MAT 2379 3X (Spring 2013)
Assignment 3 - solutions

Part I) Answer the following 2 questions **without** the use of R.

[4] 1) Let X be the weight of an Angus steer. We are told that X has a normal distribution with $\mu = 1152$ and $\sigma = 284$.

(a) We want

$$P(X < 1110) = \Phi\left(\frac{1110 - 1152}{284}\right) = \Phi(-0.15) = 0.4404$$

and

$$P(X > 1200) = 1 - \Phi\left(\frac{1200 - 1152}{284}\right) = 1 - \Phi(0.17) = 0.4325.$$

Therefore, a steer weighing more than 1200 pounds is more unusual. (*Albeit not by much.*)

(b) Solving for

$$0.9 = P(X \leq x) = \Phi\left(\frac{x - 1152}{284}\right),$$

we get $1.285 = \frac{x-1152}{284}$. Thus,

$$x = 1.285(284) + 1152 = 1516.94 \quad \text{pounds.}$$

(c) We want x such that $0.85 = P(X \geq x)$. Which is equivalent to

$$0.15 = P(X < x) = \Phi\left(\frac{x - 1152}{284}\right).$$

We get $-1.035 = \frac{x-1152}{284}$. Thus,

$$x = -1.035(284) + 1152 = 858.06 \quad \text{pounds.}$$

[8] 2) We will start by sorting the values in ascending order.

$$4.2, 4.25, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5, 5.2, 5.7.$$

(a) Let y_i be the ith order statistic.
The rank of the median is $(n+1)50\% = 13(0.4) = 6.5$. Thus, the median is

$$\tilde{x} = (1-0.5)y_6 + 0.5y_7 = 0.5(4.6) + 0.5(4.7) = 4.65.$$

The rank of the first quartile is $(n+1)25\% = 13(0.25) = 3.25$. Thus, the first quartile is

$$q_1 = (1-0.25)y_3 + 0.25y_4 = 0.75(4.3) + 0.25(4.4) = 4.325.$$

The rank of the third quartile is $(n+1)75\% = 13(0.75) = 9.75$. Thus, the first quartile is

$$q_3 = (1-0.75)y_9 + 0.75y_{10} = 0.25(4.9) + 0.75(5) = 4.975.$$

(b) The interquartile range is $\text{IQR} = q_3 - q_1 = 4.975 - 3.25 = 1.725$.

To assess if there are any outliers, we compute the upper fence

$$q_3 + 1.5\,\text{IQR} = 4.925 + 1.5\,(1.725) = 7.5125$$

and the lower fence

$$q_1 - 1.5\,\text{IQR} = 4.325 - 1.5\,(1.725) = 1.7375.$$

Since no values are outside of these fences, then there are no outliers.

Part II) Answer the following 4 questions **with** the use of R.

Remarks:

- You must provide the R commands and output that were used in answering the question.
- The R output alone is not an answer to a question. The R output is used to support your answer.
- Please do not printout your whole R session. Only provide the R commands and output that are necessary to answer the question.

[3] 3) Let X be the weight of an Angus steer. We are told that X has a normal distribution with $\mu = 1152$ and $\sigma = 284$.

(a) We want $P(X < 1110) = 0.4412$ and
$P(X > 1200) = 1 - P(X \leq 1200) = 0.4329$.
Therefore, a steer weighing more than 1200 pounds is more unusual. (*Albeit not by much.*)

R output and commands:

```
> pnorm(1110,1152,284)
[1] 0.4412158
> 1-pnorm(1200,1152,284)
[1] 0.4328928
```

(b) We want x such that $0.9 = P(X \leq x)$. So $x = 1515.961$ pounds.

R output and commands:

```
> qnorm(0.9,1152,284)
[1] 1515.961
```

(c) We want x such that $0.85 = P(X \geq x)$, which is equivalent to $0.15 = P(X < x)$. We get $x = 857.6529$ pounds.

R output and commands:

```
> qnorm(0.15,1152,284)
[1] 857.6529
```

[4] 4) Refer to Problem 16.2 in the textbook. Create a dataframe in R that contains the data from the file *ICE-SHELF.txt*.

(a) We start by assigning the data to a dataframe that we will call *table*. We display the names of the variables in the dataframe and we assign the precipitation to the variable x.

R output and commands:

```
> table = read.table(file.choose(),header=TRUE,sep="\t")
> names(table)
[1] "year"          "precipitation"
> x=table$precipitation
```

(b) The mean is $\overline{x} = 160.5484$ mm and the standard deviation is $s = 38.84914$ mm.

R output and commands:

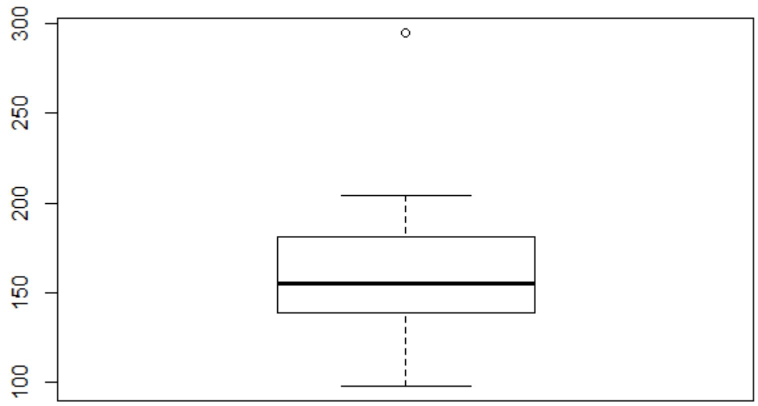
```
> mean(x)
[1] 160.5484
> sd(x)
[1] 38.84914
```

(c) The median is $\tilde{x} = 155$ mm, the first quartile is $q_1 = 138$ mm and the third quartile is $q_3 = 186$ mm.

