STAT 337

Design and Analysis of Experiments

Exercises

Editing by: Zahra Kaabi

Chapter 4

Randomized Complete Block Design

4.1. The ANOVA from a randomized complete block experiment output is shown below.

Source	DF	SS	MS	F	P
Treatment	4	1010.56	(3) ?	29.84	? (6)
Block	(1) ?	(2) ?	64.765	(5) ?	? (7)
Error	20	169.33	(4) ?		
Total	29	1503.71			

(a) Fill in the blanks. You may give bounds on the P-value.

(b) How many blocks were used in this experiment?

(c) What conclusions can you draw?

(a) - Given that $a - 1 = 4$

And $(a-1)(b-1) = 20 \Rightarrow 4(b-1) = 20 \Rightarrow b - 1 = 5$

Then $(\mathbf{1}) = b - 1 = 5$

- Given that $\text{MS}_{\text{Block}} = 64.765$

  And we know that $\text{MS}_{\text{Block}} = \frac{\text{SS}_{\text{Block}}}{b-1} \Rightarrow \text{SS}_{\text{Block}} = (b-1)\text{MS}_{\text{Block}}$

  Then $(\mathbf{2}) = \text{SS}_{\text{Block}} = (b-1)\text{MS}_{\text{Block}} = (5)(64.765) = 323.825$

- $(\mathbf{3}) = \text{MS}_{\text{Treatments}} = \frac{\text{SS}_{\text{Treatments}}}{a-1} = \frac{1010.56}{4} = 252.64$

- $(\mathbf{4}) = \text{MS}_{\text{E}} = \frac{\text{SS}_{\text{E}}}{(a-1)(b-1)} = \frac{169.33}{20} = 8.4665$

- $(\mathbf{5}) = \boldsymbol{F} = \frac{\text{MS}_{\text{Block}}}{\text{MS}_{\text{E}}} = \frac{64.765}{8.4665} = 7.65$

- $(\mathbf{6}) = \boldsymbol{p - value} = \mathbf{0}$

How to calculate p-value:

p-value= $P(F > F_{statistic})$ where, $df_1 = 4$ and $df_2 = 20$

$= P(F > 29.84)$

$= 1 - P(F < 29.84)$

$= 1 - 1$

$= 0$

Minitab: $P(F < 29.84)$

calc→probability distribution→ F

$\rightarrow cumulative\ distribution$

$\rightarrow Numerator\ degree\ of\ freedom = \mathbf{4}$

$Denominator\ degree\ of\ freedom = \mathbf{20}$

input constant=**29.84**

→OK

The answer = 1

- $(7) = p - value =$ 0.0004

p-value= $P(F > F_{statistic})$ where, $df_1 = 5$ and $df_2 = 20$

$= P(F > 7.65)$

$= 1 - P(F < 7.65)$

$= 1 -$ 0.99963127

$=$ 0.000368721 ≈ 0.0004

(b) the number of blocks= 6

(C) **Treatments:** it's clear that $p - value = 0 < \alpha = 0.05$

Then we reject $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

That means at least one of the means different to others, which means there is a different between the five different treatments in the result of the study

Blocks: it's clear that $p - value = 0.0004 < \alpha = 0.05$

Then we reject $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6$

That means at least one of the means different to others, which means there is a different between the six different Blocks in the result of the study.

4.4. Three different washing solutions are being compared to study their effectiveness in retarding bacteria growth in 5-gallon milk containers. The analysis is done in a laboratory, and only three trials can be run on any day. Because days could represent a potential source of variability, the experimenter decides to use a randomized block design. Observations are taken for four days, and the data are shown here. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

	Days				
Solution	1	2	3	4	$Y_{i.}$
1	13	22	18	39	92
2	16	24	17	44	101
3	5	4	1	22	32
$Y_{.j}$	34	50	36	105	**225**

a= 3 b=4 N=12

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^2 - \frac{Y_{..}^2}{N}$$

$$= 13^2 + 16^2 + 5^2 + 22^2 + \cdots + 1^2 + 22^2 - \frac{225^2}{12} = \mathbf{1862.25}$$

