Name : Drashti Mehta

Lab Assignment 4

(40 points)

DIRECTIONS

- ***An uploaded copy to canvas is due at Friday by 11:00 AM.***

A. From 1970 to 1976, marine biologists measured the straight carapace length of a group (n=314) of female Hawaiian green sea turtles, Cheloniamydas, nesting at the French Frigate Shoals (western Hawaiian archipelago). This group is the entire population in which we have an interest. The mean length of this group is 83.2 cm and the standard deviation is 3.95 cm. Answer the following questions by hand and using R. Please show all your work and all your commands for full credit. (10 pts total)

1.What is the probability that an individual female turtle sampled at random from this population will be 84 cm or longer? (2 pts)

By Hand: ______________________________

X = 84, mu = 83.2, sigma = 3.95.
P(X>84) = P({X - mu}/sigma) = 84 - 83.2 / 3.95 = 0.2025
Z value is 0.5793, thus subtract it from 1 > 1 - 0.5793 = 0.42

In R: ______________________________

px<-pnorm(84, mean = 83.2, sd = 3.95, lower.tail = FALSE); px; py<-1-px; py > Ans = 0.4197506

2. What is the probability that an individual female turtle sampled at random from this population will be 92 cm or shorter? (2 pts)

By Hand: ______________________________

P(X< 92) = P({X - mu}/sigma) < 92 - 83.2 / 3.95 = 2.2278
Z value is 0.9868

In R: ______________________________

> pnorm(92, mean = 83.2, sd = 3.95, lower.tail = TRUE)

> Ans: 0.9870547

3. What is the probability that any individual female turtle selected at random has a carapace length between 78 cm and 96 cm? (2 pts)

By Hand: ____

P(78 < X < 96) = P({78-83.2/3.95} < X-mu/sigma < 96-83.2/3.95) = -1.3164 < Z < 3.24
= P(Z<3.24) – P(Z<-1.3164)
= 0.9994 - 0.0951 = 0.9043

In R: ____

> pnorm(96, mean = 83.2 , sd = 3.95, lower.tail=TRUE) - pnorm(78, mean = 83.2 , sd = 3.95, lower.tail=TRUE)
> 0.9053928

4.What carapace length represents the 97.5th percentile. (2 pts)

By Hand: ____

> X = mu + $Z_{0.975}$ (sigma) = 83.2 + normsinv(0.975) * 3.95 = 83.2 + (1.96) 3.95 = 90.942

In R: ____

> qnorm(.975, mean = 83.2 , sd = 3.95, lower.tail=TRUE)
> Ans: 90.94186

5. What carapace range represents the 88% of the population. (2 pts)

By Hand: ____

For 88% population
X = mu + $Z_{0.06}$(sigma) = 83.2 + (-1.56 * 3.95) = 77.04 >>> means lower 6%
X = mu + $Z_{0.06}$(sigma) = 83.2 + (1.56 * 3.95) = 89.36 >>> means higher 6%
Hence, range: 77.04 < X < 89.36

In R: ____

> qnorm(0.8810,mean=83.2,sd=3.95)
[1] 87.861
> qnorm(0.1190,mean=83.2,sd=3.95)

[1] 78.539

B. In 1960, the mean age of 452 English Settlers in Jamestown was 26.3 years with a standard deviation of 10.6 years. Answer the following questions by hand and using R. Assume this data set is normally distributed. (8 pts total)

1. What is the probability of selecting from this population a settler that is older than 28 years? (2 pts)

By Hand: ____

X = 28, Mu = 26.3, sigma = 10.6
P(X > 28) = P(Z > X-mu/sigma) = P(Z > 28-26.3 / 10.6) = P(Z>0.16) = 0.5636

In R: ____

> pnorm(28, mean = 26.3, sd = 10.6, lower.tail = TRUE)
> Ans: 0.5637081

2. What is the probability of selecting from this population a settler that is between 28 and 34? (2 pts)

By Hand: ____

P(28 < X < 34) = P({28 – 26.3/10.6} < X-mu/sigma < 34 – 26.3/10.6) = 0.16 < Z < 0.73
= P(Z<0.73) – P(Z<0.16)
= 0.2673 – 0.0636 = 0.2037

