Homework 02 Answer Sheet

Psych 10C
Due: Sunday, October 9th (by 11:59pm PT)

Zhenze Zhang

Submission Details

- Download *HW02AnswerSheet.Rmd* from the Canvas course space and open it RStudio.
- Enter your name in the *author* field at the top of the document.
- Complete the assignment by entering your answers in your *HW02AnswerSheet.Rmd* document.
- Once you have completed the assignment, click the *Knit* button to turn your completed answer document into a pdf file.
- Submit your HW02AnswerSheet.pdf file only (no other formats are acceptable) before the assignment's deadline.

Problems

For each problem, show/describe all of your work.

Problem #1 (2 points)

The portfolios of wealthy people over the age of 50 produce yearly retirement incomes which are normally distributed with mean equal to $125,000 and standard deviation of 25,000. Describe the mean and standard deviation (standard error) of the distribution of the mean of samples of size 16 from this population.

ANSWER:

```
# Since the distribution of sample is normally distributed, the mean of the distribution of sample mean

# Since the sample is normally distributed, the standard error of distribution of sample means = sd of
25000/sqrt(16)
```

```
## [1] 6250
```

Problem #2 (3 points)

For the distribution described in Problem #1, use RStudio to find the probability of selecting a sample of 16 wealthy individuals whose portfolios produce a mean retirement income

(a) exceeding $135,000

ANSWER:

```
pnorm(135000,125000, 6250, lower.tail = FALSE)
```

```
## [1] 0.05479929
```

(b) less than $115,000

ANSWER:

```
pnorm(115000, 125000, 6250)
```

```
## [1] 0.05479929
```

(c) less than $110,000 or more than $140,000

ANSWER:

```
(Below<- pnorm(110000, 125000, 6250))
```

```
## [1] 0.008197536
```

```
(Exceed<- pnorm(140000, 125000, 6250, lower.tail = FALSE))
```

```
## [1] 0.008197536
```

```
(Overall<- Below+Exceed)
```

```
## [1] 0.01639507
```

Problem #3 (3 points)

A consumer protection group regularly monitors gas prices in southern California. One day (several years ago), they randomly sampled the prices for regular unleaded gasoline at 20 gas stations. Those prices were:

```
3.95, 4.07, 4.14, 4.04, 4.00, 3.93, 3.98, 4.01, 3.90, 4.03,
4.13, 3.88, 4.03, 4.10, 3.81, 4.18, 3.88, 4.02, 4.13, 4.21
```

(a) In RStudio, create a variable to hold the sampled gas prices and enter the data (use the c() function to create a vector, rather than creating a data frame)

ANSWER:

```
GPrice<- c(3.95, 4.07, 4.14, 4.04, 4.00, 3.93, 3.98, 4.01, 3.90, 4.03,4.13, 3.88, 4.03, 4.10, 3.81, 4.18
```

(b) Calculate and print the mean of the sample. Use the cat() function for your output.

ANSWER:

```
cat("SMean:", mean(GPrice))
```

```
## SMean: 4.021
```

(c) Calculate and print the standard deviation of the sample. Use the cat() function for your output.

ANSWER:

```
cat("SamSD:", sd(GPrice))
```

```
## SamSD: 0.1082346
```

Problem #4 (4 points)

Consider a normal population with mean = 82 and standard deviation = 12.

(a) If a random sample of size 9 is selected, what is the probability that the sample mean will lie between 80.8 and 83.2 (use RStudio to compute your answer)?

ANSWER:

```
#The sd of the distribution of the sample mean = 12/3=4
12/sqrt(9)
```

```
## [1] 4
```

```
#The mean of the distribution of the sample mean = 82

(p1<- pnorm(83.2, 82, 4))
```

```
## [1] 0.6179114
```

```
(p2<- pnorm(80.8, 82, 4))
```

```
## [1] 0.3820886
```

```
(pBetween<- p1-p2)
```

```
## [1] 0.2358228
```

(b) Generate 10 random samples of size 9 from the distribution and calculate the mean for each (create a variable to contain the calculated means). Print out the variable containing your sample means. Based on your answer to part (a), about how many of the sample means would you expect to fall between 80.8 and 83.2? How many of the sample means DID fall between 80.8 and 83.2 (use the sum() function to count this number)?

ANSWER: (create and print out variable containing sample means)

```
SampleMeans<- c()
for(i in 1:10){
  SampleMeans<- append(SampleMeans, mean(rnorm(9, 82, 12)))
}
SampleMeans
```

```
##  [1] 77.66661 83.46550 84.82087 76.22284 77.26614 82.24041 86.46561 84.53236
##  [9] 78.77574 82.29010
```

```
sum(SampleMeans> 80.8 & SampleMeans<83.2)
```

```
## [1] 2
```

ANSWER: (About how many would you expect to fall between 80.8 and 83.2?)

