MATH 204: PRINCIPLES OF STATISTICS 2

WINTER 2013

MATH 204: PRINCIPLES OF STATISTICS 2

WINTER 2013

Instructor :	David A. Stephens (Burnside 1225)
Email :	d.stephens@math.mcgill.ca
Office Hours :	Monday 15:30-17:00
Tutorial :	Monday & Wednesday 14:35 - 15:25 PM, Burnside Hall 1B39
Teaching Assistant :	Orla Murphy
Textbook :	*Statistics* (10th, 11th or 12th Edition) by J. T. McClave and T. Sincich (Chapters 10-14).
Web Site :	http://www.math.mcgill.ca/dstephens/204/

SYLLABUS

1 **ANALYSIS OF VARIANCE: COMPARING MORE THAN TWO MEANS**

- 1.1 Designed Experiments
- 1.2 Randomized Designs
- 1.3 Multiple Comparison of Means
- 1.4 Randomized Block Designs
- 1.5 Factorial Experiments

2 **LINEAR REGRESSION MODELLING**

- 2.1 Simple Linear Regression
  - 2.1.1 Probability Models
  - 2.1.2 Least-Squares Fitting
  - 2.1.3 Model Assumptions
  - 2.1.4 Parameter Estimation and Testing
  - 2.1.5 The Correlation Coefficient
  - 2.1.6 Prediction
  - 2.1.7 Polynomial Regression
- 2.2 Multiple Linear Regression
  - 2.2.1 Multiple Regression Models
  - 2.2.2 Model Building and Checking
  - 2.2.3 Stepwise Model Selection
  - 2.2.4 Residual Analysis
  - 2.2.5 Pitfalls of Regression Modelling

3 **NON-PARAMETRIC STATISTICS**

- 3.1 Distribution-Free Tests
- 3.2 Single Population Tests
- 3.3 Comparing Two Populations: Independent Samples
- 3.4 Comparing Two Populations: Dependent Samples
- 3.5 Comparing Three or More Populations
- 3.6 Rank Correlation
- 3.7 Simulation-based Testing: Permutation Tests

EVALUATION

Please note that the method of evaluation for this class will be **on the following basis only**†:

Coursework Assignments	Three assignments, from week of 21st January 2013
Mid-Term	Week of 18th February 2013
Final	Closed book (with formula sheet)

Final mark for course: the larger of

$$15\ \%\ \text{Coursework} + 15\ \%\ \text{Mid-Term} + 70\ \%\ \text{Final}$$

and

$$15\ \%\ \text{Coursework} + 85\ \%\ \text{Final}$$

NOTES:

† There will no opportunity for a make-up Mid-Term if this examination is missed, and no make-up work in place of any aspect of the course assessment.

Late work may be penalized by loss of marks.

- **Academic Integrity:**

  McGill University values academic integrity. Therefore all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures (see

  `http://www.mcgill.ca/integrity/`

  for more information).

- In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.
- In the event of extraordinary circumstances beyond the University's control, the content and/or evaluation scheme in this course is subject to change.

David A. Stephens.
January 7, 2013

UNDERSTANDING THE ANOVA F-STATISTIC

Suppose that we have $k = 3$ treatment groups in a Completely Randomized Design, with sample sizes $n_1 = n_2 = n_3 = 6$. Suppose first that the treatment means are all equal to zero, that is

$$\mu_1 = \mu_2 = \mu_3 = 0$$

and that the treatment group variance parameter σ^2 is equal to 1. A typical data set is displayed below:

							$\overline{x}_i$	s_i^2
TMT 1	-0.88	0.24	-0.46	0.78	-0.47	-0.38	-0.195	0.358
TMT 2	-0.75	0.11	0.64	1.98	-1.03	1.84	0.465	1.611
TMT 3	1.38	1.20	0.42	0.05	-1.29	-0.04	0.287	0.939

yielding $\overline{x} = 0.186$, and

$$s_P^2 = \frac{1}{n-k}\sum_{i=1}^{k}(n_i - 1)s_i^2 = 0.969.$$

For these data, we have using the definitions from lectures

$$SST = \sum_{i=1}^{k} n_i(\overline{x}_i - \overline{x})^2 = 1.399 \qquad SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)^2 = 14.539$$

and

$$SS = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x})^2 = 15.938$$

so that the equation $SS = SST + SSE$ holds. For the F-statistic, we have

$$F = \frac{MST}{MSE} = \frac{SST/(k-1)}{SSE/(n-k)} = \frac{1.399/2}{14.539/15} = 0.722$$

To complete the test, we compare this with the $1 - \alpha$ probability point of the Fisher-F distribution with $(k - 1, n - k) = (2, 15)$ degrees of freedom. With $\alpha = 0.05$, from the tables in McClave and Sincich, we see that

$$F_\alpha(2, 15) = 3.68$$

and we **do not reject** the ANOVA F-test null hypothesis

$$H_0 \quad : \quad \mu_1 = \mu_2 = \mu_3.$$

This is the **correct** conclusion, as in fact all the true treatment means are zero. Thus a **small** value of the test statistic F supports H_0.

Now suppose that, in fact,

$$\mu_1 = 0 \qquad \mu_2 = 10 \qquad \mu_3 = 20.$$

The equivalent data set to the one above but with the treatment means changed in this way takes the form

							$\overline{x}_i$	s_i^2
TMT 1	-0.88	0.24	-0.46	0.78	-0.47	-0.38	-0.195	0.358
TMT 2	9.25	10.11	10.64	11.98	8.97	11.84	10.465	1.611
TMT 3	21.38	21.20	20.42	20.05	18.71	19.96	20.287	0.939

yielding $\overline{x} = 10.186$, and

$$s_P^2 = \frac{1}{n-k}\sum_{i=1}^{k}(n_i - 1)s_i^2 = 0.969.$$

Note that the sample means have changed accordingly, but that the sample variances **have not changed at all**. On further calculation, we have

$$SST = 1259.199 \qquad SSE = 14.539 \qquad SS = 1273.738$$

so that

$$MST = \frac{1259.199}{2} = 629.600 \qquad MSE = \frac{14.539}{15} = 0.969$$

so that

$$F = \frac{629.600}{0.969} = 649.570.$$

We again compare this with $F_\alpha(2, 15) = 3.68$ (the critical value, C_R), and notice that the F statistic is **much larger** than this critical value. The test statistic thus lies within the rejection region, and hence we **reject $\mathbf{H_0}$**.

This example illustrates that SST measures the variability **between** means across the treatment groups, whereas SSE measures the variability **within** treatment groups, allowing for the possibility that the treatment means may be different. The quantity SS measures the total amount of variability; in the first example $SS = SST + SSE$ gives

$$15.938 = 1.399 + 14.539$$

so most of the variability is contributed by SSE, whereas in the second example, we have

$$1273.738 = 1259.199 + 14.539$$

and most of the variability is contributed by SST.

USING THE FISHER-F TABLES

Tables in McClave and Sincich contain information on the $1-\alpha$ probability points for the Fisher-F distribution for $\alpha = 0.1, 0.05, 0.025$ and 0.01 respectively, and for different values of the **degrees of freedom** parameters. The values in the body of the table are the numbers x which solve the equation

$$\Pr[F > x] = \alpha$$

when the statistic F has a Fisher-F distribution with ν_1 and ν_2 degrees of freedom, written

$$F \sim \text{Fisher-F}(\nu_1, \nu_2)$$

where ν_1 and ν_2 are whole numbers greater than zero.

The table on the reverse of this sheet is the Fisher-F table for $\alpha = 0.05$, equivalent to the table of McClave and Sincich. We read ν_1 from the **column** and ν_2 from the **row**. For example,

- if $\nu_1 = 10$ and $\nu_2 = 4$, we know from the table that

$$\Pr[\, F > \mathbf{5.96} \,] = 0.05$$

- if $\nu_1 = 6$ and $\nu_2 = 18$, we know from the table that

$$\Pr[\, F > \mathbf{2.66} \,] = 0.05$$

- if $\nu_1 = 20$ and $\nu_2 = 20$, we know from the table that

$$\Pr[\, F > \mathbf{2.12} \,] = 0.05$$

The Fisher-F distribution is a non-symmetric probability distribution with a specific property that allows the tables in McClave and Sincich to tabulate only the **right-hand tail** of the distribution. If we need to look up the left-hand tail, we can use the fact that if $F \sim \text{Fisher-F}(\nu_1, \nu_2)$, and $0 < p < 1$

$$\Pr[\, F > x \,] = p \qquad \Longrightarrow \qquad \Pr[\, 1/F \leq 1/x \,] = p$$

so that

$$\Pr[\, 1/F > 1/x \,] = 1 - p.$$

But it transpires that

$$F \sim \text{Fisher-F}(\nu_1, \nu_2) \qquad \Longrightarrow \qquad \frac{1}{F} \sim \text{Fisher-F}(\nu_2, \nu_1).$$

Therefore to look up the **left-tail** α probability point for the $F \sim \text{Fisher-F}(\nu_1, \nu_2)$ distribution, we look up the **right-tail** $1-\alpha$ probability point for the Fisher-F(ν_2, ν_1) distribution, and then take the reciprocal. For example,

- if $F \sim \text{Fisher-F}(10, 4)$, we use tables to discover that as

$$F_{0.05}(4, 10) = 3.48$$

it follows that

$$\Pr[\, F \leq 1/\mathbf{3.48} \,] = \Pr[\, F \leq \mathbf{0.29} \,] = 0.05$$

giving the $\alpha = 0.05$ (left-tail) probability point of the Fisher-F$(10, 4)$ distribution as 0.29.

ONE-WAY ANOVA: WORKED EXAMPLE

A standard model of memory is that the degree to which the subject remembers verbal material is a function of the degree to which it was processed when it was initially presented.

Reference: Craik, F. I. M. and Lockhart, R. S. (1972). Levels of Processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.

Experiment: Fifty subjects aged between 55 and 65 years were randomly assigned to one of five groups which carried out different memory tasks. The five groups included

- The Counting group was asked to read through a list of words and simply count the number of letters in each word.
- The Rhyming group was asked to read each word and think of a word that rhymed with it.
- The Adjective group had to process the words to the extent of giving an adjective that could reasonably be used to modify each word on the list.
- The Imagery group was instructed to try to form vivid images of each word.
- The Intentional group was told to read through the list and to memorize the words for later recall.

After subjects had gone through the list of 27 items three times, they were given a sheet of paper and asked to write down all the words they could remember. The response data were the number of words recalled by each individual in each group, and are presented below:

Counting	Rhyming	Adjective	Imagery	Intentional
9	7	11	12	10
8	9	13	11	19
6	6	8	16	14
8	6	6	11	5
10	6	14	9	10
4	11	11	23	11
6	6	13	12	14
5	3	13	10	15
7	8	10	19	11
7	7	11	11	11

Research question: Does the level of processing required when material is processed affect how much material is remembered ?

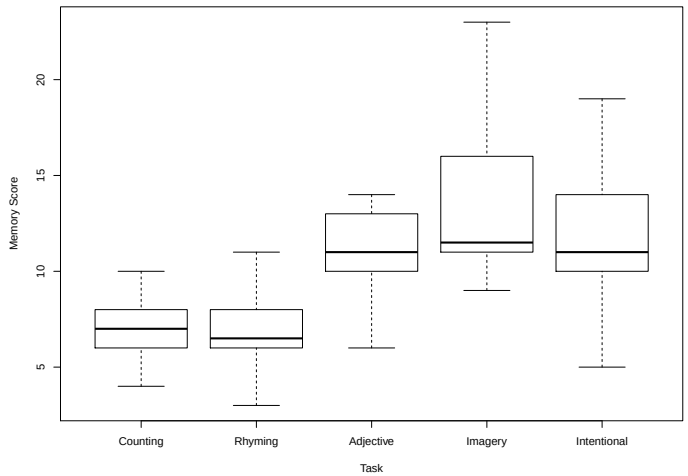


Figure 1: `Memory` data: Boxplot of number of errors made by memory task

Listing 1: Memory task data: Levene's Test and ANOVA

```
1  > leveneTest(Memory ~ Task,center='mean',data=Memory)
2  Levene's Test for Homogeneity of Variance (center = "mean")
3        Df F value  Pr(>F)
4  group  4  2.5287 0.05356 .
5        45
6  ---
7  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
8
9  > leveneTest(Memory ~ Task,center='median',data=Memory)
10 Levene's Test for Homogeneity of Variance (center = "median")
11       Df F value Pr(>F)
12 group  4  0.8932 0.4759
13       45
14
15 > summary(aov(Memory ~ Task,data=Memory))
16             Df Sum Sq Mean Sq F value   Pr(>F)
17 Task         4  351.5   87.88   9.085 1.82e-05 ***
18 Residuals   45  435.3    9.67
19 ---
20 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
21 > summary(lm(Memory ~ Task,data=Memory))
22
23 Call:
24 lm(formula = Memory ~ Task, data = Memory)
25
26 Residuals:
27    Min     1Q Median     3Q    Max
28  -7.00  -1.85  -0.45   2.00   9.60
29
30 Coefficients:
31                 Estimate Std. Error t value Pr(>|t|)
32 (Intercept)       7.0000     0.9835   7.117 6.83e-09 ***
33 TaskRhyming      -0.1000     1.3909  -0.072 0.943004
34 TaskAdjective     4.0000     1.3909   2.876 0.006138 **
35 TaskImagery       6.4000     1.3909   4.601 3.43e-05 ***
36 TaskIntentional   5.0000     1.3909   3.595 0.000802 ***
37 ---
38 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
39
40 Residual standard error: 3.11 on 45 degrees of freedom
41 Multiple R-squared: 0.4468,     Adjusted R-squared: 0.3976
42 F-statistic: 9.085 on 4 and 45 DF,  p-value: 1.815e-05
```

Conclusions: The Fisher-F statistic in this ANOVA is given on lines 17 and 42 as 9.085. This is to be compared with the Fisher$(k-1, n-k)$ distribution; here $k-1$ is the degrees of freedom for treatment (here for factor `Task`), and $n-k$ is the residual degrees of freedom. The values are given on lines 17 and 18 in the `DF` column as 4 and 45. The result of the test is given by the p-value on lines 17 and 42, which is 1.815e-05; hence we **reject** the null hypothesis that there is no difference between treatments at the $\alpha = 0.05$ significance level. Levene's test (lines 1-13) suggests that there is no strong evidence that the group variances are different.

ANOVA FOR THE COMPLETELY RANDOMIZED DESIGN: EXAMPLES

EXAMPLE Milk protein level by Diet: data set `Milk`.

Listing 2: Milk data: Summary Statistics

```
> tapply(Milk$protein,Milk$Diet,my.summary)
$barley
    n    Mean   StDev  StdErr   Lower   Upper  Min  Max
1 425 3.53193 0.31921 0.01548 3.50158 3.56228 2.64 4.47

$`barley+lupins`
    n    Mean   StDev  StdErr   Lower   Upper Min  Max
1 459 3.42969 0.30234 0.01411 3.40204 3.45735 2.7 4.59

$lupins
    n    Mean   StDev  StdErr   Lower   Upper  Min  Max
1 453 3.31238 0.33709 0.01584 3.28134 3.34343 2.45 4.32

> my.summary(Milk$protein)
     n    Mean   StDev  StdErr   Lower   Upper  Min  Max
1 1337 3.42245 0.33175 0.00907 3.40466 3.44023 2.45 4.59
```

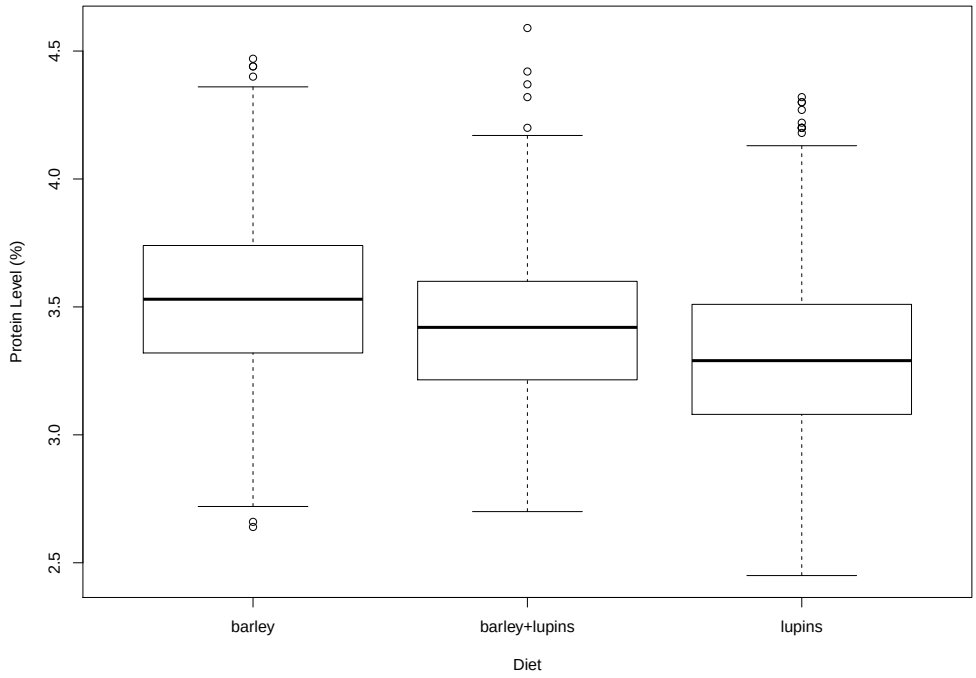


Figure 2: `Milk` data: Boxplot of Protein level by Diet

Listing 3: Milk data: Levene's Test

```
> library(car)
> leveneTest(protein ~ Diet,center='mean',data=Milk)
Levene's Test for Homogeneity of Variance (center = "mean")
        Df F value Pr(>F)
group    2  1.8377 0.1596
      1334
```

```
> leveneTest(protein~Diet,center='median',data=Milk)
Levene's Test for Homogeneity of Variance (center = "median")
        Df F value Pr(>F)
group    2  1.7344 0.1769
      1334
```

Listing 4: Milk data: ANOVA

```
> #Total
> summary(aov(protein~1,data=Milk))
              Df Sum Sq Mean Sq F value Pr(>F)
Residuals   1336    147  0.1101
> anova(lm(protein~1,data=Milk))
Analysis of Variance Table

Response: protein
            Df Sum Sq Mean Sq F value Pr(>F)
Residuals 1336 147.04 0.11006
>
> #Fit and Error (Residual)
> summary(aov(protein~Diet,data=Milk))
              Df Sum Sq Mean Sq F value Pr(>F)
Diet           2  10.61   5.303   51.85 <2e-16 ***
Residuals   1334 136.43   0.102
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(lm(protein~Diet,data=Milk))
Analysis of Variance Table

Response: protein
            Df  Sum Sq Mean Sq F value    Pr(>F)
Diet         2  10.606  5.3030  51.851 < 2.2e-16 ***
Residuals 1334 136.432  0.1023
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(lm(protein~Diet,data=Milk))

Call: lm(formula = protein ~ Diet, data = Milk)

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        3.53193    0.01551 227.681  < 2e-16 ***
Dietbarley+lupins -0.10223    0.02153  -4.749 2.27e-06 ***
Dietlupins        -0.21955    0.02160 -10.166  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3198 on 1334 degrees of freedom
Multiple R-squared: 0.07213,    Adjusted R-squared: 0.07074
F-statistic: 51.85 on 2 and 1334 DF,  p-value: < 2.2e-16
```

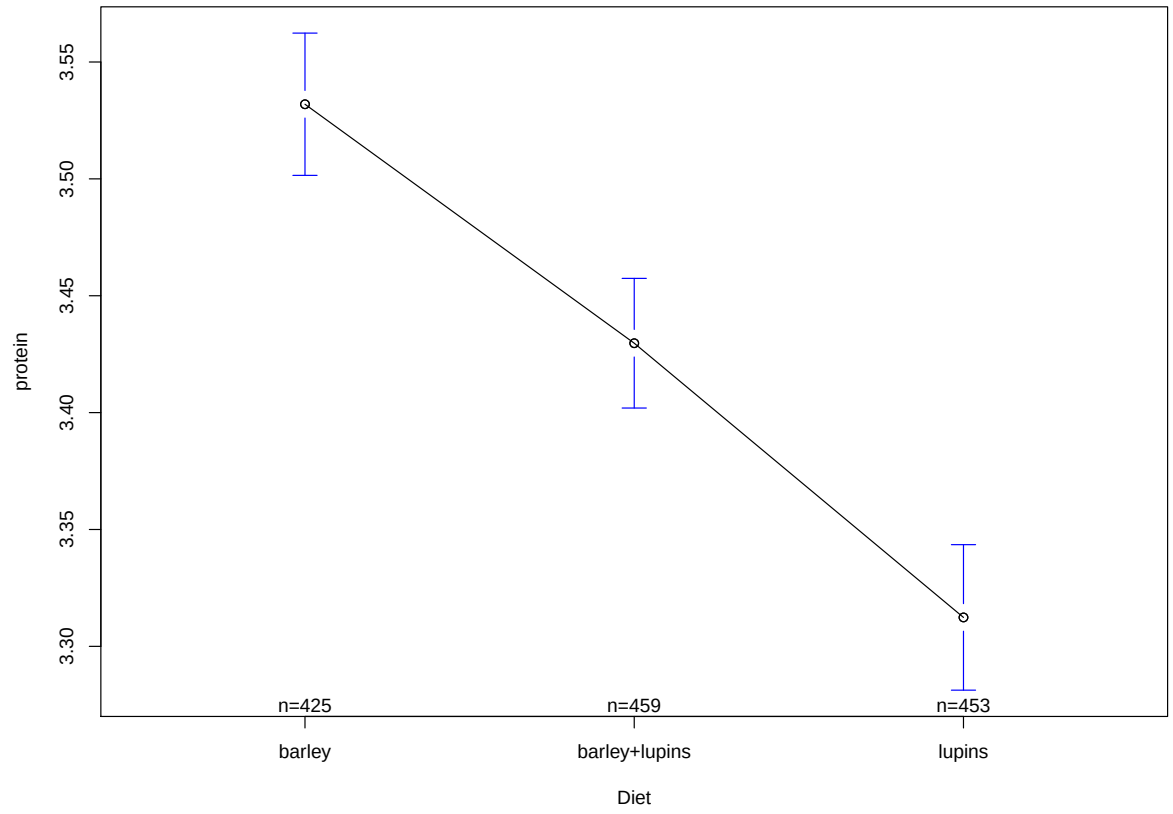


Figure 3: `Milk` data: Means plot of Protein level by Diet, with 95 % confidence intervals for means

EXAMPLE Anxiety level by Dose Level: data set `Pull`

Listing 5: Anxiety data: Summary

```
> tapply(Pull$Pull,Pull$Dose,my.summary)
$`0Unit`
  n  Mean   StDev  StdErr    Lower    Upper  Min  Max
1 5 28.86 2.83161 1.26633 26.37799 31.34201 26.2 33.5

$`1Unit`
  n  Mean   StDev  StdErr    Lower    Upper  Min  Max
1 5 25.04 2.42343 1.08379 22.91577 27.16423 22.8 27.7

$`2Unit`
  n Mean   StDev  StdErr    Lower    Upper  Min  Max
1 5 22.5 3.36378 1.50433 19.55152 25.44848 19.3 27.8

$`3Unit`
  n Mean  StDev  StdErr    Lower    Upper  Min  Max
1 5 22.3 1.9685 0.88034 20.57453 24.02547 19.6 23.9

> my.summary(Pull$Pull)
   n   Mean   StDev  StdErr    Lower    Upper  Min  Max
1 20 24.675 3.67364 0.82145 23.06495 26.28505 19.3 33.5
```

Listing 6: Anxiety data: Levene's Test

```
> #Levene's Test for Equality of variances
>
> require(car)
> leveneTest(Pull ~ Dose,center='mean',data=Pull)
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value Pr(>F)
group  3  0.2948 0.8286
      16
> leveneTest(Pull ~ Dose,center='median',data=Pull)
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  3  0.1955 0.8979
      16
```

