STAT 5309 – SP 2022

LAB 1

****

****CONTENTS: Set up data - t-test - variance test**

DUE: Thurs, Feb 3

A. PRACTICE

1. Set up and explore data

One factor of 2 levels: Mortar under Modified and Unmodified. Test to see if the factor is "significant".

Data: Tension Bond Strength.

Modified	Unmodified
16.85	16.62
16.40	16.75
17.21	17.37
16.35	17.12
16.52	16.98
17.04	16.87
16.96	17.34
17.15	17.02
16.59	17.08
16.57	17.27

#read each column into a vector; convert matrix into data frame

y1<- c(16.85,16.40,17.21,16.35,16.52,17.04,16.96,17.15,16.59,16.57)

y2 <- c(16.62,16.75,17.37,17.12,16.98,16.87,17.34,17.02,17.08,17.27)

data <- data.frame(cbind(y1,y2))

names(data) <-c("Modified", "Unmodified")   #change column names

attach(data)

```
> data
  Modified Unmodified
1    16.85      16.62
```

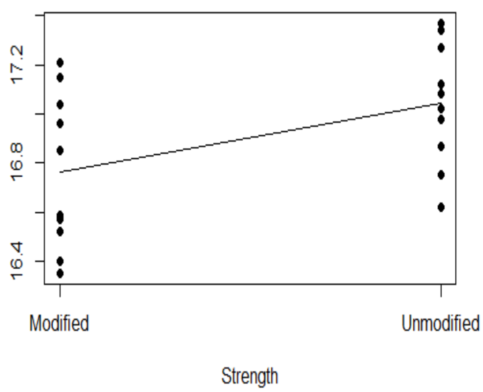
```
2      16.40      16.75
3      17.21      17.37
4      16.35      17.12
5      16.52      16.98
6      17.04      16.87
7      16.96      17.34
8      17.15      17.02
9      16.59      17.08
10     16.57      17.27
```

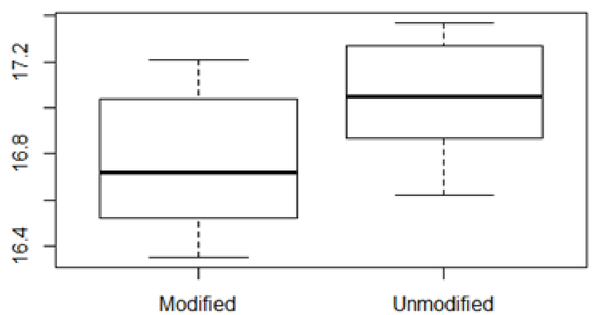
means <- apply(data,2,mean)   #2 for column mean, 1 for row mean

```
> means
  Modified Unmodified
    16.764     17.042
```

#----------------stripchart(); boxplot()--------------------

stripchart(data, xlab="Strength", vertical=TRUE,pch=16)   #Stripchart
lines(means)





boxplot(data)   #boxplot

##another way to set up data
methods<- rep(c("Modified", "Unmodified"), each=10)
strength<- c(y1,y2)

# combining vectors gives a matrix, not a data frame.
data.1 <-data.frame(cbind(methods, strength))

```
> data.1
      methods  strength
1    Modified     16.85
2    Modified      16.4
3    Modified     17.21
4    Modified     16.35
```

```
5    Modified    16.52
6    Modified    17.04
7    Modified    16.96
8    Modified    17.15
9    Modified    16.59
10   Modified    16.57
11 Unmodified    16.62
12 Unmodified    16.75
13 Unmodified    17.37
14 Unmodified    17.12
15 Unmodified    16.98
16 Unmodified    16.87
17 Unmodified    17.34
18 Unmodified    17.02
19 Unmodified    17.08
   20 Unmodified    17.27
```

Perform tests to compare means; variances

2. Two-sample Z-test(if variances are known)

```
> z.test(y1,y2, sigma.x=.1, sigma.y=.05, conf.level=.95)

        Two-sample z-Test

data:  y1 and y2
z = -7.863, p-value = 3.75e-15
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3472952 -0.2087048
sample estimates:
mean of x mean of y
   16.764    17.042
```

Note: sigmas mostly are unknown.

