Name: Drashti Mehta

Lab Assignment 5

(40 points)

DIRECTIONS

- ***An uploaded copy is due the Friday after assignment at 12PM.***

A. Read in the first <u>50</u> samples of mosquito-fish lengths from the data from canvas (unknownData.txt – Lab 4 folder), create a subset of this data using R commands. Assume this is your random sample from the population. Calculate a confidence interval from a normal distribution for the following exercises. In this case assume the population standard deviation is known (4.82mm) but the population mean is unknown. Do this question in <u>R only</u>. (10 pts total).

1. Calculate the 50% confidence interval. (2pts)

```
> View(mosFish.txt)
> head(mosFish.txt,50)
> mos<-c(head(mosFish.txt))
> mean(Lengths.of.Mosquitofish..mm.)
> qnorm(0.75)
[1] 0.6744898
> x<-0.67*(4.82/sqrt(50))
> x
[1] 0.4567061
> mean=21.83
> mean+x
[1] 22.28671
> mean-x
[1] 21.37329
```

2. Calculate the 90% confidence interval. (2pts)

```
> qnorm(0.9495)
[1] 1.640025
> mean=21.83
> x<-mean-1.64*(4.82/sqrt(50))
> x
[1] 20.71209
> x<-mean+1.64*(4.82/sqrt(50))
> x [1] 22.94791
```

3. Calculate the 95% confidence interval. (2pts)

```
> qnorm(0.9750)
[1] 1.959964
> mean=21.83
> x<-mean-1.95*(4.82/sqrt(50))
> x [1] 20.50078
> x<-mean+1.95*(4.82/sqrt(50))
> x
[1] 23.15922
```

4. What happens to the width of the confidence interval as we go from 50% to 90% to 95% confidence? (1pt)

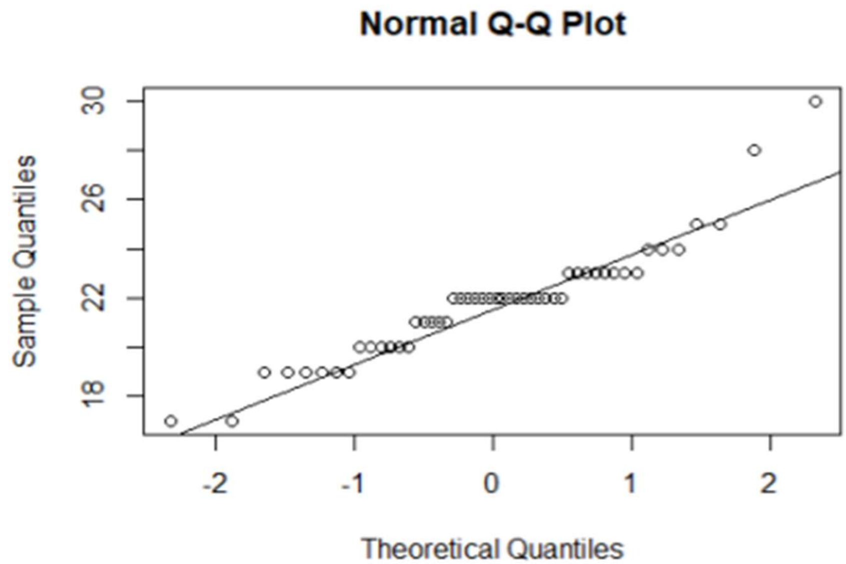
The width of the confidence intervals is gradually increasing from 50% to 90% to 95%.

	+Mean	-Mean
50%	22.28671	21.37329
90%	22.94791	20.71209
95%	23.1592	20.5007

5. Use R to construct a probability plot of the data subset (2 pts).

>qqnorm(mosFish.txt[1:50,])

>qqline(mosFish.txt[1:50,])



6. Determine if this variable is approximately normally distributed using R (1 pt).

We can determine from the probability plot that the variable is normally distributed.

B. A population of red-bellied snakes is known to have a ratio of grey color morph to red color morph of 45:55 respectively. Use the normal approximation of the binomial distribution to solve the following exercises. Please show all your work by hand and in R for full credit. (9 pts total)

1. What is the probability of selecting from this population a random sample of 25 snakes containing 4 or fewer **grey** morph individuals. (3 pts)

$\mu = 25*0.45=11.25$

$\sigma=\sqrt{25*0.45(1-0.45)}$

$= 2.48746$

$P(x \leq 4) = p(z \leq 4.1-11.25/2.48746)$

$=p(z \leq -2.88)$

$=0.0020$

In R:

```
> pnorm(4.1,11.25,2.48)
[1] 0.001969134
```

2. What is the probability of selecting from this population a random sample of 20 snakes containing 5 or fewer **red** morph individuals. (3 pts)

$\mu = 20*0.55=11$

$\sigma=\sqrt{20*0.55(1-0.55)}$

$=2.22485$

$P(x \leq 5)=p(z \leq 5.1-11/2.224)$

$=p(z \leq -2.65)$

$=0.0040$

In R:

```
> pnorm(5.1,11,2.22)
[1] 0.003934289
```

3. What is the probability of selecting a random sample of 45 snakes containing between 15 and 30 **grey** morph individuals. (3 pts)

