STAT 5309 -SP 2022

LAB 2

****

****CONTENTS: One factor design – Residuals – Multiple comparisons(Fisher LSD, Tukey HSD)**

***DUE: Thurs, Feb 10**

A. PRACTICE

1. One factor

Set up data and explore data

Suppose **one factor** has 5 levels A,B,C,D,E (factor levels or treatment levels)
A <- c(7, 7, 15, 11, 9)
B <- c(12, 17, 12, 18, 18)
C <- c(14, 18, 18, 19, 19)
D <- c(19, 25, 22, 19, 23)
E <- c(7, 10, 11, 15, 11)

```
temperature <- c(A,B,C,D,E)

trt<- rep(c("A", "B","C", "D","E"), each=5)
trt

Temperature <- data.frame(cbind(trt, temperature))

> Temperature
   trt temperature
1    A           7
2    A           7
3    A          15
4    A          11
5    A           9
6    B          12
7    B          17
8    B          12
9    B          18
10   B          18
11   C          14
12   C          18
13   C          18
14   C          19
15   C          19
16   D          19
17   D          25
18   D          22
19   D          19
20   D          23
21   E           7
22   E          10
23   E          11
24   E          15
25   E          11
```

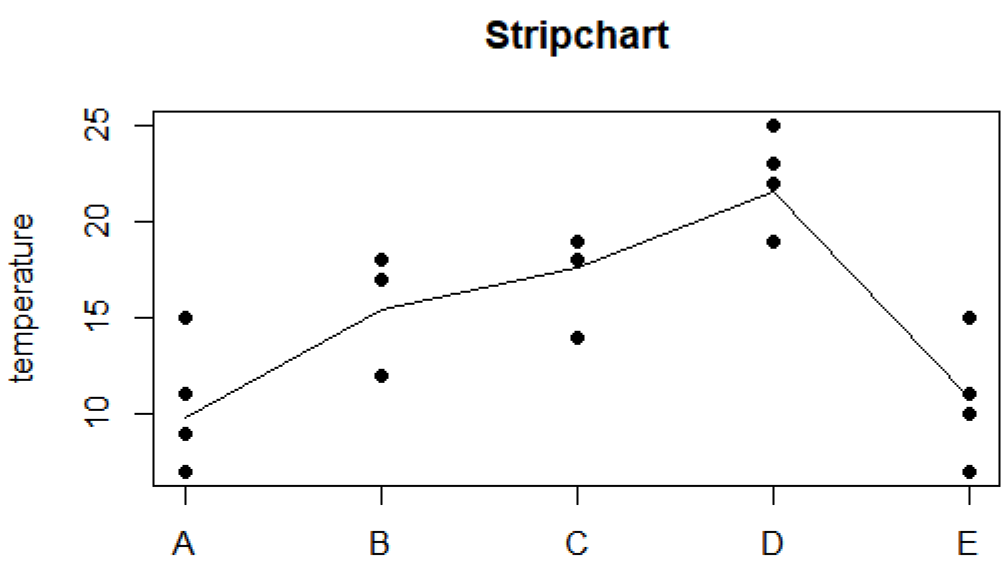
attach(Temperature)
trt<- factor(trt)   #make trt into a factor

Explore data: tapply()

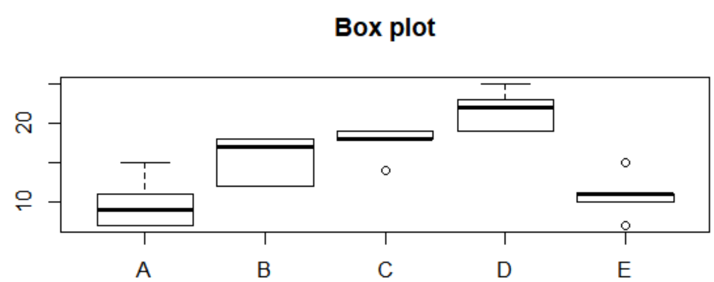
```
> trt.means <- tapply(temperature, trt, mean)
> trt.means
   A    B    C    D    E 
 9.8 15.4 17.6 21.6 10.8 
```

