



```

> remove("Type")
> RCB DHway.lay
  Units Permutation Location Strips Type
1     1           2       1      1    D
2     2           3       1      2    A
3     3           4       1      3    B
4     4           1       1      4    C
5     5           8       2      1    C
6     6           7       2      2    D
7     7           5       2      3    B
8     8           6       2      4    A
9     9           9       3      1    A
10    10          12       3      2    C
11    11          10       3      3    D
12    12          11       3      4    B

```

*The order obtained is the same as in the table below.*

Suppose that the trial is conducted and the amount of wear after one year is measured. The data obtained is given in the following table.

		Strip			
		1	2	3	4
Location	1	D	A	B	C
	2	C	D	B	A
	3	A	C	D	B
		40.2	50.0	38.0	49.7
Location	1	48.5	32.8	39.3	42.7
	2	51.9	53.5	51.1	46.3
	3				

Combine the data frame `RCBDHway.lay` with a vector `Wear`, that contains the values from the above table, to form a new data frame named (or copy using `RCBDHway.dat`). Use R to obtain boxplots for Locations and Types and to perform an analysis of variance of the data, including diagnostic checking.

*The output from the commands to analyze the data is as follows:*

```

> RCB DHway.dat <- data.frame(RCB DHway.lay,
+                             Wear=c(40.2,50.0,38.0,49.7,48.5,32.8,39.3,42.7,51.9,53.5,51.1,46.3))
> attach(RCB DHway.dat)
> boxplot(split(Wear, Location), xlab="Location", ylab="Wear in first year")
> boxplot(split(Wear, Type), xlab="Type", ylab="Wear in first year")
> RCB DHway.aov <- aov(Wear ~ Location + Type + Error(Location/Strips),
+                     RCB DHway.dat)
> summary(RCB DHway.aov)

```

```

Error: Location
      Df Sum Sq Mean Sq
Location 2 199.452  99.726

```

```

Error: Location:Strips
      Df Sum Sq Mean Sq F value Pr(>F)
Type    3 205.273  68.424  5.8067 0.03304
Residuals 6  70.702  11.784
> #Compute Location F and p
> Location.F <- 99.726/11.784
> Location.p <- 1-pf(Location.F, 2, 6)
> data.frame(Location.F,Location.p)
  Location.F Location.p
1    8.462831 0.01792619

```

```

> tukey.lsd(RCBDHway.aov, RCBDHway.dat, error.term="Location:Strips")
$Tukey.SS
[1] 20.68332

$Tukey.F
[1] 2.067573

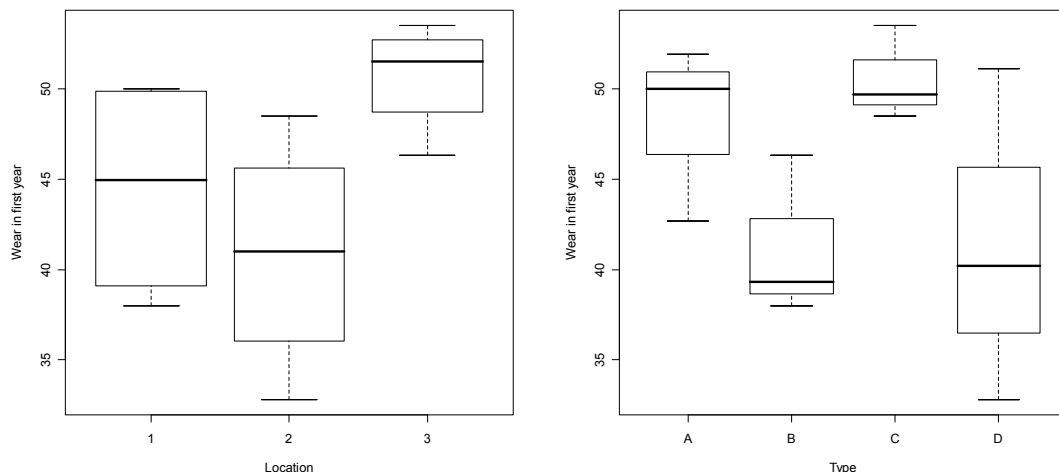
$Tukey.p
[1] 0.2099755

$Devn.SS
[1] 50.01835

> #
> # Diagnostic checking
> #
> res <- resid.errors(RCBDHway.aov)
> fit <- fitted.errors(RCBDHway.aov)
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)