Zahra Kaabi

$$SS_{Treatment} = \frac{1}{b}\sum_{i=1}^{a} Y_{i.}{}^{2} - \frac{Y_{..}{}^{2}}{N}$$

$$= \frac{1}{4}(92^2 + 101^2 + 32^2) - \frac{225^2}{12} = \mathbf{703.5}$$

$$SS_{Block} = \frac{1}{a}\sum_{j=1}^{b} Y_{.j}{}^{2} - \frac{Y_{..}{}^{2}}{N}$$

$$= \frac{1}{3}(34^2 + 50^2 + 36^2 + 105^2) - \frac{225^2}{12} = \mathbf{1106.92}$$

$$SS_E = SS_T - SS_{Treatment} - SS_{Block}$$

$$= 1862.25 - 703.5 - 1106.9 = \mathbf{51.83}$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}}$	$a-1$	$\frac{SS_{\text{Treatments}}}{a-1}$	$\frac{MS_{\text{Treatments}}}{MS_E}$
Blocks	SS_{Blocks}	$b-1$	$\frac{SS_{\text{Blocks}}}{b-1}$	
Error	SS_E	$(a-1)(b-1)$	$\frac{SS_E}{(a-1)(b-1)}$	
Total	SS_T	$N-1$		

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	703.5	2	351.75	40.72
Blocks	1106.9	3	368.97	
Error	51.83	6	8.64	
Total	1862.25	11		

We study here if there is a different between the Three different washing solutions in retarding bacteria growth in five-gallon milk containers.

(1) $H_0: \mu_1 = \mu_2 = \mu_3$ VS H_1: at least one of the means different to others

(2) The test statistic $F_0 = \frac{MS_{treatment}}{MS_E} = \frac{351.75}{8.64} = 40.72$

(3) The Critical value $= F_{\alpha, a-1, ((a-1)(b-1)} = F_{0.05,2,6} = 5.143$

(4) since $F_0 = 40.72 > 5.143 = F_{0.05,2,6}$ we reject $H_0: \mu_1 = \mu_2 = \mu_3$

That means there is a different between the three different washing solutions in retarding bacteria growth in five-gallon milk containers.

We will see if there is a different between the days in retarding bacteria growth in five-gallon milk containers.

(1) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4$ VS H_1: at least one of the means different to others

(2) The test statistic $F_0 = \frac{MS_{Block}}{MS_E} = \frac{368.97}{8.64} = 42.7$

(3) The Critical value $= F_{\alpha, b-1, ((a-1)(b-1)} = F_{0.05,3,6} = 4.75706$

(4) since $F_0 = 42.7 > 4.75706 = F_{0.05,3,6}$ we reject $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4$

That means there is a different between the days in retarding bacteria growth in five-gallon milk containers.

Using Minitab to Perform a RCBD

solution	day	obs
1	1	13
1	2	22
1	3	18
1	4	39
2	1	16
2	2	24
2	3	17
2	4	44
3	1	5
3	2	4
3	3	1
3	4	22

- Choose Stat > ANOVA >General Linear Model.
- In Response, enter. **Obs**
- In Model, enter **Solution day**
- Click **OK**

The output

General Linear Model: obs versus solution; day

```
Factor    Type   Levels  Values
solution  fixed       3  1; 2; 3
day       fixed       4  1; 2; 3; 4
```

Analysis of Variance for obs, using Adjusted SS for Tests

```
Source    DF   Seq SS   Adj SS   Adj MS      F      P
solution   2   703.50   703.50   351.75  40.72  0.000
day        3  1106.92  1106.92   368.97  42.71  0.000
Error      6    51.83    51.83     8.64
Total     11  1862.25


S = 2.93920   R-Sq = 97.22%   R-Sq(adj) = 94.90%


Unusual Observations for obs

Obs     obs     Fit  SE Fit  Residual  St Resid
```

```
  9   5.0000   0.5833   2.0783     4.4167       2.13 R
R denotes an observation with a large standardized residual.
```

Extra exercise:

Estimate the missing value

	Treatment				
Rep	A	B	C	D	$Y_{.j}$
1	9	11	3	7	30
2	8	13	x	10	31
3	7	12	8	4	31
$Y_{i.}$	24	36	11	21	92