Exploring plots

stripchart(temperature ~ trt, vertical=TRUE,pch=16, main="Stripchart")
lines(trt.means)




```
> boxplot(temperature ~trt, main="Box plot")
```



2. Linear model: aov(); lm()

```
>model<- aov(temperature ~ trt)

> summary.aov(model)
            Df Sum Sq Mean Sq F value   Pr(>F)    
trt          4  475.8  118.94   14.76 9.13e-06 ***
Residuals   20  161.2    8.06                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Notes: (i) degree of freedom: df(trt) = a-1 (4)

(ii) df (Residuals) = N- a (20)

(iii) df(SST) = N-1

(iv) F-value = 14.76; P-value=9.13*10^-6

Reject Ho: $\mu_1 = .. = \mu_5$

```
> summary.lm(model)

Call:
aov(formula = temperature ~ trt)

Residuals:
   Min     1Q Median     3Q    Max 
  -3.8   -2.6    0.4    1.4    5.2 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)    9.800      1.270   7.719 2.02e-07 ***
trtB           5.600      1.796   3.119 0.005409 ** 
trtC           7.800      1.796   4.344 0.000315 ***
trtD          11.800      1.796   6.572 2.11e-06 ***
trtE           1.000      1.796   0.557 0.583753    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.839 on 20 degrees of freedom
Multiple R-squared:  0.7469,	Adjusted R-squared:  0.6963 
F-statistic: 14.76 on 4 and 20 DF,  p-value: 9.128e-06
```

Notes:

(i) Equation of the linear model:

$$temperature = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$
$$= 9.8 + 5.6 trtB + 7.8 trtC + 11.8 trtD + 1 trtE$$

trtB, trtC, trtD, trtE are **Indicator variables**, created by R.

(ii) These ndicator variable takes values { 0,1}. For example: find the mean response for factor level B, by setting *trtB=1, trtC=trtD=trtE =0*, give average for treatment B,etc.
For mean response for factor A; set *all trtB =trtC =trt D= trtE=0*
So temperature (A) =9.8
Temperature (B) = 9.8 +5.6(1) = 15.4

The control (or reference) treatment is Treatment A. R chooses control by alphabetical order.
Can change the reference level by
trt <- relevel(trt, ref=" B") # B is the control level now

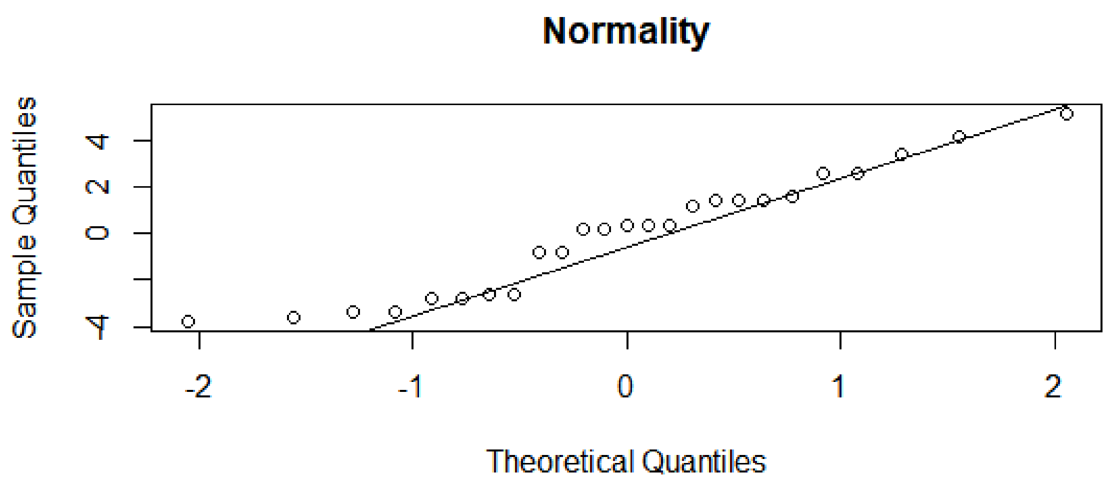
```
> trt.means
   A    B    C    D    E
 9.8 15.4 17.6 21.6 10.8


