Biostatistics Exam 1

Name: __________________

For all problems below, show how to calculate each quantity for full credit:

Suppose you make the observations X as shown in a single vector here:

1. From the data in vector X (see R output), show how to calculate the following descriptive variables.

a. sample variance s^2:

b. sample median

c. standard error of the mean:

d. upper and lower bounds for the 95% Confidence Interval for the mean.

e. What's the sample variance if the sample X is transformed by the linear transformation:

$Y = aX+b$ where $a = 3$ and $b = 4$?

f. What's the sample variance if the data in X are standardized? Explain.

2. Considering the same sample X as above, test the hypothesis that the mean of the population from which it is drawn is equal to 2.0 (two-tailed test) at type I error rate of $\alpha = 0.05$. ***Show your work.***

Assumptions:

Hypotheses:

Test Statistic:

Critical Value of the Test:
(choose appropriate value from R output)

Decision Rule:

Result:

3. Suppose you collect the following *paired* samples A & B (see R output):

a. What does it mean to have *paired* data?

b. What test(s) might you employ to see if the *means* or *medians* are different? ***Explain when you would use each.***

4. Suppose the samples V and W (see R output) are derived from *non-paired* populations with *unequal* variances, test the hypothesis that the mean of population V is *greater than* the mean of population W, at $\alpha = 0.05$. *Show your work*.

Assumptions:

Hypotheses:

Test Statistic:

Probability Using Satterthwaite's degrees of freedom = 5.4 :
(choose appropriate value from R script)

Decision Rule:

Result:

Assuming instead a two-sided test, what formula would you use to calculate an absolute value for Critical Values?

Using absolute value for critical values |C| = 2.5, what's the formula for the Lower Bound of the CI?

5. For each of the statistical problems below involving orthodontics measurments made for male and female patients, vectors C & D (see R output), please provide information on:

A. *assumptions* - Necessary conditions making the test valid.
B. *hypotheses tested* - The null and alternative hypotheses.
C. t*est statistic*
D *Decision Rule*
E *Result*

a. Test for equal variance between populations of males versus females, with Type I error rate $\alpha = 0.01$:

A

B

C

D

E

b. Based on your results in a, test whether males and females have the same or different distance measurements,with Type I error rate $\alpha = 0.05$:

A

B

C

D

E

6. QQ plot results for vectors C & D are displayed in the R report sheet.

a. What do the X &Y axes in the qqplot mean?

b. Interpret the results of the qqplot.

c. Take a look at the results of wilcox.test() in the R sheet. What conclusions do you draw from this test?

d. What's the meaning of the reported value W = 749?

7. In the general logic of probability, consider the following probabilities for events A and B:

$P(A) = 0.3$

$P(B) = 0.8$

a. What's the probability of the Union of events A & B, assuming events A and B are potentially co-occurring and *independent?*

b. When does the *Law of Multiplied Probabilities* for events A & B apply?

