STAT 5309- SP 2022

Lab 4-B

***CONTENTS: 1 blocking factor – 2 blocking factors (Latin squares)**

***Due:**

A. PRACTICE

BLOCKING FACTOR

(i) **Nuisance factors**: Blocking is a technique to deal with **nuisance factors**. A nuisance factor is a factor that probably has some effect on the response, but it's of no interest. Examples: batches of raw material, operators, machines, test equipments, times (shifts, days, etc.), experimental units. They are supposed to be unknowiningly different. These nuisance factors are **known or unknown.**

(ii) The variability of nuisance factor are present in the calculation of Sum Squares, which can't be taken out, but preferred to be reduced to a minimum.

(iii) To be effective, the Blocking factor should be : **significant as a whole ;** No interaction with Factor of treatment

1. 1 Blocking factor

Data: Rat Behavior. 50 observationse

library(daewr)

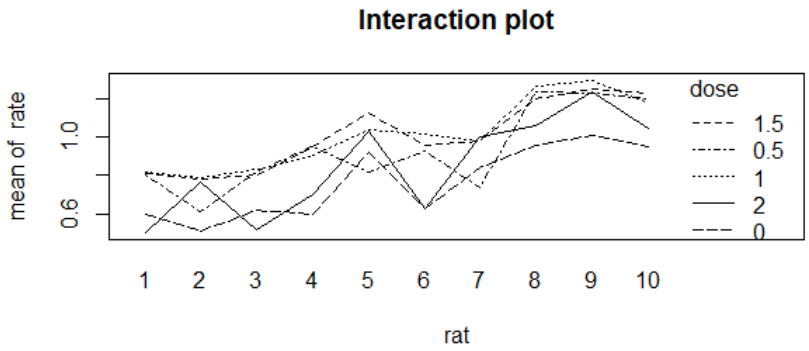
data(drug)

head(drug)

```
> drug
   rat dose rate
1    1    0 0.60
2    1  0.5 0.80
3    1    1 0.82
4    1  1.5 0.81


5    1    2 0.50
6    2    0 0.51
7    2  0.5 0.61
8    2    1 0.79
9    2  1.5 0.78
10   2    2 0.77

46  10    0 0.95
47  10  0.5 1.20
48  10    1 1.18
49  10  1.5 1.23
50  10    2 1.05
```

Note: There are 10 rats (10 levels); dose: 5 doses(5levels); Replicate n=1

Interaction between Rat and Dose



Note: Its hard to tell by plot.

Regression model with Interaction term

```
> mod <- aov(rate ~ rat*dose)
> anova(mod)
Analysis of Variance Table

Response: rate
          Df  Sum Sq  Mean Sq F value Pr(>F)
rat        9 1.66846 0.185384
dose       4 0.46021 0.115052
rat:dose  36 0.30055 0.008349
Residuals  0 0.00000
```

Note: Since n=1 (single replicate) ; $MSE = \frac{SSE}{ab(n-1)}$, MSE can't be calculated!
We can assume no significant between dose and rat and we can *avoid this situation*.

Test the Interaction term: n=1 : Tukey1df(); package:daewr

```
> data.1 <- data.frame(rate, rat, dose)
> Tukey1df(data.1)
Source           df     SS       MS        F      Pr>F
A                 9   1.6685   0.1854
B                 4   0.4602   0.1151
Error            36   0.3006   0.1336
NonAdditivity     1   0.0018   0.0018    0.21    0.6522
Residual         35   0.2988   0.0085
```

Remark: Interaction term is NOT significant!

 # Linear model without Interaction (there should no interaction between rat and dose as it is supposed to be)

drug.mod<- aov(rate ~ rat +dose, data=drug) # rat is used as a Blocking factor

summary.aov(drug.mod)

```
            Df Sum Sq Mean Sq F value   Pr(>F)
rat          9 1.6685 0.18538   22.20 3.75e-12 ***
dose         4 0.4602 0.11505   13.78 6.53e-07 ***
Residuals   36 0.3006 0.00835
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Notes: MSE is available now. F-test and P-value are calculated.

P-value = `3.75e-12.` Rat is a significant blocking factor in this design. **Blocking is effective.**

Note: Without using a blocking factor (which is rat)

drug.mod2 <- aov(rate ~ dose)
summary(drug.mod2)

```
            Df Sum Sq Mean Sq F value Pr(>F)
dose         4 0.4602 0.11505   2.629 0.0466 *
Residuals   45 1.9690 0.04376
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Remark: MSE now becomes much larger, which is not good. MSE(drug.mod1) = 0.00835. MSE(drug.mod2) =0.04376 ,

Blocking factor (rat) **reduces** MSE significantly, which is 81%.

```
> (0.00835-0.04376 )/0.04376
[1]

0.8091865
```

2. 1 Blocking factor + 2 Treatment factors

Data bha, (mouse liver enzyme experiment).