Note:  from the output,Intercept =9.8
```

Fitted values

```
> temperature.aov$fitted.values
   1    2    3    4    5    6    7    8    9   10   11   12   13   14
 9.8  9.8  9.8  9.8  9.8 15.4 15.4 15.4 15.4 15.4 17.6 17.6 17.6 17.6
  15   16   17   18   19   20   21   22   23   24   25
17.6 21.6 21.6 21.6 21.6 21.6 10.8 10.8 10.8 10.8 10.8
```

3. Check residuals (Model assumptions)

(a) Normality
res<- temperature.aov$residuals
qqnorm(res, main="Normality")
qqline(res)



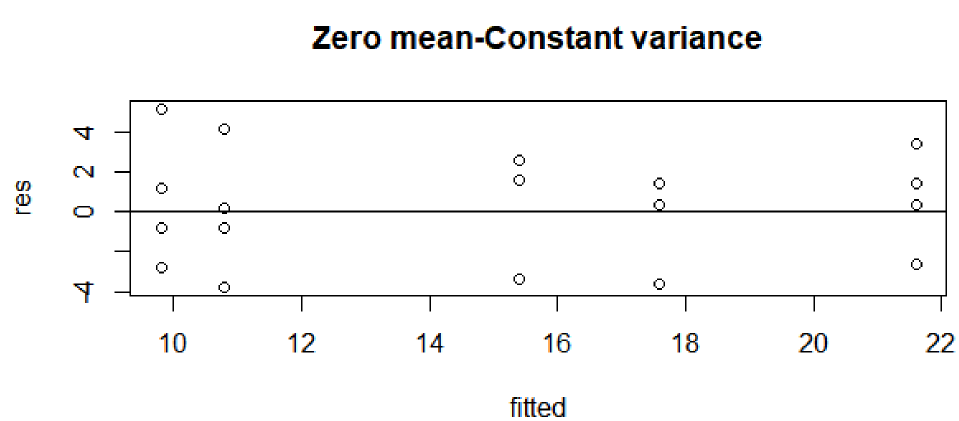

```
> shapiro.test(res)

        Shapiro-Wilk normality test

data:  res
W = 0.94387, p-value = 0.1818
```

Note: Null: Normal.

Support Null

(b) Zero mean/Constant variance

fitted<- temperature.aov$fitted.values
plot(fitted,res, main="Zero mean-Constant variance") #Check zero mean /constant variance
abline(h=0) #horizontal line through 0



(C) Equal variances

```
> bartlett.test(res~ trt)

        Bartlett test of homogeneity of variances

data:  res by trt
Bartlett's K-squared = 0.93309, df = 4, p-value = 0.9198
```

Note: Null: Variances are equal. Support Null

(d) Independence residuals: dwtest()

```
> dwtest(temperature.aov, alternative="two.sided")      #can use durbin.wat
son()

        Durbin-Watson test

data:  temperature.aov
DW = 2.4022, p-value = 0.8385
alternative hypothesis: true autocorrelation is not 0
```

Note: NULL: Independence (Autocorrelation is 0)

4. MULTIPLE COMPARISONS:

(a) Fisher LSD ; pairwise comparisons

Fisher LSD (LSD: Least Significant Difference).

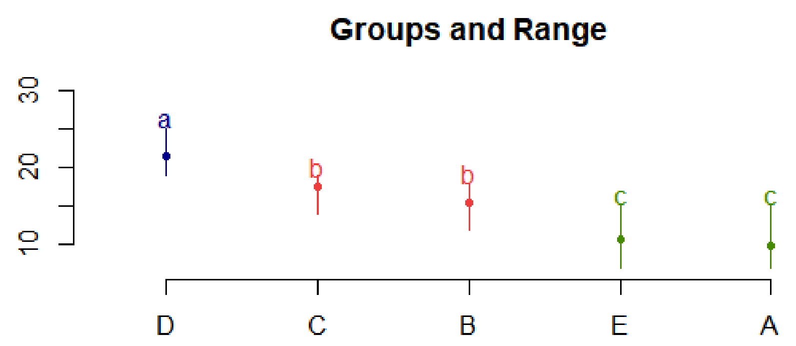
$$LSD = t_{\frac{\alpha}{2},N-a}\sqrt{\frac{2MSE}{n}} \quad .$$

Compare $|y_{i.} - y_{j.}|$ against the LSD.