Here a=4 and b=3

$$x = \frac{ay'_{i.} + by'_{.j} - y'_{..}}{(a-1)(b-1)}$$

$$x = \frac{4(11)+3(31)-92}{(2)(3)} = 7.5$$

HW 4.3

The Latin Square Design

4.23. 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	Operator 2	Operator 3	Operator 4	$Y_{i..}$
1	$C = 10$	$D = 14$	$A = 7$	$B = 8$	39
2	$B = 7$	$C = 18$	$D = 11$	$A = 8$	44
3	$A = 5$	$B = 10$	$C = 11$	$D = 9$	35
4	$D = 10$	$A = 10$	$B = 12$	$C = 14$	46
$Y_{..k}$	32	52	41	39	164

$Y_{...}$

N=16 and P=4

Latin letter	A	B	C	D	TOTAL
Treatment total $Y_{.j.}$	$Y_{.1.}$ **30**	$Y_{.2.}$ **37**	$Y_{.3.}$ **53**	$Y_{.4.}$ **44**	$Y_{...}$ **164**

$$SS_T = \sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{K}^{P} Y_{ijK}{}^{2} - \frac{Y_{..}^{2}}{N}$$

$$= 10^2 + 14^2 + 7^2 + 8^2 + \cdots + 12^2 + 14^2 - \frac{164^2}{16} = 153$$

$$SS_{Treatment} = \frac{1}{P}\sum_{J=1}^{P} Y_{.j.}{}^{2} - \frac{Y_{..}^{2}}{N}$$

$$= \frac{1}{4}(30^2 + 37^2 + 53^2 + 44^2) - \frac{164^2}{16} = 72.5$$

$$SS_{Rows} = \frac{1}{p}\sum_{i=1}^{p} Y_{i..}{}^{2} - \frac{Y_{..}^{2}}{N}$$

$$= \frac{1}{4}(39^2 + 44^2 + 35^2 + 46^2) - \frac{164^2}{16} = \mathbf{18.5}$$

$$SS_{Columns} = \frac{1}{p}\sum_{k=1}^{p} Y_{..k}{}^{2} - \frac{Y_{..}^{2}}{N}$$

$$= \frac{1}{4}(32^2 + 52^2 + 41^2 + 39^2) - \frac{164^2}{16} = 51.5$$

$$SS_E = SS_T - SS_{Treatment} - SS_{Rows} - SS_{Columns}$$

$$= 153 - 72.5 - 18.5 - 51.5 = \mathbf{10.5}$$

Analysis of Variance for the Latin Square Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}} = \frac{1}{p}\sum_{j=1}^{p} y_{.j.}^2 - \frac{y_{...}^2}{N}$	$p-1$	$\frac{SS_{\text{Treatments}}}{p-1}$	$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}$
Rows	$SS_{\text{Rows}} = \frac{1}{p}\sum_{i=1}^{p} y_{i..}^2 - \frac{y_{...}^2}{N}$	$p-1$	$\frac{SS_{\text{Rows}}}{p-1}$	
Columns	$SS_{\text{Columns}} = \frac{1}{p}\sum_{k=1}^{p} y_{..k}^2 - \frac{y_{...}^2}{N}$	$p-1$	$\frac{SS_{\text{Columns}}}{p-1}$	
Error	SS_E (by subtraction)	$(p-2)(p-1)$	$\frac{SS_E}{(p-2)(p-1)}$	
Total	$SS_T = \sum_i\sum_j\sum_k y_{ijk}^2 - \frac{y_{...}^2}{N}$	p^2-1		

Analysis of Variance for the Latin Square Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	72.5	3	24.17	13.81
Rows	18.5	3	6.17	
Columns	51.5	3	17.17	
Error	10.5	6	1.75	
Total	153	15		

(1) $H_0: \mu_{T_1} = \mu_{T_2} = \mu_{T_3} = \mu_{T_4}$ VS H_1: at least one of the means different to others

(2) The test statistic $F_0 = \frac{MS_{treatment}}{MS_E} = \frac{24.17}{1.75} = 13.81$

(3) The Critical value $= \boldsymbol{F}_{\boldsymbol{\alpha,P-1,(P-1)(P-2)}} = \boldsymbol{F}_{\boldsymbol{0.05,3,6}} = 4.76$

(4) since $F_0 = 13.81 > 4.76 = \boldsymbol{F}_{\boldsymbol{0.05,3,6}}$ we reject $H_0: \mu_{T_1} = \mu_{T_2} = \mu_{T_3} = \mu_{T_4}$

That means there is a different between the four assembly methods (*A, B, C, D*) on the assembly time for a color television component.