In R: ____

> pnorm(34,mean = 26.3, sd = 10.6) - pnorm(28,mean = 26.3, sd = 10.6)
>Ans: 0.2024997

3. What is the probability of selecting from this population a settler that is between 10 and 26? (2pts)

By Hand: ____

P(10 < X < 26) = P({10 – 26.3/10.6} < X-mu/sigma < 26 – 26.3/10.6) = -1.54 < Z < -0.03
= P(Z<-1.54) – P(Z<-0.03)
=0.4382 – 0.012 = 0.4262

In R: ____

> pnorm(26,mean = 26.3, sd = 10.6) - pnorm(10,mean = 26.3, sd = 10.6)
> Ans: 0.4266541

4. What age represent the 50th percentile? (2pts)

By Hand: ____

> P(X > 28) = P(Z > X-mu/sigma) = 50%
= P(Z > X-26.3 / 10.6) = 0.5
Z value for 0.5 is 0
> X = mu + Z(sigma) = 26.3 + (0)(10.6) = 26.3

In R: ____

C. Please provide all your R commands and output in R only for this question (10 pts total).

1) Attach the data on lengths of mosquito catching fish from online (unknownData.txt) and write it to a file named mosFish.txt, read the data back in as mosFish.data (2 pts).

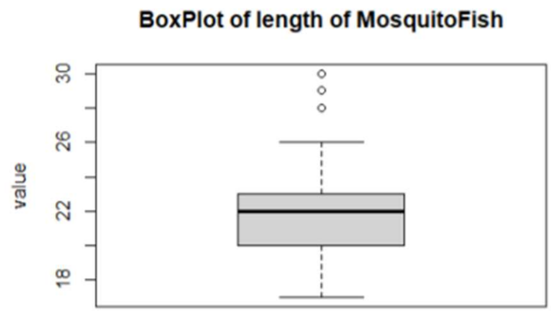
In R: ____

```
> setwd("C:/Users/dmehta12/Desktop/Drashti/Stats/Lab4")
> mosFish.txt<-read.delim("unknownData.txt",header=TRUE,sep="\t")
> mosFish.txt
> write.table(mosFish.txt,"mosFish.data")
> mosFish.txt<-read.table("mosFish.data",header=TRUE,sep="\t")
> attach(mosFish.data)
```

2) Use R to construct a boxplot of the data (2 pts), how many outliers are there? (1 pt)

In R: ____

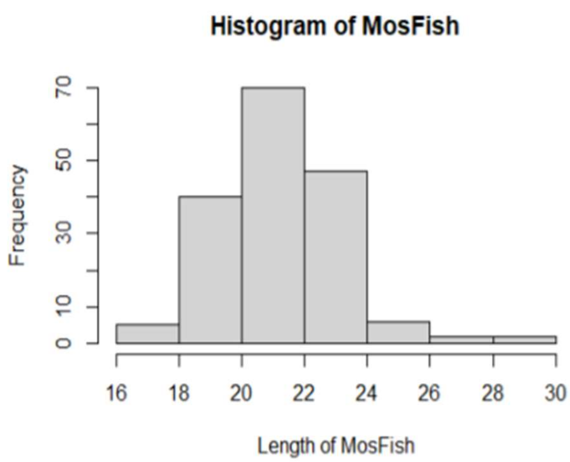
> boxplot(mosFish.data$Lengths.of.Mosquitofish..mm.,ylab="value",main="BoxPlot of length of MosquitoFish")
From the figure, we can see that there are 3 outliners in this boxplot.



3) Use R to construct a histogram, label your histogram x-axis, and provide an informative title (2 pts).