- 0.2358*10= 2.358, therefore, I would expect about 2 or 3 sample means fall between 80.8 and 83.2.

ANSWER: (How many of the sample means DID fall in this range?)

```
sum(SampleMeans> 80.8 & SampleMeans<83.2)
```

```
## [1] 2
```

(c) With a random sample of size 25, what is the probability that the sample mean will fall between 80.2 and 83.2 (use RStudio to compute your answer)?

ANSWER: Notice that we can use the same code as in part (a), simply with a different sample size!

```
#The sd of the distribution of the sample mean = 12/5=2.4
12/sqrt(25)
```

```
## [1] 2.4
```

```
#The mean of the distribution of the sample mean = 82

(pH<- pnorm(83.2, 82, 2.4))
```

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

```
(pL<- pnorm(80.8, 82, 2.4))
```

```
## [1] 0.3085375
```

```
(pB<- pH-pL)
```

```
## [1] 0.3829249
```

(d) Generate 10 random samples of size 25 from the distribution and calculate the mean for each (create a variable to contain the calculated means). Print out the variable containing your sample means. For a sample size of 25, about how many of the sample means would you expect to fall between 80.8 and 83.2? How many of the sample means DID fall between 80.8 and 83.2 (use the sum() function to count this number)?

ANSWER: (create and print out variable containing sample means)

```
SampleMeans2<- c()
for (i in 1:10) {
  SampleMeans2<- append(SampleMeans2, mean(rnorm(25, 82, 12)))
}
SampleMeans2
```

```
##  [1] 80.84014 82.48822 81.41968 80.88841 79.92889 83.17706 78.99006 84.67797
##  [9] 79.79685 82.59552
```

ANSWER: (About how many would you expect to fall between 80.8 and 83.2?)
Since 0.3829249*10=3.8, I would expect 3 or 4 sample means fall between 80.8 and 83.2.

ANSWER: (How many of the sample means DID fall in this range?)

```
sum(SampleMeans2> 80.8 & SampleMeans2<83.2)
```

```
## [1] 6
```

Problem #5

Consider a population having a discrete uniform distribution that places a probability of 0.1 on each of the integers 0 through 9. The mean of this distribution is 4.5 and the standard deviation is 2.872.

(a) Calculate the mean and standard error for the distribution of sample means for sample size 30 for this population.

ANSWER:

```
#The sample size is 30, which means even though the original population is not normal, we can consider
mean<-4.5
sd<- 2.872/sqrt(30)
sd
```

```
## [1] 0.5243531
```

(b) In RStudio, create a variable containing the means of 50 random samples of size 30 from this population (in your answer, include the code you used to create the variable, but do not print out the listing of the 50 sample means). Remember that the rdunif() function requires the purrr library.

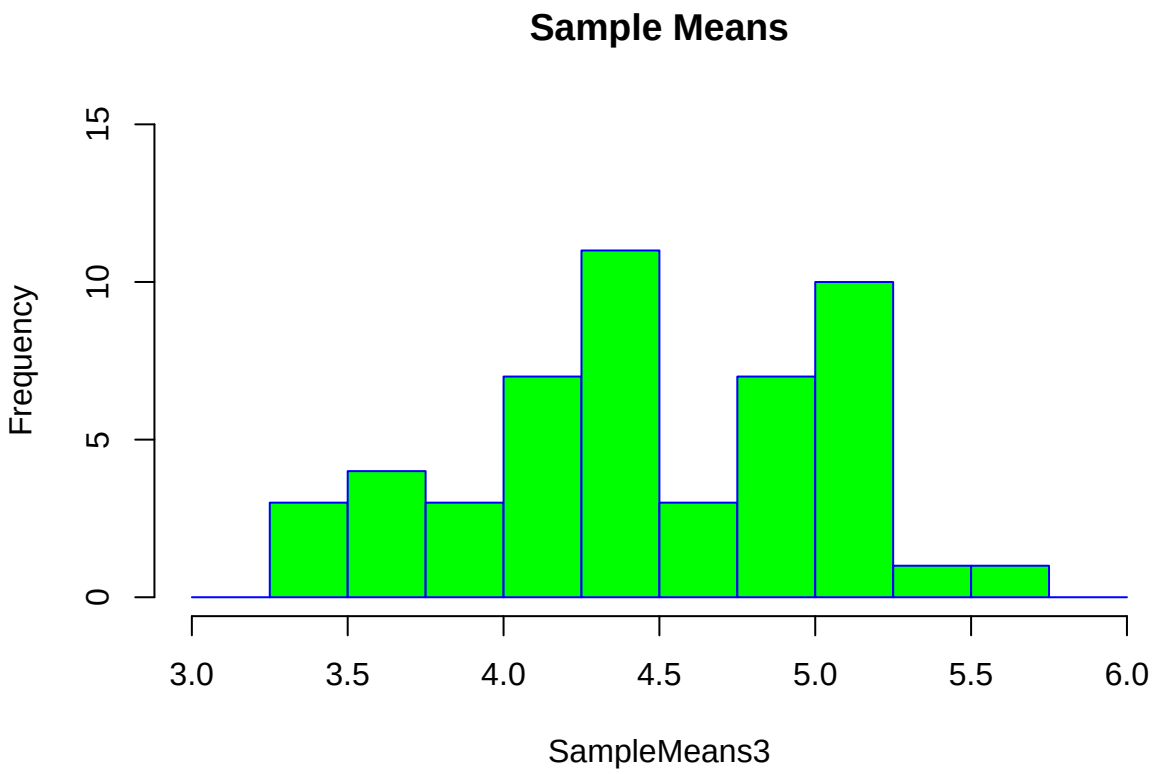
ANSWER:

```
library(purrr)
SampleMeans3<- c()
for(i in 1:50){
  SampleMeans3<- append(SampleMeans3, mean(rdunif(30, 0, 9)))
}
```