3. Two-sample t-test (Equal /Unequal Variance)

(if population variances are unknown, believed to be equal ; or tested to be equal)

##t-test
t.test(Modified,Unmodified,var.equal=TRUE   # if set up as data

attach(data.1)
t.test(strength ~ methods, var.equal=TRUE)   # if set up as data.1

```
> t.test(strength ~ methods, var.equal=TRUE)       # if set up as data.1

        Two Sample t-test

data:  strength by methods
t = -2.1869, df = 18, p-value = 0.0422
```

```
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.54507339 -0.01092661
sample estimates:
  mean in group Modified mean in group Unmodified
                  16.764                    17.042
```

Note: P-values=0.0422 < alpha=.05.

We reject Null, at the 95% significant level(2 means are not equal)

```
> t.test(strength ~ methods, var.equal=FALSE)

        Welch Two Sample t-test

data:  strength by methods
t = -2.1869, df = 17.025, p-value = 0.043
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.546174139 -0.009825861
sample estimates:
  mean in group Modified mean in group Unmodified
                  16.764                    17.042
```

4. **Paired t-test: t.test()**

Data: Hardness.

A rod with a pointed tip is pushed into a metal specimen with a known force. Two different tips (Tip 1 and Tip 2) are used, the hardness of the metal is determined (by measuring the depression caused by the tips).

Specimen	Tip 1	Tip 2
	7 , 3 , 3, 4, 8, 3, 2, 9, 5, 4	6, 3, 5, 3, 8, 2, 4, 9, 4, 5

```
x1<- c(7 , 3 , 3, 4, 8,  3, 2, 9, 5, 4)
x2 <-c(6, 3, 5, 3, 8, 2, 4, 9, 4, 5)
measurement<- c(x1, x2)
type <- rep(c("tip1","tip2"), each=10)

hardness<- data.frame(cbind(type, measurement))
hardness
> hardness
   type measurement
1  tip1           7
2  tip1           3
3  tip1           3
```

```
4  tip1           4
5  tip1           8
6  tip1           3
7  tip1           2
8  tip1           9
9  tip1           5
10 tip1           4
11 tip2           6
12 tip2           3
13 tip2           5
14 tip2           3
15 tip2           8
16 tip2           2
17 tip2           4
18 tip2           9
19 tip2           4
20 tip2           5

> t.test(measurement ~ type, pair=TRUE)

        Paired t-test

data:  measurement by type
t = -0.26414, df = 9, p-value = 0.7976
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9564389  0.7564389
sample estimates:
mean of the differences
                   -0.1
```

Note: p-value=0.7976 > alpha=.05.
Do not reject Ho(ie, support Ho): 2 tips have equal average measurements

5. **Variance test (test whether variances are equal): var.test()**

```
> var.test(y1,y2)

        F test to compare two variances

data:  y1 and y2
F = 1.6293, num df = 9, denom df = 9, p-value = 0.4785
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4046845 6.5593806
sample estimates:
ratio of variances
          1.629257
```

6. Bartlett.test(): to test variance, for more than 2 samples. However it still works for 2 samples

```
> bartlett.test(data)

        Bartlett test of homogeneity of variances

data:  data
Bartlett's K-squared = 0.50292, df = 1, p-value = 0.4782
```