```

*The boxplots generated by these commands are as follows:*



*In this case the boxplots are not very informative because each box is based on very few observations. In particular, the Types boxplot would seem to indicate much greater variability for the type B paving. However, each of the Type boxplots is based on only 3 observations, and these plot as the ends of the box and the median. Consequently, the apparent increase variance is for type B is more likely the result of one outlier, such as the very low value which is the lowest value in the data.*

*The analysis of variance constructed from the above output follows.*

**Step 1: Set up hypotheses**

- a)  $H_0: \tau_A = \tau_B = \tau_C = \tau_D$  (or  $\mathbf{X}_T \boldsymbol{\tau}$  not required in model)  
 $H_1$ : not all population Type means are equal

- b)  $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8$  (or  $\mathbf{X}_L\boldsymbol{\beta}$  not required in model)  
 $H_1$ : not all population Location means are equal

Set  $\alpha = 0.05$ .

Step 2: Calculate test statistics

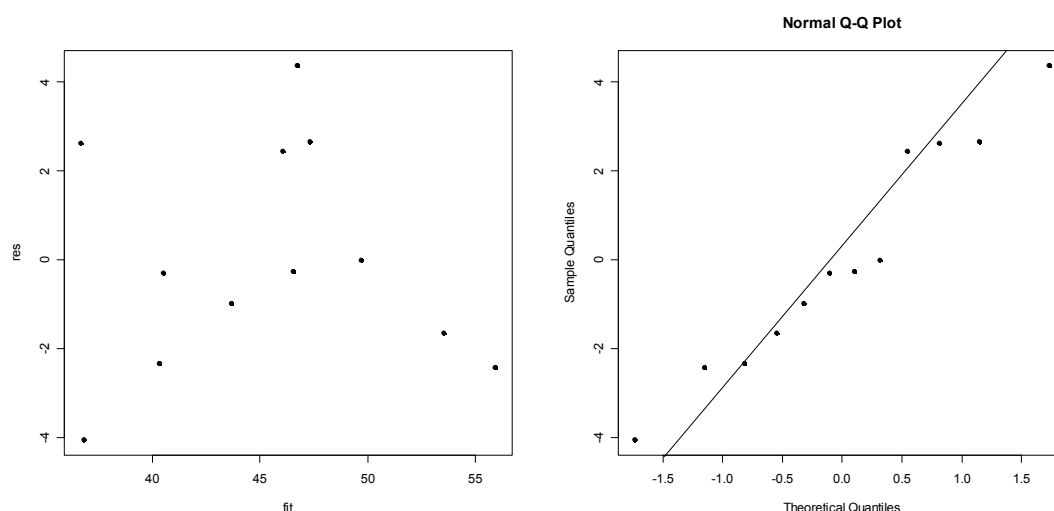
The analysis of variance table for the RCBD is:

Source	df	SSq	MSq	E[MSq]	F	Prob
Locations	2	199.45	99.73	$\sigma^2 + f_L(\boldsymbol{\psi})$	8.46	0.0179
Strips[Locations] Type	9 3	205.27	68.42	$\sigma^2 + f_T(\boldsymbol{\psi})$	5.81	0.0330
Residual	6	70.70	11.78	$\sigma^2$		
Nonadditivity	1	20.68	20.68		2.07	0.2100
Deviations	5	50.02	10.00			
Total	11	67.322				

Step 3: Decide between hypotheses

It would appear that there are significant differences between the Locations ( $p = 0.0179$ ) and the Types ( $p = 0.0330$ ) so that the additive expectation model  $\boldsymbol{\psi}_{L+T} = \mathbf{X}_L\boldsymbol{\beta} + \mathbf{X}_T\boldsymbol{\tau}$  appears to best describe the data.

