ANOVA (Analysis of Variance)

Population 1	Population 2		Population J
mean μ_1	mean μ_2		mean μ_J
std. dev. σ	std. dev. σ	. . .	std. dev. σ
$\Downarrow$	$\Downarrow$		$\Downarrow$
$y_{11}, y_{21}, \dots, y_{n_1 1}$	$y_{12}, y_{22}, \dots, y_{n_2 2}$		$y_{1J}, y_{2J}, \dots, y_{n_J J}$
$\bar{y}_1,\ s_1^2$	$\bar{y}_2,\ s_2^2$		$\bar{y}_J,\ s_J^2$

H_0: $\mu_1 = \mu_2 = \dots = \mu_J$ $\qquad$ H_1: not all of the μ_j are equal.

(We want to test *simultaneously* for differences among the means of all J populations.)

<u>Example</u>: There is a great deal of interest in comparing the mileage rating of different makes of automobiles. Other things being equal, customers buy car that gets the best mileage. Suppose we wish to compare the mean fuel consumption for $J = 3$ different makes of automobile, Car 1, Car 2, and Car 3. Suppose 5 cars of each make are selected randomly and the gasoline mileage (in miles per gallon) is recorded for each car. Suppose that

$$\bar{y}_1 = 19.0, \qquad \bar{y}_2 = 20.4, \qquad \bar{y}_3 = 21.8.$$

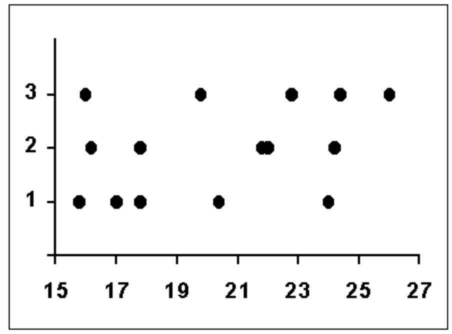
Is there enough evidence that the mean fuel consumption is not the same for Car 1, Car 2 and Car 3?

Data Set A:

	Car 1	Car 2	Car 3
	18.2	19.8	21.2
	19.4	21.0	21.8
	19.6	20.0	22.4
	19.0	20.8	22.0
	18.8	20.4	21.6
Sample mean	19.0	20.4	21.8
Sample std. dev.	0.5477	0.5099	0.4472
Sample variance	0.300	0.260	0.200

Data Set B:

	Car 1	Car 2	Car 3
	17.0	24.2	26.0
	20.4	22.0	19.8
	24.0	17.8	24.4
	15.8	16.2	16.0
	17.8	21.8	22.8
Sample mean	19.0	20.4	21.8
Sample std. dev.	3.26	3.29	3.97
Sample variance	10.66	10.84	15.76



The Analysis of Variance Idea:

Analysis of Variance (ANOVA) compares the variation due to specific sources (between groups) with the variation among individuals who should be similar (within groups). In particular, ANOVA tests whether several populations have the same mean by comparing how far apart the sample means are with how much variation there is within the samples.

ANOVA Assumptions:

- We have ***J* independent simple random samples**, one from each of *J* populations.

- The jth population has a **normal distribution** with unknown mean μ_j. The means may be different in the different populations. The ANOVA F test statistic tests the null hypothesis that all of the populations have the same mean:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_J$$

$$H_1: \text{not all of the } \mu_j \text{ are equal.}$$

- All of the populations have the **same standard deviation σ**, whose value is unknown.

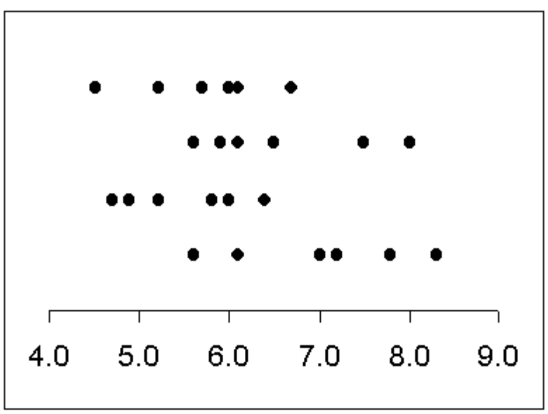
  The results of ANOVA F test are approximately correct when the largest sample standard deviation is no more than twice as large as the smallest sample standard deviation.