Listing 7: Anxiety data: ANOVA

```

> #Fit and Error (Residual)
> summary(aov(Pull ~ Dose,data=Pull))
            Df Sum Sq Mean Sq F value  Pr(>F)
Dose         3  140.1   46.70   6.423 0.00462 **
Residuals   16  116.3    7.27
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(lm(Pull ~ Dose,data=Pull))
Analysis of Variance Table

Response: Pull
          Df Sum Sq Mean Sq F value   Pr(>F)
Dose       3 140.09  46.698  6.4231 0.004623 **
Residuals 16 116.32   7.270
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(lm(Pull ~ Dose,data=Pull))

Call:lm(formula = Pull ~ Dose, data = Pull)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   28.860      1.206  23.934 5.91e-14 ***
Dose1Unit     -3.820      1.705  -2.240  0.03964 *
Dose2Unit     -6.360      1.705  -3.730  0.00183 **
Dose3Unit     -6.560      1.705  -3.847  0.00142 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.696 on 16 degrees of freedom
Multiple R-squared: 0.5463,     Adjusted R-squared: 0.4613
F-statistic: 6.423 on 3 and 16 DF,  p-value: 0.004623
```

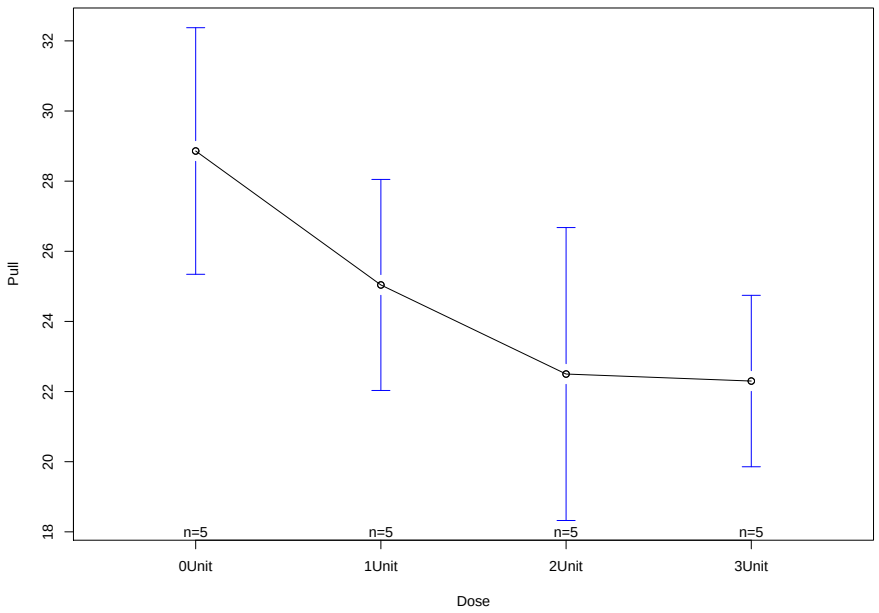


Figure 4: `Pull` data: Means plot of Anxiety level by Dose level, with 95 % confidence intervals for means

Listing 8: Anxiety data: Post-hoc tests

```
> summary(aov(Pull ~ Dose,data=Pull))
            Df Sum Sq Mean Sq F value  Pr(>F)
Dose         3  140.1   46.70   6.423 0.00462 **
Residuals   16  116.3    7.27
---

> pairwise.t.test(Pull$Pull, Pull$Dose, p.adj = "none")

        Pairwise comparisons using t tests with pooled SD

data:  Pull$Pull and Pull$Dose

      0Unit  1Unit  2Unit
1Unit 0.0396 -      -
2Unit 0.0018 0.1558 -
3Unit 0.0014 0.1277 0.9081

P value adjustment method: none

> pairwise.t.test(Pull$Pull, Pull$Dose, p.adj = "bonferroni")

        Pairwise comparisons using t tests with pooled SD

data:  Pull$Pull and Pull$Dose

      0Unit  1Unit  2Unit
1Unit 0.2378 -      -
2Unit 0.0110 0.9349 -
3Unit 0.0085 0.7660 1.0000

P value adjustment method: bonferroni
```

Listing 9: Anxiety data: Post-hoc tests using Tukey's method

```
> TukeyHSD(aov(Pull ~ Dose,data=Pull))
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Pull ~ Dose, data = Pull)

$Dose
              diff        lwr       upr     p adj
1Unit-0Unit -3.82  -8.698941  1.058941 0.1546383
2Unit-0Unit -6.36 -11.238941 -1.481059 0.0088837
3Unit-0Unit -6.56 -11.438941 -1.681059 0.0069985
2Unit-1Unit -2.54  -7.418941  2.338941 0.4661560
3Unit-1Unit -2.74  -7.618941  2.138941 0.4026925
3Unit-2Unit -0.20  -5.078941  4.678941 0.9993976
```

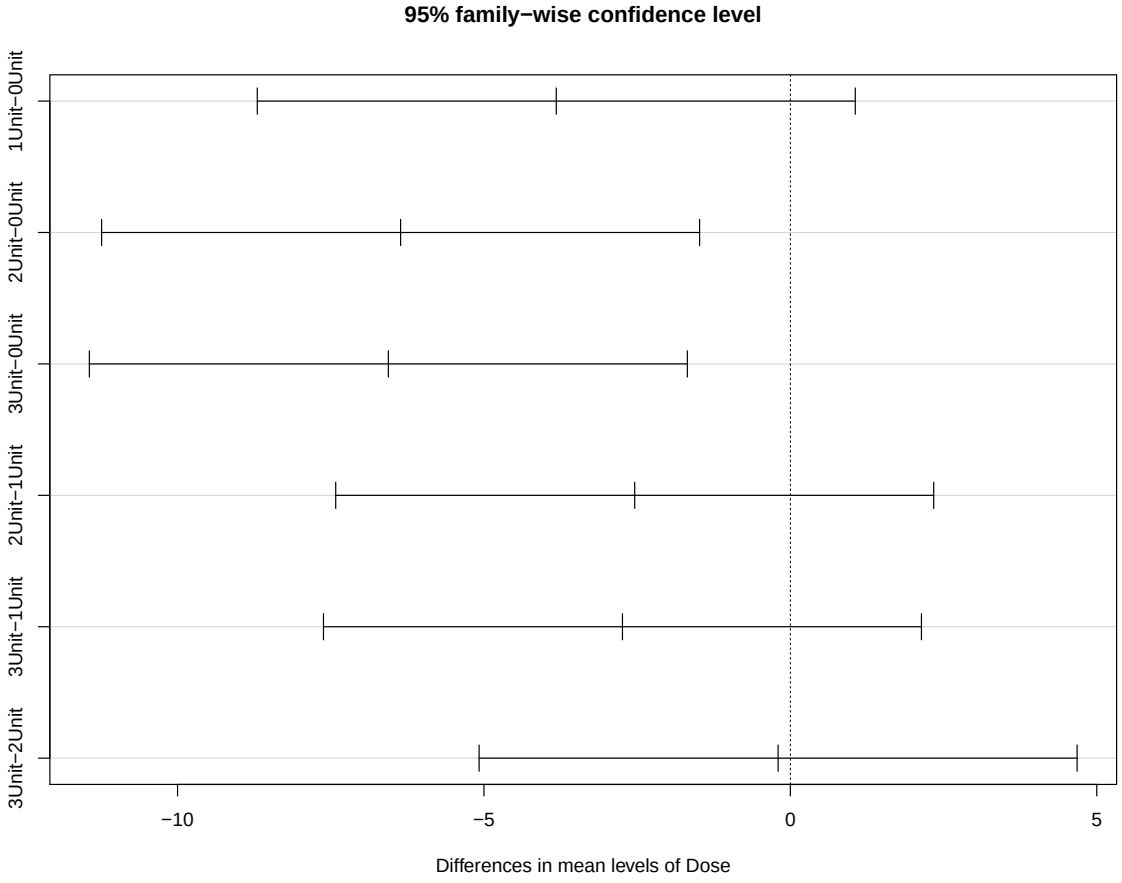


Figure 5: `Pull` data: All pairwise comparisons of dose levels, with adjusted 95 % confidence intervals for means

EXAMPLE Gut permeability by AIDS diagnosis: data set `AIDS`

Listing 10: Gut permeability data: Levene's Test and ANOVA

```
> leveneTest(mann ~ diag,center='mean',data=AIDS)
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value Pr(>F)
group  3  0.8661 0.4643
      55
> leveneTest(mann ~ diag,center='median',data=AIDS)
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  3   0.733 0.5368
      55

> summary(aov(mann ~ diag,data=AIDS))
             Df Sum Sq Mean Sq F value Pr(>F)
diag          3     49   16.34   0.595  0.621
Residuals    55   1510   27.45

> summary(lm(mann ~ diag,data=AIDS))

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.33119    1.02760  11.027  1.5e-15 ***
diagARC     -2.48934    2.23117  -1.116    0.269
diagHIV+    -1.62076    2.23117  -0.726    0.471
diagHIV-     0.06581    1.58144   0.042    0.967
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.24 on 55 degrees of freedom
Multiple R-squared: 0.03144,    Adjusted R-squared: -0.02139
F-statistic: 0.595 on 3 and 55 DF,  p-value: 0.6209
```

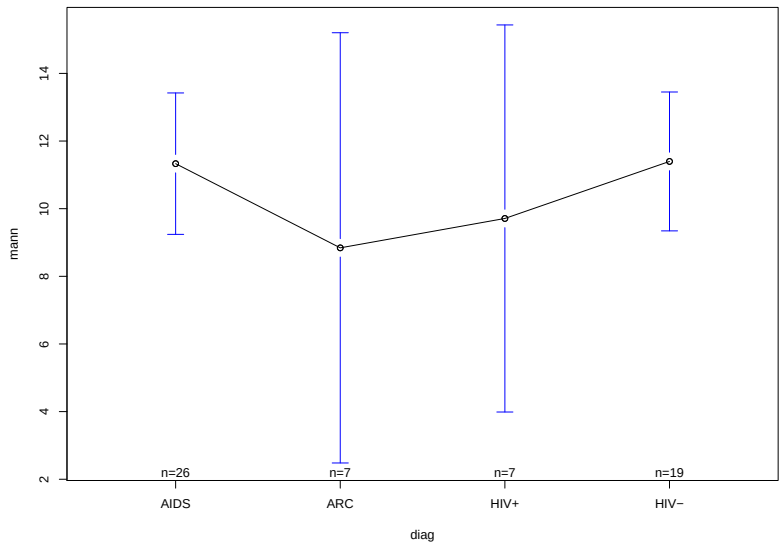


Figure 6: `AIDS` data: Means plot of gut permeability of mannitol by diagnosis, with 95 % confidence intervals for means

EXAMPLE Bacteria level in batches of milk: data set `MilkBac`

Listing 11: Milk bacteria level data: Levene's Test and ANOVA

```
> leveneTest(BacLevel ~ Batch,center='mean',data=MilkBac)
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value Pr(>F)
group  4  0.3843 0.8177
      25

> leveneTest(BacLevel ~ Batch,center='median',data=MilkBac)
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  4  0.2426 0.9114
      25

> summary(aov(BacLevel ~ Batch,data=MilkBac))
            Df Sum Sq Mean Sq F value  Pr(>F)
Batch        4  803.0  200.75   9.008 0.00012 ***
Residuals   25  557.2   22.29
---

> summary(lm(BacLevel ~ Batch,data=MilkBac))

            Estimate Std. Error t value Pr(>|t|)
(Intercept)   23.833      1.927  12.366 3.76e-12 ***
Batch2       -10.500      2.726  -3.852 0.000723 ***
Batch3       -12.167      2.726  -4.464 0.000150 ***
Batch4       -14.667      2.726  -5.381 1.39e-05 ***
Batch5        -6.000      2.726  -2.201 0.037162 *
---
Residual standard error: 4.721 on 25 degrees of freedom
Multiple R-squared: 0.5904,     Adjusted R-squared: 0.5248
F-statistic: 9.008 on 4 and 25 DF,  p-value: 0.0001197
```

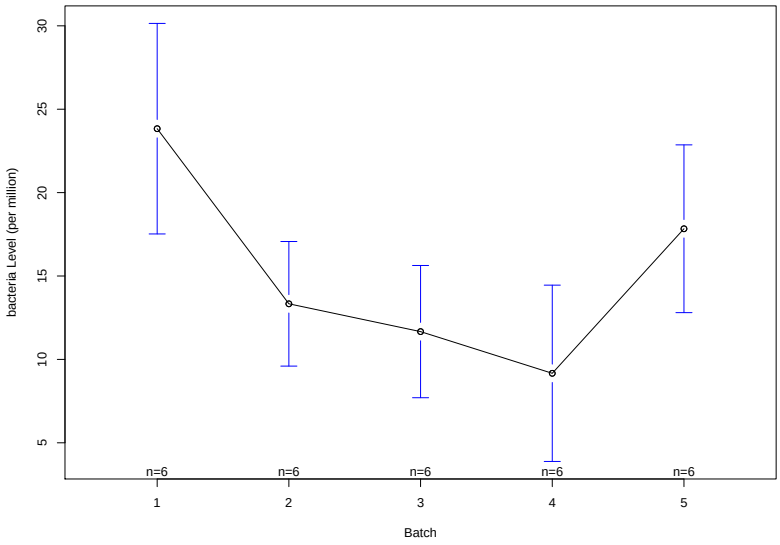


Figure 7: `MilkBac` data: Means plot of bacteria level by batch, with 95 % confidence intervals for means

EXAMPLE Post-traumatic stress disorder: data set `PTSD`
Data Source: Foa, E. B., Rothbaum, B. O., Riggs, D. S., & Murdock, T. B. (1991) Treatment of post traumatic stress disorder in rape victims: A comparison between cognitive-behavioral procedures and counseling. *Journal of Consulting and Clinical Psychology*, 59, 715-723.

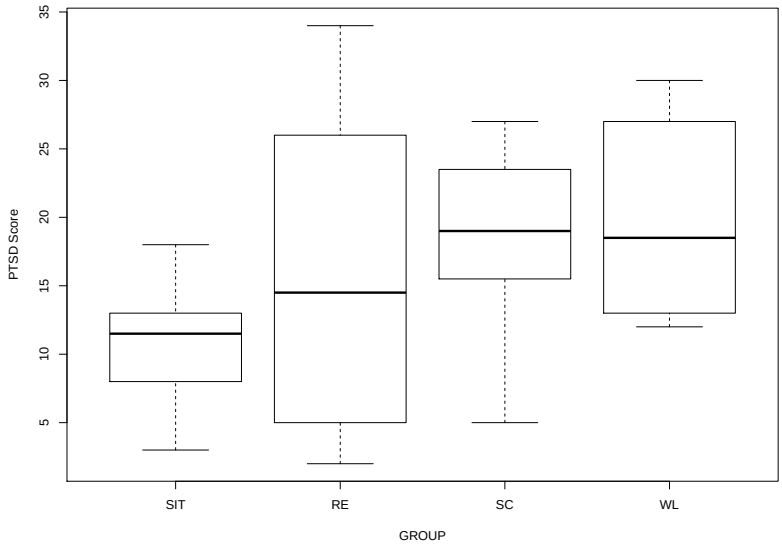


Figure 8: `PTSD` data: Boxplot of PTSD score by therapy

Listing 12: PTSD data: Levene's Test and ANOVA

```
> leveneTest(PTSD ~ GROUP,center='mean',data=PTSD)
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value   Pr(>F)
group  3   6.633 0.000934 ***
      41

> leveneTest(PTSD ~ GROUP,center='median',data=PTSD)
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value   Pr(>F)
group  3  6.1713 0.001465 **
      41

> summary(aov(PTSD ~ GROUP,data=PTSD))
            Df Sum Sq Mean Sq F value Pr(>F)
GROUP        3  507.8  169.28   3.046 0.0394 *
Residuals   41 2278.7   55.58

> summary(lm(PTSD ~ GROUP,data=PTSD))
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   11.071      1.992   5.557 1.84e-06 ***
GROUPRE        4.329      3.087   1.402  0.16835
GROUPSC        7.019      3.004   2.337  0.02441 *
GROUPWL        8.429      3.087   2.731  0.00928 **
---
Residual standard error: 7.455 on 41 degrees of freedom
Multiple R-squared: 0.1822,     Adjusted R-squared: 0.1224
F-statistic: 3.046 on 3 and 41 DF,  p-value: 0.03936
```

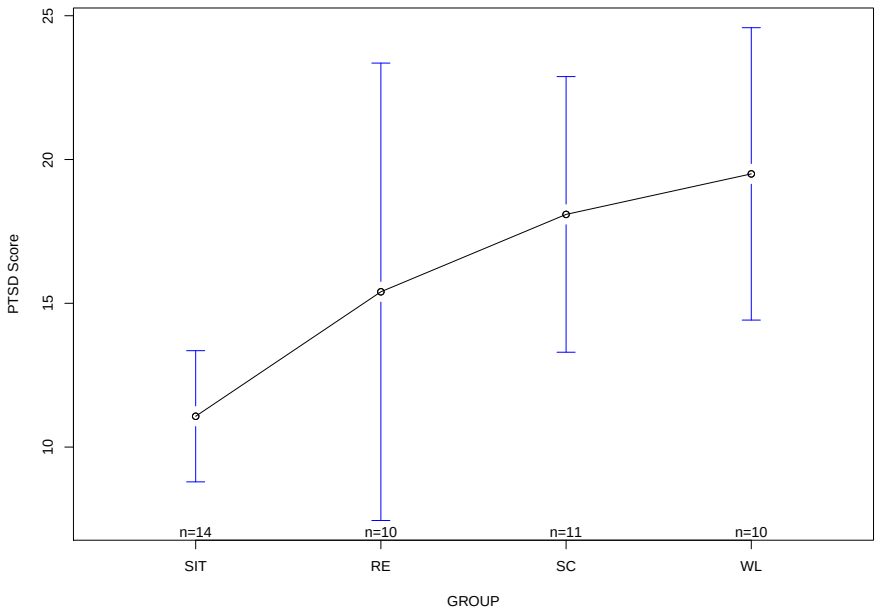


Figure 9: `PTSD` data: Means plot of PTSD score by therapy, with 95 % confidence intervals for means

Listing 13: PTSD data: Post-hoc tests

```
> summary(aov(PTSD ~ GROUP, data=PTSD))
             Df Sum Sq Mean Sq F value Pr(>F)
GROUP         3  507.8  169.28   3.046 0.0394 *
Residuals    41 2278.7   55.58

> pairwise.t.test(PTSD$PTSD, PTSD$GROUP, p.adj = "none")

        Pairwise comparisons using t tests with pooled SD

data:  PTSD$PTSD and PTSD$GROUP

   SIT    RE     SC
RE 0.1683 -      -
SC 0.0244 0.4135 -
WL 0.0093 0.2258 0.6676

P value adjustment method: none

> pairwise.t.test(PTSD$PTSD, PTSD$GROUP, p.adj = "bonferroni")

        Pairwise comparisons using t tests with pooled SD

data:  PTSD$PTSD and PTSD$GROUP

   SIT   RE    SC
RE 1.000 -     -
SC 0.146 1.000 -
WL 0.056 1.000 1.000

P value adjustment method: bonferroni
```

Listing 14: PTSD data: Post-hoc tests using Tukey's method

```
> TukeyHSD(aov(PTSD ~ GROUP,data=PTSD))
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = PTSD ~ GROUP, data = PTSD)

$GROUP
            diff        lwr      upr     p adj
RE-SIT 4.328571 -3.9364832 12.59363 0.5051618
SC-SIT 7.019481 -1.0234398 15.06240 0.1062481
WL-SIT 8.428571  0.1635168 16.69363 0.0441454
SC-RE  2.690909 -6.0311179 11.41294 0.8417772
WL-RE  4.100000 -4.8272793 13.02728 0.6118884
WL-SC  1.409091 -7.3129361 10.13112 0.9725237
```

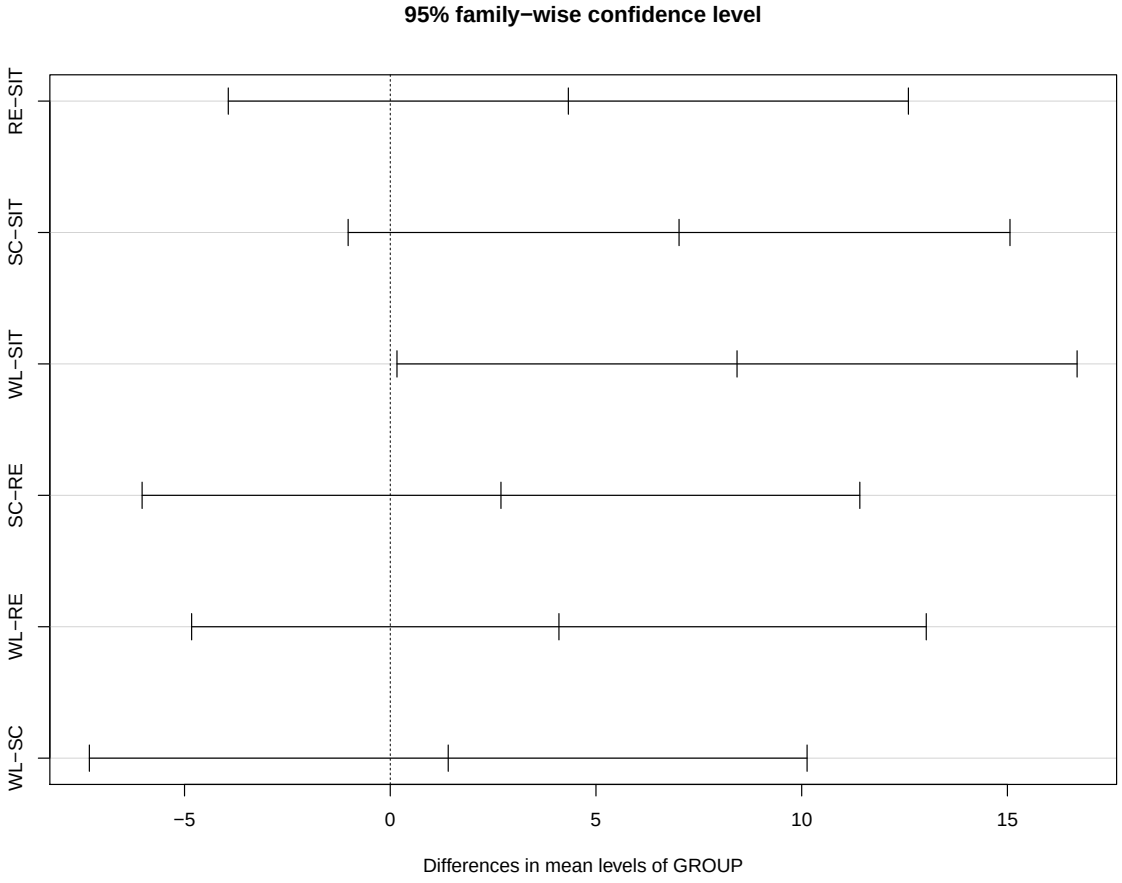


Figure 10: `PTSD` data: All pairwise comparisons of dose levels, with adjusted 95 % confidence intervals for means

RANDOMIZED BLOCK DESIGNS AND THE ANOVA F-TEST

Consider a **randomized block design** (RBD) with k treatments and b blocks. Assume that each block has k experimental units, and that one unit is assigned to each treatment. Let x_{ij} be the measured response for the experimental unit from block j in treatment i and

- sample mean for **treatment** i

$$\overline{x}_i = \frac{1}{b}\sum_{j=1}^{b} x_{ij} \qquad i = 1, \ldots, k$$

- sample mean for block j

$$\overline{x_j^{(\mathrm{B})}} = \frac{1}{k}\sum_{i=1}^{k} x_{ij} \qquad j = 1, \ldots, b$$

- overall sample mean

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{b} x_{ij}$$

- Sum of Squares for Treatments (SST)

$$SST = \sum_{i=1}^{k} b(\overline{x}_i - \overline{x})^2$$

- Sum of Squares for Blocks (SSB)

$$SSB = \sum_{j=1}^{b} k(\overline{x_j^{(\mathrm{B})}} - \overline{x})^2$$

- Overall Sum of Squares (SS)

$$SS = \sum_{i=1}^{k}\sum_{j=1}^{b}(x_{ij} - \overline{x})^2$$

The following decomposition holds

$$SS = SST + SSB + SSE \qquad \therefore \qquad SSE = SS - SST - SSB$$

For testing

$$\begin{array}{lcl} H_0 & : & \mu_1 = \cdots = \mu_k \\ H_a & : & \text{At least two treatment means different} \end{array}$$

in an RBD, the test statistic is

$$F = \frac{MST}{MSE}$$

where

$$MST = \frac{SST}{k-1} \qquad MSE = \frac{SSE}{n-b-k+1}$$

If H_0 is **true**, then $F \sim$ Fisher-F$(k-1, n-b-k+1)$, and the rejection region for the test with significance level α is

$$F > F_\alpha(k-1, n-b-k+1)$$

where $F_\alpha(\nu_1, \nu_2)$ is the $1-\alpha$ percentage point of the Fisher-F distribution with ν_1 and ν_2 degrees of freedom.