Between 15 and 30 grey morph
$\mu= 45*0.45=20.25$
$\sigma=\sqrt{45*0.45(1-0.45)}$
$=3.33728$
$P(15<x<30)=p(15-20.25/3.33728)<(x- \mu/ \sigma)<(30-20.25/3.33728)$
$=p(-1.57<x<2.92)$
$=p(z<2.92)-p(z<-1.57)$
$=0.9982-0.0582$
$=0.94$
In R:

```
> pnorm(30,20.25,3.337)-pnorm(15,20.25,3.337)
[1] 0.9404316
```

C. We found the probability that the sample mean length of 10 random bluegill sunfish falls between 110mm and 133mm in the lab. In this case assume the population standard deviation is 35.5mm and the population mean is 125mm. Use the Central Limit Theorem to solve the following exercises. Please show all your work by hand and in R in questions for full credit. (5 pts total)

1. Find the probability that the sample mean length of 20 random bluegill sunfish falls between 110 and 133? (2pts)

$Z = x - \mu / \sigma/\sqrt{n}$

$P(110<x<133) = 110 - 125/35.5/\sqrt{20} < x-\mu/\sigma < 133-125/35.5\sqrt{20}$

$= p(-1.889<z<1.007)$

$=p(z<1.007)-p(z<-1.889)$

$=0.8413-0.0301$

$=0.8112$

In R:

```
> pnorm(133,125,35.5/sqrt(20))-pnorm(110,125,35.5/sqrt(20))
[1] 0.8138226
```

2. Find the probability that the sample mean length of 30 random bluegill sunfish falls between 110 and 133? (2pts)

$Z = x - \mu / \sigma/\sqrt{n}$

$P(110<x<133) = 110 - 125/35.5/\sqrt{30} < x-\mu/\sigma < 133-125/35.5\sqrt{30}$

$= p(-2.314<z<1.234)$

$=p(z<1.234)-p(z<-2.314)$

$=0.8907-0.0104$

$=0.8803$

In R:

```
> pnorm(133,125,35.5/sqrt(30))-pnorm(110,125,35.5/sqrt(30))
[1] 0.8811302
```

3. What happens to the standard error as the sample size increases from n=20 to n=30? (1pts)

Standard error = σ/√n

=> 35.5/√20 = 7.938.

=> 35.5/√30 = 6.4813

* Standard error is decreasing as the sample size increases.

D. We sampled of 14 iguanas and their weights (in grams). In this case assume the population standard deviation and the population mean is unknown and the data is normally distributed. Show your work in R only (10 pts total).

1450,1550,2200,1400,1650,2000,2435,1550,1050,2300,2000,2750,1800,2650

First, we will calculate the mean and standard deviation using R.

```
> iguanas<-c(1450,1550,2200,1400,1650,2000,2435,1550,1050,2300,2000,2750,1800,2650 )
> mean(iguanas)
[1] 1913.214
> sd(iguanas)
[1] 506.2137
```

1. Compute the 95% confidence interval (2 pt).

×±z∝/2*σ/(√n)

= 1913.214±1.96*506.21/√14

= [2178.38][1648.04]

In R:

```
> x<-1913.214+1.96*(506.21/sqrt(14))
> x
[1] 2178.383
> x<-1913.214-1.96*(506.21/sqrt(14))
> x
[1] 1648.045
```

2. Compute the 80% confidence interval (2 pt).

×±z∝/2*σ/(√n)

= 1913.214±0.6*506.21/√14

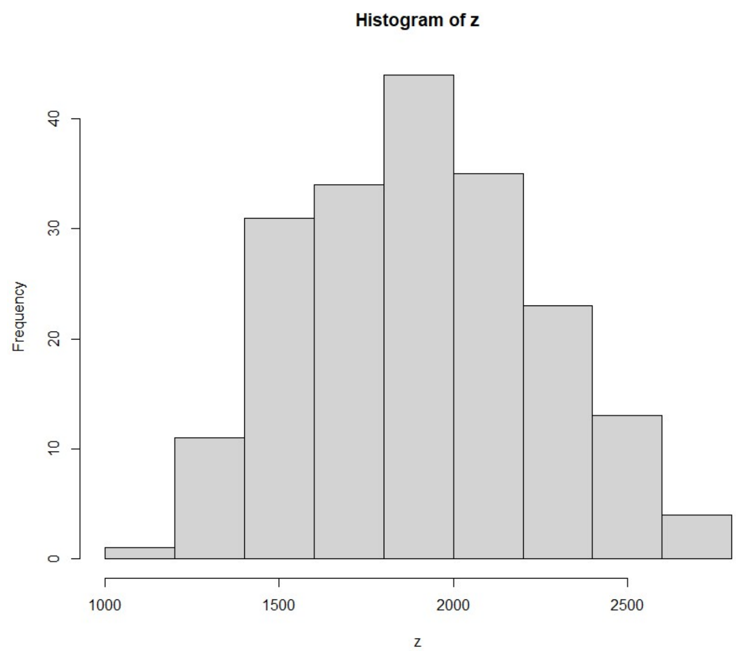
= [1994.38][1832.03]