There is some suggestion of curvature in the residuals-versus-fitted-values plot (below). However, Tukey's test for transformable nonadditivity is not significant ( $p = 0.2100$ ) and so we conclude that there is no significant curvature in the residuals. One problem here is that there residuals are based on only 6 degrees of freedom and so are limited for checking the assumptions. The normal probability plot displays an approximately linear trend and so is satisfactory.



**IV.4** An experiment was conducted to study the effects of temperature on the life (in hours) of a component. An RCBD was employed with five ovens forming the blocks. Four temperatures were randomly assigned to four runs within each oven. The following results were recorded:

		Temperature (degrees)			
		200	300	400	500
Oven	I	340	324	307	274
	II	361	338	312	281
	III	346	328	298	276
	IV	358	332	315	285
	V	343	321	294	269

What is the response variable for this experiment? *Life*

There are two factors other than Temperature in this experiment. What are they?

*Oven and Runs*

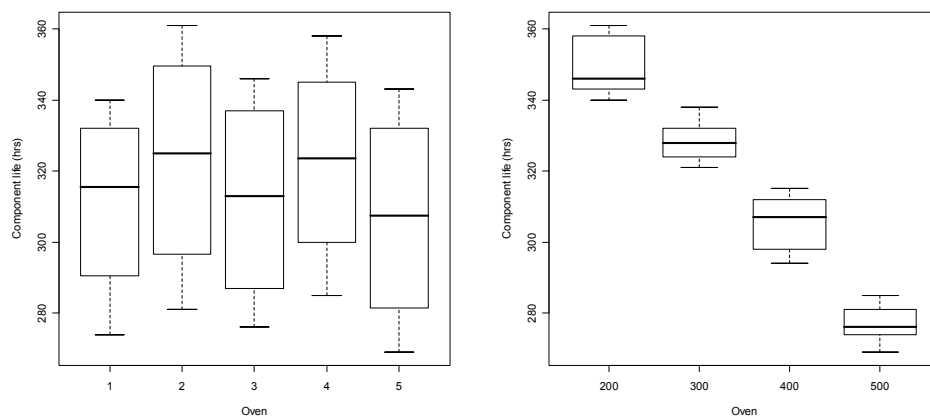
Use `fac.gen` to generate these two factors in a data frame.

Then add Temperature and Life to the data frame and produce the boxplots for an initial exploration of the data.

Assuming Oven is random, perform an analysis of variance on the data using R, including diagnostic checking and an appropriate examination of mean differences. In particular what temperature would you recommend be used with this component.

*The following expressions produce the boxplots given below.*

```
> RCBDComponent.dat <- fac.gen(list(Oven = 5, Run = 4))
> RCBDComponent.dat$Temperature <- factor(rep(c(200,300,400,500), times = 5))
> RCBDComponent.dat$Life <- c(340,324,307,274,361,338,312,281,346,328,
+                             298,276,358,332,315,285,343,321,294,269)
> attach(RCBDComponent.dat)
> boxplot(split(Life, Oven), xlab="Oven", ylab="Component life (hrs)")
> boxplot(split(Life, Temperature), xlab="Oven", ylab="Component life (hrs)")
```



*There appears to be little difference between the ovens. The mean life appears to steadily decrease with increasing temperature.*

*As the factor Temperature is quantitative, polynomials will be investigated for describing the trend in the temperature means. The expressions for doing this are shown below.*

```
> #
> # Set up to fit polynomials
> #
> Temperature.lev <- c(200,300,400,500)
> RCBDDComponent.dat$Temperature <- ordered(RCBDDComponent.dat$Temperature,
+                                           levels=Temperature.lev)
> contrasts(RCBDDComponent.dat$Temperature) <- contr.poly(4,
scores=Temperature.lev)
> contrasts(RCBDDComponent.dat$Temperature)
      .L      .Q      .C
200 -0.6708204  0.5 -0.2236068
300 -0.2236068 -0.5  0.6708204
400  0.2236068 -0.5 -0.6708204
500  0.6708204  0.5  0.2236068
> RCBDDComponent.aov <- aov(Life ~ Temperature + Error(Oven/Run),
RCBDDComponent.dat)
>
> summary(RCBDDComponent.aov,
+         split = list(Temperature = list(L = 1, Q = 2, Dev=3)))
```