ANOVA table:

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*Test Statistic*
Source	**SS**	**DF**	**MS**	**F**
Between	SSB	$J-1$	$\frac{\text{SSB}}{J-1}$	$\frac{\text{MSB}}{\text{MSW}}$
Within	SSW	$N-J$	$\frac{\text{SSW}}{N-J}$	
Total	SSTot	$N-1$		

1. Six samples of each of four types of cereal grain grown in a certain region were analyzed to determine thiamin content, resulting in the following data (μg/g):

Wheat	5.2	4.5	6.0	6.1	6.7	5.7
Barley	6.5	8.0	6.1	7.5	5.9	5.6
Maize	5.8	4.7	6.4	4.9	6.0	5.2
Oats	8.3	6.1	7.8	7.0	5.6	7.2



Does this data suggest that at least two of the grains differ with respect to true average thiamin content? Use a level $\alpha = 0.05$ test.

	n_j	$\bar{y}_j$	s_j	s_j^2
Wheat	6	5.7	0.7668	0.588
Barley	6	6.6	0.9508	0.904
Maize	6	5.5	0.6693	0.448
Oats	6	7.0	1.0139	1.028

$$y_{ij} = \bar{y} + (\bar{y}_j - \bar{y}) + (y_{ij} - \bar{y}_j)$$

$$\begin{bmatrix} 5.2 & 4.5 & 6.0 & 6.1 & 6.7 & 5.7 \\ 6.5 & 8.0 & 6.1 & 7.5 & 5.9 & 5.6 \\ 5.8 & 4.7 & 6.4 & 4.9 & 6.0 & 5.2 \\ 8.3 & 6.1 & 7.8 & 7.0 & 5.6 & 7.2 \end{bmatrix}$$

$$= \begin{bmatrix} 6.2 & 6.2 & 6.2 & 6.2 & 6.2 & 6.2 \\ 6.2 & 6.2 & 6.2 & 6.2 & 6.2 & 6.2 \\ 6.2 & 6.2 & 6.2 & 6.2 & 6.2 & 6.2 \\ 6.2 & 6.2 & 6.2 & 6.2 & 6.2 & 6.2 \end{bmatrix}$$

$$+ \begin{bmatrix} -0.5 & -0.5 & -0.5 & -0.5 & -0.5 & -0.5 \\ 0.4 & 0.4 & 0.4 & 0.4 & 0.4 & 0.4 \\ -0.7 & -0.7 & -0.7 & -0.7 & -0.7 & -0.7 \\ 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \end{bmatrix}$$

$$+ \begin{bmatrix} -0.5 & -1.2 & 0.3 & 0.4 & 1.0 & 0.0 \\ -0.1 & 1.4 & -0.5 & 0.9 & -0.7 & -1.0 \\ 0.3 & -0.8 & 0.9 & -0.6 & 0.5 & -0.3 \\ 1.3 & -0.9 & 0.8 & 0.0 & -1.4 & 0.2 \end{bmatrix}$$

$$(y_{ij} - \bar{y}) \quad = \quad (\bar{y}_j - \bar{y}) \quad + \quad (y_{ij} - \bar{y}_j)$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n_j}\left(y_{ij} - \bar{y}\right)^2 \quad = \quad \sum_{j=1}^{J}\sum_{i=1}^{n_j}\left(\bar{y}_j - \bar{y}\right)^2 \quad + \quad \sum_{j=1}^{J}\sum_{i=1}^{n_j}\left(y_{ij} - \bar{y}_j\right)^2$$

SS Total	SS Between	SS Within
	$\sum_{j=1}^{J} n_j\left(\bar{y}_j - \bar{y}\right)^2$	$\sum_{j=1}^{J}\left(n_j - 1\right)s_j^2$
$N - 1$	$J - 1$	$N - J$
degrees of freedom	degrees of freedom	degrees of freedom

a) How many different groups (treatments) are there?