(c) Use the hist() function to create a frequency histogram of the sample means, using the following breaks: 3.0, 3.25, 3.5, 3.75, 4.0, 4.25, 4.5, 4.75, 5.0, 5.25, 5.5, 5.75, 6.0.

ANSWER:

```
hist(SampleMeans3, main= "Sample Means", col = "Green",
     border = "blue", ylab = "Frequency", xlim = c(3,6), ylim=c(0,15),
     breaks = c(3.0, 3.25, 3.5, 3.75, 4.0, 4.25,
                4.5, 4.75, 5.0, 5.25, 5.5, 5.75, 6.0))
```



(d) Calculate the mean and standard deviation of the sample means.

ANSWER:

```
(mean(SampleMeans3))
```

```
## [1] 4.499333
```

```
(sd(SampleMeans3))
```

```
## [1] 0.5658253
```

Problem #6 (2 points)

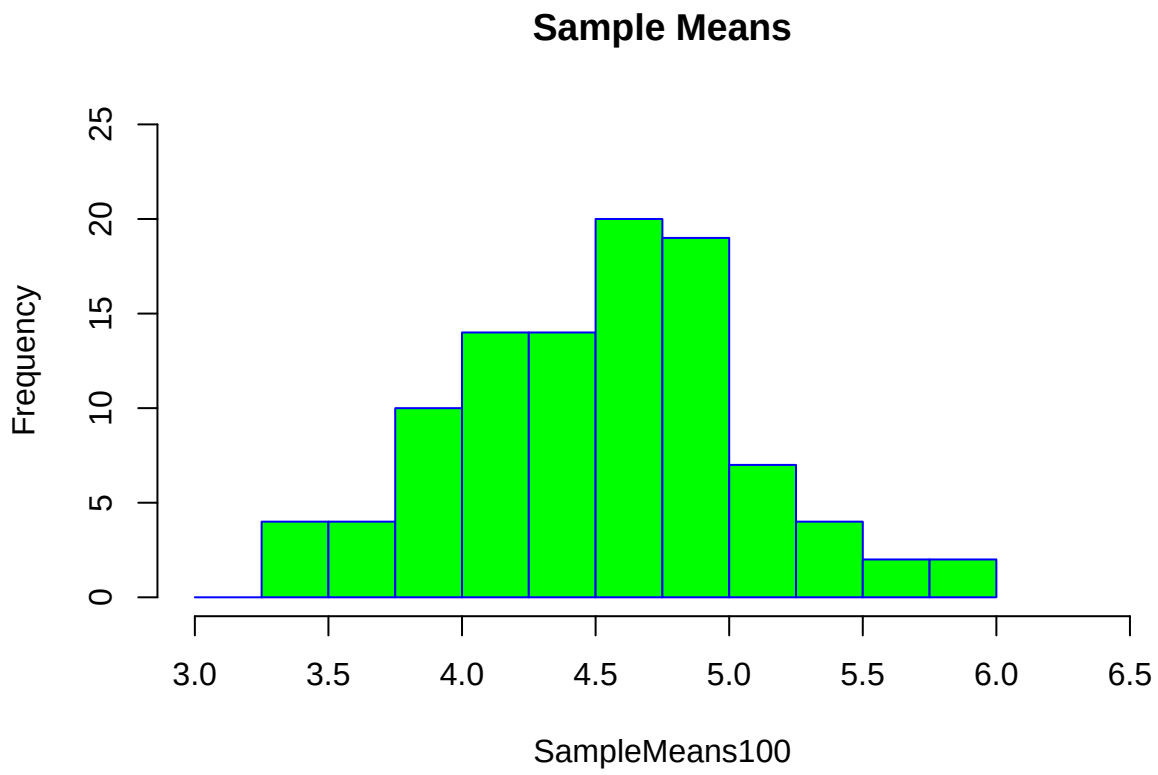
Repeat problem 5 (parts b through d) using 100 random samples of size 30 from the distribution. Select your own appropriate breaks for the histogram.