7. **Compare 2 means if distribution are not known: wilcox.test()**

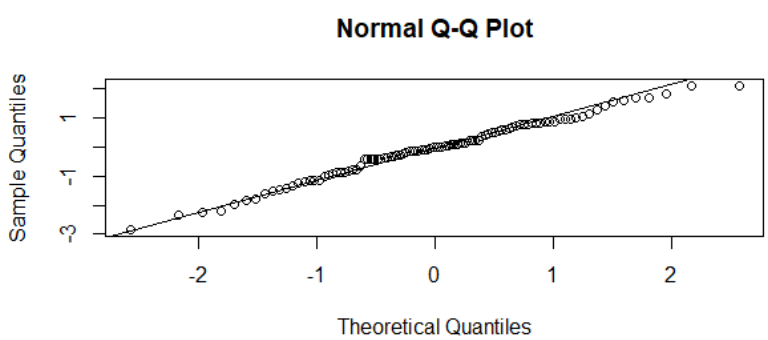
```
> wilcox.test(y1, y2)

        Wilcoxon rank sum exact test

data:  y1 and y2
W = 24, p-value = 0.05243
alternative hypothesis: true location shift is not equal to 0
```

B. Extra

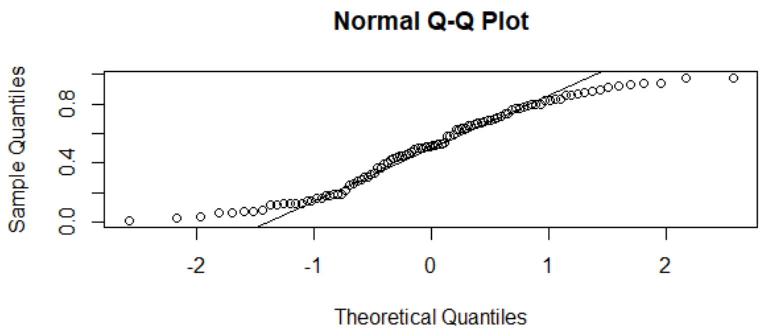
1. Assumptions for all tests: Populations are NORMAL

sample1<- rnorm(100) # sample of 100 from Normal population

qqnorm(sample1)
qqline(sample1)



sample2 <-runif(100) # Sample of 100 from uniform population)
qqnorm(sample2)
qqline(sample2)



2. F distribution: compare 2 variances

1. Chisquare variable

$$\frac{SSX}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2\,(n-1);$$

2. F variable: Assume Ho: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$F_0 = \frac{S_1^2}{S_2^2} = \left[\frac{(n_1-1)S_1^2}{\sigma^2} \,/\, (n_1-1)\right] \,/\, \left[\frac{(n_2-1)S_2^2}{\sigma^2} \,/ (n_2-1)\right]$$

$$= \frac{\left(\frac{SSX_1}{\sigma^2}\right)/(n_1-1)}{\left(\frac{SSX_2}{\sigma^2}\right)/(n_2-1)} = \left(\frac{\chi^2_{n_1-1}}{n_1-1}\right) / \left(\frac{\chi^2_{n_2-1}}{n_2-1}\right) \sim F_{n_1-1,n_2-1}$$

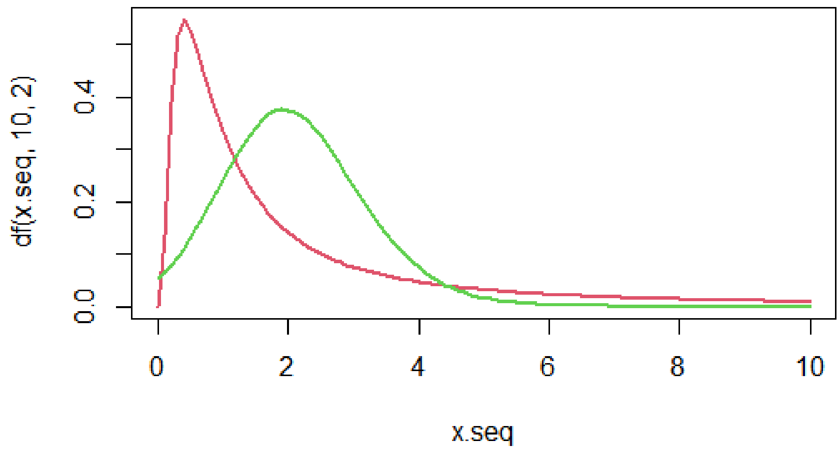
$$F_o = \frac{S_1^2}{S_2^2} \sim F_{n_1-1,n_2-1}$$