b) What was the total sample size, N?

c) Compute the overall average $\overline{y}$.

$$\overline{y} = \frac{n_1 \cdot \overline{y}_1 + n_2 \cdot \overline{y}_2 + \ldots + n_J \cdot \overline{y}_J}{N}$$

d) Compute SSB.

$$\text{SSB} = n_1 \cdot \left(\overline{y}_1 - \overline{y}\right)^2 + n_2 \cdot \left(\overline{y}_2 - \overline{y}\right)^2 + \ldots + n_J \cdot \left(\overline{y}_J - \overline{y}\right)^2$$

e) Find the number of degrees of freedom that is associated with SSB.

f) Compute MSB.

$$\text{MSB} = \frac{\text{SSB}}{J - 1}$$

g) Compute SSW.

$$\text{SSW} = \left(n_1 - 1\right)\cdot s_1^2 + \left(n_2 - 1\right)\cdot s_2^2 + \ldots + \left(n_J - 1\right)\cdot s_J^2$$

h) Find the number of degrees of freedom that is associated with SSW.

i) Compute MSW.

$$\text{MSW} = \frac{\text{SSW}}{N - J}$$

j) Compute SSTot.

SSTot = SSB + SSW

k) Find the number of degrees of freedom that is associated with SSTot.

l) Compute the value of the test statistic F.

$$\text{F} = \frac{\text{MSB}}{\text{MSW}}$$

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$. H_1: not all of the μ_j are equal.

Reject H_0 if $F > F_\alpha(J-1, N-J)$, where

$F_\alpha(J-1, N-J)$ is the critical value of F distribution with probability α lying to its right,

degrees of freedom in the numerator = $J - 1$.

degrees of freedom in the denominator = $N - J$.

m) What is the critical value, $F_\alpha(J-1, N-J)$, at $\alpha = 0.05$?

$\alpha = 0.05$, $J - 1 = 3$, $N - J = 20$.

$F_\alpha(J-1, N-J) = F_{0.05}(3, 20) =$ __________.

n) What is your conclusion regarding the null hypothesis at $\alpha = 0.05$?

o) What is the critical value, $F_\alpha(J-1, N-J)$, at $\alpha = 0.01$?

$\alpha = 0.01$, $J - 1 = 3$, $N - J = 20$.

$F_\alpha(J-1, N-J) = F_{0.01}(3, 20) =$ __________.

p) What is your conclusion regarding the null hypothesis at $\alpha = 0.01$?

$J = \mathbf{4}$. $\qquad N = n_1 + n_2 + \ldots + n_J = 6 + 6 + 6 + 6 = \mathbf{24}$.

$$\bar{y} = \frac{n_1 \cdot \bar{y}_1 + n_2 \cdot \bar{y}_2 + \ldots + n_J \cdot \bar{y}_J}{N} = \frac{6 \cdot 5.7 + 6 \cdot 6.6 + 6 \cdot 5.5 + 6 \cdot 7.0}{24} = \mathbf{6.2}.$$

$$\text{SSB} = n_1 \cdot \left(\bar{y}_1 - \bar{y}\right)^2 + n_2 \cdot \left(\bar{y}_2 - \bar{y}\right)^2 + \ldots + n_J \cdot \left(\bar{y}_J - \bar{y}\right)^2$$

$$= 6 \cdot (5.7 - 6.2)^2 + 6 \cdot (6.6 - 6.2)^2 + 6 \cdot (5.5 - 6.2)^2 + 6 \cdot (7.0 - 6.2)^2 = \mathbf{9.24}.$$

$$\text{MSB} = \frac{\text{SSB}}{J - 1} = \frac{9.24}{3} = \mathbf{3.08}.$$

$$\text{SSW} = \left(n_1 - 1\right) \cdot s_1^2 + \left(n_2 - 1\right) \cdot s_2^2 + \ldots + \left(n_J - 1\right) \cdot s_J^2$$

$$= 5 \cdot 0.588 + 5 \cdot 0.904 + 5 \cdot 0.448 + 5 \cdot 1.028 = \mathbf{14.84}.$$

$$\text{MSW} = \frac{\text{SSW}}{N - J} = \frac{14.84}{20} = \mathbf{0.742}.$$

SSTot = SSB + SSW = **24.08**.