8. A genetic disease occurs in the population at large in 2% of an island population. If a test procedure for the disease has a sensitivity of 90% and a False Positive rate (1-specificity) of 1%, what's the Bayesian probability for a particular patient given a positive result for the test procedure? ***Be sure to fill out the appropriate Bayesian chart completely, and show how to calculate the final result.***

```
> #BIOSTATISTICS EXAM 1
>
> #QUESTION 1:
> X=c(1,2,6,4,2)
> mean(X)
[1] 3
> var(X)
[1] 4
>
> #QUESTION 2:
> A=c(6,6,7,5)
> B=c(5,6,5,6)

> #PROBABILITY DISTRIBUTION TABLES:
> alpha=0.05
>
> dchisq(alpha,4)
[1] 0.01219137
> pchisq(alpha,4)
[1] 0.0003073402
> qchisq(alpha,4)
[1] 0.710723
> dchisq(alpha,5)
[1] 0.001450062
> pchisq(alpha,5)
[1] 2.920954e-05
> qchisq(alpha,5)
[1] 1.145476
>
> dt(alpha,4)
[1] 0.3744147
> pt(alpha,4)
[1] 0.5187402
> qt(alpha,4)
[1] -2.131847
> dt(alpha,5)
[1] 0.3790378
> pt(alpha,5)
[1] 0.5189709
> qt(alpha,5)
[1] -2.015048
>
> df(alpha,4,5)
[1] 0.2346973
> pf(alpha,4,5)
[1] 0.006224779
> qf(alpha,4,5)
[1] 0.1598451
>
> dchisq(alpha/2,4)
[1] 0.006172361
> pchisq(alpha/2,4)
[1] 7.7477e-05
> qchisq(alpha/2,4)
[1] 0.4844186
> dchisq(alpha/2,5)
[1] 0.0005191228
> pchisq(alpha/2,5)
[1] 5.20982e-06
> qchisq(alpha/2,5)
[1] 0.8312116
>
> dt(alpha/2,4)
[1] 0.3748536
> pt(alpha/2,4)
[1] 0.5093738
> qt(alpha/2,4)
[1] -2.776445
> dt(alpha/2,5)
[1] 0.3794644
> pt(alpha/2,5)
[1] 0.509489
> qt(alpha/2,5)
[1] -2.570582
>
> df(alpha/2,4,5)
[1] 0.1280641
> pf(alpha/2,4,5)
[1] 0.001649186
> qf(alpha/2,4,5)
[1] 0.1067866
>
> dchisq(1-alpha,4)
[1] 0.1476977
> pchisq(1-alpha,4)
[1] 0.08271954
> qchisq(1-alpha,4)
[1] 9.487729
> dchisq(1-alpha,5)
[1] 0.07657453
> pchisq(1-alpha,5)
[1] 0.03350303
> qchisq(1-alpha,5)
[1] 11.0705
>
> dt(1-alpha,4)
[1] 0.2254952
> pt(1-alpha,4)
[1] 0.8020464
> qt(1-alpha,4)
[1] 2.131847
> dt(1-alpha,5)
[1] 0.2307469
> pt(1-alpha,5)
[1] 0.8071313
> qt(1-alpha,5)
[1] 2.015048
>
> df(1-alpha,4,5)
[1] 0.417931
> pf(1-alpha,4,5)
[1] 0.4939578
> qf(1-alpha,4,5)
[1] 5.192168
>
> dchisq(1-alpha/2,4)
[1] 0.1497015
> pchisq(1-alpha/2,4)
[1] 0.08643718
> qchisq(1-alpha/2,4)
[1] 11.14329
> dchisq(1-alpha/2,5)
[1] 0.07862799
> pchisq(1-alpha/2,5)
[1] 0.03544311
> qchisq(1-alpha/2,5)
[1] 12.8325
>
> dt(1-alpha/2,4)
[1] 0.220055
> pt(1-alpha/2,4)
[1] 0.8076157
> qt(1-alpha/2,4)
[1] 2.776445
> dt(1-alpha/2,5)
[1] 0.2251936
> pt(1-alpha/2,5)
[1] 0.8128304
> qt(1-alpha/2,5)
[1] 2.570582
>
> df(1-alpha/2,4,5)
[1] 0.4076643
> pf(1-alpha/2,4,5)
[1] 0.5042773
> qf(1-alpha/2,4,5)
[1] 7.387886
>
```

```
> #QUESTION 3:
> library(car)
> A=c(8,6,7,5)
> B=c(7,6,7,6)
>
>
> #QUESTION 4:
> V=c(9,8,8,7)
> W=c(7,8,7,8)
> mean(V)
[1] 8
> mean(W)
[1] 7.5
> var(V)
[1] 0.6666667
> var(W)
[1] 0.3333333
> alpha=0.05
> #Satterthwaite's degree of freedom (dS):
> dS=5.4
> 2*pt(1,dS)
[1] 1.640009
> 2*(1-pt(1,dS))
[1] 0.3599907
> pt(1,dS)
[1] 0.8200047
> 1-pt(1,dS)
[1] 0.1799953
>
> #QUESTION 5:
> setwd("~/DATA/Models")
> DATA=read.table("Orthodont.txt")
> attach(DATA)
> C=distance[Sex=="Male"]
> D=distance[Sex=="Female"]
> detach(DATA)
> C
 [1] 26.0 25.0 29.0 31.0 21.5 22.5 23.0 26.5 23.0 22.5 24.0 27.5 25.5 27.5 26.5 27.0 20.0 23.5 22.5 26.0
[21] 24.5 25.5 27.0 28.5 22.0 22.0 24.5 26.5 24.0 21.5 24.5 25.5 23.0 20.5 31.0 26.0 27.5 28.0 31.0 31.5
[41] 23.0 23.0 23.5 25.0 21.5 23.5 24.0 28.0 17.0 24.5 26.0 29.5 22.5 25.5 25.5 26.0 23.0 24.5 26.0 30.0
[61] 22.0 21.5 23.5 25.0
> D
 [1] 21.0 20.0 21.5 23.0 21.0 21.5 24.0 25.5 20.5 24.0 24.5 26.0 23.5 24.5 25.0 26.5 21.5 23.0 22.5 23.5
[21] 20.0 21.0 21.0 22.5 21.5 22.5 23.0 25.0 23.0 23.0 23.5 24.0 20.0 21.0 22.0 21.5 16.5 19.0 19.0 19.5
[41] 24.5 25.0 28.0 28.0
> length(DATA$distance)
[1] 108
>
> var.test(C,D,alternative="two.sided",confidence.level=0.99)

       F test to compare two variances

data:  C and D
F = 1.4627, num df = 63, denom df = 43, p-value = 0.1883
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.8277844 2.5098586
sample estimates:
ratio of variances
           1.46267
```