EXAMPLE

Data: Measurements were made on the amount of sulphur (in parts per million) in soil samples using four different solvents. The soil samples were collected from five different geographical locations in Florida, USA, and represented different soil types.

The **response variable** sulphur level. The single **factor** is the *solvent* and there are $k = 4$ **factor levels**:

1. Calcium Chloride ($CaCl_2$)
2. Ammonium Acetate (NH_4OAc)
3. Mono-Calcium Phosphate ($Ca(H_2P\,O_4)_3$)
4. Water (H_2O)

The *soil* types determine the $b = 5$ **blocks**

1. Troup, Jackson Co. (*Paleudults* soil)
2. Lakeland, Walton Co. (*Quartzipsamments* soil)
3. Leon, Duval Co. (*Haplaquads* soil)
4. Chipley, Jackson Co. (*Quartzipsamments* soil)
5. Norfolk, Alachua Co. (*Paleudults* soil)

The data observed in the study were as follows:

	Block				
Treatment	Troup	Lakeland	Leon	Chipley	Norfolk
$CaCl_2$	5.07	3.31	2.54	2.34	4.71
NH_4OAc	4.43	2.74	2.09	2.07	5.29
$Ca(H_2P\,O_4)_3$	7.09	2.32	1.09	4.38	5.70
H_2O	4.48	2.35	2.70	3.85	4.98

Listing 15: Soil data: ANOVA

```
1 > summary(aov(sulphur ~ solvent+soil,data=Soil))
2             Df Sum Sq Mean Sq F value   Pr(>F)
3 solvent      3   1.62   0.540   0.673 0.585130
4 soil         4  33.96   8.491  10.568 0.000663 ***
5 Residuals   12   9.64   0.803
6 ---
7 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The this is the standard ANOVA table for the randomized block design. As expected, there is a significant difference between **blocks** (line 4, $F = 10.568$, p-value=0.001), but **no significant difference** between **treatments** (line 3, $F = 0.673$, p-value=0.585).

THE NEED FOR BLOCKING IN AN RBD ANALYSIS

Consider the following response data: five measurements collected in three treatment groups:

—	1	2	3	4	5
Group 1	-20.88	-4.76	-0.46	10.78	-10.47
Group 2	-15.75	0.11	5.64	16.98	-6.03
Group 3	-8.62	6.20	10.42	20.05	-1.29

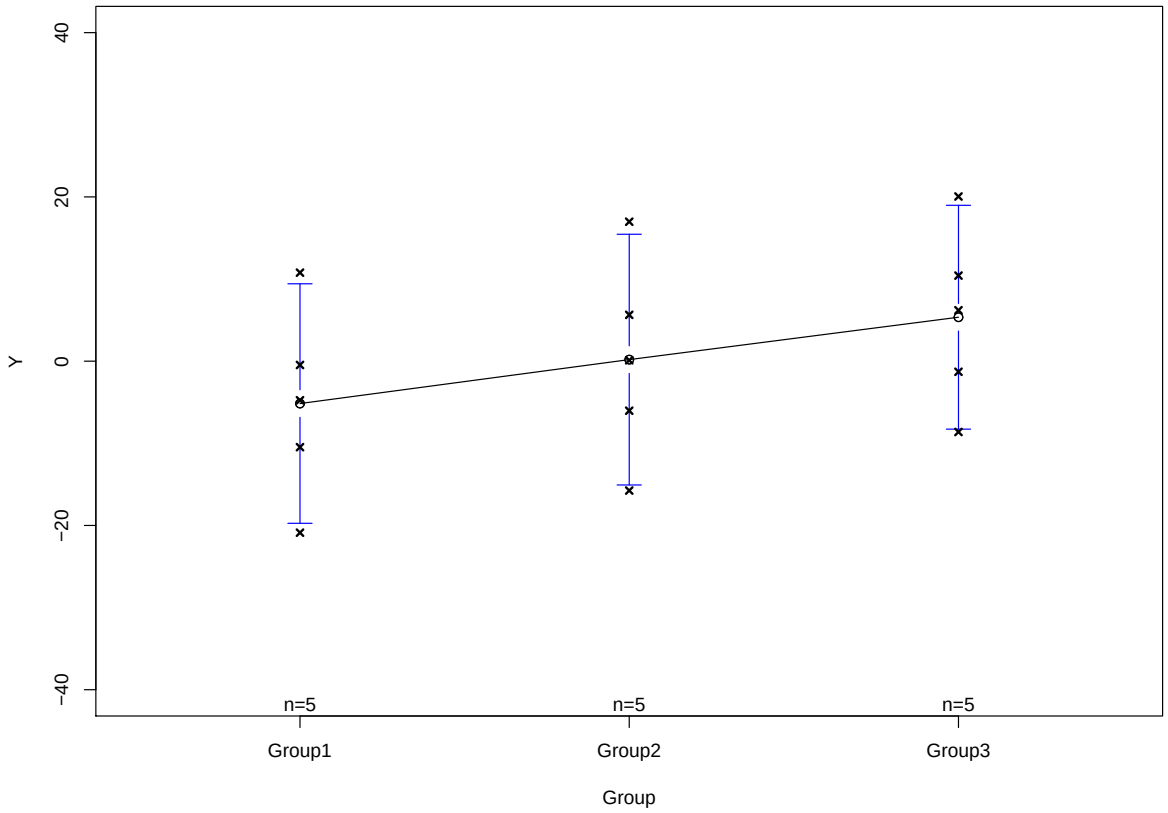


Figure 11: RBD Example: Raw data

Listing 16: RBD data: ANOVA ignoring blocking

```
> Y<-c(-20.88,-4.76,-0.46,10.78,-10.47,
+ -15.75,0.11,5.64,16.98,-6.03,
+ -8.62,6.20,10.42,20.05,-1.29)
>
> Group<-gl(3,5,labels=c("Group1","Group2","Group3"))
> Block<-gl(5,1,15,labels=c(1:5))

> summary(aov(Y~Group))
            Df Sum Sq Mean Sq F value Pr(>F)
Group        2  276.2   138.1   1.012  0.393
Residuals   12 1637.8   136.5
```

It seems from this analysis that there is no difference between groups; however, this is not the case if the block structure is taken into account.

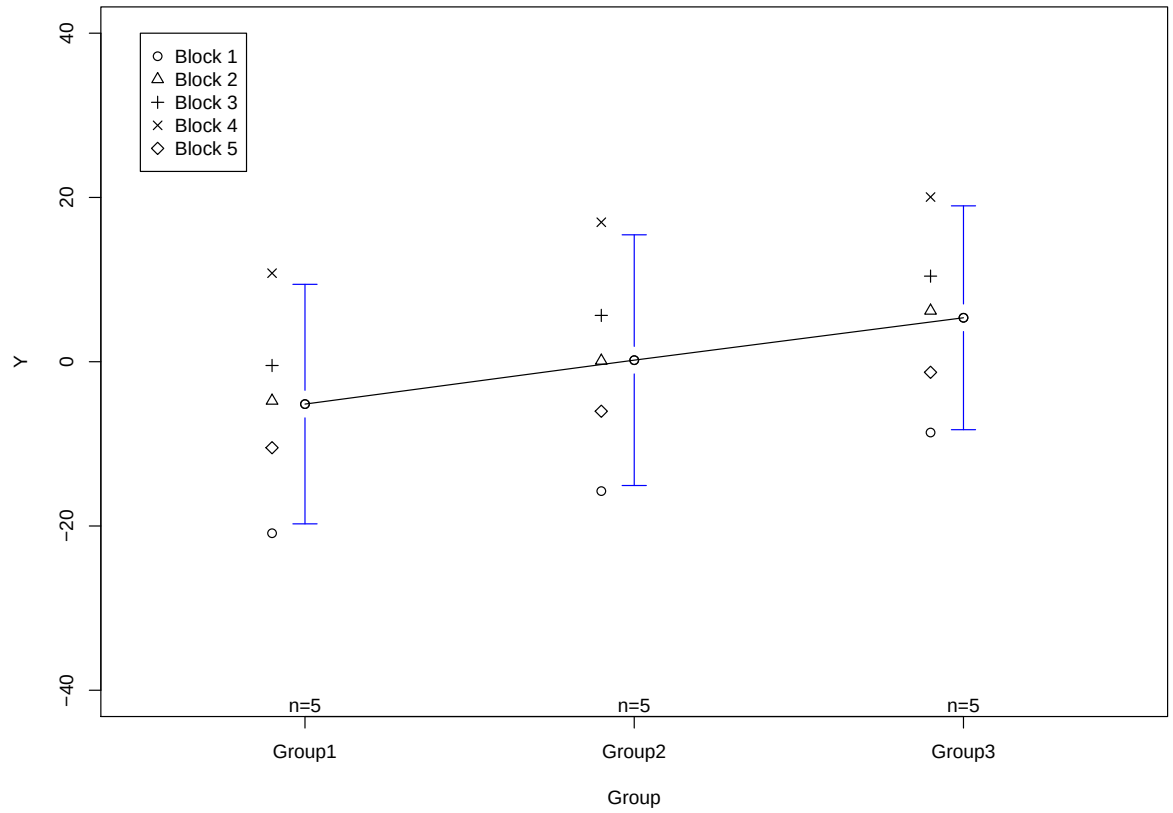


Figure 12: RBD Example: Raw data broken down by block

Listing 17: RBD data: ANOVA with blocking

```
> summary(aov(Y~Group+Block))
            Df Sum Sq Mean Sq F value   Pr(>F)
Group        2  276.2   138.1   178.8 2.29e-07 ***
Block        4 1631.6   407.9   528.2 1.01e-09 ***
Residuals    8    6.2     0.8
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In fact, there is hidden structure in the data. If this structure is taken into account, evidence that the treatment means are significantly different is uncovered. The reason that the CRD and one-way ANOVA do not discover this is that they assume that the variability can be decomposed as

$$SS = SST + SSE$$

whereas in fact

$$SS = SST + SSB + SSE$$

that is, the CRD assumes that the random variability that is observed is MUCH LARGER than it actually is. Once the variation due to BLOCKS is taken into account, the ANOVA-F test result for TREATMENTS becomes significant.

RANDOMIZED COMPLETE BLOCK DESIGNS WITH BALANCED REPLICATION

Consider a **randomized block design** (RBD) with k treatments and b blocks, and r **replications**, giving $n = rbk$ observations in total. Let x_{ijt} be the tth replicated observation in the (i,j)th treatment/block combination.

- sample mean for **treatment** i

$$\overline{x}_i = \frac{1}{br}\sum_{j=1}^{b}\sum_{t=1}^{r} x_{ijt} \qquad i = 1, \ldots, k$$

- sample mean for **block** j

$$\overline{x_j^{(\mathrm{B})}} = \frac{1}{kr}\sum_{i=1}^{k}\sum_{t=1}^{r} x_{ijt} \qquad j = 1, \ldots, b$$

- sample mean for replicates in (i,j)th **treatment/block** combination

$$\overline{x}_{ij} = \frac{1}{r}\sum_{t=1}^{r} x_{ijt} \qquad i = 1, \ldots, k,\ j = 1, \ldots, b$$

- overall sample mean

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{b}\sum_{t=1}^{r} x_{ijt}$$

- Sum of Squares for Treatments (SST)

$$SST = \sum_{i=1}^{k} br(\overline{x}_i - \overline{x})^2$$

- Sum of Squares for Blocks (SSB)

$$SSB = \sum_{j=1}^{b} kr(\overline{x_j^{(\mathrm{B})}} - \overline{x})^2$$

- Sum of Squares for Interaction (SSI)

$$SSI = \sum_{i=1}^{k}\sum_{j=1}^{b} r(\overline{x}_{ij} - \overline{x}_i - \overline{x_j^{(\mathrm{B})}} + \overline{x})^2$$

- Overall Sum of Squares (SS)

$$SS = \sum_{i=1}^{k}\sum_{j=1}^{b}\sum_{t=1}^{r}(x_{ijt} - \overline{x})^2$$

The following decomposition holds

$$SS = SST + SSB + SSI + SSE \qquad \therefore \qquad SSE = SS - SST - SSB - SSI$$

Define

$$MST = \frac{SST}{k-1} \qquad MSB = \frac{SSB}{b-1} \qquad MSI = \frac{SSI}{(k-1)(b-1)}$$

and

$$MSE = \frac{SSE}{n-bk}$$

HYPOTHESIS TESTING

- For testing for a **TREATMENT** effect, use

$$F = \frac{MST}{MSE}$$

Under the assumption of **NO TREATMENT EFFECT**, then

$$F \sim \text{Fisher-F}(k-1, n-bk)$$

which defines the rejection region and p-value in the usual way.

- For testing for a **BLOCK** effect, use

$$F = \frac{MSB}{MSE}$$

Under the assumption of **NO BLOCK EFFECT**, then

$$F \sim \text{Fisher-F}(b-1, n-bk)$$

- For testing for an **INTERACTION**, use

$$F = \frac{MSI}{MSE}$$

Under the assumption of **NO INTERACTION**, then

$$F \sim \text{Fisher-F}((k-1)(b-1), n-bk)$$

RANDOMIZED COMPLETE BLOCK DESIGNS WITH BALANCED REPLICATION: EXAMPLE

Data: Measurements were made on the lifetimes of batteries (in hours) for three battery types constructed from different materials, to investigate the effect of operating temperature on lifetime. It was believed before the experiment that the battery types were likely to behave differently in the experiment.

The **response variable** is lifetime. The single **factor** is the *temperature* and there are $k = 3$ **factor levels**:

1. 15 Celsius
2. 70 Celsius
3. 125 Celsius

The *material* types determine the $b = 3$ **blocks**

1. Lead
2. Acetate
3. Nickel Cadmium

$r = 4$ replicate measurements were made, so that $n = 3 \times 3 \times 4 = 36$ data were obtained in total.

The data observed in the study were as follows:

	Block		
Treatment	Lead	Acetate	Nickel Cadmium
15	130,155,74,180	150,188,159,126	138,119,168,160
70	34,40,80,75	126,122,106,115	174,120,150,139
120	20,70,82,58	25,70,58,45	96,104,82,60

Listing 18: Batteries data: ANOVA with interaction

```
 1 > summary(aov(life ~ temp*material,data=Batteries))
 2               Df Sum Sq Mean Sq F value   Pr(>F)
 3 temp           2  39083   19542  29.344 1.69e-07 ***
 4 material       2  10633    5317   7.983  0.00189 **
 5 temp:material  4   9438    2359   3.543  0.01897 *
 6 Residuals     27  17981     666
 7 ---
 8 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 9
10 > leveneTest(life ~ temp*material,data=Batteries,center='mean')
11 Levene's Test for Homogeneity of Variance (center = "mean")
12       Df F value Pr(>F)
13 group  8  1.0586 0.4196
14       27
15
16 >leveneTest(life ~ temp*material,data=Batteries,center='median')
17 Levene's Test for Homogeneity of Variance (center = "median")
18       Df F value Pr(>F)
19 group  8  0.9362 0.5036
20       27
```

There is a **significant difference** between **blocks** (materials - see line 4, `material`, $F = 7.983$, p-value=0.00189), a **significant difference** between **treatments** (temperatures - see line 3, `temp`, $F = 29.344$, p-value$1.69e - 07$), and also a significant interaction (line 5, `temp*material`, $F = 3.543$, p-value=0.01897). Levene's test reveals that there is no evidence to suspect that the population variances are different:

BALANCED COMPLETE FACTORIAL DESIGNS

Consider a **factorial design** (FD) with two factors A and B, with levels $1, \ldots, a$ and $1, \ldots, b$ respectively, yielding a total of $k = ab$ factor combinations (treatments), and suppose that there are r **replications** in each treatment, giving $n = rab$ observations in total. Let x_{ijt} be the tth replicated observation in the (i, j)th factor-level combination.

- sample mean for Factor A level i

$$\overline{x}_{i.} = \frac{1}{br} \sum_{j=1}^{b} \sum_{t=1}^{r} x_{ijt} \qquad i = 1, \ldots, a$$

- sample mean for Factor B level j

$$\overline{x}_{.j} = \frac{1}{ar} \sum_{i=1}^{a} \sum_{t=1}^{r} x_{ijt} \qquad j = 1, \ldots, b$$

- sample mean for replicates in (i, j)th factor combination

$$\overline{x}_{ij} = \frac{1}{r} \sum_{t=1}^{r} x_{ijt} \qquad i = 1, \ldots, a, \ j = 1, \ldots, b$$

- overall sample mean

$$\overline{x}_{..} = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{t=1}^{r} x_{ijt}$$

- Sum of Squares for Treatments due to factor A (SST_{A})

$$SST_A = \sum_{i=1}^{a} br(\overline{x}_{i.} - \overline{x}_{..})^2$$

- Sum of Squares for Treatments due to factor B (SST_{B})

$$SST_B = \sum_{j=1}^{b} ar(\overline{x}_{.j} - \overline{x}_{..})^2$$

- Sum of Squares for Interaction (SSI_{AB})

$$SSI_{AB} = \sum_{i=1}^{a} \sum_{j=1}^{b} r(\overline{x}_{ij} - \overline{x}_{i.} - \overline{x}_{.j} + \overline{x}_{..})^2$$

- Overall Sum of Squares (SS)

$$SS = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{t=1}^{r} (x_{ijt} - \overline{x}_{..})^2$$

The following decomposition holds

$$SS = SST_A + SST_B + SSI_{AB} + SSE \qquad \therefore \qquad SSE = SS - SST_A - SST_B - SSI_{AB}$$

Define

$$MST_A = \frac{SST_A}{a-1} \qquad MST_B = \frac{SST_B}{b-1} \qquad MSI_{AB} = \frac{SSI_{AB}}{(a-1)(b-1)}$$

and

$$MSE = \frac{SSE}{n-ab}$$

HYPOTHESIS TESTING

- For testing for a **FACTOR A** effect, use

$$F = \frac{MST_A}{MSE}$$

Under the assumption of **NO FACTOR A EFFECT**, then

$$F \sim \text{Fisher-F}(a-1, n-ab)$$

which defines the rejection region and p-value in the usual way.

- For testing for a **FACTOR B** effect, use

$$F = \frac{MST_B}{MSE}$$

Under the assumption of **NO FACTOR B EFFECT**, then

$$F \sim \text{Fisher-F}(b-1, n-ab)$$

- For testing for an **INTERACTION**, use

$$F = \frac{MSI_{AB}}{MSE}$$

Under the assumption of **NO INTERACTION**, then

$$F \sim \text{Fisher-F}((a-1)(b-1), n-ab)$$

Note: The only difference between a randomized block design and a factorial design is that in the block design, one of the factors is known or strongly believed to have a significant effect on the response. The method of analysis for interaction and no interaction models are identical.

EXAMPLE Butterfat data (Sokal, R. R. and Rohlf F. J. (1981). *Biometry*, 2nd edition)
The data give the average butterfat content (percentages) for random samples of twenty cows (ten two-year old and ten mature (greater than four years old)) from each of five breeds. The data are from Canadian records of pure-bred dairy cattle. There are 100 observations on two age groups (two year old and mature) and five breeds.

The **response variable** is butterfat level. **Factor A** is the *age* and there are $a = 2$ **factor levels**:

1. Mature
2. Two years

Factor B is the *breed* and there are $b = 5$ **factor levels**:

1. Ayrshire
2. Canadian
3. Guernsey
4. Holstein-Fresian
5. Jersey

$r = 2$ replicate measurements were made, so that $n = 2 \times 5 \times 2 = 20$ data were obtained in total.

Results:

Listing 19: `Butterfat` data: Levene's Test

```
> leveneTest(butterfat ~ breed*age,data=Butterfat,center='mean')
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value   Pr(>F)
group  9  2.7108 0.007686 **
      90
>
>leveneTest(butterfat ~ breed*age,data=Butterfat,center='median')
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value  Pr(>F)
group  9  1.8709 0.06644 .
      90
```

Listing 20: `Butterfat` data: ANOVA

```
> m3<-aov(butterfat ~ breed*age,data=Butterfat)
> summary(m3)
            Df Sum Sq Mean Sq F value Pr(>F)
breed        4  34.32   8.580  49.565 <2e-16 ***
age          1   0.27   0.274   1.580  0.212
breed:age    4   0.51   0.128   0.742  0.566
Residuals   90  15.58   0.173
---

> m2<-aov(butterfat ~ breed+age,data=Butterfat)
> summary(m2)
            Df Sum Sq Mean Sq F value Pr(>F)
breed        4  34.32   8.580  50.115 <2e-16 ***
age          1   0.27   0.274   1.598  0.209
Residuals   94  16.09   0.171
---

> m1<-aov(butterfat ~ breed,data=Butterfat)
> summary(m1)
            Df Sum Sq Mean Sq F value Pr(>F)
breed        4  34.32   8.580    49.8 <2e-16 ***
Residuals   95  16.37   0.172
---

> m0<-aov(butterfat ~ 1,data=Butterfat)
> anova(m0,m1,m2,m3)
Analysis of Variance Table

Model 1: butterfat ~ 1
Model 2: butterfat ~ breed
Model 3: butterfat ~ breed + age
Model 4: butterfat ~ breed * age
  Res.Df    RSS Df Sum of Sq       F Pr(>F)
1     99 50.689
2     95 16.368  4    34.321 49.5651 <2e-16 ***
3     94 16.094  1     0.274  1.5801 0.2120
4     90 15.580  4     0.514  0.7421 0.5658
```

1. **Interaction model**: First note that the Levene test **REJECTS** the null hypothesis of equal group variances ($p = 0.008$), so the following ANOVA results are questionable. However, the p-value is not too small, so we proceed but with caution.

   There is a **significant difference** due to Factor B (`breed`, $F = 49.565$, p-value < 0.001), but there is no effect of Factor A (`age`, F=1.580, $p = 0.212$), and no significant interaction (`age:breed`, $F = 0.742$, $p = 0.566$.

2. **Factor B only**: If we omit the Factor A and interaction term, and fit model `m1`, we confirm the strong effect of Factor B (line 21, $F = 49.8, p < 0.000$), and then can estimate the Factor B treatment means. Note how the error degrees of freedom changes when terms in the model are omitted.