$$\text{F} = \frac{\text{MSB}}{\text{MSW}} = \frac{3.08}{0.742} = \mathbf{4.151}.$$

ANOVA table:

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*Test Statistic*
Source	**SS**	**DF**	**MS**	**F**
Between	9.24	3	3.08	4.151
Within	14.84	20	0.742	
Total	24.08	23		

$\text{F}_{0.05}(3, 20) = 3.10$. $\qquad$ Reject H_0 at $\alpha = 0.05$.

$\text{F}_{0.01}(3, 20) = 4.94$. $\qquad$ Accept H_0 at $\alpha = 0.01$.

Practice Problem

Construct the ANOVA tables and perform the ANOVA F test at $\alpha = 0.05$ for Data Set A and Data Set B.

Answers:

Data Set A:

Source	SS	DF	MS	F
Between	19.6	2	9.8	38.68421
Within	3.04	12	0.253333	
Total	22.64	14		

$F_{0.05}(2, 12) = 3.88.$ **Reject H_0.**

Data Set B:

Source	SS	DF	MS	F
Between	19.6	2	9.8	0.78905
Within	149.04	12	12.42	
Total	168.64	14		

$F_{0.05}(2, 12) = 3.88.$ **Do NOT Reject H_0.**

Population 1	Population 2
mean μ_1	mean μ_2
std. dev. σ	std. dev. σ
$\Downarrow$	$\Downarrow$
$y_{11}, y_{21}, \ldots, y_{n_1 1}$	$y_{12}, y_{22}, \ldots, y_{n_2 2}$
$\bar{y}_1,\ s_1^2$	$\bar{y}_2,\ s_2^2$

Assumptions:

1) Two independent samples.
2) Both populations are normal.
3) The population standard deviations are equal.

$$s_{\text{pooled}}^2 = \frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{n_1 + n_2 - 2}$$

A confidence interval for $\mu_1 - \mu_2$ is

$$\left(\bar{y}_1 - \bar{y}_2\right) \pm t_{\alpha/2} \cdot s_{\text{pooled}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$n_1 + n_2 - 2$ degrees of freedom

To test $\quad H_0: \mu_1 - \mu_2 = \delta_0$

Test Statistic: $\quad t = \dfrac{\left(\bar{y}_1 - \bar{y}_2\right) - \delta_0}{s_{\text{pooled}} \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$.

$n_1 + n_2 - 2$ degrees of freedom

1. Assume that the distributions of Y_1 and Y_2 are $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively. Given the $n_1 = 6$ observations of Y_1,

65, 68, 67, 66, 71, 68

and the $n_2 = 8$ observations of Y_2,

68, 70, 67, 71, 70, 72, 73, 69

test $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$. Use a 5% level of significance. What can you say about the p-value of this test?

$$\bar{y}_1 = \frac{405}{6} = 67.5 \qquad \bar{y}_2 = \frac{560}{8} = 70$$

$$s_1^2 = \frac{21.5}{5} = 4.3 \qquad s_2^2 = \frac{28}{7} = 4$$

$$s_{\text{pooled}}^2 = \frac{(6-1)\cdot 4.3 + (8-1)\cdot 4}{6+8-2} = 4.125$$

Test Statistic: $t = \dfrac{(67.5-70)-0}{\sqrt{4.125\cdot\left(\frac{1}{6}+\frac{1}{8}\right)}} = -2.279.$