Error: Oven

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	4	845.30	211.33		

Error: Oven:Run

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Temperature	3	14610.6	4870.2	365.4934	4.698e-12
Temperature: L	1	14544.4	14544.4	1091.5092	3.748e-13
Temperature: Q	1	64.8	64.8	4.8630	0.04769
Temperature: Dev	1	1.4	1.4	0.1081	0.74802
Residuals	12	159.9	13.3		

```
> #Compute Oven F and p
> Oven.F <- 211.33/13.3
> Oven.p<- 1-pf(Oven.F, 4, 12)
> data.frame(Oven.F,Oven.p)
      Oven.F      Oven.p
1 15.88947 9.704087e-05
> #
> # Diagnostic checking
> #
```

```

> res <- resid.errors(RCBDDComponent.aov)
> fit <- fitted.errors(RCBDDComponent.aov)
>
> plot(fit, res, pch=16)
> qqnorm(res, pch = 16)
> qqline(res)
> tukey.lsd(RCBDDComponent.aov, RCBDDComponent.dat, error.term="Oven:Run")
** Warning - there appears to be extremely little non-linear variation so that
the values for Tukey.SS are unstable and the results below may be unreliable.
Only use if at least two non-interacting factors above the same Residual
in the analysis.
$Tukey.SS
[1] 0.742649

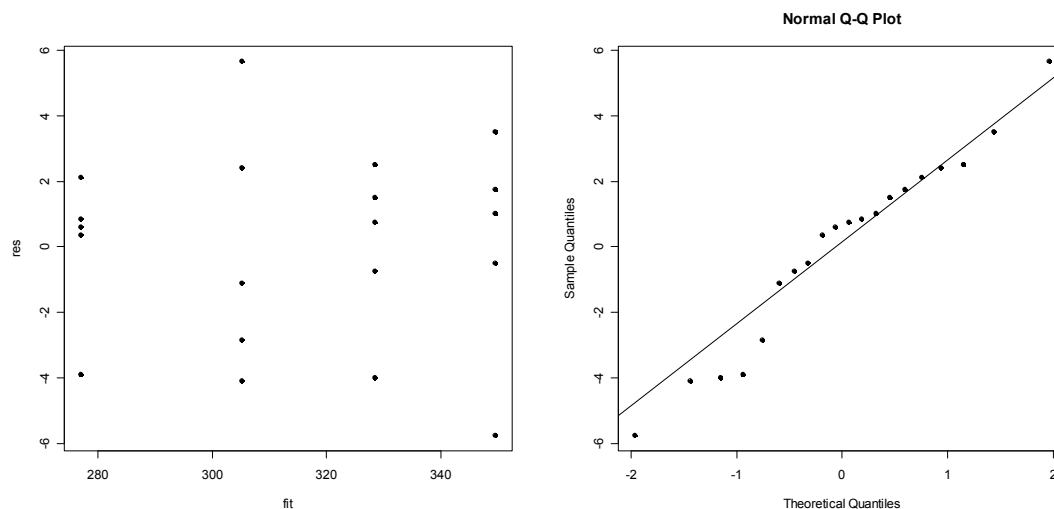
$Tukey.F
[1] 0.05132743

$Tukey.p
[1] 0.8249244

$Devn.SS
[1] 159.1574

> #
> # get fitted equation
> #
> Te <- as.vector(Temperature)
> Te <- as.numeric(Te)
> Te2 <- Te * Te
> RCBDDComponent.lm <- lm(Life ~ Te + Te2)
> coef(RCBDDComponent.lm)
(Intercept)      Te      Te2
  379.72000   -0.11520   -0.00018
> #
> # plot means and fitted line
> #
> RCBDDComponent.tab <- model.tables(RCBDDComponent.aov, type="means")
> Temperature.Mean <- RCBDDComponent.tab$tables$Temperature
> plot(x=Temperature.lev, y=Temperature.Mean, xlab="Temperature", ylab="Wear")
> RCBDDComponent.coef <- coef(RCBDDComponent.lm)
> tempx <- seq(200, 500, 5)
> Temperature.Fit <- RCBDDComponent.coef[[1]] + RCBDDComponent.coef[[2]]*tempx +
+ RCBDDComponent.coef[[3]]*tempx*tempx
> lines(x=tempx, y=Temperature.Fit, type="l")