3. The `anova` function can be used to summarize the fits of all the models (lines 26-37).

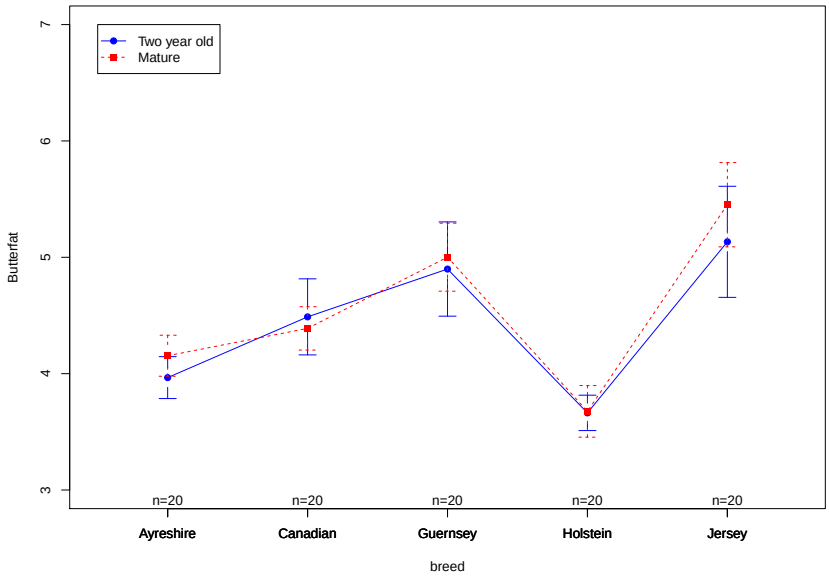


Figure 13: Butterfat data: Means plot by breed

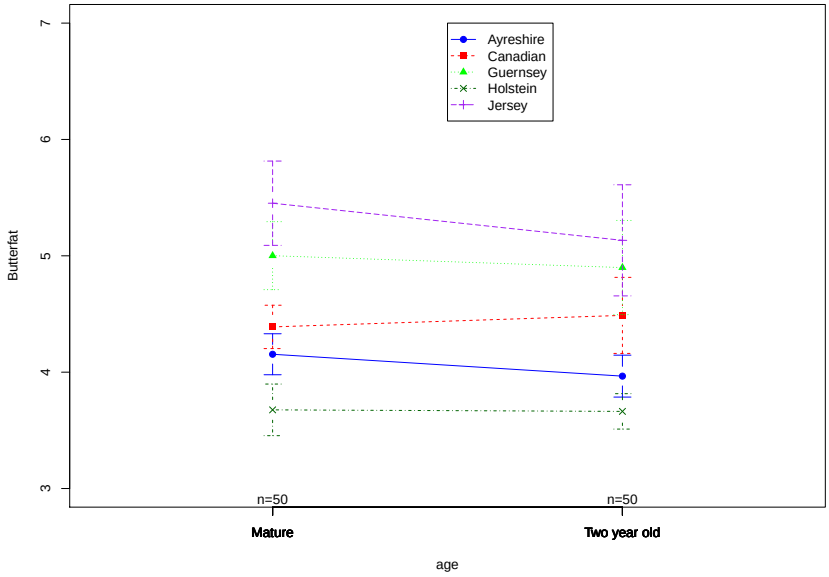


Figure 14: Butterfat data: Means plot by age

EXAMPLE `Lyrics` data (McClave and Sincich, *Statistics*)
The effect of violent song lyrics on the aggression level of listeners is to be investigated. Two songs (classified as *Violent* and *Non-Violent*) were played to two groups (or "pools") of students, one volunteer group and one group drawn from a psychology class. The students then rated the songs lyrical content, and from this (by means of a word-association test), the aggression level of the students was computed.

The **response variable** is aggression level. **Factor A** is the *song* and there are $a = 2$ **factor levels**:

1. Violent
2. Non-violent

Factor B is the *pool* and there are $b = 2$ **factor levels**:

1. Volunteer
2. Psychology class

$r = 15$ replicate measurements were made, so that $n = 2 \times 2 \times 15 = 60$ data were obtained in total.

Listing 21: `Lyrics` data: Levene's Test

```
> leveneTest(m1,center='mean')
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value Pr(>F)
group  1   0.017 0.8966
      58
> leveneTest(m1,center='median')
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.0082 0.9282
      58
> leveneTest(m3,center='mean')
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value Pr(>F)
group  3  0.3294 0.8041
      56
> leveneTest(m3,center='median')
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  3  0.2256 0.8782
      56
```

First note that the Levene test **DOES NOT REJECT** the null hypothesis of equal group variances for the model that fits the `SONG` factor only, nor for the full factorial model that fits both main effects and interaction.

Listing 22: `Lyrics` data: ANOVA

```
> m3<-aov(SCORE~SONG*POOL,data=Lyrics)
> summary(m3)
            Df Sum Sq Mean Sq F value   Pr(>F)
SONG         1  5.891   5.891  26.114 4.03e-06 ***
POOL         1  0.131   0.131   0.579    0.450
SONG:POOL    1  0.353   0.353   1.563    0.216
Residuals   56 12.632   0.226

> m2<-aov(SCORE~SONG+POOL,data=Lyrics)
> summary(m2)
            Df Sum Sq Mean Sq F value   Pr(>F)
SONG         1  5.891   5.891  25.859 4.26e-06 ***
POOL         1  0.131   0.131   0.574    0.452
Residuals   57 12.985   0.228

> m1<-aov(SCORE~SONG,data=Lyrics)
> summary(m1)
            Df Sum Sq Mean Sq F value   Pr(>F)
SONG         1  5.891   5.891   26.05 3.85e-06 ***
Residuals   58 13.115   0.226

> m0<-aov(SCORE~1,data=Lyrics)
> anova(m0,m1,m2,m3)
Analysis of Variance Table

Model 1: SCORE ~ 1
Model 2: SCORE ~ SONG
Model 3: SCORE ~ SONG + POOL
Model 4: SCORE ~ SONG * POOL
  Res.Df    RSS Df Sum of Sq       F    Pr(>F)
1     59 19.006
2     58 13.115  1    5.8907 26.1144 4.031e-06 ***
3     57 12.985  1    0.1307  0.5793    0.4498
4     56 12.632  1    0.3527  1.5634    0.2164
```

Results:

1. **Interaction model**: There is a **significant difference** due to Factor A (song, $F = 26.114$, p-value < 0.001), but there is no effect of Factor B (pool, line 5, $F = 0.579$, $p = 0.450$), and no significant interaction (`song:pool`, line 6, $F = 1.563$, $p = 0.216$.
2. Fits of the main-effects model `m2` (Factor A and Factor B but no interaction), and the Factor A only model `m1` confirm the results.

EXAMPLE `Gravel` data
A company produces gravel from a number of quarries and in each quarry there are morning and afternoon shifts of workers. The company wishes to know whether there are differences in the quantity of gravel produced from these quarries and gathers the following data on the amount of gravel produced by each shift in one week (in tonnes). It can be assumed that the week being studied was a typical week, and that there was no systematic differences due to different workers etc.

The **response variable** is amount of gravel produced. **Factor A** is the *shift* and there are $a = 2$ **factor levels**:

1. AM
2. PM

Factor B is the *quarry* and there are $b = 4$ **factor levels**:

1. A
2. B
3. C
4. D

$r = 5$ replicate measurements were made, so that $n = 2 \times 4 \times 5 = 40$ data were obtained in total.

Listing 23: `Gravel` data: ANOVA

```
> leveneTest(m3,center='median')
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  7  0.1679 0.9899
      32

> m3<-aov(amount ~ shift*quarry,data=Gravel)
> summary(m3)
              Df Sum Sq Mean Sq F value   Pr(>F)
shift          1  135.8  135.79  13.667 0.000814 ***
quarry         3  596.0  198.68  19.996 1.74e-07 ***
shift:quarry   3   32.7   10.92   1.099 0.363982
Residuals     32  317.9    9.94

> m2<-aov(amount ~ shift+quarry,data=Gravel)
> summary(m2)
             Df Sum Sq Mean Sq F value   Pr(>F)
shift         1  135.8  135.79   13.55 0.000776 ***
quarry        3  596.0  198.68   19.83  1.1e-07 ***
Residuals    35  350.7   10.02

> m1<-aov(amount ~ quarry,data=Gravel)
> m0<-aov(amount ~ 1,data=Gravel)
> anova(m0,m1,m2,m3)
Analysis of Variance Table

Model 1: amount ~ 1
Model 2: amount ~ quarry
Model 3: amount ~ shift + quarry
Model 4: amount ~ shift * quarry
  Res.Df     RSS Df Sum of Sq       F    Pr(>F)
1     39 1082.52
2     36  486.48  3    596.04 19.9964 1.742e-07 ***
3     35  350.69  1    135.79 13.6670 0.0008139 ***
4     32  317.94  3     32.75  1.0986 0.3639819
```

1. **Interaction model**: First note that the Levene test **DOES NOT REJECT** the null hypothesis of equal group variances. There is a **significant difference** due to Factor A (`shift`, line 10, $F = 13.667$, $p = 0.001$), and due to Factor B (`quarry`, line 11, $F = 19.996$, $p < 0.001$), but **no significant interaction** (`shift:quarry`, $F = 1.099$, line 12, $p = 0.364$.

2. **Factor A and B only**: If we omit the interaction term, and refit the model, we confirm the strong effect of both factors (`shift`, line 18, $F = 13.55, p = 0.001$, `quarry`, line 19, $F = 19.83, p < 0.001$. The conclusion is that there is a difference between the two levels of factor *shift* and the four levels of factor *quarry*, but that there is no interaction, that is, the difference between morning and afternoon shift is the same in each block; this is depicted in the Marginal Means plot.

   Note again how the error degrees of freedom changes when terms in the model are omitted.

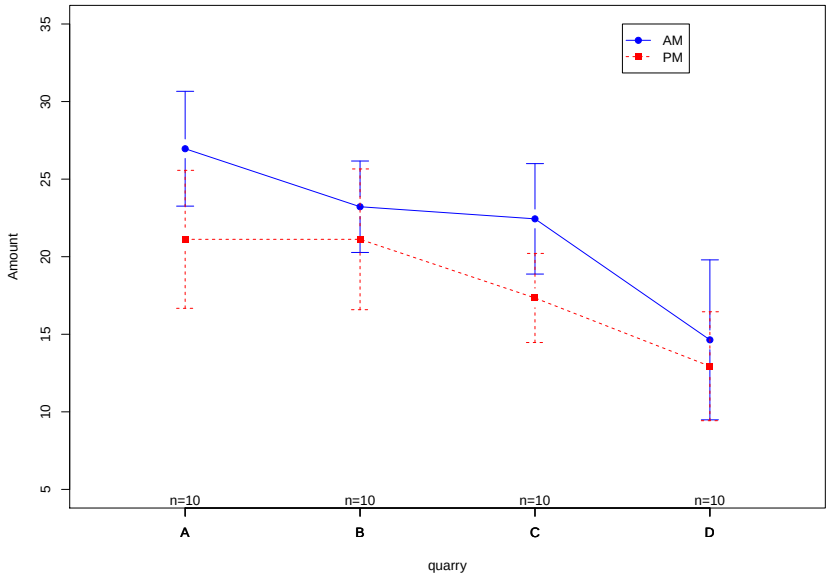


Figure 15: Gravel data: Means plot by quarry

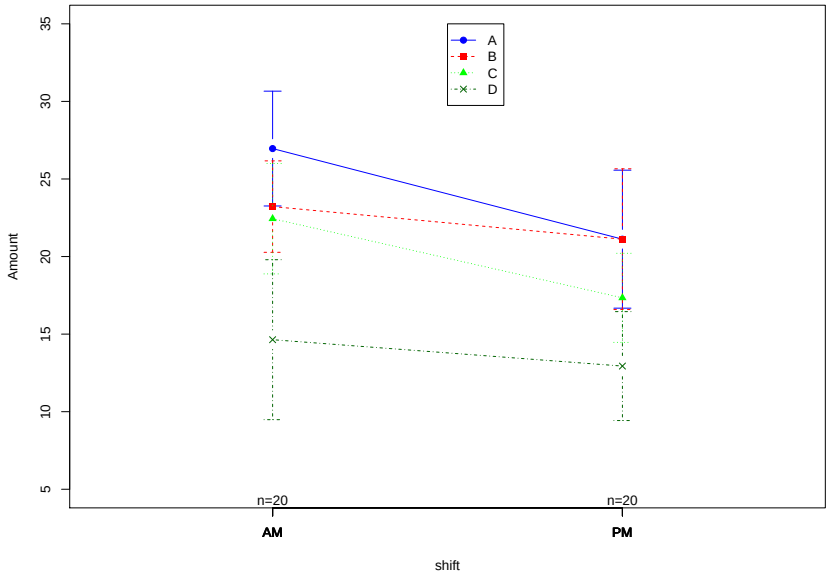


Figure 16: Gravel data: Means plot by shift

SIMPLE LINEAR REGRESSION

We consider the model for response variable, Y, as a function of the predictor, X, observed to take the value x. Specifically we consider the model

$$Y = \beta_0 + \beta_1 x + \epsilon$$

where β_0 and β_1 are the **intercept** and **slope** parameters respectively, and ϵ is a random variable with expectation zero and variance σ^2. In this model

$$E[Y|X = x] = \beta_0 + \beta_1 x.$$

To estimate the parameters β_0 and β_1 from data $(x_i, y_i), i = 1, \ldots, n$, we use the **least-squares** criterion, and choose the values $\widehat{\beta}_0$ and $\widehat{\beta}_1$ to minimize the **sum of squared errors**

$$SSE(\beta_0, \beta_1) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

It can be shown that the parameter estimates depend on the following sample summary statistics:

- Sample mean of x values:
$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$
- Sample mean of y values:
$$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$
- Sum of Squares SS_{xx}:
$$SS_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2$$
- Sum of Squares SS_{xy}:
$$SS_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

The **least-squares estimates** are:

$$\widehat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \qquad \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

yielding **fitted-values**

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$$

and **residual errors** (or **residuals**)

$$\widehat{e}_i = y_i - \widehat{y}_i.$$

An estimate of the **residual error variance** is given by

$$\widehat{\sigma}^2 = \frac{SSE(\widehat{\beta}_0, \widehat{\beta}_1)}{n - 2}$$

EXAMPLE Blood Viscosity and Packed Cell Volume
The following data are measurements of packed cell volume (PCV) and blood viscosity in samples taken from 32 hospital patients. We wish to model viscosity (y) as a function of PCV (x).

Reference: Begg, C. B. and Hearns, J. B. (1966) Components of Blood Viscosity. The relative contributions of haematocrit, plasma fibrinogen and other proteins, *Clinical Science*, **31**, 87-92.

Unit	PCV x	Viscosity y	Unit	PCV x	Viscosity y	Unit	PCV x	Viscosity y	Unit	PCV x	Viscosity y
1	40.00	3.71	9	46.75	4.14	17	51.25	4.68	25	49.50	5.12
2	40.00	3.78	10	48.00	4.20	18	50.25	4.73	26	56.00	5.15
3	42.50	3.85	11	46.00	4.20	19	49.00	4.87	27	50.00	5.17
4	42.00	3.88	12	47.00	4.27	20	50.00	4.94	28	47.00	5.18
5	45.00	3.98	13	43.25	4.27	21	50.00	4.95	29	53.25	5.38
6	42.00	4.03	14	45.00	4.37	22	49.00	4.96	30	57.00	5.77
7	42.50	4.05	15	50.00	4.41	23	50.50	5.02	31	54.00	5.90
8	47.00	4.14	16	45.00	4.64	24	51.25	5.02	32	54.00	5.90

- Sample mean of x values: $\overline{x} = 47.938$; sample mean of y values: $\overline{y} = 4.646$
- Sums of Squares

$$SS_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2 = 615.75 \qquad SS_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 75.386$$

Thus

$$\widehat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = 0.122 \qquad \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} = -1.223$$

The estimate of the residual error variance is

$$\widehat{\sigma}^2 = \frac{SSE(\widehat{\beta}_0, \widehat{\beta}_1)}{n-2} = \frac{2.721}{30} = 0.091$$

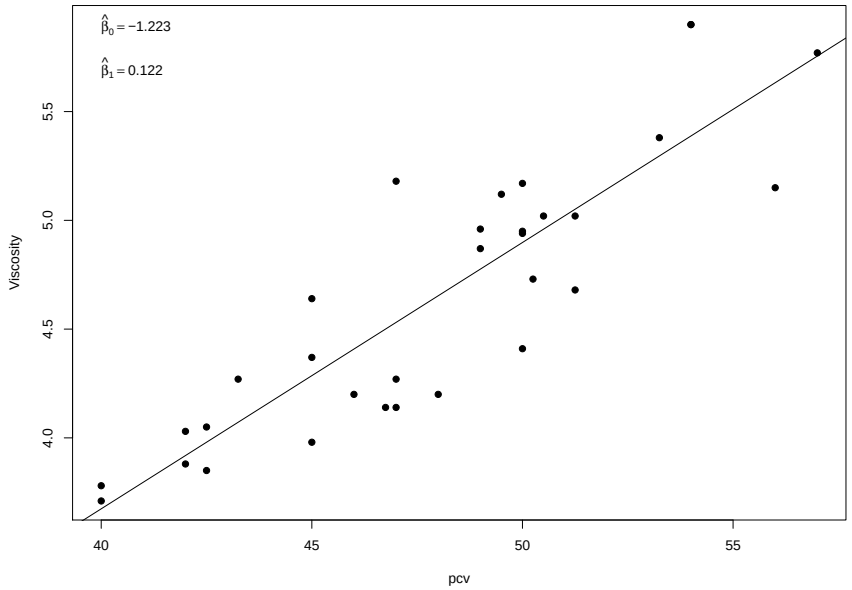


Figure 17: Viscosity data: straight-line fit

Listing 24: `Viscosity` data: regression

```
> cor.test(Viscosity$pcv,Viscosity$viscosity)

        Pearson's product-moment correlation

data:  Viscosity$pcv and Viscosity$viscosity
t = 10.0878, df = 30, p-value = 3.731e-11
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7643504 0.9395865
sample estimates:
     cor
0.878818

> m0.visc<-lm(viscosity~pcv,data=Viscosity)
> summary(m0.visc)

Call:lm(formula = viscosity ~ pcv, data = Viscosity)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.22336    0.58422  -2.094   0.0448 *
pcv          0.12243    0.01214  10.088 3.73e-11 ***
---

Residual standard error: 0.3012 on 30 degrees of freedom
Multiple R-squared: 0.7723,     Adjusted R-squared: 0.7647
F-statistic: 101.8 on 1 and 30 DF,  p-value: 3.731e-11
```

EXAMPLE Coleman Report Data
Data were collected at 20 US schools, and used to examine the relationship between performance of students in the school in a verbal reasoning test and the socioeconomic status of the catchment area.

Reference: Mosteller and Tukey (1977) *Data Analysis and Regression*

School	Status	Score	School	Status	Score
	x	y		x	y
1	7.20	37.01	11	-12.86	23.30
2	-11.71	26.51	12	0.92	35.20
3	12.32	36.51	13	4.77	34.90
4	14.28	40.70	14	-0.96	33.10
5	6.31	37.10	15	-16.04	22.70
6	6.16	33.90	16	10.62	39.70
7	12.70	41.80	17	2.66	31.80
8	-0.17	33.40	18	-10.99	31.70
9	9.85	41.01	19	15.03	43.10
10	-0.05	37.20	20	12.77	41.01

Results:
Here, to test for significant correlation, we use the test statistic

$$t = \frac{r}{\sqrt{(1-r^2)/(n-2)}} = \frac{0.927}{\sqrt{(1-0.927^2)/(20-2)}} = 10.499$$

which we must compare against the Student$(n-2) \equiv$ Student(18) distribution. For a two-tailed test at the significance level $\alpha = 0.05$, the critical values are $C_R = \pm 2.101$ (McClave and Sincich t-tables), so the hypothesis H_0 that the true correlation is zero is **rejected**.

Listing 25: `Coleman` data: correlation test

```
> cor.test(Coleman$status,Coleman$testscore)

        Pearson's product-moment correlation

data:  Coleman$status and Coleman$testscore
t = 10.4991, df = 18, p-value = 4.199e-09
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8218255 0.9712072
sample estimates:
      cor
0.9271611
```

The regression analysis is contained below:

Listing 26: `Coleman` data: regression

```
> m0.cole<-lm(testscore ~ status,data=Coleman)
> summary(m0.cole)

Call:
lm(formula = testscore ~ status, data = Coleman)

Residuals:
    Min      1Q  Median      3Q     Max
-8.9412 -1.2190 -0.0547  1.9510  6.9895

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -50.6816     5.1929   -9.76  1.3e-08 ***
status        1.5342     0.1461   10.50  4.2e-09 ***
---

Residual standard error: 3.705 on 18 degrees of freedom
Multiple R-squared: 0.8596,     Adjusted R-squared: 0.8518
F-statistic: 110.2 on 1 and 18 DF,  p-value: 4.199e-09

> summary(aov(m0.cole))
            Df Sum Sq Mean Sq F value  Pr(>F)
status       1 1513.2  1513.2   110.2 4.2e-09 ***
Residuals   18  247.1    13.7
---
```

EXAMPLE Hooker's Temperature and Pressure Data
The following data record the boiling point temperature (in degrees Celsius) of water under different atmospheric pressures. The data were collected in a Himalayan expedition by botanist Joseph Hooker.

Reference: Forbes, J. (1957). Further experiments and remarks on the measurement of heights by boiling point of water. *Transactions of the Royal Society of Edinburgh*, 21, 235-243.

y	x	y	x	y	x	y	x
210.8	29.211	196.4	21.928	189.5	18.869	184.1	16.817
210.2	28.559	196.3	21.654	188.8	18.356	183.2	16.385
208.4	27.972	195.6	21.605	188.5	18.507	182.4	16.235
202.5	24.697	193.4	20.480	185.7	17.267	181.9	16.106
200.6	23.726	193.6	20.212	186.0	17.221	181.9	15.928
200.1	23.369	191.4	19.758	185.6	17.062	181.0	15.919
199.5	23.030	191.1	19.490	184.1	16.959	180.6	15.376
197.0	21.892	190.6	19.386	184.6	16.881		

Here, to test for significant correlation, we use the test statistic

$$t = \frac{r}{\sqrt{(1-r^2)/(n-2)}} = \frac{0.996}{\sqrt{(1-0.996^2)/(31-2)}} = 59.143$$

which we must compare against the Student$(31-2) \equiv$ Student(29) distribution. For a two-tailed test at the significance level $\alpha = 0.05$, the critical values are $C_R = \pm 2.045$ (McClave and Sincich t-tables), so the hypothesis H_0 that the true correlation is zero is **rejected**.