In R:

```
> x<-mean(iguanas)+0.6*(sd(iguanas)/sqrt(14))
> x
[1] 1994.389
> x<-mean(iguanas)-0.6*(sd(iguanas)/sqrt(14))
> x
[1] 1832.04
```

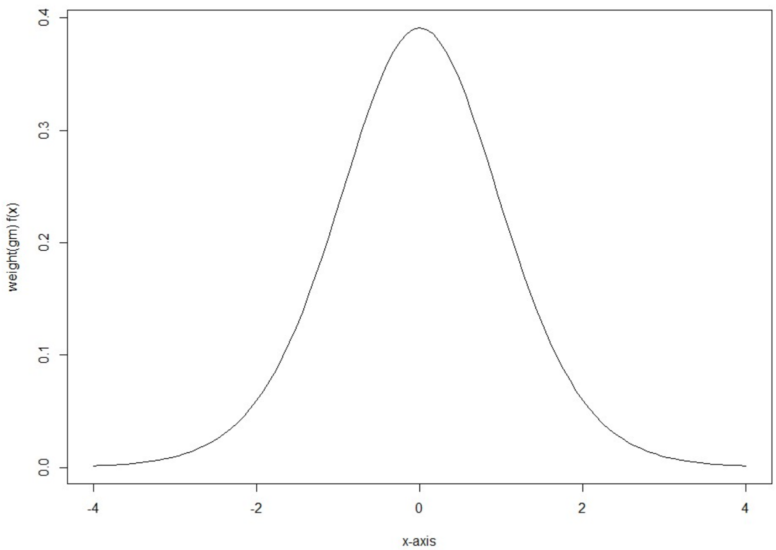
3. In R, use a 'for' loop to generate a sampling distribution of size 2 from this data set (2 pt). Plot this in a histogram (2 pt).

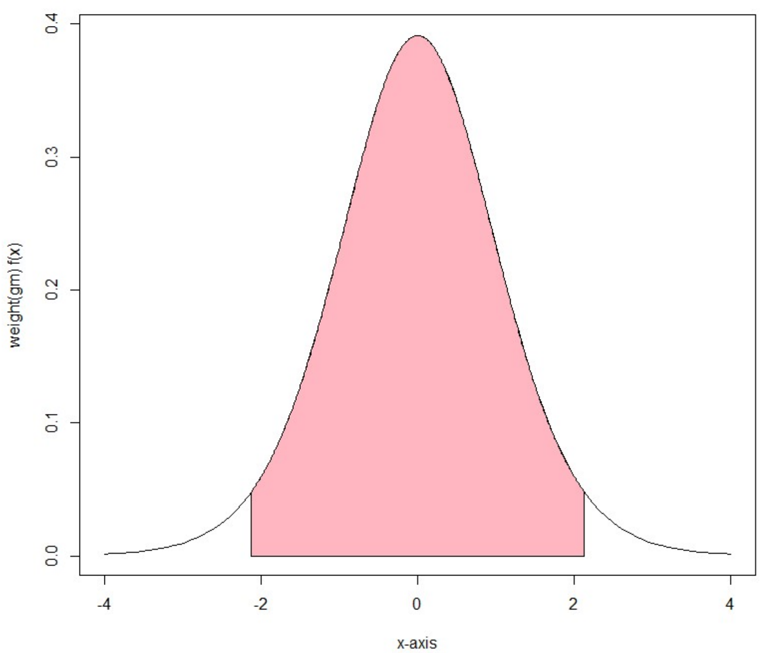
```
> x<-c(1450,1550,2200,1400,1650,2000,2435,1550,1050,2300,2000,2750,1800,2650 )
> y<-x
> z<-c()
> for(n in x) {
+ for(m in y) {
+ z<-c(z,(n+m)/2)
+
+ }}
> hist(z)
```



4. In R, plot a t distribution with 13 degrees of freedom. Color the 95% confidence interval. (2 pts)

```
> f<-function(x){dt(x,13)}
> plot(f,-4,4,xlab="x-axis",ylab="weight(gm) f(x)")
> x<-seq(qt(0.025, 15),qt(0.975, 15),0.01)
> polygon(c(qt(0.025, 15),x,qt(0.975, 15)), c(0,f(x),0), col="lightpink")
```





E. <u>In your words</u> define degrees of freedom (3 pts total).

The number of observations that can vary freely when estimating parameters are called the degrees of freedom (df). In other words, it is the number of ways by which the dynamic system can be changed without any violation of constraint imposed on it. As the value of df increases, sample size will also increase along-with the measurement accuracy.

F. <u>In your words</u> define p-value (3 pts total).

---

p-value describes the level of significance of a statistical test. It assumes that null hypothesis is true which means the probability that the result obtained could have happened by chance.

1. If the p-value $\leq \alpha$, then reject Ho.
2. If the p-value $\geq \alpha$, then fail to reject Ho.