = c(0, 35), col = "skyblue",
     xlab = "Score", main = "Statistics Quiz Scores")
```



This distribution has a negative skew.

C8

The quartiles are the quantiles at 25%, 50%, 75%, and 100%.

```
# Quartiles for 'anx_base' and 'anx_pre'
quantile(dataset$Anx_base)
```

```
##   0%  25%  50%  75% 100%
##   10   16   18   20   39
```

```
quantile(dataset$Anx_pre)
```

```
##   0%  25%  50%  75% 100%
##    8   14   19   25   39
```

C10

Request the following percentiles for the variables hr_base and hr_pre: 15, 30, 42.5, 81, and 96.

```
# Particular percentiles for 'Hr_base' and 'Hr_pre'
quantile(dataset$Hr_base, c(.15, .30, .425, .81, .96))
```

```
##   15%   30% 42.5%   81%   96%
##    69    71    72    76    78
```

```r
quantile(dataset$Hr_pre, c(.15, .30, .425, .81, .96))
```

```
##   15%   30% 42.5%   81%   96%
## 68.85 71.00 73.00 78.00 83.00
```