In R: ____

> hist(mosFish.data$Lengths.of.Mosquitofish..mm.,main="Histogram of MosFish",xlab="Length of MosFish")



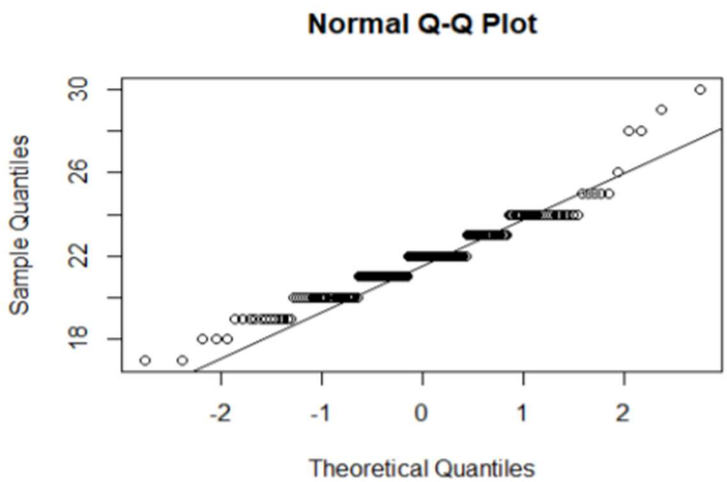
4) Comment on if this variable is approximately normally distributed using a probability plot and shapiro testing (2 pts).

In R: ____

> qqnorm(mosFish.txt$Lengths.of.Mosquitofish..mm.)
> qqline(mosFish.txt$Lengths.of.Mosquitofish..mm.)
> shapiro.test(mosFish.txt$Lengths.of.Mosquitofish..mm.)

Shapiro-Wilk normality test

data: mosFish.txt$Lengths.of.Mosquitofish..mm.
W = 0.94348, p-value = 2.448e-06
This is normally distributed but in the end it is a little right skewed

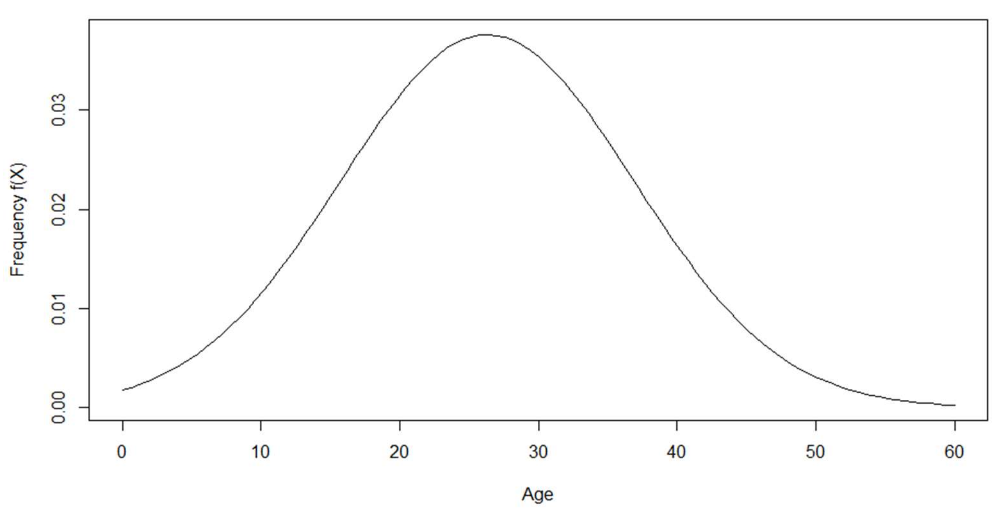


D. Plot a normal distribution in R only (6 pts total).

1) Plot a normal distribution using the data from question B (2 pts), label x and y-axis (1 pt).