```
>library(agricolae)
>MSerror <-8.06
> Fisher<- LSD.test(model, "trt", MSerror, console=T)

Study: model ~ "trt"

LSD t Test for temperature

Mean Square Error:  8.06

trt,  means and individual ( 95 %) CI
```

```
  temperature      std r        LCL      UCL Min Max
A         9.8 3.346640 5  7.151566 12.44843   7  15
B        15.4 3.130495 5 12.751566 18.04843  12  18
C        17.6 2.073644 5 14.951566 20.24843  14  19
D        21.6 2.607681 5 18.951566 24.24843  19  25
E        10.8 2.863564 5  8.151566 13.44843   7  15

Alpha: 0.05 ; DF Error: 20
Critical Value of t: 2.085963

least Significant Difference: 3.745452

Treatments with the same letter are not significantly different.

  temperature groups
D        21.6      a
C        17.6      b
B        15.4      b
E        10.8      c
A         9.8      c
```

```
> plot(Fisher)
```



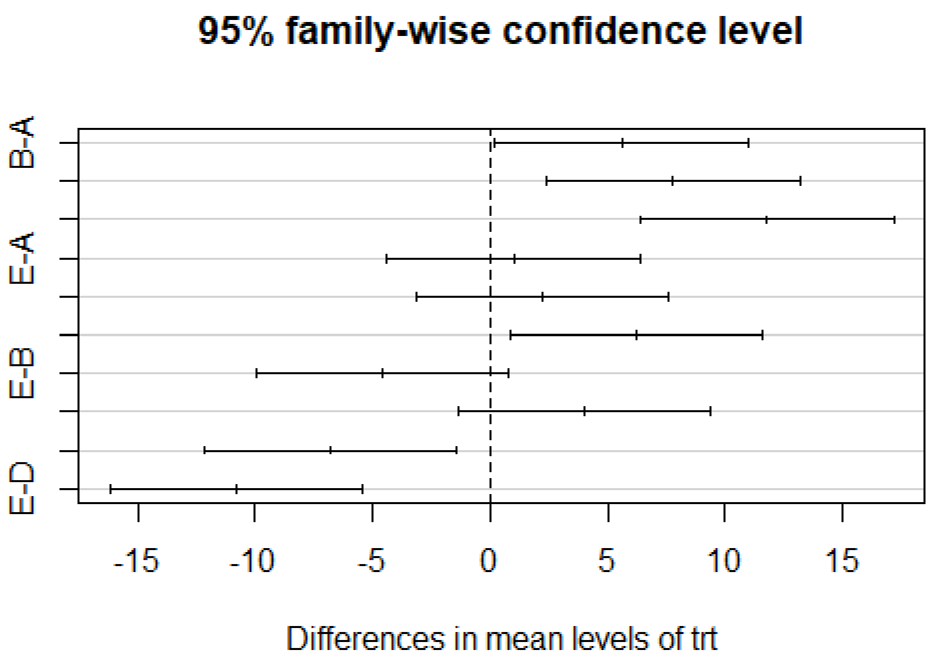
(b) **TukeyHSD() test**

```
> Tukey <- TukeyHSD(model, "trt")
> Tukey
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = temperature ~ trt)

$trt
     diff        lwr       upr     p adj
B-A   5.6  0.2270417 10.9729583 0.0385024
C-A   7.8  2.4270417 13.1729583 0.0025948
D-A  11.8  6.4270417 17.1729583 0.0000190
E-A   1.0 -4.3729583  6.3729583 0.9797709
```

```
C-B    2.2  -3.1729583  7.5729583 0.7372438
D-B    6.2   0.8270417 11.5729583 0.0188936
E-B   -4.6  -9.9729583  0.7729583 0.1162970
D-C    4.0  -1.3729583  9.3729583 0.2101089
E-C   -6.8 -12.1729583 -1.4270417 0.0090646
E-D  -10.8 -16.1729583 -5.4270417 0.0000624
```

NOTE: Either look at the p values which are greaer than alpha=.05 . Alternately, can check if the CI=(lwr, upr) contains 0.