$n_1 + n_2 - 2 = 12$ degrees of freedom

The t Distribution

r	$t_{0.40}$	$t_{0.25}$	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055

Rejection Region: Reject H_0 if $t < -t_{0.025}(12)$ or $t > t_{0.025}(12)$

$t_{0.025}(12) = 2.179.$ **Reject H_0 at $\alpha = 0.05$.**

OR

-2.681	$<$	-2.279	$<$	-2.179
$-t_{0.01}(12)$	$<$	t	$<$	$-t_{0.025}(12)$
0.01	$<$	one tail	$<$	0.025

p-value = two tails

0.02 < p-value < 0.05. **Reject H_0 at $\alpha = 0.05$.**

$J = 2.$ $\qquad N = n_1 + n_2 = 6 + 8 = 14.$

$$\bar{y} = \frac{n_1 \cdot \bar{y}_1 + n_2 \cdot \bar{y}_2 + \ldots + n_J \cdot \bar{y}_J}{N} = \frac{6 \cdot 67.5 + 8 \cdot 70}{14} = \frac{965}{14} = 68.92857.$$

$$\text{SSB} = n_1 \cdot \left(\bar{y}_1 - \bar{y}\right)^2 + n_2 \cdot \left(\bar{y}_2 - \bar{y}\right)^2 + \ldots + n_J \cdot \left(\bar{y}_J - \bar{y}\right)^2$$
$$= 6 \cdot (67.5 - 68.92857)^2 + 8 \cdot (70 - 68.92857)^2 = 21.42857.$$

$$\text{MSB} = \frac{\text{SSB}}{J-1} = \frac{21.42857}{1} = 21.42857.$$

$$\text{SSW} = \left(n_1 - 1\right) \cdot s_1^2 + \left(n_2 - 1\right) \cdot s_2^2 + \ldots + \left(n_J - 1\right) \cdot s_J^2$$
$$= 5 \cdot 4.3 + 7 \cdot 4 = 49.5.$$

$$\text{MSW} = \frac{\text{SSW}}{N-J} = \frac{49.5}{12} = 4.125. \qquad [= s_{\text{pooled}}^2]$$

$$\text{SSTot} = \text{SSB} + \text{SSW} = 70.92857.$$

Test Statistic: $\qquad \text{F} = \dfrac{\text{MSB}}{\text{MSW}} = \dfrac{21.42857}{4.125} = \mathbf{5.1948}.$ $\qquad [= t^2]$

ANOVA table:

Source	SS	DF	MS	F
Between	21.42857	1	21.42857	**5.1948**
Within	49.5	12	4.125	
Total	70.92857	13		

Critical Value: $\qquad \text{F}_{0.05}(1, 12) = 4.75.$ $\qquad [= (t_{0.025}(12))^2]$

Reject $\mathbf{H_0}$ at $\boldsymbol{\alpha = 0.05}$.

$F_{0.05}(1, 12) = 4.75 \;<\; 5.1948 \;<\; 9.33 = F_{0.01}(1, 12)$.

0.05 > p-value > 0.01. **Reject H_0 at $\alpha = 0.05$.**

In general,

$J = 2$. $N = n_1 + n_2$.

$$\bar{y} = \frac{n_1 \cdot \bar{y}_1 + n_2 \cdot \bar{y}_2}{N} = \frac{n_1 \cdot \bar{y}_1 + n_2 \cdot \bar{y}_2}{n_1 + n_2}.$$

$$\text{MSB} = \frac{\text{SSB}}{J-1} = n_1 \cdot \left(\bar{y}_1 - \bar{y}\right)^2 + n_2 \cdot \left(\bar{y}_2 - \bar{y}\right)^2$$

$$= n_1 \cdot \left(\bar{y}_1 - \frac{n_1 \cdot \bar{y}_1 + n_2 \cdot \bar{y}_2}{n_1 + n_2}\right)^2 + n_2 \cdot \left(\bar{y}_2 - \frac{n_1 \cdot \bar{y}_1 + n_2 \cdot \bar{y}_2}{n_1 + n_2}\right)^2$$

$$= n_1 \cdot \left(\frac{n_2 \cdot \left(\bar{y}_1 - \bar{y}_2\right)}{n_1 + n_2}\right)^2 + n_2 \cdot \left(\frac{n_1 \cdot \left(\bar{y}_2 - \bar{y}_1\right)}{n_1 + n_2}\right)^2$$