(b) **ANSWER:**

```
library(purrr)
SampleMeans100<- c()
for (i in 1:100){
  SampleMeans100<- append(SampleMeans100, mean(rdunif(30,0,9)))
}
```

(c)

```
hist(SampleMeans100, main= "Sample Means", col = "Green", border = "blue",
     ylab = "Frequency", xlim = c(3,6.5), ylim=c(0,25),
     breaks = c(3.0, 3.25, 3.5, 3.75, 4.0, 4.25, 4.5, 4.75,
                5.0, 5.25, 5.5, 5.75, 6.0))
```



(d)

```
(mean(SampleMeans100))
```

```
## [1] 4.522667
```

```
(sd(SampleMeans100))
```

```
## [1] 0.5445141
```

Problem #7 (2 points)

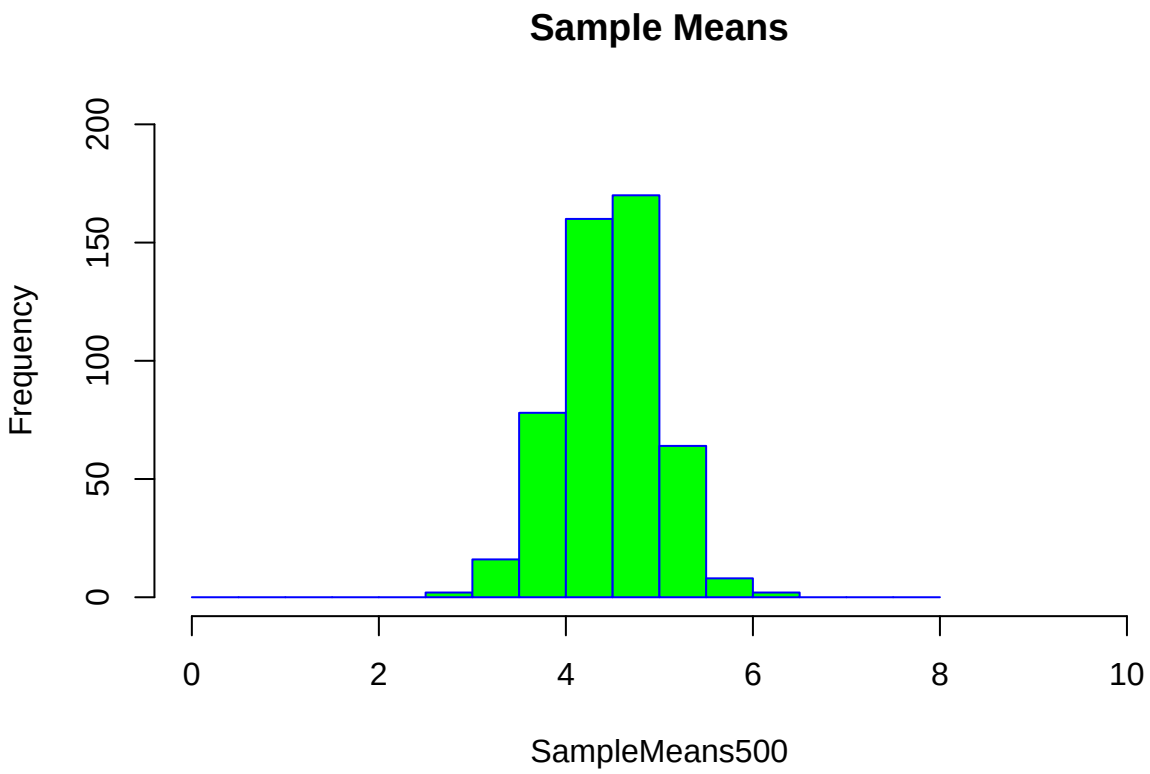
Repeat problem 5 (parts b through d) using 500 random samples of size 30 from the distribution. Select your own appropriate breaks for the histogram.

ANSWER: (b)

```
library(purrr)
SampleMeans500<- c()
for(i in 1:500){
  SampleMeans500<- append(SampleMeans500, mean(rdunif(30, 0, 9)))
}
```

(c)

```
hist(SampleMeans500, main= "Sample Means", col = "Green",
     border = "blue", ylab = "Frequency", xlim = c(0,10), ylim=c(0,200),
     breaks = c(0,0.5,1,1.5,2,2.5,3,3.5,4,4.5,5,5.5,6,6.5,7,7.5,8))
```



(d)

```
(mean(SampleMeans500))
```

```
## [1] 4.476733
```

```
(sd(SampleMeans500))
```

```
## [1] 0.5199502
```