Listing 27: `Hooker` data: correlation test and regression

```
> cor.test(Hooker$pressure,Hooker$temp)

        Pearson's product-moment correlation

data:  Hooker$pressure and Hooker$temp
t = 59.1431, df = 29, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9913779 0.9980339
sample estimates:
      cor
0.9958803

> m0.hooker<-lm(temp~pressure,data=Hooker)
> summary(m0.hooker)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 146.67290    0.77641  188.91   <2e-16 ***
pressure      2.25260    0.03809   59.14   <2e-16 ***

Residual standard error: 0.806 on 29 degrees of freedom
Multiple R-squared: 0.9918,     Adjusted R-squared: 0.9915
F-statistic:  3498 on 1 and 29 DF,  p-value: < 2.2e-16

> summary(aov(m0.hooker))
            Df Sum Sq Mean Sq F value Pr(>F)
pressure     1 2272.5  2272.5    3498 <2e-16 ***
Residuals   29   18.8     0.6
```

POLYNOMIAL REGRESSION: QUADRATIC MODEL

Listing 28: `Hooker` data: quadratic regression

```
> m0.hooker<-lm(temp~pressure,data=Hooker)
> summary(m0.hooker)

Call:lm(formula = temp ~ pressure, data = Hooker)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 146.67290    0.77641  188.91   <2e-16 ***
pressure      2.25260    0.03809   59.14   <2e-16 ***

Residual standard error: 0.806 on 29 degrees of freedom
Multiple R-squared: 0.9918,     Adjusted R-squared: 0.9915
F-statistic:  3498 on 1 and 29 DF,  p-value: < 2.2e-16

> m1.hooker<-lm(temp~pressure+I(pressure^2),data=Hooker)
> summary(m1.hooker)

Call:lm(formula = temp ~ pressure + I(pressure^2), data = Hooker)

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   126.701623   2.112363  59.981  < 2e-16 ***
pressure        4.157627   0.199069  20.885  < 2e-16 ***
I(pressure^2)  -0.043754   0.004552  -9.612 2.29e-10 ***

Residual standard error: 0.3956 on 28 degrees of freedom
Multiple R-squared: 0.9981,     Adjusted R-squared: 0.998
F-statistic:  7307 on 2 and 28 DF,  p-value: < 2.2e-16

> anova(m0.hooker,m1.hooker)
Analysis of Variance Table

Model 1: temp ~ pressure
Model 2: temp ~ pressure + I(pressure^2)
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     29 18.8404
2     28  4.3817  1    14.459 92.394 2.288e-10 ***
---
```

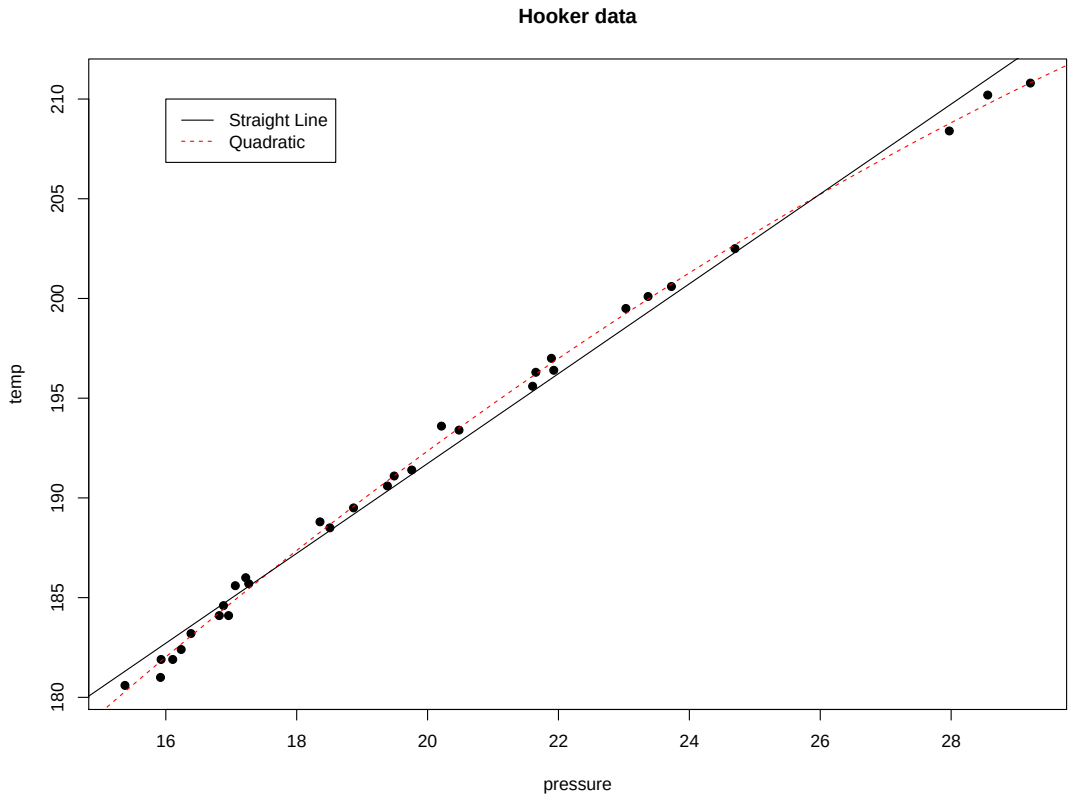


Figure 18: Hooker data: straight-line and quadratic fits

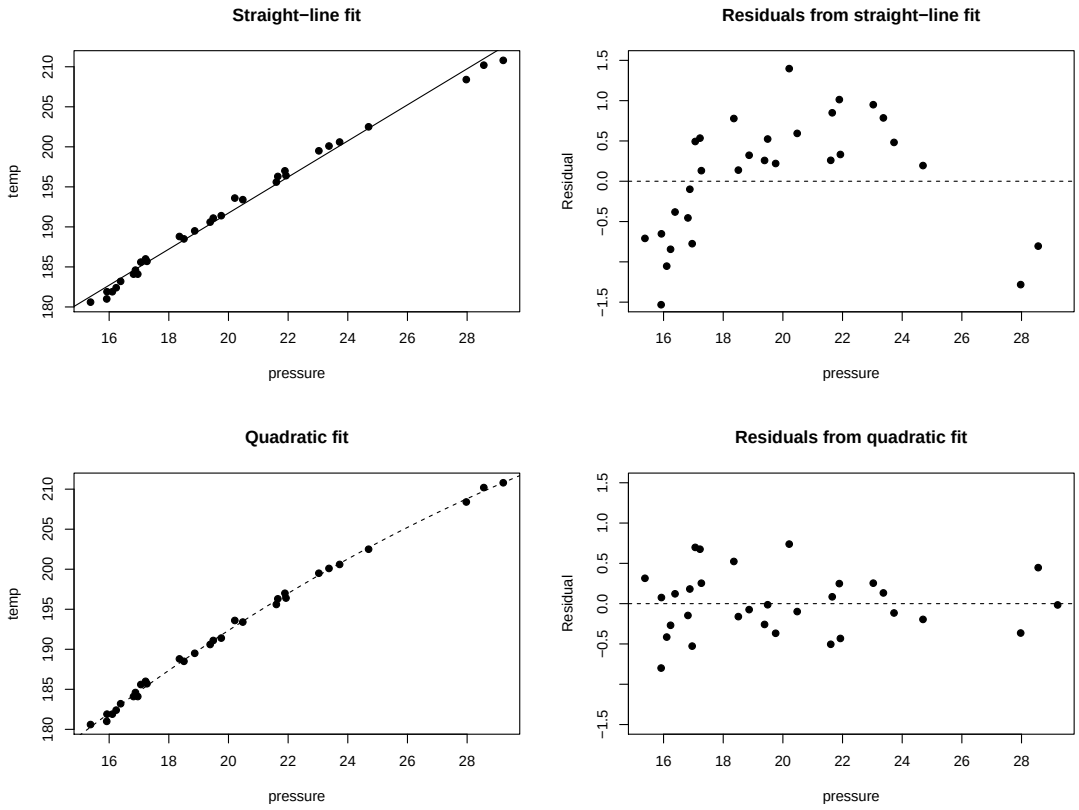


Figure 19: Hooker data: residuals from straight-line and quadratic fits

MATRIX MANIPULATIONS FOR ESTIMATION (MATERIAL NOT EXAMINABLE)

An $r \times c$ **matrix** A is a rectangular arrangement of numbers with r rows and c columns;

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2c} \\ \vdots & \vdots & \vdots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rc} \end{bmatrix}$$

Recall the following rules for manipulating matrices:

- **Transpose:** the transpose operator $^\intercal$ means "flipping" a $r \times c$ matrix into a $c \times r$ matrix. That is

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2c} \\ \vdots & \vdots & \vdots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rc} \end{bmatrix} \quad \Longleftrightarrow \quad A^\intercal = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{r1} \\ a_{12} & a_{22} & \cdots & a_{r2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1c} & a_{2c} & \cdots & a_{rc} \end{bmatrix}$$

  For example, if $r = 2$ and $c = 4$

$$A = \begin{bmatrix} 5 & -4 & 0 & 1 \\ 3 & 5 & -2 & 0 \end{bmatrix} \quad \Longleftrightarrow \quad A^\intercal = \begin{bmatrix} 5 & 3 \\ -4 & 5 \\ 0 & -2 \\ 1 & 0 \end{bmatrix}$$

  A square matrix A is termed **symmetric** if $A = A^\intercal$.

- **Matrix Multiplication:** If A and B are two matrices, where A is a $r_1 \times c$ matrix and B is a $c \times r_2$ matrix, then the product $A.B$ (also written AB) is an $r_1 \times r_2$ matrix, with (i, j)th element

$$\sum_{k=1}^{c} a_{ik} b_{kj} \qquad i = 1, \ldots, r_1, \; j = 1, \ldots, r_2.$$

  For example,

$$\begin{bmatrix} 5 & -4 & 0 & 1 \\ 3 & 5 & -2 & 0 \end{bmatrix} \begin{bmatrix} 3 & 3 & -3 \\ -1 & 2 & -2 \\ 0 & -2 & 0 \\ -5 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 5 & -6 \\ 4 & 23 & -19 \end{bmatrix}$$

  That is, for the first entry in the result matrix, we multiply the **first row** of the first matrix by the **first column** of the second matrix:

$$(5 \times 3) + (-4 \times -1) + (0 \times 0) + (1 \times -5) = 15 + 4 - 5 = 14$$

  Note that for matrix multiplication to work, we need the first matrix to have the same number of columns as the number of rows in the second matrix. If this holds, the matrices are termed **conformable**. In general, for rectangular matrices

$$A.B \neq B.A \qquad \text{and} \qquad A.B.C = A.(B.C) = (A.B).C$$

- **Matrix Identity :** A square $k \times k$ with ones along the main diagonal, and zeros elsewhere, is termed the **identity** matrix, and denoted I_k

$$I_k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \qquad \text{so that} \qquad I_k.A = A \qquad \text{for any } k \times k \text{ matrix } A$$

- **Matrix Inversion :** A square $k \times k$ matrix A has an **inverse**, denoted A^{-1} if

$$A.A^{-1} = A^{-1}.A = I_k$$

MATRICES IN LINEAR REGRESSION

- $n \times 1$ vector $\underset{\sim}{y} = [y_1, \ldots, y_n]^\intercal$
- $n \times 2$ matrix $\boldsymbol{X}$ given by

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}^\intercal$$

- 2×1 Parameter estimate vector $\underset{\sim}{\widehat{\beta}} = \left[\widehat{\beta}_0, \widehat{\beta}_1\right]^\intercal$

It can be shown that

$$\underset{\sim}{\widehat{\beta}} = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \underset{\sim}{y}$$

The other quantities of interest in statistical inference for the simple linear regression are also available in matrix form.

- SSE:

$$SSE = S(\underset{\sim}{\widehat{\beta}}) = (\underset{\sim}{y} - \boldsymbol{X}\underset{\sim}{\widehat{\beta}})^\intercal (\underset{\sim}{y} - \boldsymbol{X}\underset{\sim}{\widehat{\beta}})$$

- Residual error variance estimate, $\widehat{\sigma}^2$:

$$\widehat{\sigma}^2 = \frac{S(\underset{\sim}{\widehat{\beta}})}{n-2} = \frac{1}{n-2}(\underset{\sim}{y} - \boldsymbol{X}\underset{\sim}{\widehat{\beta}})^\intercal (\underset{\sim}{y} - \boldsymbol{X}\underset{\sim}{\widehat{\beta}})$$

- Variance/Standard Errors of the Parameter estimates:

$$Var[\underset{\sim}{\widehat{\beta}}] = \widehat{\sigma}^2 (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}$$

  This is a 2×2 matrix, with diagonal entries equal to the squared estimated standard errors for $\widehat{\beta}_0$ and $\widehat{\beta}_1$, $s^2_{\widehat{\beta}_0}$ and $s^2_{\widehat{\beta}_1}$ respectively.
- Fitted-values :

$$\underset{\sim}{\widehat{y}} = \boldsymbol{X}\underset{\sim}{\widehat{\beta}} = \boldsymbol{X}(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \underset{\sim}{y} = \boldsymbol{H}\underset{\sim}{y}$$

  say, where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal$
- Residuals:

$$\underset{\sim}{\widehat{e}} = \underset{\sim}{y} - \underset{\sim}{\widehat{y}} = \underset{\sim}{y} - \boldsymbol{H}\underset{\sim}{y} = (I_n - \boldsymbol{H})\underset{\sim}{y}$$

- Prediction: if $\boldsymbol{x}_p = [1, x_p]^\intercal$, then the prediction is at the value x_p is

$$y_p = \boldsymbol{x}_p^\intercal \underset{\sim}{\widehat{\beta}}$$

  and the prediction error variances are

$$\begin{aligned} \text{Expected Value} &: \widehat{\sigma}^2 \boldsymbol{x}_p^\intercal (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{x}_p \\ \text{Individual Value} &: \widehat{\sigma}^2 (1 + \boldsymbol{x}_p^\intercal (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{x}_p) \end{aligned}$$

MULTIPLE LINEAR REGRESSION

EXAMPLE: BLOOD VISCOSITY AND PACKED CELL VOLUME

The following blood viscosity data studied earlier are a good example of where multiple regression could be used. Recall that the data blood viscosity in samples taken from 32 hospital patients. We wish to model viscosity (y) as a function three covariates

- Packed Cell Volume (PCV), x_1.
- Plasma Fibrinogen, x_2.
- Plasma Protein, x_3.

Unit	Viscosity y	PCV x_1	Plasma Fib. x_2	Plasma Pro. x_3
1	3.71	40.00	344	6.27
2	3.78	40.00	330	4.86
3	3.85	42.50	280	5.09
4	3.88	42.00	418	6.79
5	3.98	45.00	774	6.40
6	4.03	42.00	388	5.48
7	4.05	42.50	336	6.27
8	4.14	47.00	431	6.89
9	4.14	46.75	276	5.18
10	4.20	48.00	422	5.73
11	4.20	46.00	280	5.89
12	4.27	47.00	460	6.58
13	4.27	43.25	412	5.67
14	4.37	45.00	320	6.23
15	4.41	50.00	502	4.99
16	4.64	45.00	550	6.37
17	4.68	51.25	414	6.40
18	4.73	50.25	304	6.00
19	4.87	49.00	472	5.94
20	4.94	50.00	728	5.16
21	4.95	50.00	716	6.29
22	4.96	49.00	400	5.96
23	5.02	50.50	576	5.90
24	5.02	51.25	354	5.81
25	5.12	49.50	392	5.49
26	5.15	56.00	352	5.41
27	5.17	50.00	572	6.24
28	5.18	47.00	634	6.50
29	5.38	53.25	458	6.60
30	5.77	57.00	1070	4.82
31	5.90	54.00	488	5.70
32	5.90	54.00	488	5.70

We consider four analyses:

$$\begin{aligned}
\textbf{Multiple regression :} \quad y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \\
\textbf{Regression on } \boldsymbol{x_1}\textbf{ :} \quad y &= \beta_0 + \beta_1 x_1 + \epsilon \\
\textbf{Regression on } \boldsymbol{x_2}\textbf{ :} \quad y &= \beta_0 + \beta_2 x_2 + \epsilon \\
\textbf{Regression on } \boldsymbol{x_3}\textbf{ :} \quad y &= \beta_0 + \beta_3 x_3 + \epsilon
\end{aligned}$$

Listing 29: `Viscosity` data: multiple regression

```
> mv.visc<-lm(viscosity~pcv+plasfib+plaspro,data=Viscosity)
> summary(mv.visc)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.3782383  0.8966650  -1.537    0.136
pcv          0.1168232  0.0136089   8.584  2.5e-09 ***
plasfib      0.0004019  0.0003505   1.147    0.261
plaspro      0.0400364  0.0971527   0.412    0.683

Residual standard error: 0.3037 on 28 degrees of freedom
Multiple R-squared: 0.7839,     Adjusted R-squared: 0.7607
F-statistic: 33.86 on 3 and 28 DF,  p-value: 1.876e-09

> m1.visc<-lm(viscosity~pcv,data=Viscosity)
> summary(m1.visc)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.22336    0.58422  -2.094   0.0448 *
pcv          0.12243    0.01214  10.088 3.73e-11 ***

Residual standard error: 0.3012 on 30 degrees of freedom
Multiple R-squared: 0.7723,     Adjusted R-squared: 0.7647
F-statistic: 101.8 on 1 and 30 DF,  p-value: 3.731e-11

> m2.visc<-lm(viscosity~plasfib,data=Viscosity)
> summary(m2.visc)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.8708133  0.2924499  13.236 4.64e-14 ***
plasfib     0.0016595  0.0005892   2.816   0.0085 **

Residual standard error: 0.5613 on 30 degrees of freedom
Multiple R-squared: 0.2091,     Adjusted R-squared: 0.1828
F-statistic: 7.932 on 1 and 30 DF,  p-value: 0.008504

> m3.visc<-lm(viscosity~plaspro,data=Viscosity)
> summary(m3.visc)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.2961     1.1743   4.510 9.25e-05 ***
plaspro      -0.1104     0.1983  -0.556    0.582

Residual standard error: 0.6279 on 30 degrees of freedom
Multiple R-squared: 0.01022,    Adjusted R-squared: -0.02278
F-statistic: 0.3096 on 1 and 30 DF,  p-value: 0.582
```

Listing 30: `Viscosity` data: multiple regression comparisons

```
> anova(m1.visc,mv.visc)
Analysis of Variance Table

Model 1: viscosity ~ pcv
Model 2: viscosity ~ pcv + plasfib + plaspro
  Res.Df    RSS Df Sum of Sq    F Pr(>F)
1     30 2.7208
2     28 2.5825  2   0.13835 0.75 0.4816

> anova(m2.visc,mv.visc)
Analysis of Variance Table

Model 1: viscosity ~ plasfib
Model 2: viscosity ~ pcv + plasfib + plaspro
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     30 9.4513
2     28 2.5825  2    6.8688 37.237 1.293e-08 ***

> anova(m3.visc,mv.visc)
Analysis of Variance Table

Model 1: viscosity ~ plaspro
Model 2: viscosity ~ pcv + plasfib + plaspro
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     30 11.8283
2     28  2.5825  2    9.2458 50.123 5.593e-10 ***
```

FACTOR PREDICTOR REGRESSION USING DUMMY VARIABLES

We can fit a factor predictor using **dummy variables**.

Suppose that a **factor predictor**, X, takes L levels, indexed by $l = 1, 2, \ldots, L$. We proceed as follows:

1. Define L **new** "dummy" variables $X_1, \ldots, X_L$, where, for $l = 1, \ldots, L$,

$$X_l = \begin{cases} 1 & \text{if } X = l \\ 0 & \text{if } X \neq l \end{cases}$$

2. Fit the multiple regression model with $L - 1$ of the dummy variables as continuous covariates, that is,

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{L-1} x_{L-1} + \epsilon_i$$

   Note that we cannot include all of $X_1, X_2, \ldots, X_L$if we have an intercept β_0 in the model; we omit X_L and regard L as the baseline group.

The estimates, standard errors etc. from this model are identical to those obtained using the *General Linear Model* analysis.

EXAMPLE Diabetes Data Set
The data set `Diabetes` has three subgroups defined by different patient characteristics. Thus $L = 3$. A subset of the data are displayed below, with the new variables X_1, X_2 and X_3 defined as above. They can be computed using the

Compute

pulldown menu, or entered by hand.

ID	glutest y	group x	Dummy 1 x_1	Dummy 2 x_2	Dummy 3 x_3
1	356	3	0	0	1
2	289	3	0	0	1
3	319	3	0	0	1
4	356	3	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮
87	503	2	0	1	0
88	540	2	0	1	0
89	469	2	0	1	0
90	486	2	0	1	0
⋮	⋮	⋮	⋮	⋮	⋮
113	1468	1	1	0	0
114	1487	1	1	0	0
115	714	1	1	0	0
116	1470	1	1	0	0

Listing 31: `Diabetes` data: regression in different parameterizations

```
> summary(lm(glutest ~ group,data=Diabetes))

Call:
lm(formula = glutest ~ group, data = Diabetes)

Residuals:
    Min      1Q  Median      3Q     Max
-489.37  -43.97   -1.67   39.30  492.63

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                1027.37      25.74   39.91   <2e-16 ***
groupChemically Diabetic   -533.43      35.38  -15.08   <2e-16 ***
groupNormal                -677.40      30.68  -22.08   <2e-16 ***

Residual standard error: 145.6 on 141 degrees of freedom
Multiple R-squared: 0.7771,     Adjusted R-squared: 0.7739
F-statistic: 245.8 on 2 and 141 DF,  p-value: < 2.2e-16

> summary(lm(glutest ~ -1+group,data=Diabetes))

Call:
lm(formula = glutest ~ -1 + group, data = Diabetes)

Residuals:
    Min      1Q  Median      3Q     Max
-489.38  -43.97   -1.67   39.30  492.62

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
groupOvert Diabetic        1027.38      25.74   39.91   <2e-16 ***
groupChemically Diabetic    493.94      24.27   20.35   <2e-16 ***
groupNormal                 349.97      16.70   20.95   <2e-16 ***

Residual standard error: 145.6 on 141 degrees of freedom
Multiple R-squared: 0.9455,     Adjusted R-squared: 0.9443
F-statistic: 815.5 on 3 and 141 DF,  p-value: < 2.2e-16
```

Listing 32: `Diabetes` data: redefining the baseline

```

Diabetes.copy<-Diabetes
Diabetes.copy$group2<-factor(Diabetes.copy$group,
            c("Normal","Chemically Diabetic","Overt Diabetic"))
```

FACTOR PREDICTOR REGRESSION

We need to take some care when combining factor predictors and covariates in the regression model. Suppose that we have only two predictors

- A **covariate**, x_1
- A **factor predictor**, x_2, now taking L levels, with the levels being indexed by $l = 1, 2, \ldots, L$.

We want to build a model that takes into account both x_1 and x_2.

EXAMPLE Binary Factor $L = 2$
Suppose that factor predictor x_2 takes two levels, labelled 0 and 1, that identify two data subgroups. Five models can be considered, that correspond to different straight-line models

- **MODEL 0 : Same intercept, slope zero**, in the two subgroups
- **MODEL 1 : Different intercept, slope zero**, in the two subgroups
- **MODEL 2 : Same intercept, same non-zero slope**, in the two subgroups
- **MODEL 3 : Different intercept, same non-zero slope**, in the two subgroups
- **MODEL 4 : Different intercept, different non-zero slopes**, in the two subgroups

We can write out the models in terms of the usual slope and intercept parameters. The general model can be written

$$y = \begin{cases} \beta_{00} + \beta_{01}x_1 + \epsilon & \text{GROUP 0} \quad (l = 0) \\ \beta_{10} + \beta_{11}x_1 + \epsilon & \text{GROUP 1} \quad (l = 1) \end{cases}$$

- **MODEL 0 :** $\beta_{00} = \beta_{10} = \beta_0,\ \beta_{01} = \beta_{11} = 0$
- **MODEL 1 :** $\beta_{00} \neq \beta_{10},\ \beta_{01} = \beta_{11} = 0$
- **MODEL 2 :** $\beta_{00} = \beta_{10} = \beta_0,\ \beta_{01} = \beta_{11} = \beta_1 \neq 0$
- **MODEL 3 :** $\beta_{00} \neq \beta_{10},\ \beta_{01} = \beta_{11} = \beta_1 \neq 0$
- **MODEL 4 :** $\beta_{00} \neq \beta_{10},\ \beta_{01} \neq \beta_{11}$

The numbers of parameters, p, in each model are as follows:

MODEL 0	:	$p = 1$	β_0
MODEL 1	:	$p = 2$	β_{00}, β_{10}
MODEL 2	:	$p = 2$	$\beta_{00} = \beta_{01} = \beta_0, \beta_{10} = \beta_{11} = \beta_1$
MODEL 3	:	$p = 3$	$\beta_{00}, \beta_{10}, \beta_1$
MODEL 4	:	$p = 4$	$\beta_{00}, \beta_{10}, \beta_{10}, \beta_{11}$

The default parameterization used by `R` is different from the one described above. As in ANOVA, `R` takes a **baseline** group, and looks for **differences** in the parameters compared to the baseline group. The baseline group is taken to be the **first** listed subgroup for the factor predictor; in the binary example above, the baseline group would be Group 0.

The interaction model is therefore written

$$y = [\beta_0 + (1 - x_2)\delta_{10}] + [(\beta_1 + (1 - x_2)\delta_{11})x_1] + \epsilon$$

- δ_{10} is the **change in intercept** from Group 0 to Group 1
- δ_{11} is the **change in slope** from Group 1 to Group 0

EXAMPLE Diabetes Data Set
The data in the data set `Diabetes` contain information on 68 diabetes patients falling into two clinically different categories (overt and chemical diabetics) and 76 normal controls. Measurements of plasma glucose in blood samples when fasting and in a dietary test are recorded.