Using Minitab to Perform the Latin Square Design

order of assembly	operator	position	obs
1	1	C	10
1	2	D	14
1	3	A	7
1	4	B	8
2	1	B	7
2	2	C	18
2	3	D	11
2	4	A	8
3	1	A	5
3	2	B	10
3	3	C	11
3	4	D	9
4	1	D	10
4	2	A	10
4	3	B	12
4	4	C	14

- Choose Stat > ANOVA >General Linear Model> Fit General Linear Model.
- In Response, enter. **Obs**
- In Model, enter **order of assembly , operator, position**
- Click **OK**

General Linear Model: obs versus order of assembly; operator; position

$H_0: \mu_{T_1} = \mu_{T_2} = \mu_{T_3} = \mu_{T_4}$ VS

H_1: at least one of the means different to others means

Since p-value $= 0.004 < 0.05 = \alpha$

$\Rightarrow$ We reject $H_0: \mu_{T_1} = \mu_{T_2} = \mu_{T_3} = \mu_{T_4}$, That means there is a different between the four assembly methods (*A, B, C, D*) on the assembly time for a color television component.

```
Method

Factor coding  (-1; 0; +1)


Factor Information

Factor             Type    Levels  Values
order of assembly  Random       4  1; 2; 3; 4
operator           Random       4  1; 2; 3; 4
position           Fixed        4  A; B; C; D
```

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
order of assembly	3	18.50	6.167	3.52	0.089
operator	3	51.50	17.167	9.81	0.010
position	3	72.50	24.167	13.81	0.004
Error	6	10.50	1.750		
Total	15	153.00			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.32288	93.14%	82.84%	51.20%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	10.250	0.331	30.99	0.000	
order of assembly					
1	-0.500	0.573	-0.87	0.416	*
2	0.750	0.573	1.31	0.238	*
3	-1.500	0.573	-2.62	0.040	*
operator					
1	-2.250	0.573	-3.93	0.008	*
2	2.750	0.573	4.80	0.003	*
3	0.000	0.573	0.00	1.000	*
position					
A	-2.750	0.573	-4.80	0.003	1.50
B	-1.000	0.573	-1.75	0.131	1.50
C	3.000	0.573	5.24	0.002	1.50

```
Regression Equation

obs = 10.250 - 0.500 order of assembly_1 + 0.750 order of assembly_2
      - 1.500 order of assembly_3 + 1.250 order of assembly_4 - 2.250 operator_1
      + 2.750 operator_2 + 0.000 operator_3 - 0.500 operator_4 - 2.750 position_A
      - 1.000 position_B + 3.000 position_C + 0.750 position_D

Equation treats random terms as though they are fixed.


Expected Mean Squares, using Adjusted SS
```

	Source	Expected Mean Square for Each Term
1	order of assembly	(4) + 4.0000 (1)
2	operator	(4) + 4.0000 (2)
3	position	(4) + Q[3]
4	Error	(4)

Error Terms for Tests, using Adjusted SS

	Source	Error DF	Error MS	Synthesis of Error MS
1	order of assembly	6.00	1.7500	(4)
2	operator	6.00	1.7500	(4)
3	position	6.00	1.7500	(4)

Variance Components, using Adjusted SS

Source	Variance	% of Total	StDev	% of Total
order of assembly	1.10417	16.46%	1.05079	40.57%
operator	3.85417	57.45%	1.96320	75.80%
Error	1.75	26.09%	1.32288	51.08%
Total	6.70833		2.59005	

HW : 4.22

Balanced Incomplete Block Designs (BIBD)

4.40 An engineer is studying the mileage performance characteristics of five types of gasoline additives. In the road test he wishes to use cars as blocks; however, because of a time constraint, he must use an incomplete block design. He runs the balanced design with the five blocks that follow. Analyze the data from this experiment (use □α =.05) and draw conclusions.