Description: mouse liver enzyme experiment . 16 observations.

library(daewr)

data(bha)

```
> bha
   block strain   treat    y
1      1    A/J treated 18.7
2      1    A/J control  7.7
3      2    A/J treated 16.7
4      2    A/J control  6.4
5      1 129Ola treated 17.9
6      1 129Ola control  8.4
7      2 129Ola treated 14.4
8      2 129Ola control  6.7
9      1    NIH treated 19.2
10     1    NIH control  9.8
11     2    NIH treated 12.0
12     2    NIH control  8.1
13     1 BALB/c treated 26.3
14     1 BALB/c control  9.7
15     2 BALB/c treated 19.8
16     2 BALB/c control  6.0
```

Block: a factor with 2 levels:(1, 2).

Strain: a factor with 4 levels(A/J, 129O1a, NIH, BALB/c.)

Treat: a factor with 2 levels: (treated, control); **y** : response

attach(bha)

Regression model

bha.mod <- aov(y ~ block +strain *treat, data=bha) # consider interaction between strain and treat

summary.aov(bha.mod)

```
             Df Sum Sq Mean Sq F value   Pr(>F)
block         1   47.6    47.6  18.372  0.00363 **
strain        3   33.0    11.0   4.240  0.05274 .
treat         1  422.3   422.3 162.961 4.19e-06 ***
strain:treat  3   40.3    13.4   5.189  0.03368 *
Residuals     7   18.1     2.6
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Remark: The blocking factor is significant.

Treatment factors: Strain, Treat, and Interaction are significant.

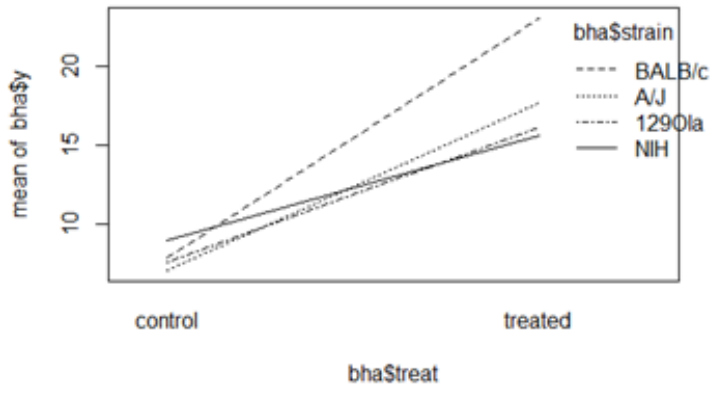
#-----Model without Blocking factor

```
bha.mod1 <- aov(y ~ strain*treat, data=bha)

anova(bha.mod1)

Analysis of Variance Table

Response: y
             Df Sum Sq Mean Sq F value    Pr(>F)
strain        3  32.96   10.99  1.3369    0.3290
treat         1 422.30  422.30 51.3828 9.538e-05 ***
strain:treat  3  40.34   13.45  1.6362    0.2566
Residuals     8  65.75    8.22
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Remark: The MSE without Blocking is Higher than with Blocking. Also, Strain is NOT significant. That shows, proper blocking should be used to produce correct results.

 # Interaction plot

```
interaction.plot(treat, strain, y)    #interaction between treat and strain
```



Tukey1df(): test significance of interaction if single replicate n=1

Note: $MSE = \frac{SSE}{ab(n-1)}$, MSE can’t be calculated!

#--------Power/Sample size: power.anova.test()----------

Can use power.anova.test() to calculate the power or number of replicates, for Factor of Treaments

trt.means <- tapply(rate, dose, mean)
trt.means

```
    0   0.5     1   1.5     2
0.764 0.934 1.014 1.009 0.850
```

MSE <- 0.00835

```
> power.anova.test(groups=5, n=2, between.var=var(trt.means), within.var=MS
E, sig.level=.05,power=NULL)

     Balanced one-way analysis of variance power calculation

         groups = 5
              n = 2
    between.var = 0.0115052
     within.var = 0.00835
      sig.level = 0.05
          power = 0.3780748

NOTE: n is number in each group
```

3. 2 Blocking factors design (LATIN SQUARES)

- Latin Squares: are used to control **two (2) sources of nuisance variability (**2 nuisance factors and 1 Treatment factor). Assumption is that the three factors (treatments, 2 nuisance factors) **do not interact.** If violated, the Latin square design will not produce valid results

Example (data entered manually)

Suppose we have 5 different operators and 5 different batches of raw materials. In addition, there are 5 treatments (5 types of fertilizers A-E, for example). Obviously, operators and batches of materials are considered blocking factors.