The objective is to predict the the test glucose levels from the fasting glucose levels in the three subgroups, and to find out if there is any significant difference between the subgroups.

In this analysis, there is a single response variable, one covariate and one factor predictor:

- y : **glutest**, the test glucose level
- x_1 : covariate **glufast**, the fasting glucose level
- x_2 : factor predictor **group**, the diabetes group
  - GROUP 1: Overt Diabetic
  - GROUP 2: Chemical Diabetic
  - GROUP 3: Normal Patients

  After recoding, the data set `Diabetes.copy` has levels
  - GROUP 1: Normal Patients
  - GROUP 2: Chemical Diabetic
  - GROUP 3: Overt Diabetic

Listing 33: `Diabetes` data: factor and covariate regression

```
> m.Diabetes<-lm(log(glutest) ~ log(glufast)*group2,data=Diabetes.copy)

> summary(aov(m.Diabetes))
                     Df Sum Sq Mean Sq  F value   Pr(>F)
log(glufast)          1 24.766  24.766 2593.831  < 2e-16 ***
group2                2  2.266   1.133  118.641  < 2e-16 ***
log(glufast):group2   2  0.155   0.077    8.099 0.000472 ***
Residuals           138  1.318   0.010

> summary(m.Diabetes)

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                             4.5036   0.5588   8.060 3.25e-13 ***
log(glufast)                            0.2991   0.1239   2.414 0.017085 *
group2Chemically Diabetic              -1.4355   0.9579  -1.499 0.136243
group2Overt Diabetic                   -2.0371   0.6193  -3.289 0.001275 **
log(glufast):group2Chemically Diabetic  0.3819   0.2098   1.820 0.070891 .
log(glufast):group2Overt Diabetic       0.5349   0.1337   4.001 0.000103 ***

Residual standard error: 0.09771 on 138 degrees of freedom
Multiple R-squared: 0.9538,     Adjusted R-squared: 0.9521
F-statistic: 569.5 on 5 and 138 DF,  p-value: < 2.2e-16
```

The first ANOVA table demonstrates that there **is** a significant interaction between the covariate and the factor predictor ($F = 8.099$, p-value 0.000472). This means that there is a **significantly different slope** in **at least two of the three** subgroups.

The second table gives the slope and intercept parameters in the three groups. The `R` parameterization is not directly in terms of the slopes and intercepts, but looks at **differences** from baseline subgroup, Group 1. For example, the Group 3 (Overt Diabetic) intercept and slope are, respectively,

$$\text{INTERCEPT} : 4.504 + (-2.037) = 2.467 \qquad \text{SLOPE} : 0.2991 + 0.5349 = 0.8340.$$

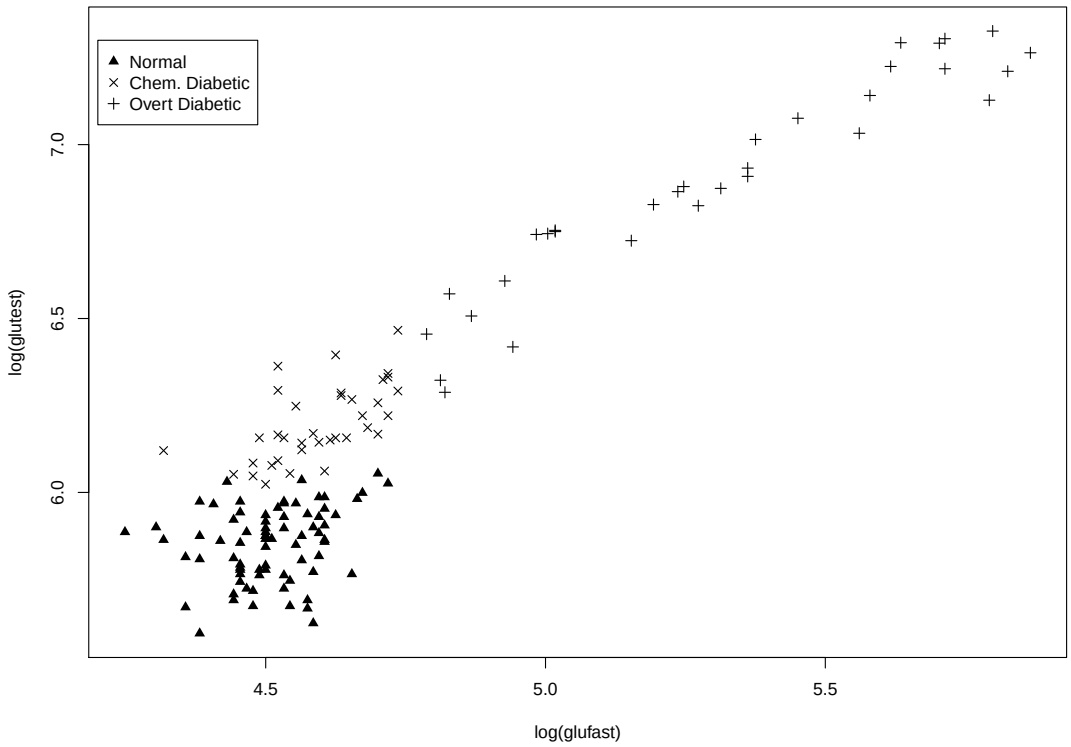


Figure 20: `Diabetes` data: raw data broken down by group

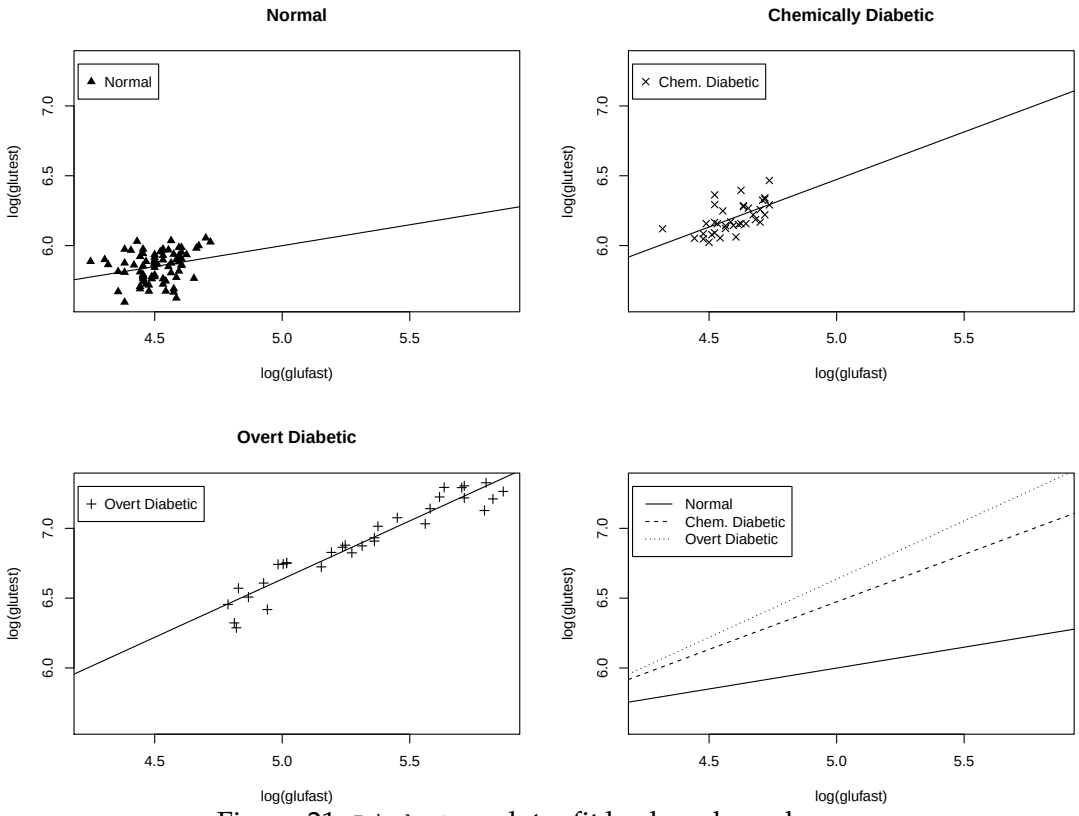


Figure 21: `Diabetes` data: fit broken down by group

Listing 34: `Diabetes` data: test to omit the interaction

```
> drop1(m.Diabetes,test="F")
Single term deletions

Model:
log(glutest) ~ log(glufast) * group2
                     Df Sum of Sq    RSS     AIC F value    Pr(>F)
<none>                             1.3176 -663.93
log(glufast):group2  2   0.15467 1.4723 -651.95  8.0994 0.0004721 ***
```

The models that can be fitted are as follows:

MODEL 0 : Null model
MODEL 1 : Effect of `group` only
MODEL 2 : Effect of `log(glufast)` only
MODEL 3 : Additive main effect of `group` and `log(glufast)`
MODEL 4 : Main effects plus interaction

Listing 35: `Diabetes` data: possible models

```
> m0.diabetes<-lm(log(glutest)~1,data=Diabetes.copy)
> m1.diabetes<-lm(log(glutest)~group2,data=Diabetes.copy)
> m2.diabetes<-lm(log(glutest)~log(glufast),data=Diabetes.copy)
> m3.diabetes<-lm(log(glutest)~group2+log(glufast),data=Diabetes.copy)
> m4.diabetes<-lm(log(glutest)~group2*log(glufast),data=Diabetes.copy)

> summary(aov(m0.diabetes))
             Df Sum Sq Mean Sq F value Pr(>F)
Residuals   143   28.5  0.1993

> summary(aov(m1.diabetes))
             Df Sum Sq Mean Sq F value Pr(>F)
group2        2  24.34   12.17   412.6 <2e-16 ***
Residuals   141   4.16    0.03

> summary(aov(m2.diabetes))
              Df Sum Sq Mean Sq F value Pr(>F)
log(glufast)   1 24.766  24.766   940.8 <2e-16 ***
Residuals    142  3.738   0.026

> summary(aov(m3.diabetes))
              Df Sum Sq Mean Sq F value Pr(>F)
group2         2 24.344  12.172  1157.4 <2e-16 ***
log(glufast)   1  2.688   2.688   255.6 <2e-16 ***
Residuals    140  1.472   0.011

> summary(aov(m4.diabetes))
                     Df Sum Sq Mean Sq  F value   Pr(>F)
group2                2 24.344  12.172 1274.815  < 2e-16 ***
log(glufast)          1  2.688   2.688  281.484  < 2e-16 ***
group2:log(glufast)   2  0.155   0.077    8.099 0.000472 ***
Residuals           138  1.318   0.010
```

EXAMPLE Potato damage data

Listing 36: `Potato` data: model selection

```
> m.potato<-lm(leak~potato*regime*temp,data=Potato)
> summary(m.potato)

Coefficients:
                                            Estimate Std. Error tval Pr(>|t|)
(Intercept)                                    3.872   3.851   1.005 0.318333
potatoVariety 2                               18.200   4.532   4.016 0.000152 ***
regimeCold Room                               -1.533   4.584  -0.334 0.739129
temp -8 C                                      2.058   5.447   0.378 0.706734
potatoVariety 2:regimeCold Room              -18.121   6.108  -2.967 0.004171 **
potatoVariety 2:temp -8 C                      8.190   6.409   1.278 0.205668
regimeCold Room:temp -8 C                      6.585   6.446   1.022 0.310646
potatoVariety 2:regimeCold Room:temp -8 C -9.444   8.611  -1.097 0.276701

Residual standard error: 8.612 on 67 degrees of freedom
Multiple R-squared: 0.6402,     Adjusted R-squared: 0.6026
F-statistic: 17.03 on 7 and 67 DF,  p-value: 1.015e-12

> summary(aov(m.potato))
                   Df Sum Sq Mean Sq F value   Pr(>F)
potato              1   3420    3420  46.118 3.58e-09 ***
regime              1   1952    1952  26.317 2.67e-06 ***
temp                1   1256    1256  16.930 0.000108 ***
potato:regime       1   2090    2090  28.180 1.35e-06 ***
potato:temp         1     29      29   0.389 0.534719
regime:temp         1      7       7   0.092 0.763128
potato:regime:temp  1     89      89   1.203 0.276701
Residuals          67   4969      74
---

> drop1(m.potato,test="F")
Single term deletions

Model:
leak ~ potato * regime * temp
                   Df Sum of Sq    RSS    AIC F value Pr(>F)
<none>                          4968.9 330.51
potato:regime:temp  1    89.198 5058.1 329.84  1.2027 0.2767
> #Can drop 3-way interaction

> m1.potato<-update(m.potato, ~.-potato:regime:temp)
>
> drop1(m1.potato,test="F")
Single term deletions

Model:
leak ~ potato + regime + temp + potato:regime + potato:temp +
    regime:temp
               Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>                      5058.1 329.84
potato:regime  1   2093.16 7151.2 353.82 28.1402 1.329e-06 ***
potato:temp    1     35.45 5093.5 328.37  0.4765    0.4923
```

```
regime:temp    1      6.79 5064.9 327.94  0.0913    0.7635
---
> #Can drop 2-way interaction regime:temp

> m2.potato<-update(m1.potato, ~ .-regime:temp)
> drop1(m2.potato,test="F")
Single term deletions

Model:
leak ~ potato + regime + temp + potato:regime + potato:temp
              Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>                     5064.9 327.94
potato:regime  1   2094.58 7159.4 351.90 28.5351 1.118e-06 ***
potato:temp    1     28.88 5093.7 326.37  0.3935    0.5326
> #Can drop 2-way interaction potato:temp

> m3.potato<-update(m2.potato, ~ .-potato:temp)
> drop1(m3.potato,test="F")
Single term deletions

Model:
leak ~ potato + regime + temp + potato:regime
              Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>                     5093.7 326.37
temp           1    1225.7 6319.5 340.54  16.844 0.0001082 ***
potato:regime  1    2089.9 7183.7 350.16  28.721 1.015e-06 ***

> summary(m3.potato)

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                       0.8573     2.8719   0.299 0.766194
potatoVariety 2                  22.2955     3.1742   7.024 1.13e-09 ***
regimeCold Room                   1.7709     3.1920   0.555 0.580818
temp -8 C                         8.0874     1.9705   4.104 0.000108 ***
potatoVariety 2:regimeCold Room -22.8538     4.2644  -5.359 1.02e-06 ***

Residual standard error: 8.53 on 70 degrees of freedom
Multiple R-squared: 0.6312,     Adjusted R-squared: 0.6101
F-statistic: 29.95 on 4 and 70 DF,  p-value: 1.591e-14

> summary(aov(m3.potato))
              Df Sum Sq Mean Sq F value   Pr(>F)
potato         1   3420    3420   47.00 2.29e-09 ***
regime         1   1952    1952   26.82 2.05e-06 ***
temp           1   1256    1256   17.25 9.10e-05 ***
potato:regime  1   2090    2090   28.72 1.02e-06 ***
Residuals     70   5094      73
```

EXAMPLE Task Distraction data

Listing 37: `TaskDistract` data: model selection

```
> m.Task<-lm(Errors~Group*Task*Distract,data=TaskDistract)
> summary(aov(m.Task))
                     Df Sum Sq Mean Sq F value   Pr(>F)
Group                 2    355     177   3.664  0.02860 *
Task                  2  28662   14331 296.236  < 2e-16 ***
Distract              1   5747    5747 118.799  < 2e-16 ***
Group:Task            4   1627     407   8.406 5.37e-06 ***
Group:Distract        2    496     248   5.128  0.00733 **
Task:Distract         2   2643    1322  27.320 1.84e-10 ***
Group:Task:Distract   4    143      36   0.739  0.56749
Residuals           117   5660      48

> drop1(m.Task,test="F")
Single term deletions

Model:
Errors ~ Group * Task * Distract
                    Df Sum of Sq    RSS    AIC F value Pr(>F)
<none>                           5660.0 540.35
Group:Task:Distract  4    142.92 5802.9 535.71  0.7386 0.5675
>
> #Can drop 3-way interaction
>
> m1.Task<-update(m.Task, ~.-Group:Task:Distract)
> drop1(m1.Task,test="F")
Single term deletions

Model:
Errors ~ Group + Task + Distract + Group:Task + Group:Distract +
    Task:Distract
               Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>                      5802.9 535.71
Group:Task      4   1824.55 7627.5 564.62  9.5111 1.014e-06 ***
Group:Distract  2    414.36 6217.3 541.03  4.3200   0.01541 *
Task:Distract   2   2643.28 8446.2 582.39 27.5582 1.373e-10 ***

>
> #Maybe drop Group:Distract
> m2.Task<-update(m1.Task, ~.-Group:Distract)
> drop1(m2.Task,test="F")
Single term deletions

Model:
Errors ~ Group + Task + Distract + Group:Task + Task:Distract
              Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>                     6217.3 541.03
Group:Task     4      1754 7971.3 566.57   8.675 3.349e-06 ***
Task:Distract  2      2725 8942.3 586.09  26.955 1.960e-10 ***
```

Listing 38: `TaskDistract` data: final model

```
> summary(m1.Task)


Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     7.677871   8.626059   0.890 0.375190
GroupDel. Smoker              -20.445431   8.963488  -2.281 0.024299 *
GroupAct. Smoker                0.661859   8.751091   0.076 0.939837
TaskCogTask                   -33.547944   8.457839  -3.966 0.000124 ***
TaskDrivSim.                   -2.549221   9.595242  -0.266 0.790941
Distract                        0.015496   0.075933   0.204 0.838638
GroupDel. Smoker:TaskCogTask   13.687948   3.592782   3.810 0.000220 ***
GroupAct. Smoker:TaskCogTask   13.923409   3.649372   3.815 0.000216 ***
GroupDel. Smoker:TaskDrivSim.  -1.570455   3.659027  -0.429 0.668541
GroupAct. Smoker:TaskDrivSim.  -7.424159   3.728939  -1.991 0.048739 *
GroupDel. Smoker:Distract       0.184095   0.077112   2.387 0.018518 *
GroupAct. Smoker:Distract       0.001493   0.074665  -0.020 0.984085
TaskCogTask:Distract            0.455290   0.071333   6.383 3.35e-09 ***
TaskDrivSim.:Distract           0.025430   0.081050   0.314 0.754249

Residual standard error: 6.925 on 121 degrees of freedom
Multiple R-squared: 0.872,      Adjusted R-squared: 0.8582
F-statistic:  63.4 on 13 and 121 DF,  p-value: < 2.2e-16

> summary(aov(m1.Task))
               Df Sum Sq Mean Sq F value   Pr(>F)
Group           2    355     177   3.696  0.02766 *
Task            2  28662   14331 298.818  < 2e-16 ***
Distract        1   5747    5747 119.834  < 2e-16 ***
Group:Task      4   1627     407   8.479 4.59e-06 ***
Group:Distract  2    496     248   5.172  0.00699 **
Task:Distract   2   2643    1322  27.558 1.37e-10 ***
Residuals     121   5803      48

> m3.Task<-lm(Errors~Group+Task+Distract,data=TaskDistract)
> summary(m3.Task)

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             -28.83049    4.66915  -6.175 7.99e-09 ***
GroupDelayed Smoker       4.46773    1.91971   2.327   0.0215 *
GroupActive Smoker        4.40068    1.90928   2.305   0.0228 *
TaskCognitive Task       27.51022    1.91803  14.343  < 2e-16 ***
TaskDriving Simulation   -1.84144    1.91603  -0.961   0.3383
Distract                  0.31619    0.03775   8.375 8.00e-14 ***

Residual standard error: 9.051 on 129 degrees of freedom
Multiple R-squared: 0.7669,     Adjusted R-squared: 0.7578
F-statistic: 84.86 on 5 and 129 DF,  p-value: < 2.2e-16
```

PREDICTION FROM GENERAL LINEAR MODELS

Consider the analysis presented in the Liting 38, which includes in lines 4-19 a summary of parameter estimates obtained for the Task Distraction data. The model being fitted is

$$\texttt{Group} + \texttt{Task} + \texttt{Distract} + \texttt{Group.Task} + \texttt{Task.Distract}$$

where

- `Group` is a factor predictor with three levels
- `Task` is a factor predictor with three levels
- `Distract` is a covariate

We know that `Group` and `Task` are factor predictors as the entries in the parameter estimates table appear in the form displaying a parameter estimate for each level. etc. `R` takes the first level for each factor to be the baseline level; in this case, each factor has three levels, so the baseline treatment group is the one where

$$\texttt{Group} =' \texttt{NonSmoker}' \quad \text{and} \quad \texttt{Task} =' \texttt{Pattern Recognition}'$$

To make a prediction from this model, we need to combine the parameter estimates. We do this by considering the model formula, and extracting one numerical value for each term, remembering to also add in the intercept.

Suppose we want to make predictions for

- **Prediction A:** Subject in Group "`NonSmoker`", doing Task "Driving Simulation", with Distraction level 180.
- **Prediction B:** Subject in Group "`Delayed Smoker`", doing Task "Cognitive Task", with Distraction level 180.
- **Prediction C:** Subject in Group "`Active Smoker`", doing Task "`Pattern Recognition`", with Distraction level 100.

We can go through term by term in the model, as in the following table :

Term	Prediction A	Prediction B	Prediction C
Intercept	7.678	7.678	7.678
`Group`	0.000	-20.445	0.662
`Task`	-2.549	-33.548	0.000
`Distract`	0.015×180	0.015×180	0.015×100
`Group. Task`	0.000	13.688	0.000
`Task. Distract`	0.025×180	0.455×180	0.000×100
Total	12.495	85.251	9.740

This can be achieved in `R` using the `predict` function, and a `newdata` data frame that specifies the values at which the predictions are made. We may compute confidence intervals for the predicted value excluding or including the residual variation.

Listing 39: `TaskDistract` data: predictions

```
> newTask<-data.frame(Group=c("NonSmoker","Delayed Smoker","Active Smoker"),
+      Task=c("Driving Simulation","Cognitive Task","Pattern Recognition"),
+      Distract=c(180,180,100))
> predict(m1.Task,newdata=newTask,interval="confidence")
        fit       lwr      upr
1 12.495286  2.714065 22.27651
2 85.251144 76.356984 94.14530
3  9.740084  5.745879 13.73429
> predict(m1.Task,newdata=newTask,interval="prediction")
        fit       lwr       upr
1 12.495286 -4.346402  29.33697
2 85.251144 68.908676 101.59361
3  9.740084 -4.540107  24.02028
```

Notes:

- when you include covariate terms, remember to multiply the coefficient by the value at which you are predicting; here we multiply the coefficients involving `Distract` by 180 (in Predictions A and B) and 100 (in Prediction C).
- this approach of going through term by term works for all models. For example, if we fitted the model

  $$\texttt{Group} + \texttt{Task} + \texttt{Distract}$$

  which is model `m3.Task` on line 35 of Listing 38, and obtained the parameter estimates

Listing 40: `TaskDistract` data: additive model

```
> summary(m3.Task)

Call:lm(formula = Errors ~ Group + Task + Distract, data = TaskDistract)

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)            -28.83049    4.66915  -6.175 7.99e-09 ***
GroupDelayed Smoker      4.46773    1.91971   2.327   0.0215 *
GroupActive Smoker       4.40068    1.90928   2.305   0.0228 *
TaskCognitive Task      27.51022    1.91803  14.343  < 2e-16 ***
TaskDriving Simulation  -1.84144    1.91603  -0.961   0.3383
Distract                 0.31619    0.03775   8.375 8.00e-14 ***

Residual standard error: 9.051 on 129 degrees of freedom
Multiple R-squared: 0.7669,     Adjusted R-squared: 0.7578
F-statistic: 84.86 on 5 and 129 DF,  p-value: < 2.2e-16
```

then for a prediction in Group '`Delayed Smoker`', Task '`Driving Simulation`' and Distraction 140, we would have

$$-28.830 + 4.466 + (-1.841) + (0.316 \times 140) = 18.035$$

For Predictions A, B and C, we have

Listing 41: `TaskDistract` data: predictions for different model

```
> predict(m3.Task,newdata=newTask,interval="confidence")
        fit       lwr      upr
1 26.242471 19.990763 32.49418
2 60.061855 54.066673 66.05704
3  7.189297  3.607416 10.77118
> predict(m3.Task,newdata=newTask,interval="prediction")
        fit        lwr      upr
1 26.242471   7.274087 45.21085
2 60.061855  41.176465 78.94724
3  7.189297 -11.073934 25.45253
```

Note that predictions from the full factorial model

`Group*Task*Distract`

(fitted as m.Task in Listing 37) do not differ greatly from the predictions in the selected model presented in Listing 38)

Listing 42: `TaskDistract` data: predictions from full model

```
> predict(m.Task,newdata=newTask,interval="confidence")
        fit        lwr      upr
1 12.888733 -0.1956475 25.97311
2 88.540812 77.8687809 99.21284
3  8.468974  4.1220082 12.81594
> predict(m.Task,newdata=newTask,interval="prediction")
        fit       lwr       upr
1 12.888733 -6.109705  31.88717
2 88.540812 71.115777 105.96585
3  8.468974 -5.975254  22.91320
```

- Making predictions near and beyond the end of the Distraction range might lead to inappropriate (for example, negative-valued) predictions. Here the minimum value of Distraction is 64, the maximum is 191.