	Car					
Additive	**1**	**2**	**3**	**4**	**5**	$y_{i.}$
1		17	14	13	12	56
2	14	14		13	10	51
3	12		13	12	9	46
4	13	11	11	12		47
5	11	12	10		8	41
$y_{.j}$	50	54	48	50	39	$y_{..} = 241$

a=5 b=5 k=4 =r=4 N=ak=bk=5(4)=20

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^{2} - \frac{Y_{..}^{2}}{N}$$

$$= 17^2 + 14^2 + 13^2 + 12^2 + \cdots + 10^2 + 8^2 - \frac{241^2}{20} = 2981 - \frac{241^2}{20} = 76.95$$

$$SS_{Block} = \frac{1}{k}\sum_{j=1}^{b} Y_{.j}{}^{2} - \frac{Y_{..}{}^{2}}{N}$$

$$= \frac{1}{4}(50^2 + 54^2 + 48^2 + 50^2 + 39^2) - \frac{241^2}{20} = 31.2$$

$$SS_{Treatment(adjusted)} = \frac{k\sum_{i=1}^{a} Q_i{}^2}{\lambda a}$$

$$\lambda = \frac{r(k-1)}{a-1} = \frac{4(4-1)}{5-1} = 3$$

$$Q_i = y_{i.} - \sum_{j=1}^{b} n_{ij} y_{.j} \quad i = 1, 2, \dots, a$$

$$Q_1 = 56 - \frac{1}{4}(54 + 48 + 50 + 39) = 33/4 = 8.25$$

$$Q_2 = 51 - \frac{1}{4}(50 + 54 + 50 + 39) = 11/4 = 2.75$$

$$Q_3 = 46 - \frac{1}{4}(50 + 48 + 50 + 39) = -3/4 = -0.75$$

$$Q_4 = 47 - \frac{1}{4}(50 + 54 + 48 + 50) = -7/2 = -3.5$$

$$Q_5 = 41 - \frac{1}{4}(50 + 54 + 48 + 39) = -\frac{27}{4} = -6.75$$

Then,

$$SS_{Treatment(adjusted)} = \frac{k\sum_{i=1}^{a} Q_i{}^2}{\lambda a} = 536/15 = 35.73$$