$$= n_1 \cdot \frac{n_2^2 \cdot \left(\bar{y}_1 - \bar{y}_2\right)^2}{\left(n_1 + n_2\right)^2} + n_2 \cdot \frac{n_1^2 \cdot \left(\bar{y}_2 - \bar{y}_1\right)^2}{\left(n_1 + n_2\right)^2}$$

$$= n_1 n_2 \cdot \frac{\left(\bar{y}_1 - \bar{y}_2\right)^2}{\left(n_1 + n_2\right)} = \frac{\left(\bar{y}_1 - \bar{y}_2\right)^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\text{MSW} = \frac{\text{SSW}}{N-J} = \frac{\left(n_1 - 1\right) \cdot s_1^2 + \left(n_2 - 1\right) \cdot s_2^2}{n_1 + n_2 - 2} = s_{\text{pooled}}^2$$

$$F = \frac{\text{MSB}}{\text{MSW}} = \frac{\left(\bar{y}_1 - \bar{y}_2\right)^2}{s_{\text{pooled}}^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = t^2.$$

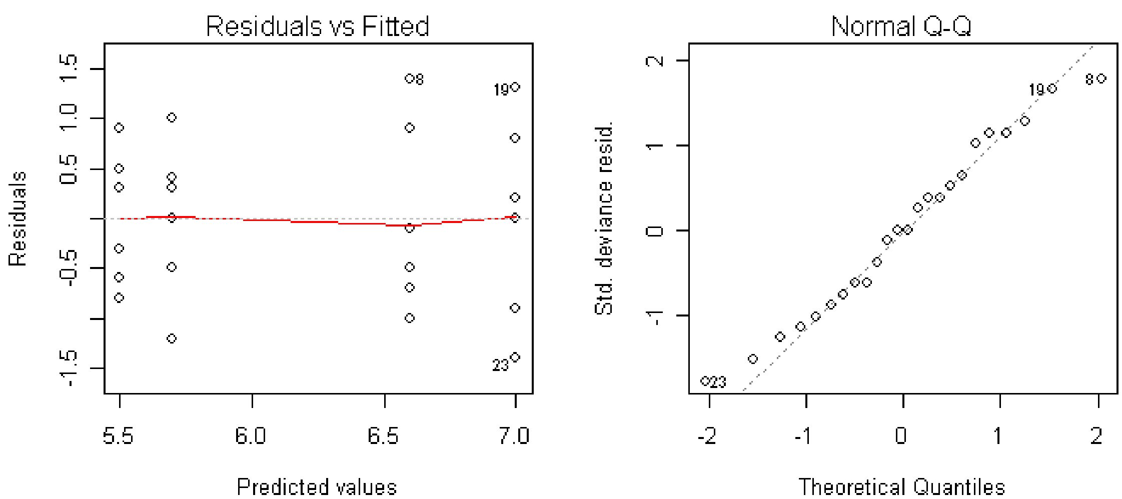
```
> y1 = c(65,68,67,66,71,68)
> y2 = c(68,70,67,71,70,72,73,69)
>
> t.test(y1,y2,alternative=c("two.sided"),var.equal=TRUE)

        Two Sample t-test

data:  y1 and y2
t = -2.2792, df = 12, p-value = 0.04174
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.8898756 -0.1101244
sample estimates:
mean of x mean of y
     67.5      70.0

>
> y = c(y1,y2)
> pop = c(rep(1,6),rep(2,8))
>
> result = glm(y ~ factor(pop))
> summary(aov(result))
            Df Sum Sq Mean Sq F value Pr(>F)
factor(pop)  1  21.43  21.429   5.195 0.0417 *
Residuals   12  49.50   4.125
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

STAT 420

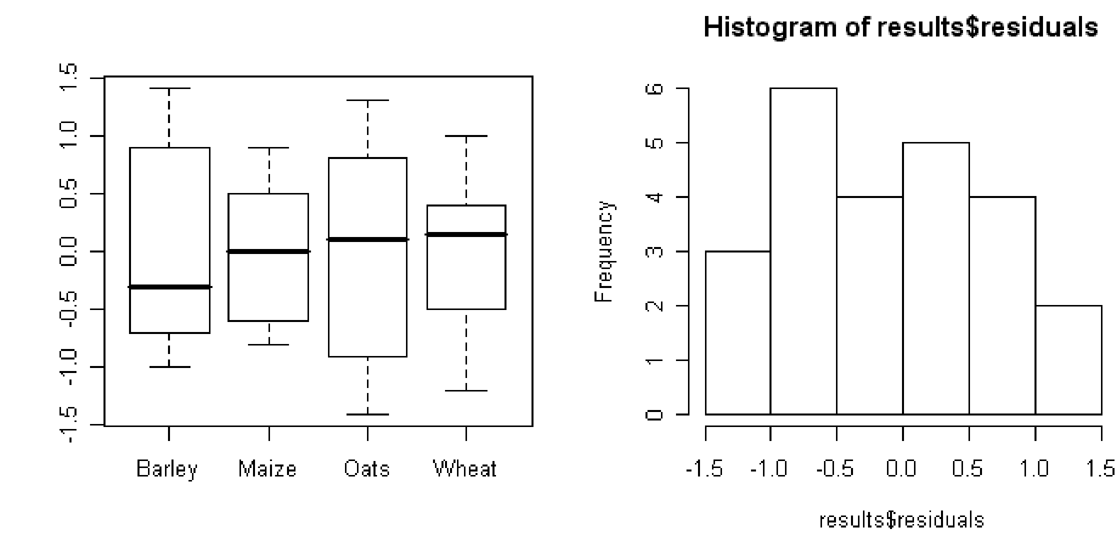
```
> Wheat <- c(5.2,4.5,6.0,6.1,6.7,5.7)
> Barley <- c(6.5,8.0,6.1,7.5,5.9,5.6)
> Maize <- c(5.8,4.7,6.4,4.9,6.0,5.2)
> Oats <- c(8.3,6.1,7.8,7.0,5.6,7.2)