SUMMARY OF ISSUES IN ANOVA, REGRESSION AND LINEAR MODELLING

1 **Model Assumptions :** The key model assumption is that the residual (measurement) errors are independent and identically distributed Normal random quantities. If this assumption is not met, then none of the hypothesis tests based on the Student and Fisher-F distributions are valid.

The validity of this model assumption can be checked by the inspection of the *residuals*, $\widehat{e}_i$, or *standardized residuals*, $\widehat{z}_i$ where

$$\widehat{e}_i = y_i - \widehat{y}_i \qquad \widehat{z}_i = \frac{\widehat{e}_i}{s} = \frac{y_i - \widehat{y}_i}{s}$$

for $i = 1, \ldots, n$. Plots of the residuals can be used to check for

(i) *Normality* To check the normality of the residuals, a *histogram* or *probability plot* can be used. A probability plot is a plot constructed using the **observed** (standardized) residuals and their **theoretical** counterparts **assuming a normal model**. Points in such a plot should lie on a straight-line with slope one; any deviation from this may indicate deviation from normality.

The examples below are from (a) the Viscosity vs PCV data, and (b) the straight-line and (c) the quadratic model analysis of the Hooker data. In the (a), the probability plot indicates that the residual variance is larger in the middle compared to the of the residual range. In (b), the points in the probability plot do not lie on the straight-line, so again a deviation from normality is indicated. In (c), the plot indicates normality of the residuals.

Listing 43: P-P plots and Q-Q plots

```
m0.visc<-lm(viscosity~pcv,data=Viscosity)
m0.hooker<-lm(temp~pressure,data=Hooker)
m1.hooker<-lm(temp~pressure+I(pressure^2),data=Hooker)

#P-P plots
library(e1071)
probplot(res.visc)
title('P-P plot for Viscosity residuals')
probplot(res.hooker0)
title('P-P plot for Linear fit to Hooker data')
probplot(res.hooker1)
title('P-P plot for Quadratic fit to Hooker data')

#Q-Q plots
qqnorm(res.visc,main='Q-Q plot for Viscosity residuals')
qqline(res.visc)
qqnorm(res.hooker0,main='Q-Q plot for Linear fit to Hooker data')
qqline(res.hooker0)
qqnorm(res.hooker1,main='Q-Q plot for Quadratic fit to Hooker data')
qqline(res.hooker1)
```

The difference between P-P plots and Q-Q plots is that the P-P plots use the observed cumulative probability (that is, plots the observed residuals against the **rank** of the observed residuals divided by $n+1$), whereas Q-Q plots plot theoretical quantiles against observed quantiles from the residual distribution(that is, the $i/(n+1)$ quantiles for $i = 1, \ldots, n$ against the observed residuals sorted into ascending order).

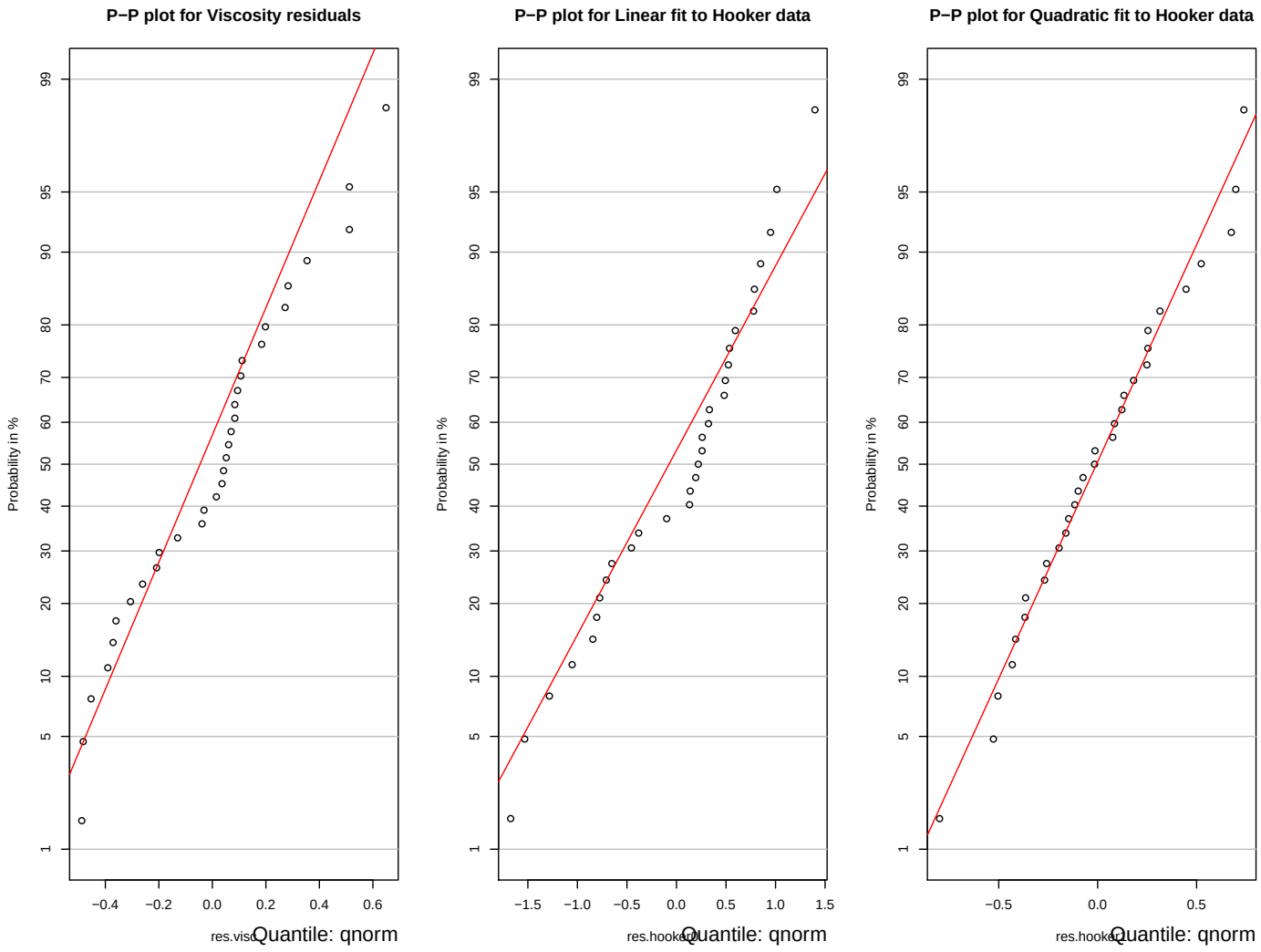


Figure 22: P-P plots

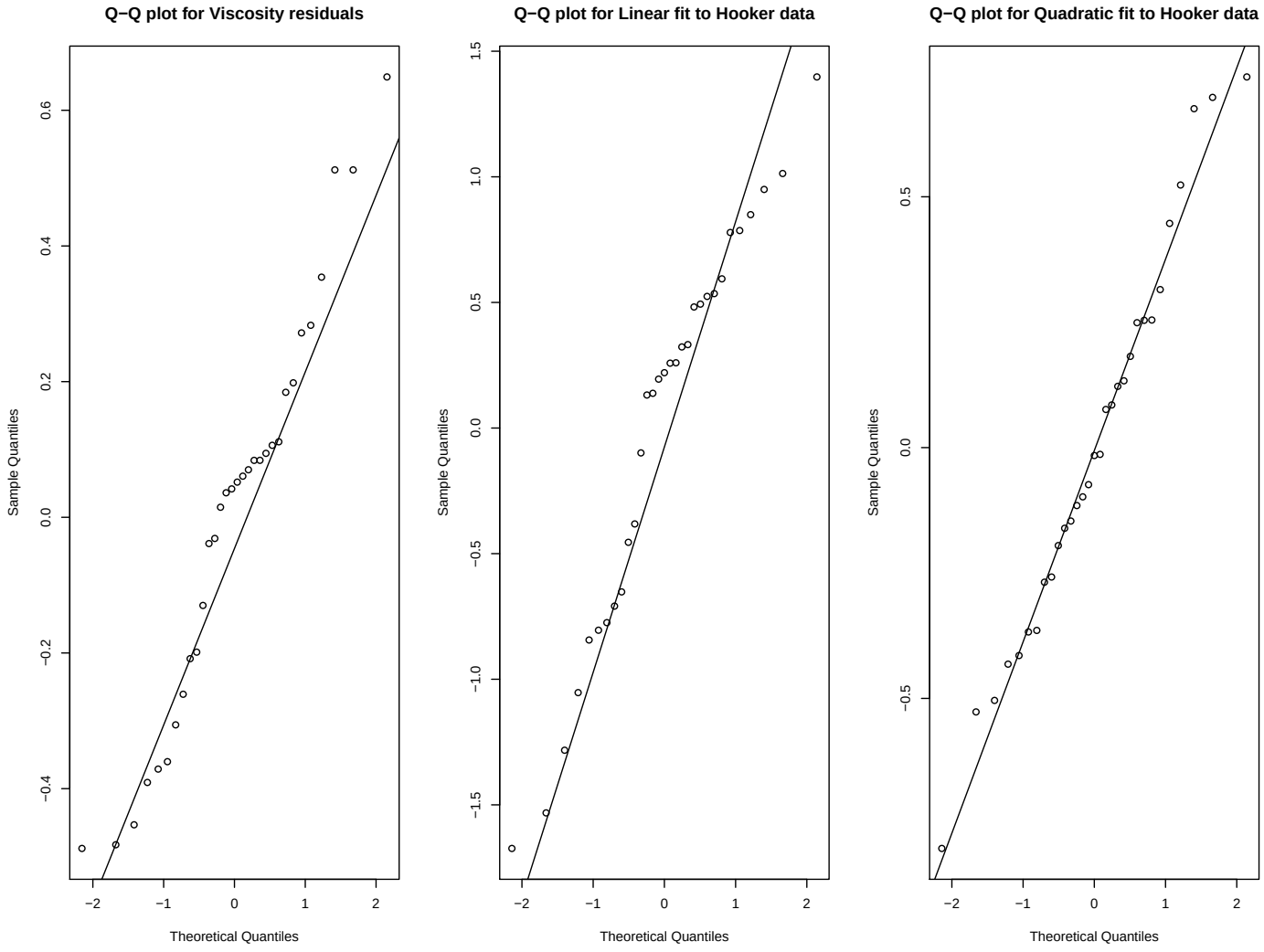


Figure 23: Q-Q plots

(ii) *Dependence* on the covariates

(iii) *Constant variance*

(iv) *Outliers*: An *outlier* is an response value that gives rise to a residual which is large in magnitude, indicating that the fit of the model is poor for that data point.

Outliers can significantly alter the fit of a model, and the parameter estimates. If an outlier is suspected, then careful consideration should be given to omitting that data point from the analysis (see below; estimates (standard errors) change in the presence of an outlier).

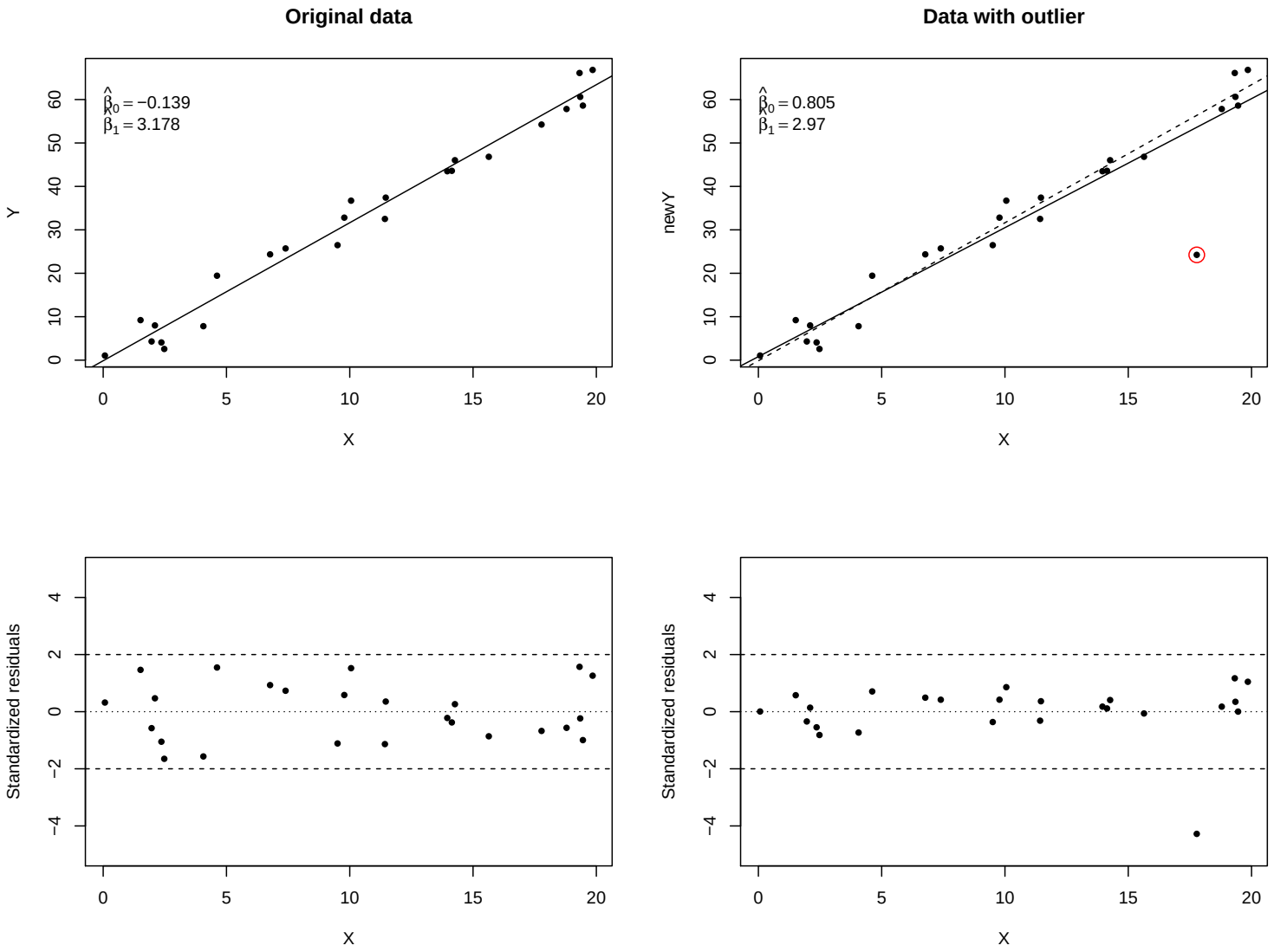


Figure 24: The influence of an outlier on fit.

2 **Data Transformations :** The response variable and continuous covariates can be **transformed** (using log or square-root transformation say) to improve the fit of the model, or to make the model assumptions more appropriate.

3 **Model Selection :** Model selection by means of stepwise selection and sequential ANOVA-F testing can be an effective way of finding the important explanatory variables and interactions. However it must be carried out with care. In general, we aim to select the **simplest model** that provides an

adequate fit to the data. The goodness of fit measures R^2 and adjusted R^2 statistics can provide a final assessment of model adequacy.

4 **Multicollinearity :** *Multicollinearity* is the term to describe dependence between the covariates used in a regression model. If the covariates are highly correlated, then the estimated coefficients for those covariates in a multiple regression need careful interpretation.

If two covariates are highly correlated, then if one is a useful predictor of the response, the other will likely appear to be a useful predictor as well, that is if one estimated coefficient is significantly different from zero, then the other will be also. However, in a multiple regression model with both covariates included, it might be that neither coefficient is significantly different from zero.

5 **Predicting outside the Range of the Covariates :** In a regression model, the fitted parameters reflect relationships and dependencies in the *observed* data. The model can be used for prediction, but is only likely to be reliable if the prediction is carried out at x values within the range of the observed xs.

For example, in a simple linear regression, if x takes values on the range $(0, 100)$, predictions at new x values within this range will be reliable, but predictions at, say, $x = 200$ will be much less reliable.

CHI-SQUARED TESTS FOR CATEGORICAL DATA

In a **multinomial** experiment, the independent experimental units are classified to one of k categories determined by the levels of a discrete factor. Let $n_1, n_2, \ldots, n_k$ be the counts of the numbers of experimental units in the k categories, where $n_1 + n_2 + \cdots + n_k = n$..

The probability that an experimental unit is classified to category i is p_i, for $i = 1, \ldots, k$, so that

$$p_1 + p_2 + \cdots + p_k = 1.$$

- The **one-way** classification table can be displayed as follows:

Category	1	2	$\cdots$	k
Count	n_1	n_2	$\cdots$	n_k
Probability	p_1	p_2	$\cdots$	p_k

  We can test a hypothesis H_0 that fully specifies $p_1, \ldots, p_k$, for example

$$H_0 : p_1 = p_1^{(0)}, p_2 = p_2^{(0)}, \ldots, p_k = p_k^{(0)}$$

  so that, for $k = 3$, we might have

$$H_0 : p_1 = p_2 = p_3 = 1/3 \qquad \text{or} \qquad H_0 : p_1 = 1/2, p_2 = p_3 = 1/4.$$

  We use the test statistic

$$X^2 = \sum_{i=1}^{k} \frac{\left(n_i - np_i^{(0)}\right)^2}{np_i^{(0)}} = \sum_{i=1}^{k} \frac{(\text{Observed Count in Cell } i - \text{Expected Count in Cell } i)^2}{\text{Expected Count in Cell } i}$$

  We sometimes write $\widehat{n}_i = np_i^{(0)}$. If H_0 is true, $X^2 \approx$ Chi-squared$(k-1)$.

- The **two-way** classification table can also be constructed to represent the cross-classification for two discrete factors A and B with r and c levels respectively.

		Factor B			
		1	2	$\cdots$	c
Factor A	1	n_{11}	n_{12}	$\cdots$	n_{1c}
	2	n_{21}	n_{22}	$\cdots$	n_{2c}
	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	r	n_{r1}	n_{r2}	$\cdots$	n_{rc}

  To test the hypothesis

$$H_0 \ : \ \text{Factor A and Factor B levels are assigned independently}$$

  we use the same test statistic that can be rewritten

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - \widehat{n}_{ij})^2}{\widehat{n}_{ij}}$$

  where

$$\widehat{n}_{ij} = \frac{n_{i.} n_{.j}}{n} \qquad n_{i.} = \sum_{j=1}^{c} n_{ij} \qquad n_{.j} = \sum_{i=1}^{r} n_{ij}.$$

  The terms $n_{i.}$ and $n_{.j}$ are the row and column totals for row i and column j respectively. If H_0 is true

$$X^2 \approx \text{Chi-squared}((r-1)(c-1))$$

EXAMPLE DNA Sequence Data
The counts of the numbers of nucleotides (A,C,G,T) in the DNA sequence of the cancer-related gene BRCA 2 are presented in the table below.

Category	1	2	3	4	Total
Nucleotide	A	C	G	T	
Count	38514	24631	25685	38249	127079

so that $k = 4$. To test the hypothesis

$$H_0 \ : \ p_1 = p_2 = p_3 = p_4 = 1/4$$

We use the one-way table chi-squared test: here

$$\widehat{n}_i = np_i^{(0)} = \frac{127079}{4} = 31769.75$$

so the test statistic is

$$\begin{aligned} X^2 &= \frac{(38514 - 31769.75)^2}{31769.75} + \frac{(24631 - 31769.75)^2}{31769.75} + \frac{(25685 - 31769.75)^2}{31769.75} + \frac{(38249 - 31769.75)^2}{31769.75} \\ &= 5522.597 \end{aligned}$$

We compare this with the Chi-squared$(k-1) \equiv$ Chi-squared(3) distribution. From McClave and Sincich tables,

$$\text{Chisq}_{0.05}(3) = 7.815 < X^2$$

so H_0 is **rejected**.

EXAMPLE Eye and Hair Colour Data
The table below contains counts of the number of people in a study with a combination of eye and hair colour.

		Hair				
		Black	Brunette	Red	Blonde	$n_{i.}$
Eyes	Brown	68	119	26	7	220
	Blue	20	84	17	94	215
	Hazel	15	54	14	10	93
	Green	5	29	14	16	64
	$n_{.j}$	108	286	71	127	592

so $r = c = 4$. To test the hypothesis

$$H_0 \ : \ \text{Eye and Hair colour are assigned independently}$$

we use the X^2 statistic

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - \widehat{n}_{ij})^2}{\widehat{n}_{ij}}$$

Here, for example, for $i = 2$ and $j = 3$

$$\widehat{n}_{23} = \frac{n_{2.} \times n_{.3}}{n} = \frac{215 \times 71}{592} = 25.785.$$

In fact, on complete calculation, we find that

$$X^2 = 138.2898.$$

We compare this with the Chi-squared$((r-1)(c-1)) \equiv$ Chi-squared(9) distribution. From McClave and Sincich,

$$\text{Chisq}_{0.05}(9) = 16.919 < X^2$$

so H_0 is **rejected**

Listing 44: Chi-squared testing

```
> #DNA data Chisq Test
> DNA.data<-c(38514,24631,25685,38249)
> DNA.fitted<-rep(1/4,4)*sum(DNA.data)
>
> (Xsq<-sum((DNA.data-DNA.fitted)^2/DNA.fitted))
[1] 5522.597
> (pval<-1-pchisq(Xsq,df=3))
[1] 0
>
> #DNA data Chisq Test: alternative null
> DNA.data<-c(38514,24631,25685,38249)
> DNA.fitted<-c(0.3,0.2,0.2,0.3)*sum(DNA.data)
>
> (Xsq<-sum((DNA.data-DNA.fitted)^2/DNA.fitted))
[1] 31.49232
> (pval<-1-pchisq(Xsq,df=3))
[1] 6.695129e-07
>
>
> ##########################################################
> #Hair & Eye colour
>
> xtabs(count ~ eye+hair,data=HairEyes)
       hair
eye     Black Brown Red Blond
  Brown    68   119  26     7
  Blue     20    84  17    94
  Hazel    15    54  14    10
  Green     5    29  14    16

> chisq.test(xtabs(count ~ eye+hair,data=HairEyes))

        Pearson's Chi-squared test

data:  xtabs(count ~ eye + hair, data = HairEyes)
X-squared = 138.2898, df = 9, p-value < 2.2e-16


> #Direct calculation
> HairEyes.table<-xtabs(count ~ eye+hair,data=HairEyes)
> HE.rowtots<-apply(HairEyes.table,1,sum)
> HE.coltots<-apply(HairEyes.table,2,sum)
>
> #Form the fitted values from the row and column totals
> HE.fitted<-outer(HE.rowtots,HE.coltots,"*")/sum(HairEyes.table)
>
> Xsq<-sum((HairEyes.table-HE.fitted)^2/HE.fitted)
>
> (pval<-1-pchisq(Xsq,df=9))
[1] 0
```

NON-PARAMETRIC STATISTICS

ONE AND TWO SAMPLE TESTS

Non-parametric tests are normally based on **ranks** of the data samples, and test hypotheses relating to **quantiles** of the probability distribution representing the population from which the data are drawn. Specifically, tests concern the **population median**, η, where

$$\Pr[\text{ Observation } \leq \eta\,] = \frac{1}{2}$$

The **sample median**, x_{MED}, is the mid-point of the sorted sample; if the data $x_1, \ldots, x_n$ are sorted into **ascending** order, then

$$x_{\text{MED}} = \begin{cases} x_m & n \text{ odd}, n = 2m+1 \\ \dfrac{x_m + x_{m+1}}{2} & n \text{ even}, n = 2m \end{cases}$$

ONE SAMPLE TEST FOR MEDIAN: THE SIGN TEST

For a single sample of size n, to test the hypothesis $\eta = \eta_0$ for some specified value η_0 we use the **Sign Test.**. The test statistic S depends on the alternative hypothesis, H_a.