In R: ____

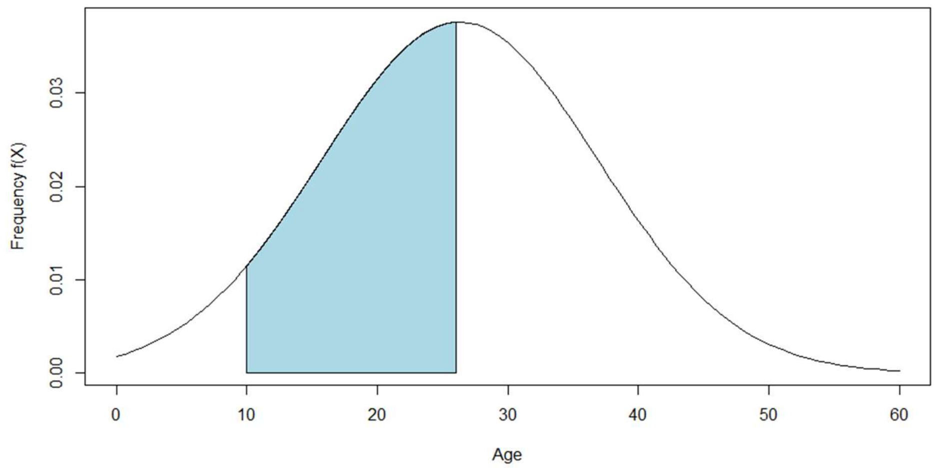
X = 28, Mu = 26.3, sigma = 10.6, Z = 0.16
> f<-function(x){dnorm(x,26.3,10.6)}
> plot(f,0,60,xlab="AGE",ylab="Frequency f(X)")



2) Shade the area described above in B3 light blue in the graph (3 pts).

In R: ___

```
> x<-seq(10,26,0.01)
> polygon(c(10,x,26),c(0,f(x),0),col="light blue")
```



E. The human population is known to have a ratio of 51% males and 49% females. 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. (6 pts total)

1. What is the probability of selecting from this population a random sample of 1000 humans containing 485 **males** or less? (3 pts)

By Hand: ___

p=0.51, q=0.49, n=1000, x=485 or less
n*q and n*p must be more than 5,
n*q=1000*0.49=490
n*p=1000*0.51=510
490 and 510 > 5

Then, we compute for the mean value and standard deviation
$\mu=n*p=1000*0.51=510$
$\sigma2=n*p*q=1000*0.51*0.49=249.9$
$\sigma=249.9=15.808$

$P(x\leq485+0.5) = P(x\leq485.5)$
Next, we find the z-score, to get it, we use the formula

$z=\sigma\mu-x$ > z=15.808510−485.5 > z=1.5498 (this is approximately 1.55)

The area for 1.55 is 0.9394 (Remember, we add 0.5 here, P(x≤485+0.5) = P(x≤485.5))
Now, we subtract 0.5 from 0.9394, and we get 0.4394.

Thus, the probability is **43.94%.**

In R: ____

2. What is the probability of selecting from this population a random sample of 1000 humans containing between 455 and 515 **females**? (3 pts)

By Hand: ____

p=0.49, q=0.51, n=1000, x=between 455 and 515

We are looking for P(455≤x≤515)
but we can also represent it as
P(x≥455)-P(x>515)

We compute for the mean value and standard deviation
$\mu=n*p$=1000∗0.49=490
$\sigma2=n*p*q$=1000∗0.49∗0.51=249.9
σ=249.9=**15.808**

we get, P(x≥455-0.5)-P(x>515+0.5) = P(x≥454.5)-P(x>515.5)
Next, we find the z-score of both, to get it, we use the formula
for P(x≥454.5)
$z=\sigma\mu-x$
z=15.808490−454.5
z=2.25

The area for P(x<454.5) with z=2.25 is **0.9878**,

Thus, P(x≥454.5)=1-P(x<454.5)=1-0.9878=0.0122
(Remember, we subtract 0.5 here P(x≥455-0.5) = P(x≥454.5))
Now, we add 0.5 from 0.0122, and we get *0.5122*.

for P(x>515.5)
$z=\sigma\mu-x$
z=15.808490−515.5
z=−1.61

In R: ____

```
> pnorm(515.1,490,15.808) -pnorm(454.9,490,15.81)
[1] 0.9306291
```