> Grain <- c(rep("Wheat",6),rep("Barley",6),rep("Maize",6),rep("Oats",6))
> Thiamin <- c(Wheat,Barley,Maize,Oats)

> Grain
[1] "Wheat"  "Wheat"  "Wheat"  "Wheat"  "Wheat"  "Wheat"  "Barley" "Barley"
 [9] "Barley" "Barley" "Barley" "Barley" "Maize"  "Maize"  "Maize"  "Maize"
[17] "Maize"  "Maize"  "Oats"   "Oats"   "Oats"   "Oats"   "Oats"   "Oats"
> Thiamin
[1] 5.2 4.5 6.0 6.1 6.7 5.7 6.5 8.0 6.1 7.5 5.9 5.6 5.8 4.7 6.4 4.9 6.0 5.2 8.3
[20] 6.1 7.8 7.0 5.6 7.2

> results <- glm(Thiamin ~ factor(Grain))
> results

Call:  glm(formula = Thiamin ~ factor(Grain))

Coefficients:
      (Intercept)  factor(Grain)Maize   factor(Grain)Oats  factor(Grain)Wheat
              6.6                -1.1                 0.4                -0.9

Degrees of Freedom: 23 Total (i.e. Null);  20 Residual
Null Deviance:      24.08
Residual Deviance: 14.84        AIC: 66.57

> summary(results)

Call:
glm(formula = Thiamin ~ factor(Grain))

Deviance Residuals:
       Min          1Q      Median          3Q         Max
-1.400e+00  -6.250e-01  -1.776e-15   5.750e-01   1.400e+00

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          6.6000     0.3517  18.768 3.62e-14 ***
factor(Grain)Maize  -1.1000     0.4973  -2.212   0.0388 *
factor(Grain)Oats    0.4000     0.4973   0.804   0.4307
factor(Grain)Wheat  -0.9000     0.4973  -1.810   0.0854 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.742)