```
> plot(Tukey)
```



(c) pairwise.t.test (); with pooled variance---------------

```
> pairwise.t.test(temp, trt)

        Pairwise comparisons using t tests with pooled SD

data:  temp and trt

  A       B      C      D
B 0.0270  -      -      -
C 0.0025  0.4694 -      -
D 2.1e-05 0.0151 0.1126 -
E 0.5838  0.0744 0.0081 6.3e-05

P value adjustment method: holm

> pairwise.t.test(temp,trt, p.adj="bonf")
```

```
       Pairwise comparisons using t tests with pooled SD

data:  temp and trt

  A       B      C      D
B 0.0541  -      -      -
C 0.0031  1.0000 -      -
D 2.1e-05 0.0251 0.3754 -
E 1.0000  0.1859 0.0116 7.0e-05

P value adjustment method: bonferroni
```

B. PROBLEMS

1. Data: Bacteria with Packages

Packaging Condition	log(count/cm^2)
Commercial plastic wrap Vacuum packaged 1% CO,40% O2, 59% N 100% CO2	7.66, 6.98, 7.80 5.26, 5.44, 5.80, 7.41, 7.33, 7.04 3.51, 2.91, 3.66

a) Set up the data frame.

(Hint: There are 12 observations . 3 observations for each factor level. Form a vector for factor levels "package". Then form a vector for response, named "logcount". Convert package to factor. Form a data frame, named "bacteria", with "package" and "logcount".

b) Perform a stripchart, with line connecting means, of logcount vs package
c) Build a linear model, using aov() response as logcount. Do a summary.lm() and summary.aov()
d) Perform a Bartlett test of equal variances.
e) Perform a multiple comparison of treatment mean, using TukeyHSD()

2. Data: Tensile strength of Portland Cement

Four different mixing techniques are used. The following data have be collected.

Mixing Technique	Tensile Strength (lb/in^2)
1	3129 3000 2865 2890
2	3200 3300 2975 3150
3	2800 2900 2985 3050
4	2600 2700 2600 2765

(a) Set up a data frame , with varibles: mixing (factor) and strength (response)
(b) Perform a stripchart. Perform a Box plot.
(c) Test the hypothesis that mixing techniques affect the strength of the cement. Use α=0.05 What test do use. Perform the test. Conclusion.

(d) Use the Fisher LSD (Least Significant Difference) $\alpha = 0.05$ to make comparison

Note: $LSD = t_{\frac{\alpha}{2},N-a}\sqrt{\frac{2MSE}{n}}$

3. Problem[Dataset 3-22]

3-6. A manufacturer of television sets is interested in the effect on tube conductivity of four different types of coating for color picture tubes. The following conductivity data are obtained:

Coating Type	Conductivity			
1	143	141	150	146
2	152	149	137	143
3	134	136	132	127
4	129	127	132	129

(a) Is there a difference in conductivity due to coating type? Use $\alpha = 0.05$.
(b) Estimate the overall mean and the treatment effects.
(c) Compute a 95 percent confidence interval estimate of the mean of coating type 4. Compute a 99 percent confidence interval estimate of the mean difference between coating types 1 and 4.
(d) Test all pairs of means using the Fisher LSD method with $\alpha = 0.05$.
(e) Use the graphical method discussed in Section 3-5.3 to compare the means. Which coating type produces the highest conductivity?
(f) Assuming that coating type 4 is currently in use, what are your recommendations to the manufacturer? We wish to minimize conductivity.

Note: Skip (e)