```



Step 1: Set up hypotheses

a)  $H_0: \tau_k - \mu - \gamma_1 x_k - \gamma_2 x_k^2 = 0$  for all  $k$  (Deviations from quadratic are zero)  
 $H_1: \tau_k - \mu - \gamma_1 x_k - \gamma_2 x_k^2 \neq 0$  for all  $k$

b)  $H_0: \gamma_2 = 0$   
 $H_1: \gamma_2 \neq 0$

c)  $H_0: \gamma_1 = 0$   
 $H_1: \gamma_1 \neq 0$

d)  $H_0: \sigma_0^2 = 0$   
 $H_1: \sigma_0^2 \neq 0$

Set  $\alpha = 0.05$ .

Step 2: Calculate test statistics

The analysis of variance table for a RCBD is:

Source	df	SSq	MSq	E[MSq]	F	Prob
Oven	4	845.30	211.32	$\sigma^2 + 4\sigma_0^2$	15.86	0.0001
Runs[Oven]	15	15555.90				
Temperature	3	14610.60	4870.20	$\sigma^2 + q_T(\psi)$	365.79	0.0000
Linear	1	14544.36	14544.36		1091.51	0.0000
Quadratic	1	64.80	64.80		4.86	0.0477
Deviations	1	1.44	1.44		0.11	0.7480
Residual	12	159.90	13.32	$\sigma^2$		
Total	19	16401.20				

Step 3: Decide between hypotheses

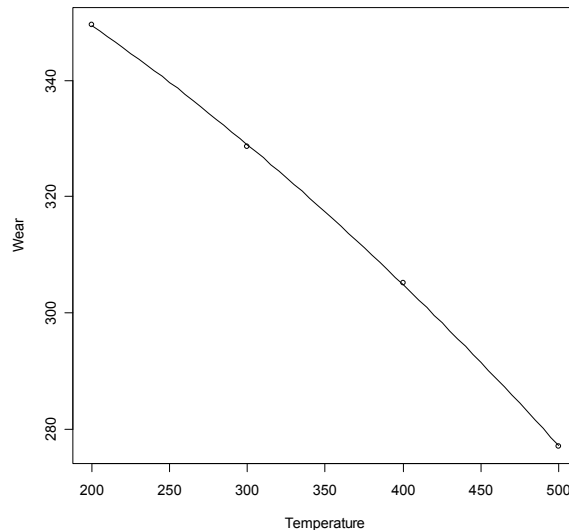
*As the Deviations term is not significant ( $p = 0.7480$ ) but the Quadratic term is significant ( $p = 0.0477$ ), the life displays a quadratic trend over time. The expectation model that best describes the data appears to be  $E[\mathbf{Y}] = \mathbf{X}_2 \boldsymbol{\theta}_2$  where  $\boldsymbol{\theta}_2' = [\mu \ \gamma_1 \ \gamma_2]$ .*

*The residuals-versus-fitted-values plot appears to be satisfactory, as does the normal probability plot. Note that Tukey's test for transformable nonadditivity is not applicable here as Oven is random and there is only the one term, Temperature, in the expectation model.*



The equation of the fitted line is:

$$E[Y_i] = 379.72 - 0.1152 \text{Temp} - 0.00018 \text{Temp}^2$$



The best temperature to use would be 200 as this would give the maximum life over the observed range of temperatures.

**IV.5** In evaluating insecticides, the numbers of living adult plum curculios emerging from separate caged areas of treated soil were observed. The results are shown in the table below.