    Null deviance: 24.08  on 23  degrees of freedom
Residual deviance: 14.84  on 20  degrees of freedom
AIC: 66.572

Number of Fisher Scoring iterations: 2
```

```
> summary(aov(results))
               Df Sum Sq Mean Sq F value   Pr(>F)
factor(Grain)  3  9.240   3.080  4.1509 0.01936 *
Residuals      20 14.840   0.742
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
```

```
> par(mfrow=c(2,2))
> plot(results)
```




```
> boxplot(results$residuals ~ Grain)
> hist(results$residuals)
```




```
> shapiro.test(results$residuals)

        Shapiro-Wilk normality test

data:  results$residuals
W = 0.9704, p-value = 0.6775
```

```
==================================================================

> maize <- c(rep(0,12),rep(1,6),rep(0,6))
> oats <- c(rep(0,18),rep(1,6))
> wheat <- c(rep(1,6),rep(0,18))
>
> maize
 [1] 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0
> oats
 [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1
> wheat
 [1] 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0


> results2 <- lm(Thiamin ~ maize + oats + wheat)
> summary(results2)

Call:
lm(formula = Thiamin ~ maize + oats + wheat)

Residuals:
       Min         1Q     Median         3Q        Max
-1.400e+00 -6.250e-01 -6.126e-16  5.750e-01  1.400e+00

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.6000     0.3517  18.768 3.62e-14 ***
maize        -1.1000     0.4973  -2.212   0.0388 *
oats          0.4000     0.4973   0.804   0.4307
wheat        -0.9000     0.4973  -1.810   0.0854 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8614 on 20 degrees of freedom
Multiple R-Squared: 0.3837,     Adjusted R-squared: 0.2913
F-statistic: 4.151 on 3 and 20 DF,  p-value: 0.01936


> anova(lm(Thiamin ~ 1), results2)
Analysis of Variance Table

Model 1: Thiamin ~ 1
Model 2: Thiamin ~ maize + oats + wheat
  Res.Df   RSS Df Sum of Sq      F  Pr(>F)
1     23 24.08
2     20 14.84  3      9.24 4.1509 0.01936 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$Y_i = \beta_0 + \beta_1 \,\mathtt{maize}_i + \beta_2 \,\mathtt{oats}_i + \beta_3 \,\mathtt{wheat}_i + \varepsilon_i, \qquad i = 1, 2, \dots, 24.$$

Barley	β_0
Maize	$\beta_0 + \beta_1$
Oats	$\beta_0 + \beta_2$
Wheat	$\beta_0 + \beta_3$

H_0: Barley = Maize = Oats = Wheat

$\Leftrightarrow \quad H_0: \beta_1 = \beta_2 = \beta_3 = 0$

==

```
> barley <- c(rep(0,6),rep(1,6),rep(0,12))
> barley
 [1] 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0

> results3 <- lm(Thiamin ~ maize + oats + wheat + barley)
> summary(results3)

Call:
lm(formula = Thiamin ~ maize + oats + wheat + barley)

Residuals:
       Min         1Q     Median         3Q        Max 
-1.400e+00 -6.250e-01 -6.126e-16  5.750e-01  1.400e+00 

Coefficients: (1 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   6.6000     0.3517  18.768 3.62e-14 ***
maize        -1.1000     0.4973  -2.212   0.0388 *  
oats          0.4000     0.4973   0.804   0.4307    
wheat        -0.9000     0.4973  -1.810   0.0854 .  
barley            NA         NA      NA       NA    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 0.8614 on 20 degrees of freedom
Multiple R-Squared: 0.3837,     Adjusted R-squared: 0.2913 
F-statistic: 4.151 on 3 and 20 DF,  p-value: 0.01936 
```

==

```
> Grain2 <- c(rep(1,6),rep(2,6),rep(3,6),rep(4,6))
> results4 <- glm(Thiamin ~ factor(Grain2))
> summary(aov(results4))
               Df Sum Sq Mean Sq F value  Pr(>F)  
factor(Grain2)  3  9.240   3.080  4.1509 0.01936 *
Residuals      20 14.840   0.742                  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
```