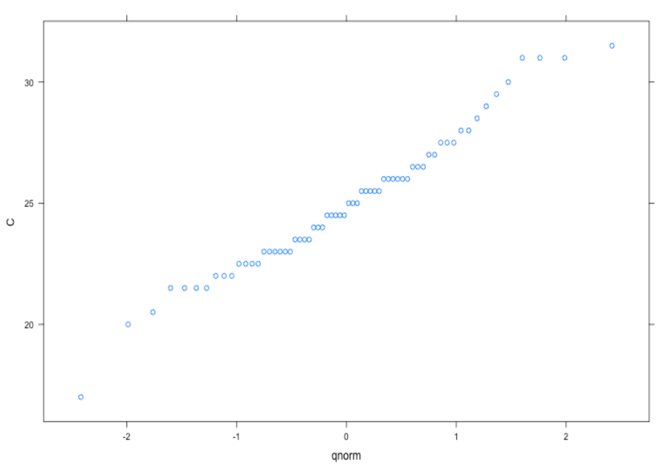
```
> leveneTest(DATA$distance,DATA$Sex,center=mean)
Levene's Test for Homogeneity of Variance (center = mean)
       Df F value Pr(>F)
group   1  1.4229 0.2356
      106
> t.test(C,D,alternative="two.sided",var.equal=TRUE,confidence.level=0.99)

       Two Sample t-test

data:  C and D
t = 4.3769, df = 106, p-value = 2.831e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.269663 3.372382
sample estimates:
mean of x mean of y
 24.96875  22.64773

> t.test(C,D,alternative="greater",var.equal=TRUE,confidence.level=0.95)

       Two Sample t-test

data:  C and D
t = 4.3769, df = 106, p-value = 1.416e-05
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.441076      Inf
sample estimates:
mean of x mean of y
 24.96875  22.64773

> t.test(C,D,alternative="less",var.equal=TRUE,confidence.level=0.99)

       Two Sample t-test

data:  C and D
t = 4.3769, df = 106, p-value = 1
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 3.200969
sample estimates:
mean of x mean of y
 24.96875  22.64773

> t.test(C,D,alternative="two.sided",var.equal=FALSE,confidence.level=0.95)

       Welch Two Sample t-test

data:  C and D
t = 4.5333, df = 102.33, p-value = 1.58e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.305529 3.336517
sample estimates:
mean of x mean of y
 24.96875  22.64773
```

```
> t.test(C,D,alternative="greater",var.equal=FALSE,confidence.level=0.99)

	Welch Two Sample t-test

data:  C and D
t = 4.5333, df = 102.33, p-value = 7.902e-06
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.471177      Inf
sample estimates:
mean of x mean of y 
 24.96875  22.64773 

> t.test(C,D,alternative="less",var.equal=FALSE,confidence.level=0.95)

	Welch Two Sample t-test

data:  C and D
t = 4.5333, df = 102.33, p-value = 1
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 3.170868
sample estimates:
mean of x mean of y 
 24.96875  22.64773 

> 
```

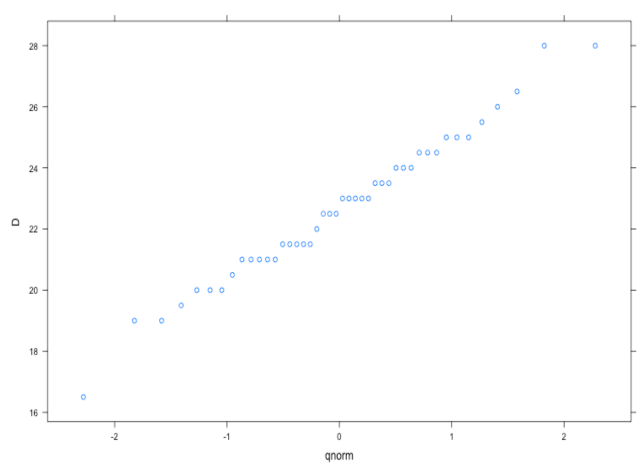
> #QUESTION 6:

```
> library(lattice)
```

qqmath(C)



qqmath(D)




```
> 
> wilcox.test(DATA$distance~DATA$Sex,paired=FALSE,exact=F,alternative="two.sided")

	Wilcoxon rank sum test with continuity correction

data:  DATA$distance by DATA$Sex
W = 749, p-value = 3.709e-05
alternative hypothesis: true location shift is not equal to 0
> 
```