And Fo ≈ 1. If Ho is not true, then Fo >> 1

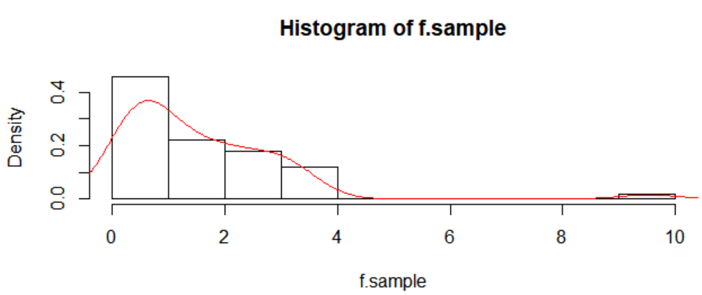
This F-test is use in comparing more than 2 levels (one factor with more than 2 levels)

 # F-distribution PDF

```
## F random sample;  rf(n, df1, df2)

x.seq <- seq(0, 10, by=0.1)
plot(x.seq, df(x.seq, 10,2), type="l",lwd=2,  col=2)
points(x.seq, dt(x.seq, 20, 2), type="l",lwd=2,  col=3)
```




```
> x.chisq <- rchisq(50, 10)
> y.chisq <- rchisq(50, 5)    # 50 random form chisquare, df=5
> f.sample <- (x.chisq/10)/(y.chisq/5)

> hist(f.sample, probability = TRUE)

> points(density(f.sample), type="l", col="red")
```



C. PROBLEMS

1. Problem

2-5. The shelf life of a carbonated beverage is of interest. Ten bottles are randomly selected and tested, and the following results are obtained:

Days	
108	138
124	163
124	159
106	134
115	139

(a) We would like to demonstrate that the mean shelf life exceeds 120 days. Set up appropriate hypotheses for investigating this claim.
(b) Test these hypotheses using $\alpha = 0.01$. What are your conclusions?

(c) Find the P-value for the test in part (b).
(d) Construct a 99 percent confidence interval on the mean shelf life.

2. Problem

2-9. Two machines are used for filling plastic bottles with a net volume of 16.0 ounces. The filling processes can be assumed to be normal, with standard deviations of $\sigma_1 = 0.015$ and $\sigma_2 = 0.018$. The quality engineering department suspects that both machines fill to the same net volume, whether or not this volume is 16.0 ounces. An experiment is performed by taking a random sample from the output of each machine.

Machine 1		Machine 2	
16.03	16.01	16.02	16.03
16.04	15.96	15.97	16.04
16.05	15.98	15.96	16.02
16.05	16.02	16.01	16.01
16.02	15.99	15.99	16.00

(a) State the hypotheses that should be tested in this experiment.
(b) Test these hypotheses using $\alpha = 0.05$. What are your conclusions?
(c) Find the P-value for this test.
(d) Find a 95 percent confidence interval on the difference in mean fill volume for the two machines.

(b) Perform a var.test() to check if 2 samples come from equal variance populations.
(e) Perform a boxplot and a stripchart with line connecting means.

3. Problem

2-11. The following are the burning times of chemical flares of two different formulations. The design engineers are interested in both the mean and variance of the burning times.

Type 1		Type 2	
65	82	64	56
81	67	71	69
57	59	83	74
66	75	59	82
82	70	65	79

(a) Test the hypothesis that the two variances are equal. Use $\alpha = 0.05$.
(b) Using the results of (a), test the hypothesis that the mean burning times are equal. Use $\alpha = 0.05$. What is the P-value for this test?
(c) Discuss the role of the normality assumption in this problem. Check the assumption of normality for both types of flares.

(a) Use var.test()