Zahra Kaabi

$$SS_E = SS_T - SS_{Treatment(adjusted)} - SS_{Block}$$

$$= 76.95 - 35.73 - 31.2 = 10.02$$

$$SS_{Treatment} = \frac{1}{r}\sum_{i=1}^{a} Y_{i.}{}^2 - \frac{Y_{..}{}^2}{N}$$

$$= \frac{1}{4}(56^2 + 51^2 + 46^2 + 47^2 + 41^2) - \frac{241^2}{20} = 31.7$$

$$SS_{Block(adjusted)} = \frac{r\sum_{j=1}^{b} Q'_j{}^2}{\lambda b}$$

$$\lambda = \frac{r(k-1)}{a-1} = \frac{4(4-1)}{5-1} = 3$$

$$Q'_i = y_{.j} - \sum_{i=1}^{a} n_{ij} y_{i.} \qquad i = 1, 2, \dots, a$$

$$Q'_1 = 50 - \frac{1}{4}(51 + 46 + 47 + 41) = 15/4 = 3.75$$

$$Q'_2 = 54 - \frac{1}{4}(56 + 51 + 47 + 41) = 21/4 = 5.25$$

$$Q'_3 = 48 - \frac{1}{4}(56 + 46 + 47 + 41) = \frac{1}{2} = 0.5$$

$$Q'_4 = 50 - \frac{1}{4}(56 + 51 + 46 + 47) = 0$$

$$Q'_5 = 39 - \frac{1}{4}(56 + 51 + 46 + 41) = -19/2 = -9.5$$

then,

$$SS_{Block(adjusted)} = \frac{r\sum_{j=1}^{b} Q'_j{}^2}{\lambda b} = 35.23$$

Zahra Kaabi

$$SS_E = SS_T - SS_{Treatment} - SS_{Block(adjusted)}$$

$$= 76.95 - 31.7 - 35.23 = 10.02$$

Analysis of Variance for the Balanced Incomplete Block Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments (adjusted)	$\dfrac{k\sum Q_i^2}{\lambda a}$	$a-1$	$\dfrac{SS_{\text{Treatments(adjusted)}}}{a-1}$	$F_0 = \dfrac{MS_{\text{Treatments(adjusted)}}}{MS_E}$
Blocks	$\dfrac{1}{k}\sum y_{.j}^2 - \dfrac{y_{..}^2}{N}$	$b-1$	$\dfrac{SS_{\text{Blocks}}}{b-1}$	
Error	SS_E (by subtraction)	$N-a-b+1$	$\dfrac{SS_E}{N-a-b+1}$	
Total	$\sum\sum y_{ij}^2 - \dfrac{y_{..}^2}{N}$	$N-1$		

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	p-value
Treatments (adjusted)	35.73	4	8.9325	9.81	0.001
Treatments (unadjusted)	31.7	4			
Blocks (unadjusted)	31.2	4			
Blocks (adjusted)	35.23	4	8.8075	9.67	0.001
Error	10.02	11	0.9109		
Total	76.95	21			

We study here if there is a different between the five types of gasoline (treatments) on the road test

(1) $H_0: \mu_{T_1} = \mu_{T_2} = \mu_{T_3} = \mu_{T_4} = \mu_{T_5}$ VS H_1: at least one of the means different to others

(2) The test statistic $F_0 = \frac{MS_{treatment((adjusted)}}{MS_E} = \frac{8.9333}{0.9109} = 9.81$

(3) The Critical value $= \boldsymbol{F_{\alpha,a-1,N-a-b+1}} = \boldsymbol{F_{0.05,4,11}}$ =3.357

(4) since $F_0 = 9.81 > 3.357 = \boldsymbol{F_{0.05,4,11}}$ we reject $H_0: \mu_{T_1} = \mu_{T_2} = \mu_{T_3} = \mu_{T_4} = \mu_{T_5}$
That means there is a different between the five types of gasoline on the road test .

We study here if there is a different between the five types of cars (Blocks) on the road test

(1) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$ VS H_1: at least one of the means different to others

(2) The test statistic $F_0 = \frac{MS_{Block((adjusted))}}{MS_E} = \frac{8.8075}{0.9109} = 9.67$

(3) The Critical value $= F_{\alpha,b-1,\boldsymbol{-a-b+1}} = \boldsymbol{F_{0.05,4,11}}$ =3.357

(4) since $F_0 = 9.67 > 3.357 = \boldsymbol{F_{0.05,4,11}}$ we reject $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$

That means there is a different between the five types of cars on the road test

Using Minitab to Perform a RCBD

Additive	Car	obs
1	2	17
1	3	14
1	4	13
1	5	12
2	1	14
2	2	14
2	4	13

2	5	10
3	1	12
3	3	13
3	4	12
3	5	9
4	1	13
4	2	11
4	3	11
4	4	12
5	1	11
5	2	12
5	3	10
5	5	8

- Choose Stat > ANOVA >General Linear Model.
- In Response, enter. **Obs**
- In Model, enter **Additive Car**
- Click **OK**

The output

```
Factor Information

Factor    Type   Levels  Values
Additive  Fixed       5  1; 2; 3; 4; 5
Car       Fixed       5  1; 2; 3; 4; 5


Analysis of Variance

Source      DF  Adj SS  Adj MS  F-Value  P-Value
  Additive   4   35.73  8.9333     9.81    0.001
  Car        4   35.23  8.8083     9.67    0.001
Error       11   10.02  0.9106
Total       19   76.95


Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0.954257  86.98%     77.52%      56.97%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant  12.050    0.213    56.47    0.000
Additive
  1        2.200    0.441     4.99    0.000  1.71
  2        0.733    0.441     1.66    0.124  1.71
  3       -0.200    0.441    -0.45    0.659  1.71
  4       -0.933    0.441    -2.12    0.058  1.71
Car
```