		Insecticide					
		Lindane	Dieldrin	Aldrin	EPN	Chlordane	Check
Block	1	14	7	6	95	37	212
	2	6	1	1	133	31	172
	3	8	0	1	86	13	202
	4	36	15	4	115	69	217

This data, including the factors, is contained in *RCBDInse.dat.rda* available from the [Statistical Modelling resources web site](#). Open this file and then perform an analysis of variance on the data using R, including diagnostic checking and the appropriate examination of mean differences.

The output from the commands to analyze the data is as follows:

```
> load("RCBDInse.dat.rda")
> attach(RCBDInse.dat)
> boxplot(split(No.Curc, Blocks), xlab="Blocks", ylab="No. Curculios")
> boxplot(split(No.Curc, Insecticide), xlab="Insecticide",
+          ylab="No. Curculios")
> #
> # analysis
> #
> RCBDInse.aov <- aov(No.Curc ~ Blocks + Insecticide + Error(Blocks/Plots),
+                    RCBDInse.dat)
> summary(RCBDInse.aov)
```

```

Error: Blocks
      Df Sum Sq Mean Sq
Blocks 3 1945.46  648.49

Error: Blocks:Plots
      Df Sum Sq Mean Sq F value    Pr(>F)
Insecticide 5 122640   24528  124.22 1.173e-11
Residuals  15   2962    197
> #Compute Blocks F and p
> Blocks.F <- 648.49/197
> Blocks.p <- 1-pf(Blocks.F, 3, 15)
> data.frame(Blocks.F,Blocks.p)
  Blocks.F Blocks.p
1 3.291827 0.04981474
> #
> # Diagnostic checking
> #
> res <- resid.errors(RCBDInse.aov)
> fit <- fitted.errors(RCBDInse.aov)
>
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> tukey.ldf(RCBDInse.aov, RCBDInse.dat, error.term="Blocks:Plots")
$Tukey.SS
[1] 5.173814

$Tukey.F
[1] 0.02449873

$Tukey.p
[1] 0.8778573

$Devn.SS
[1] 2956.618

> #
> # multiple comparisons
> #
> model.tables(RCBDInse.aov, type="means")
Tables of means
Grand mean

61.70833

  Blocks
Blocks
  1      2      3      4
61.83 57.33 51.67 76.00

  Insecticide
Insecticide
  Aldrin    Check Chlordane  Dieldrin      EPN    Lindane
  3.00    200.75    37.50     5.75    107.25    16.00
> q <- qtkey(0.95, 6, 15)
> q
[1] 4.594735

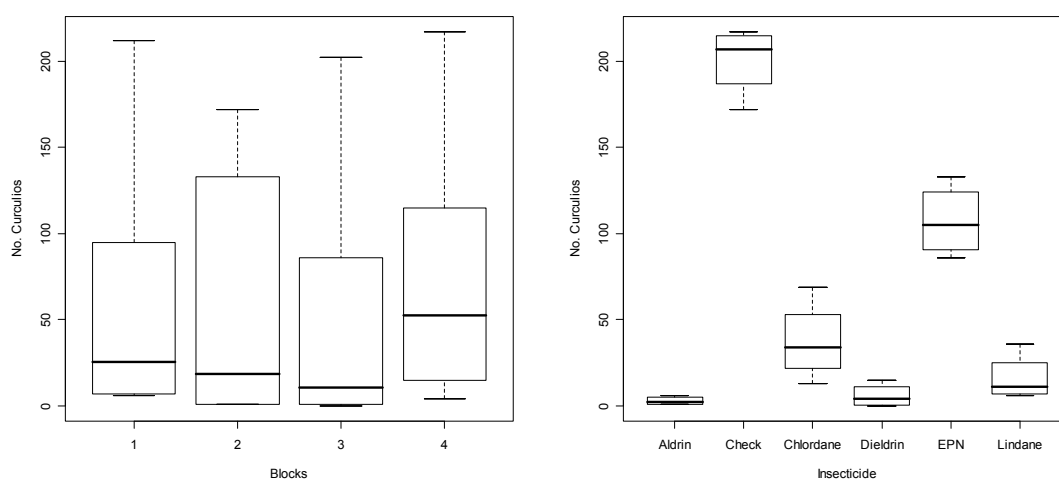
```

```

> #
> # Plotting Treat means
> #
> RCBDInse.tab <- model.tables(RCBDInse.aov, type="means")
> RCBDInse.Inse.Mean <- data.frame(Inse.lev = levels(Insecticide),
+                               Inse.Mean = as.vector(RCBDInse.tab$tables$Insecticide))
> RCBDInse.Inse.Mean <- RCBDInse.Inse.Mean[order(RCBDInse.Inse.Mean$Inse.Mean,
+                                               decreasing=TRUE),]
> RCBDInse.Inse.Mean$Inse.lev <- factor(RCBDInse.Inse.Mean$Inse.lev,
+                                       levels=RCBDInse.Inse.Mean$Inse.lev)
> barchart(Inse.Mean ~ Inse.lev, xlab="Insecticide",
+          ylab="No. Curculios", main="Fitted values for No. Curculios",
+          data=RCBDInse.Inse.Mean)