TABLE 4.9
Latin Square Design for the Rocket Propellant Problem

Batches of Raw Material	Operators				
	1	2	3	4	5
1	$A = 24$	$B = 20$	$C = 19$	$D = 24$	$E = 24$
2	$B = 17$	$C = 24$	$D = 30$	$E = 27$	$A = 36$
3	$C = 18$	$D = 38$	$E = 26$	$A = 27$	$B = 21$
4	$D = 26$	$E = 31$	$A = 26$	$B = 23$	$C = 22$
5	$E = 22$	$A = 30$	$B = 20$	$C = 29$	$D = 31$

5-by-5 Latin squares; manual setup

```
block.1 <- rep(c(1,2,3,4,5), each=5)
block.2 <- rep(c(1,2,3,4,5), time=5)
x1<- c("A", "B", "C", "D", "E")
x2<- c("B", "C"  ,"D", "E", "A")
x3<- c("C", "D"  ,"E", "A", "B")
x4<- c("D", "E"  ,"A", "B", "C")
x5<- c("E", "A"  ,"B", "C", "D")


treat <- c(x1,x2,x3,x4,x5)

design.




<- data.frame(block.1, block.2, treat)
design
> design
   block.1 block.2 treat
1        1       1     A
2        1       2     B
3        1       3     C
4        1       4     D
5        1       5     E
6        2       1     B
7        2       2     C
8        2       3     D
9        2       4     E
10       2       5     A
11       3       1     C
12       3       2     D
13       3       3     E
14       3       4     A
15       3       5     B
16       4       1     D
17       4       2     E
18       4       3     A
19       4       4     B
20       4       5     C
21       5       1     E
22       5       2     A
23       5       3     B
24       5       4     C
   25       5       5     D
```

#design.lsd() [package agricolae]

Data: suppose there are 4 weeks (Wk1-Wk4) period and 4 stores(Store1-Store4). And there are 4 treaments(A-D).

```
library(agricolae)
treat <- c("A", "B", "C", "D")
lsd <- design.lsd(treat, seed=543, serie=2)

lsd.book <- lsd$book
```

```
> lsd.book
   plots row col treat
1    101   1   1   C
2    102   1   2   A
3    103   1   3   B
4    104   1   4   D
5    201   2   1   D
6    202   2   2   B
7    203   2   3   C
8    204   2   4   A
9    301   3   1   B
10   302   3   2   D
11   303   3   3   A
12   304   3   4   C
13   401   4   1   A
14   402   4   2   C
15   403   4   3   D
16   404   4   4   B
```

```
names(lsd.book)

[1] "plots" "row"   "col"   "treat"


#can rename the rows and columns

levels(lsd.book$row) <- c("Week1", "Week2","Week3", "Week4")

levels(lsd.book$col) <- c("Store1", "Store2", "Store3", "Store4")

sales <- c(10,12,15,12,8,16,8,11,15,10,13,8,14,7,10,14)
```

```
> data <- data.frame(lsd.book,sales )
> data
   plots   row    col treat sales
1    101 Week1 Store1   C    10
2    102 Week1 Store2   A    12
3    103 Week1 Store3   B    15
4    104 Week1 Store4   D    12
5    201 Week2 Store1   D     8
```

```
6    202 Week2 Store2   B    16
7    203 Week2 Store3   C     8
8    204 Week2 Store4   A    11
9    301 Week3 Store1   B    15
10   302 Week3 Store2   D    10
11   303 Week3 Store3   A    13
12   304 Week3 Store4   C     8
13   401 Week4 Store1   A    14
14   402 Week4 Store2   C     7
15   403 Week4 Store3   D    10
16   404 Week4 Store4   B    14
```

```
> sales.aov <- aov(sales ~ row +col +treat, data=data)
> summary.aov(sales.aov)
            Df Sum Sq Mean Sq F value  Pr(>F)
row          3   4.69    1.56   0.652 0.61011
col          3   0.69    0.23   0.096 0.95964
treat        3 104.19   34.73  14.496 0.00372 **
Residuals    6  14.37    2.40
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

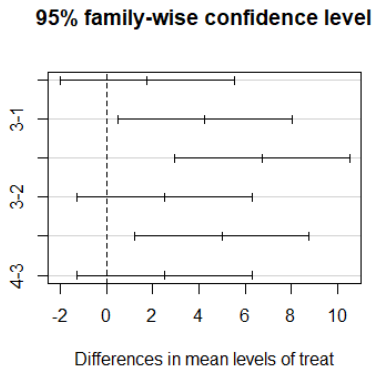
Note: the row and col means are NOT significant[which are not of interest].The trt means are significant (which is of interest)

#----------Graeco-Latin Square: 2 blockings- 2 treatments