1	1.000	0.441	2.27	0.044	1.71
2	1.400	0.441	3.18	0.009	1.71
3	0.133	0.441	0.30	0.768	1.71
4	0.000	0.441	0.00	1.000	1.71

Regression Equation

```
obs = 12.050 + 2.200 Additive_1 + 0.733 Additive_2 - 0.200 Additive_3 - 0.933 Additive_4
      - 1.800 Additive_5 + 1.000 Car_1 + 1.400 Car_2 + 0.133 Car_3 + 0.000 Car_4
      - 2.533 Car_5
```

Fits and Diagnostics for Unusual Observations

Obs	obs	Fit	Resid	Std Resid	
14	11.000	12.517	-1.517	-2.14	R

R Large residual

Nonparametric Methods in the Analysis of Variance

The Kruskal–Wallis Test

Example 3.12 page:129

EXAMPLE 3.12

The data from Example 3.1 and their corresponding ranks are shown in Table 3.20. There are ties, so we use Equation 3.67 as the test statistic. From Equation 3.67

$$S^2 = \frac{1}{19}\left[2869.50 - \frac{20(21)^2}{4}\right] = 34.97$$

and the test statistic is

$$H = \frac{1}{S^2}\left[\sum_{i=1}^{a} \frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}\right]$$

$$= \frac{1}{34.97}[2796.30 - 2205]$$

$$= 16.91$$

■ TABLE 3.20

Data and Ranks for the Plasma Etching Experiment in Example 3.1

Power							
160		180		200		220	
y_{1j}	R_{1j}	y_{2j}	R_{2j}	y_{3j}	R_{3j}	y_{4j}	R_{4j}
575	6	565	4	600	10	725	20
542	3	593	9	651	15	700	17
530	1	590	8	610	11.5	715	19
539	2	579	7	637	14	685	16
570	5	610	11.5	629	13	710	18
$R_{i.}$	17		39.5		63.5		90

Because $H > \chi^2_{0.01,3} = 11.34$, we would reject the null hypothesis and conclude that the treatments differ. (The P-value for $H = 16.91$ is $P = 7.38 \times 10^{-4}$.) This is the same conclusion as given by the usual analysis of variance F test.

power	Obs
160	**575**
160	**542**
160	**530**
160	**539**
160	**570**
180	**565**
180	**593**
180	**590**
180	**579**
180	**610**
200	**600**
200	**651**
200	**610**
200	**637**
200	**629**
220	**725**
220	**700**
220	**715**
220	**685**
220	**710**

To Perform a Kruskal-Wallis

- Click **Analyze > Nonparametric Tests > Legacy Dialogs > K Independent Samples.**
- In Response, enter **obs**
- In factors, enter **power**

Kruskal-Wallis Test: obs versus power

Descriptive Statistics

power	N	Median	Mean Rank	Z-Value
160	5	542	3.4	-3.10
180	5	590	7.9	-1.13
200	5	629	12.7	0.96
220	5	710	18.0	3.27
Overall	20		10.5	

Test

Null hypothesis	H_0: All medians are equal
Alternative hypothesis	H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	3	16.89	0.001
Adjusted for ties	3	16.91	0.001

3.51. Use the Kruskal – Wallis test for the experiment in Problem 3.23.

3.23. The effective life of insulating fluids at an accelerated load of 35 kV is being studied. Test data have been obtained for four types of fluids. The results from a completely randomized experiment were as follows:

Fluid	**Type Life (in h) at 35 kV Load**					
1	17.6	18.9	16.3	17.4	20.1	21.6
2	16.9	15.3	18.6	17.1	19.5	20.3
3	21.4	23.6	19.4	18.5	20.5	22.3
4	19.3	21.1	16.9	17.5	18.3	19.8