```

*The boxplots generated by these commands are as follows:*



*These indicate that there is not much difference between the blocks, but that there are substantial differences between the insecticides.*

*The analysis of variance constructed from the above output follows.*

**Step 1: Set up hypotheses**

a)  $H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = \tau_6$  (or  $\mathbf{X}_I\boldsymbol{\tau}$  not required in model)  
 $H_1$ : at least one pair of population Insecticide means is different

b)  $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4$  (or  $\mathbf{X}_B\boldsymbol{\beta}$  not required in model)  
 $H_1$ : at least one pair of population Block means is different

Set  $\alpha = 0.05$ .

*Step 2: Calculate test statistics*

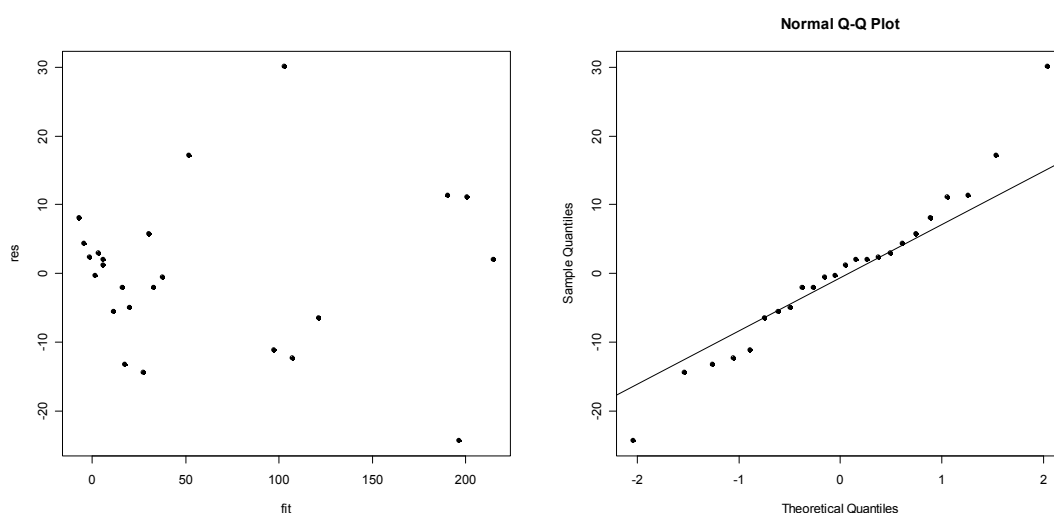
*The analysis of variance table for a RCBD is:*

Source	df	SSq	MSq	E[MSq]	F	Prob
Blocks	3	1945	648.3	$\sigma^2 + q_B(\psi)$	3.28	0.0498
Plots[Blocks]	20	125601				
Insecticide	5	122639	24527.9	$\sigma^2 + q_I(\psi)$	124.22	0.0000
Residual	15	2962	197.0	$\sigma^2$		
Non-additivity	1	5	5.2		0.02	0.8779
Deviations	14	2957	211.2			
Total	23	127546				