The boxplot for the precipitation is displayed below. We observe that there is only one outlier in the sample. The outlier is the largest value which is a total annual precipitation of 295 mm.

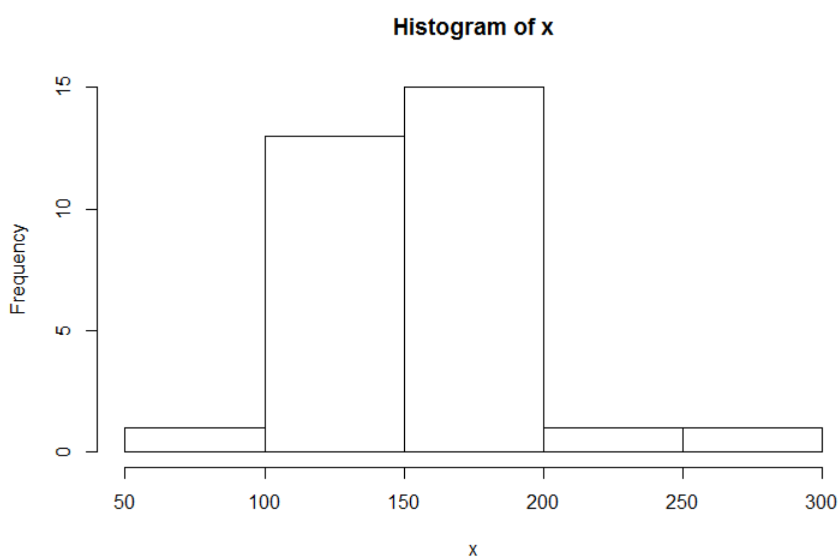
R output and commands:

```
> quantile(x,type=6)
  0%  25%  50%  75% 100%
  98  138  155  186  295
> boxplot
```



(d) The histogram for the precipitation is displayed below.
R output and commands:

```
> hist(x)
```



[4] 5) We import the data from the file and display it.

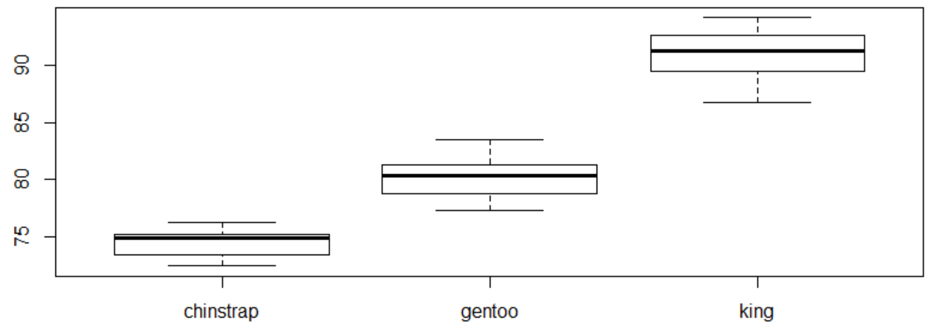
```
> table = read.table(file.choose(),header=TRUE,sep="\t")
> names(table)
[1] "king"      "gentoo"    "chinstrap"
> table
   king gentoo chinstrap
1  93.2   78.5      73.2
2  91.2   79.2      76.3
3  94.1   81.2      74.5
4  89.3   83.5      74.9
5  88.6   79.1      75.2
6  90.5   81.5      73.1
7  93.1   80.4      72.5
8  89.5   77.3      76.3
9  92.1   78.4      74.9
10 86.7   82.3      75.2
11 91.3   80.4      73.7
```

This is not a proper dataframe as we defined in class, since we have more than one observational unit per row. We should reformat the data where we have one variable for the variable length and one variable identifying the species. We import the reformated data and display the names of the variables.

```
> table = read.table(file.choose(),header=TRUE,sep="\t")
> names(table)
[1] "length"  "species"
```

We are now ready to build the side-by-side boxplots.

```
> boxplot(table$length~table$species)
```



There is more variability in the lengths of king penguins, compared with the other two species. Moreover the median length is largest for the king penguins and the median length of the gentoo penguins is larger than the median length of the chinstrap penguins.

[**4**] 6) Let X be the number of children that are deficient in vitamin D among the $n = 100$ selected children. X has a binomial distribution with $n = 100$ and $p = 0.09$.

(a) We want $P(8 \leq X \leq 13) = P(X \leq 13) - P(X \leq 7) = 0.6228$.
R output and commands:

```
> pbinom(13,100,0.09)-pbinom(7,100,0.09)
[1] 0.6227533
```

(b) We want $P(X > 20) = 1 - P(X \leq 20) = 0.000198$.
R output and commands:

```
> 1-pbinom(20,100,0.09)
[1] 0.0001983642
```

[/**27**]