```
str(design.graeco)
trt <- c("A", "B", "C", "D")
trt2 <- 1:4
graeco <- design.graeco(trt, trt2,seed=543, serie=2)
graeco$book
```

```
   plots row col trt trt2
1    101   1   1   A    1
2    102   1   2   D    4
3    103   1   3   B    3
4    104   1   4   C    2

5    201   2   1   D    3
6    202   2   2   A    2
7    203   2   3   C    1
8    204   2   4   B    4

9    301   3   1   B    2
10   302   3   2   C    3
11   303   3   3   A    4
12   304   3   4   D    1
```

```
13    401    4    1    C      4
14    402    4    2    B      1
15    403    4    3    D      2
16    404    4    4    A      3


## Pairwise comparison:   Tukey HSD
> sales.Tukey <- TukeyHSD(sales.aov, "treat")
> sales.Tukey
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = sales ~ row + col + treat, data = data)

$treat
    diff       lwr       upr     p adj
2-1 1.75 -2.0388216  5.538822 0.4445573
3-1 4.25  0.4611784  8.038822 0.0310526
4-1 6.75  2.9611784 10.538822 0.0033922
3-2 2.50 -1.2888216  6.288822 0.2037456
4-2 5.00  1.2111784  8.788822 0.0149758
4-3 2.50 -1.2888216  6.288822 0.2037456
```



. EXERCISE

1. **Problem[Dataset 4-8]**

A chemist wishes to test the effect of four chemical agents on the strength of a particular type of cloth. Because there might be variability from one bolt to another, the chemist decides to use a randomized block design, with the bolts of cloth considered as blocks. She selects five bolts and applies all four chemicals in random order to each bolt. The resulting tensile strengths follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw appropriate conclusions.

	Bolt				
Chemical	1	2	3	4	5
1	73	68	74	71	67
2	73	67	75	72	70
3	75	68	78	73	68
4	73	71	75	75	69

(a) Set up the data frame, named "chem", with "Bolt" and "Chemical" factors, "strength" as response.

(b) Any evidence that the Chemical affect Strength ?Note: Chemical is the treatment factor. Is Bolt is an effective blocking factor?
(c) Perform a TukeyHSD to compare the treatment means.
Which Chemical is the preferred(bring the highest strength)
(d) Check the assumptions of the residuals.

2. Problem [Data 4-26]

An industrial engineer is conducting an experiment on eye focus time. He is interested in the effect of the distance of the object from the eye on the focus time. Four different distances are of interest. He has five subjects available for the experiment. Because there may be differences among individuals, he decides to conduct the experiment in a randomized block design. The data obtained follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw appropriate conclusions.

	Subject				
Distance (ft)	1	2	3	4	5
4	10	6	6	6	6
6	7	6	6	1	6
8	5	3	3	2	5
10	6	4	4	2	3

(a) Set up the data frame, named "eye" , "Subject" , "Distance" are factors. "time" as response.
(b) Build a regression model, name "eye.mod" . Is "Subject" an effective locking factor?
(c) Perform TukeyHSD on "Distance". Which Distances bring the longest/ shortest focus time
(d) Calculate the sample size for power > .90, use power.anova.test().

3. Problem [Dataset 4-28] (Latin Squares)

An industrial engineer is investigating the effect of four assembly methods (A, B, C, D) on the assembly time for a color television component. Four operators are selected for the study. Furthermore, the engineer knows that each assembly method produces such fatigue that the time required for the last assembly may be greater than the time required for the first, regardless of the method. That is, a trend develops in the required assembly time. To account for this source of variability, the engineer uses the Latin square design shown below. Analyze the data from this experiment ($\alpha = 0.05$) and draw appropriate conclusions.

Order of Assembly	Operator			
	1	2	3	4
1	$C = 10$	$D = 14$	$A = 7$	$B = 8$
2	$B = 7$	$C = 18$	$D = 11$	$A = 8$
3	$A = 5$	$B = 10$	$C = 11$	$D = 9$
4	$D = 10$	$A = 10$	$B = 12$	$C = 14$

(a) Set up a data frame manually to use the Latin Square design.

Hint: Create a vector for 1st blocking factor, named "Assembly" 4 levels: 1,2,3,4.
Create a vector for 2nd blocking factor, named "Operator" 4 levels: 1,2,3,4.
Create a vector for treatment factor , named "Treatment" levels "C", "D", "A", "B", "B", "C", etc
Create a vector of response, named: "time"

(b) Build a regression model, using aov(). Do the Treatment affect the assembly time?

(c) Find the lowest assembly time.