*Step 3: Decide between hypotheses*

*There is a marked difference between the insecticides in the number of curculios emerging. Also, the Blocks source is not significant, although only just — it would seem advisable to retain the Blocks term in the model. Consequently, the expectation model that appears to best describe the data is the additive model  $\psi_{B+I} = \mathbf{X}_B\beta + \mathbf{X}_I\tau$ .*

*The residuals-versus-fitted-values plot appears to be satisfactory as there is no particular pattern in the residuals, and the normal probability plot is displaying a roughly straight line trend except for both a high and a low outlier. Perhaps these two outliers that should be further investigated to see if there is an explanation for them. Also, Tukey's test for transformable nonadditivity is not significant ( $p = 0.8779$ ).*



*The value for Tukey's HSD procedure is as follows:*

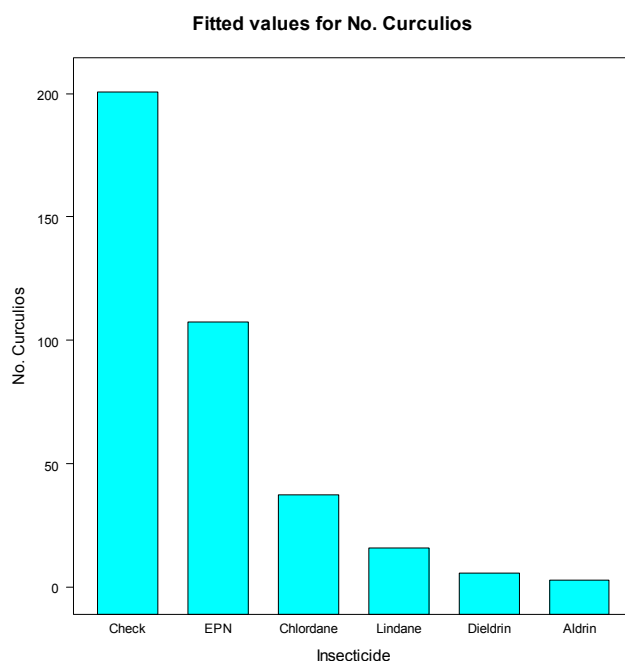
$$w(5\%) = \frac{4.594735}{\sqrt{2}} \times \sqrt{\frac{197 \times 2}{4}} = \frac{4.594735}{\sqrt{2}} \times 9.247 = 32.25$$

*The table of means is*

Insecticide						
Aldrin	Check	Chlordane	Dieldrin	EPN	Lindane	
3.00	200.75	37.50	5.75	107.25	16.00	

*To assist in interpreting this diagram a bar chart of the means is given in the following figure. This bar chart was produced by saving the Insecticide means and levels in a data.frame — these were ordered in descending order for the means.*

```
> #
> # Plotting Treat means
> #
> RCBDInse.tab <- model.tables(RCBDInse.aov, type="means")
> RCBDInse.Inse.Mean <- data.frame(Inse.lev = levels(Insecticide),
+                               Inse.Mean = as.vector(RCBDInse.tab$Insecticide))
> RCBDInse.Inse.Mean <- RCBDInse.Inse.Mean[order(RCBDInse.Inse.Mean$Inse.Mean,
+                                               decreasing=TRUE),]
> #use factor to order bars
> RCBDInse.Inse.Mean$Inse.lev <-factor(RCBDInse.Inse.Mean$Inse.lev,
+                                   levels=RCBDInse.Inse.Mean$Inse.lev)
> barchart(Inse.Mean ~ Inse.lev, xlab="Insecticide",
+          ylab="No. Curculios", main="Fitted values for No. Curculios",
+          data=RCBDInse.Inse.Mean)
```



*From the bar chart and the multiple comparisons, the following conclusions are reached:*

*The only difference between the last 4 insecticides is that the difference between Lindane and Chlordane is just significant. Check is significantly better than all the insecticides and EPN is significantly higher than the remaining four insecticides.*