Fluid							
1		2		3		4	
Y_{1j}	R_{1j}	Y_{2j}	R_{2j}	Y_{3j}	R_{3j}	Y_{4j}	R_{4j}
17.6	**8**	**16.9**	**3.5**	**21.4**	**21**	**19.3**	**13**
18.9	**12**	**15.3**	**1**	**23.6**	**24**	**21.1**	**20**
16.3	**2**	**18.6**	**11**	**19.4**	**14**	**16.9**	**3.5**
17.4	**6**	**17.1**	**5**	**18.5**	**10**	**17.5**	**7**
20.1	**17**	**19.5**	**15**	**20.5**	**19**	**18.3**	**9**
21.6	**22**	**20.3**	**18**	**22.3**	**23**	**19.8**	**16**
$R_{i.}$	67		53.5		111		68.5

N=24 a=4 $n_i = 6$ ***i=1,2,3,4***

There is a ties then we will use

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{a}\sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4}\right]$$

Zahra Kaabi

$$S^2 = \frac{1}{24-1}\left[4899.5 - \frac{24(24+1)^2}{4}\right] = 49.978$$

$$H = \frac{1}{S^2}\left[\sum_{i=1}^{a} \frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}\right]$$

$$H = \frac{1}{49.978}\left[4060.75 - \frac{24(24+1)^2}{4}\right] = 6.218$$

If

$$H > \chi^2_{\alpha,a-1}$$

Then we reject the null hypothesis

$$\chi^2_{0.05,3} = 7.8147$$

since $H = 6.218 < 7.8147 = \chi^2_{0.05,3}$ we cannot reject the null hypothesis then there is no difference between the four types of fluids

Fluid	Obs
1	17.6
1	18.9
1	16.3
1	17.4
1	20.1
1	21.6
2	16.9
2	15.3
2	18.6
2	17.1
2	19.5

2	20.3
3	21.4
3	23.6
3	19.4
3	18.5
3	20.5
3	22.3
4	19.3
4	21.1
4	16.9
4	17.5
4	18.3
4	19.8

Kruskal-Wallis Test: observation versus fluid

Descriptive Statistics

Fluid	N	Median	Mean Rank	Z-Value
1	6	18.25	11.2	-0.53
2	6	17.85	8.9	-1.43
3	6	20.95	18.5	2.40
4	6	18.80	11.4	-0.43
Overall	24		12.5	

Test

Null hypothesis H_0: All medians are equal

Alternative hypothesis H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	3	6.21	0.102
Adjusted for ties	3	6.22	0.101

Or since p-value= 0.101 >0.05 =α then we cannot reject H_0: All medians are equal then there is no difference between the four types of fluids

Zahra Kaabi

HW: Use the Kruskal – Wallis test for the experiment in Problem 3.13.

3.13: A rental car company wants to investigate whether the type of car rented affects the length of the rental period. An experiment is run for one week at a particular location, and 10 rental contracts are selected at random for each car type. The results are shown in the following table.

Type of Car	Observations									
Subcompact	3	5	3	7	6	5	3	2	1	6
Compact	1	3	4	7	5	6	3	2	1	7
Midsize	4	1	3	5	7	1	2	4	2	7
Full size	3	5	7	5	10	3	4	7	2	7

(a) Is there evidence to support a claim that the type of car rented affects the length of the rental contract? Use α =0.05

Type of car	Observations
subcompact	3
subcompact	5
subcompact	3
subcompact	7
subcompact	6
subcompact	5
subcompact	3
subcompact	2
subcompact	1
subcompact	6
compact	1
compact	3
compact	4
compact	7
compact	5
compact	6
compact	3

a=4

n=10

N=40

compact	2
compact	1
compact	7
midsize	4
midsize	1
midsize	3
midsize	5
midsize	7
midsize	1
midsize	2
midsize	4
midsize	2
midsize	7
fullsize	3
fullsize	5
fullsize	7
fullsize	5
fullsize	10
fullsize	3
fullsize	4
fullsize	7
fullsize	2
fullsize	7

Kruskal-Wallis Test: Observations versus Type of car

Descriptive Statistics

Type of car	N	Median	Mean Rank	Z-Value
compact	10	3.5	19.0	-0.47
fullsize	10	5.0	25.5	1.56
midsize	10	3.5	17.4	-0.97
subcompact	10	4.0	20.1	-0.12
Overall	40		20.5	

Test

Null hypothesis	H_0: All medians are equal
Alternative hypothesis	H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	3	2.71	0.439