(a) For **one-sided** tests, to test

$$\begin{aligned} H_0 &: \eta = \eta_0 \\ H_a &: \eta > \eta_0 \end{aligned}$$

we define test statistic S by

$$S = \text{Number of observations } \textbf{greater than } \eta_0$$

whereas to test

$$\begin{aligned} H_0 &: \eta = \eta_0 \\ H_a &: \eta < \eta_0 \end{aligned}$$

we define S by

$$S = \text{Number of observations } \textbf{less than } \eta_0$$

If H_0 is **true**, it follows that

$$S \sim \text{Binomial}\left(n, \frac{1}{2}\right)$$

The p-value is defined by

$$p = \Pr[X \geq S]$$

where $X \sim \text{Binomial}(n, 1/2)$. The rejection region for significance level α is defined implicitly by the rule

$$\text{Reject } H_0 \text{ if } \alpha \geq p.$$

The Binomial distribution is tabulated in McClave and Sincich.

(b) For a **two-sided** test,

$$\begin{aligned} H_0 &: \eta = \eta_0 \\ H_a &: \eta \neq \eta_0 \end{aligned}$$

we define the test statistic by

$$S = \max\{S_1, S_2\}$$

where S_1 and S_2 are the counts of the number of observations less than, and greater than, η_0 respectively. The p-value is defined by

$$p = 2\Pr[X \geq S]$$

where $X \sim \text{Binomial}(n, 1/2)$.

Notes :

1 The only assumption behind the test is that the data are drawn independently from a continuous distribution.

2 If any data are equal to η_0, we **discard** them before carrying out the test.

3 **Large sample approximation.** If n is large (say $n \geq 30$), and $X \sim \text{Binomial}(n, 1/2)$, then it can be shown that

$$X \dot{\sim} \text{Normal}(np, np(1-p))$$

Thus for the sign test, where $p = 1/2$, we can use the test statistic

$$Z = \frac{S - \frac{n}{2}}{\sqrt{n \times \frac{1}{2} \times \frac{1}{2}}} = \frac{S - \frac{n}{2}}{\sqrt{n} \times \frac{1}{2}}$$

and note that if H_0 is true,

$$Z \dot{\sim} \text{Normal}(0, 1).$$

so that the test at $\alpha = 0.05$ uses the following critical values

$$\begin{aligned} H_a &: \eta > \eta_0 \quad \text{then} \quad C_R = 1.645 \\ H_a &: \eta < \eta_0 \quad \text{then} \quad C_R = -1.645 \\ H_a &: \eta \neq \eta_0 \quad \text{then} \quad C_R = \pm 1.960 \end{aligned}$$

4 For the large sample approximation, it is common to make a **continuity correction**, where we replace S by $S - 1/2$ in the definition of Z

$$Z = \frac{\left(S - \frac{1}{2}\right) - \frac{n}{2}}{\sqrt{n} \times \frac{1}{2}}$$

Tables of the standard Normal distribution are given in McClave and Sincich.

TWO SAMPLE TESTS FOR INDEPENDENT SAMPLES

THE MANN-WHITNEY-WILCOXON TEST

For a two **independent** samples of size n_1 and n_2, to test the hypothesis of **equal population medians**

$$\eta_1 = \eta_2$$

we use the **Wilcoxon Rank Sum Test**, or an equivalent test, the **Mann-Whitney U Test**; we refer to this as the

Mann-Whitney-Wilcoxon (MWW) Test

By convention it is usual to formulate the test statistic in terms of the **smaller** sample size. Without loss of generality, we label the samples such that

$$n_1 > n_2.$$

The test is based on the **sum of the ranks** for the data from sample 2.

EXAMPLE $n_1 = 4, n_2 = 3$ yields the following ranked data

SAMPLE 1	0.31	0.48	1.02	3.11
SAMPLE 2	0.16	0.20	1.97	

SAMPLE	2	2	**1**	**1**	**1**	2	**1**
	0.16	0.20	**0.31**	**0.48**	**1.02**	1.97	**3.11**
RANK	1	2	**3**	**4**	**5**	6	**7**

Thus the rank sum for sample 1 is

$$R_1 = 3 + 4 + 5 + 7 = 19$$

and the rank sum for sample 2 is

$$R_2 = 1 + 2 + 6 = 9.$$

Let η_1 and η_2 denote the medians from the two distributions from which the samples are drawn. We wish to test

$$H_0 \ : \ \eta_1 = \eta_2$$

Two related test statistics can be used

- **Wilcoxon Rank Sum Statistic**
$$W = R_2$$
- **Mann-Whitney U Statistic**
$$U = R_2 - \frac{n_2(n_2+1)}{2}$$

We again consider three alternative hypotheses:

$$\begin{aligned} H_a &: \eta_1 < \eta_2 \\ H_a &: \eta_1 > \eta_2 \\ H_a &: \eta_1 = \eta_2 \end{aligned}$$

and define the rejection region separately in each case.

Large Sample Test

If $n_2 \geq 10$, a large sample test based on the Z statistic

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

can be used. Under the hypothesis $H_0 : \eta_1 = \eta_2$,

$$Z \dot{\sim} \text{Normal}(0, 1)$$

so that the test at $\alpha = 0.05$ uses the following critical values

$$\begin{array}{lll} H_a : \eta_1 > \eta_2 & \text{then} & C_R = -1.645 \\ H_a : \eta_1 < \eta_2 & \text{then} & C_R = 1.645 \\ H_a : \eta_1 \neq \eta_2 & \text{then} & C_R = \pm 1.960 \end{array}$$

Small Sample Test

If $n_1 < 10$, an **exact** but more complicated test can be used. The test statistic is R_2 (the sum of the ranks for sample 2). The null distribution under the hypothesis $H_0 : \eta_1 = \eta_2$ can be computed, but it is complicated.

The table in McClave and Sincich gives the critical values (T_L and T_U) that determine the rejection region for different n_1 and n_2 values up to 10.

- **One-sided tests:**

$$\begin{array}{lll} H_a : \eta_1 > \eta_2 & \text{Rejection Region is} & R_2 \leq T_L \\ H_a : \eta_1 < \eta_2 & \text{Rejection Region is} & R_2 \geq T_U \end{array}$$

  These are tests at the $\alpha = 0.025$ significance level.

- **Two-sided tests:**

$$H_a : \eta_1 \neq \eta_2 \quad \text{Rejection Region is} \quad R_2 \leq T_L \text{ or } R_2 \geq T_U$$

  This is a test at the $\alpha = 0.05$ significance level.

Notes :

1 The only assumption is are needed for the test to be valid is that the samples are independently drawn from two continuous distributions.

2 The sum of the ranks across **both** samples is

$$R_1 + R_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2}$$

3 If there are **ties** (equal values) in the data, then the rank values are replaced by **average** rank values.

DATA VALUE	0.16	0.20	0.31	0.31	0.48	1.97	3.11
ACTUAL RANK	1	2	3	3	5	6	7
AVERAGE RANK	1	2	3.5	3.5	5	6	7

EXAMPLE Sign Test: Water Content Example

The following data are measurements of percentage water content of soil samples collected by two experimenters. We wish to test the hypothesis

$$H_0 \ : \ \eta = 9.0$$

for each experiment.

Experimenter 1:	$n = 10$	5.5	6.0	6.5	7.6	7.6	7.7	8.0	8.2	9.1	15.1
Experimenter 2:	$n = 20$	5.6	6.1	6.3	6.3	6.5	6.6	7.0	7.5	7.9	8.0
		8.0	8.1	8.1	8.2	8.4	8.5	8.7	9.4	14.3	26.0

To perform the test, we need tables of the Binomial distribution with $p = 1/2$. The individual probabilities are given by the formula

$$\Pr[X = x] = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} \frac{1}{2^n} = \frac{n!}{x!(n-x)!} \frac{1}{2^n} \qquad x = 0, 1, \ldots, n$$

We test at the $\alpha = 0.05$ level. For the first experiment, with $n = 10$:

- For a test against the alternative hypothesis

  $$H_a \ : \ \eta > 9.0$$

  the test statistic is

  $$S = \text{Number of observations } \textbf{greater than } 9 \qquad \therefore \qquad S = 2$$

  and the p-value is

  $$p = \Pr[X \geq 2] = 1 - \Pr[X < 2] = 1 - \Pr[X = 0] - \Pr[X = 1] = 0.9893$$

  so we **do not** reject H_0 in favour of this H_a.

- For a test against the alternative hypothesis

  $$H_a \ : \ \eta < 9.0$$

  the test statistic is

  $$S = \text{Number of observations } \textbf{less than } 9 \qquad \therefore \qquad S = 8$$

  and the p-value is

  $$p = \Pr[X \geq 8] = \Pr[X = 8] + \Pr[X = 9] + \Pr[X = 10] = 0.0547$$

  so we **do not** reject H_0 in favour of this H_a.

- For a test against the alternative hypothesis

  $$H_a \ : \ \eta \neq 9.0$$

  the test statistic is

  $$S = \max\{S_1, S_2\} = \max\{2, 8\} = 8$$

  and the p-value is

  $$p = 2\Pr[X \geq 8] = 2(\Pr[X = 8] + \Pr[X = 9] + \Pr[X = 10]) = 0.1094$$

  so we **do not** reject H_0 in favour of this H_a.

For the second experiment, with $n = 20$:

- For a test against the alternative hypothesis $H_a : \eta > 9.0$, the test statistic is $S = 3$. The p-value is therefore

$$p = \Pr[X \geq 3] = 1 - \Pr[X < 3] = 1 - \Pr[X = 0] - \Pr[X = 1] - \Pr[X = 2] = 0.9998.$$

  so we **do not** reject H_0 in favour of this H_a.

- For a test against the alternative hypothesis $H_a : \eta < 9.0$, the test statistic $S = 17$. The p-value is therefore

$$p = \Pr[X \geq 17] = \Pr[X = 17] + \Pr[X = 18] + \Pr[X = 19] + \Pr[X = 20] = 0.0013.$$

  so we **do** reject H_0 in favour of this H_a.

- For a test against the alternative hypothesis $H_a : \eta \neq 9.0$, the test statistic is $S = \max\{S_1, S_2\} = \max\{3, 17\} = 17$. The p-value is therefore

$$p = 2\Pr[X \geq 17] = 2(\Pr[X = 17] + \Pr[X = 18] + \Pr[X = 19] + \Pr[X = 20]) = 0.0026.$$

  so we **do** reject H_0 in favour of this H_a.

Listing 45: Sign test: direct calculation

```
> #Test statistic
> S<-sum(Sign1$h2o>9)
> #First p-value
> 1-pbinom(S-1,nrow(Sign1),0.5)
[1] 0.9892578
> #Second p-value
> pbinom(S,nrow(Sign1),0.5)
[1] 0.0546875
> #Third p-value
> 2*(1-pbinom(max(S,nrow(Sign1)-S)-1,nrow(Sign1),0.5))
[1] 0.109375
```

Listing 46: Sign test:

```
> eta<-9
>binom.test(sum(Sign1$h2o>eta),nrow(Sign1),p=0.5,alternative="greater")

        Exact binomial test

data:  sum(Sign1$h2o > eta) and nrow(Sign1)
number of successes = 2, number of trials = 10, p-value = 0.9893
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.03677144 1.00000000
sample estimates:
probability of success
                   0.2

>binom.test(sum(Sign1$h2o>eta),nrow(Sign1),p=0.5,alternative="less")

        Exact binomial test

data:  sum(Sign1$h2o > eta) and nrow(Sign1)
number of successes = 2, number of trials = 10, p-value = 0.05469
alternative hypothesis: true probability of success is less than 0.5
95 percent confidence interval:
 0.0000000 0.5069013
sample estimates:
probability of success
                   0.2

>binom.test(sum(Sign1$h2o>eta),nrow(Sign1),p=0.5,alternative="two.sided")

        Exact binomial test

data:  sum(Sign1$h2o > eta) and nrow(Sign1)
number of successes = 2, number of trials = 10, p-value = 0.1094
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.02521073 0.55609546
sample estimates:
probability of success
                   0.2
```

Listing 47: Mann-Whitney-Wilcoxon test: direct calculation

```
> #Mann-Whitney-Wilcoxon test
> Sample1<-c(0.31,0.48,1.02,3.11)
> Sample2<-c(0.16,0.20,1.97)
> Y<-c(Sample1,Sample2)
> Samp<-rep(c(1,2),c(4,3))
> rY<-rank(Y)
> (R1<-sum(rY[Samp==1]))
[1] 19
> (R2<-sum(rY[Samp==2]))
[1] 9
```

Listing 48: Mann-Whitney-Wilcoxon test

```
> wilcox.test(Y ~ Samp)
        Wilcoxon rank sum test

data:  Y by Samp
W = 9, p-value = 0.4
alternative hypothesis: true location shift is not equal to 0

> wilcox.test(Y ~ Samp,alternative="less")
        Wilcoxon rank sum test

data:  Y by Samp
W = 9, p-value = 0.8857
alternative hypothesis: true location shift is less than 0

> wilcox.test(Y ~ Samp,alternative="greater")
        Wilcoxon rank sum test

data:  Y by Samp
W = 9, p-value = 0.2
alternative hypothesis: true location shift is greater than 0


#Using the coin library
> require(coin)
> wilcox_test(Y ~ as.factor(Samp)) #Asymptotic approximation

        Asymptotic Wilcoxon Mann-Whitney Rank Sum Test

data:  Y by as.factor(Samp) (1, 2)
Z = 1.0607, p-value = 0.2888
alternative hypothesis: true mu is not equal to 0

> wilcox_test(Y ~ as.factor(Samp),distribution="exact") #Exact null

        Exact Wilcoxon Mann-Whitney Rank Sum Test

data:  Y by as.factor(Samp) (1, 2)
Z = 1.0607, p-value = 0.4
alternative hypothesis: true mu is not equal to 0

> wilcox_test(Y ~ as.factor(Samp),distribution="approximate")
#Approximate null using permutation Monte Carlo

        Approximative Wilcoxon Mann-Whitney Rank Sum Test

data:  Y by as.factor(Samp) (1, 2)
Z = 1.0607, p-value = 0.396
alternative hypothesis: true mu is not equal to 0
```

EXAMPLE Mann-Whitney-Wilcoxon Test: Low Birthweight Example
The birthweights (in grammes) of babies born to two groups of mothers A and B are displayed below: Thus $n_1 = 9, n_2 = 8$. From this sample (which has ties, so we need to use average ranks), so that the two

Group A :	$n = 9$	2164	2600	2184	2080	1820	2496	2184	2080	2184
Group B :	$n = 8$	2576	3224	2704	2912	2444	3120	2912	3848	

statistics are $R_1 = 48, R_2 = 105$, and

$$\text{Wilcoxon} \quad W = R_2 = 105$$

$$\text{Mann-Whitney} \quad U = R_2 - \frac{n_2(n_2+1)}{2} = 105 - 36 = 69$$

- For the **small sample** test, from tables in McClave and Sincich, we find

$$T_L = 51 \qquad T_U = 93$$

Thus $W > 93$, so we

$$\begin{aligned} \textbf{Do not} \text{ reject } H_0 \text{ against } H_a &: \eta_1 > \eta_2 \quad \text{as} \quad W = R_2 > T_L \\ \textbf{Reject } H_0 \text{ against } H_a &: \eta_1 < \eta_2 \quad \text{as} \quad W = R_2 > T_U \\ \textbf{Reject } H_0 \text{ against } H_a &: \eta_1 \neq \eta_2 \quad \text{as} \quad W = R_2 > T_U \end{aligned}$$

Note that the one-sided tests are carried out at $\alpha = 0.025$, the two sided test is carried out at $\alpha = 0.05$.

- For the **large sample** test, we find

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} = 3.175$$

Thus we

$$\begin{aligned} \textbf{Do not} \text{ reject } H_0 \text{ against } H_a &: \eta_1 > \eta_2 \quad \text{as} \quad Z > C_R = -1.645 \\ \textbf{Reject } H_0 \text{ against } H_a &: \eta_1 < \eta_2 \quad \text{as} \quad Z > C_R = 1.645 \\ \textbf{Reject } H_0 \text{ against } H_a &: \eta_1 \neq \eta_2 \quad \text{as} \quad Z > C_{R_2} = 1.960 \end{aligned}$$

All tests are carried out at $\alpha = 0.05$.

Listing 49: Birthweight data: Mann-Whitney-Wilcoxon test

```
> attach(Birthweight)
> (R1<-sum(rank(BW)[gp==1]))
[1] 58
> (R2<-sum(rank(BW)[gp==2]))
[1] 95
> n1<-sum(gp==1);n2<-sum(gp==2)
> detach(Birthweight)
> W<-R2
> U<-W-0.5*n2*(n2+1)
> Z<-(U-0.5*n1*n2)/sqrt(n1*n2*(n1+n2+1)/12)
```

Listing 50: Birthweight data: Mann-Whitney-Wilcoxon test

```
> wilcox.test(BW ~ Group,data=Birthweight)
        Wilcoxon rank sum test with continuity correction

data:  BW by Group
W = 3, p-value = 0.002032
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(x = c(2164L, 2600L, 2184L, 2080L, 1820L,  :
  cannot compute exact p-value with ties

> wilcox.test(BW ~ Group,data=Birthweight,alternative="greater")
        Wilcoxon rank sum test with continuity correction

data:  BW by Group
W = 3, p-value = 0.9993
alternative hypothesis: true location shift is greater than 0

> wilcox.test(BW ~ Group,data=Birthweight,alternative="less")
        Wilcoxon rank sum test with continuity correction

data:  BW by Group
W = 3, p-value = 0.001016
alternative hypothesis: true location shift is less than 0



>wilcox_test(BW ~ Group,data=Birthweight,dist="exact",ties.method="mid-ranks")

        Exact Wilcoxon Mann-Whitney Rank Sum Test

data:  BW by Group (A, B)
Z = -3.1344, p-value = 0.0006684
alternative hypothesis: true mu is not equal to 0

>wilcox_test(BW ~ Group,data=Birthweight,dist="exact",alternative="greater")

        Exact Wilcoxon Mann-Whitney Rank Sum Test

data:  BW by Group (A, B)
Z = -3.1344, p-value = 0.9998
alternative hypothesis: true mu is greater than 0

>wilcox_test(BW ~ Group,data=Birthweight,dist="exact",alternative="less")

        Exact Wilcoxon Mann-Whitney Rank Sum Test

data:  BW by Group (A, B)
Z = -3.1344, p-value = 0.0003085
alternative hypothesis: true mu is less than 0
```

EXAMPLE Mann-Whitney-Wilcoxon Test: Treadmill Test Example
The treadmill stress test times (in seconds) of two groups of patients (disease group and healthy controls) are displayed below:

Disease :	$n = 10$	864	636	638	708	786	600	1320	750	594	750
Healthy :	$n = 8$	1014	684	810	990	840	978	1002	1110		

Thus $n_1 = 10, n_2 = 8$. From this sample (which has ties, so we need to use average ranks), we find that

$$R_1 = 70 \qquad R_2 = 101$$

so that the two statistics are

$$\text{Wilcoxon} \quad W = R_2 = 101$$

$$\text{Mann-Whitney} \quad U = R_2 - \frac{n_2(n_2+1)}{2} = 101 - 36 = 65$$

- For the **small sample** test, from tables in McClave and Sincich, we find

  $$T_L = 54 \qquad T_U = 98$$

  Thus $W > 98$, so we

  Do not reject H_0 against $H_a : \eta_1 > \eta_2$ as $W = R_2 > T_L$
  Reject H_0 against $H_a : \eta_1 < \eta_2$ as $W = R_2 > T_U$
  Reject H_0 against $H_a : \eta_1 \neq \eta_2$ as $W = R_2 > T_U$

  Again, the one-sided tests are carried out at $\alpha = 0.025$, the two sided test is carried out at $\alpha = 0.05$.

- For the **large sample** test, we find

  $$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} = 2.221$$

  Thus we

  Do not reject H_0 against $H_a : \eta_1 > \eta_2$ as $Z > C_R = -1.645$
  Reject H_0 against $H_a : \eta_1 < \eta_2$ as $Z > C_R = 1.645$
  Reject H_0 against $H_a : \eta_1 \neq \eta_2$ as $Z > C_{R_2} = 1.960$

  All tests are carried out at $\alpha = 0.05$.

Listing 51: `Treadmill` data: Mann-Whitney-Wilcoxon test

```
> wilcox.test(Time ~ Group,data=Treadmill)

        Wilcoxon rank sum test with continuity correction

data:  Time by Group
W = 65, p-value = 0.02941
alternative hypothesis: true location shift is not equal to 0

> wilcox.test(Time ~ Group,data=Treadmill,alternative="greater")

        Wilcoxon rank sum test with continuity correction

data:  Time by Group
W = 65, p-value = 0.0147
alternative hypothesis: true location shift is greater than 0

> wilcox.test(Time ~ Group,data=Treadmill,alternative="less")

        Wilcoxon rank sum test with continuity correction

data:  Time by Group
W = 65, p-value = 0.9883
alternative hypothesis: true location shift is less than 0




>wilcox_test(Time ~ as.factor(Group),data=Treadmill,dist="exact",
        ties.method="mid-ranks")

        Exact Wilcoxon Mann-Whitney Rank Sum Test

data:  Time by as.factor(Group) (1, 2)
Z = 2.2225, p-value = 0.02491
alternative hypothesis: true mu is not equal to 0

>wilcox_test(Time ~ as.factor(Group),data=Treadmill,dist="exact",
        alternative="greater")

        Exact Wilcoxon Mann-Whitney Rank Sum Test

data:  Time by as.factor(Group) (1, 2)
Z = 2.2225, p-value = 0.0125
alternative hypothesis: true mu is greater than 0

>wilcox_test(Time ~ as.factor(Group),data=Treadmill,dist="exact",
        alternative="less")

        Exact Wilcoxon Mann-Whitney Rank Sum Test

data:  Time by as.factor(Group) (1, 2)
Z = 2.2225, p-value = 0.9891
alternative hypothesis: true mu is less than 0
```

TWO DEPENDENT SAMPLES: WILCOXON SIGNED RANK TEST

Data collected from the same experimental units are in general **dependent**. For example, if data are collected on two occasions (time 1 and time 2, or before and after treatment) from the same n individuals, then the resulting data samples $(y_{11}, \ldots, y_{n1})$ and $(y_{12}, \ldots, y_{n2})$ are dependent. Such data are often referred to as **paired**. We wish to test whether there is a significant change across the two measurements.

For a **parametric** test, we typically assume that the within-individual differences

$$x_i = y_{i1} - y_{i2} \qquad i = 1, \ldots, n$$

are **Normally** distributed, and test the hypothesis that the mean difference μ is zero

$$H_0 \ : \ \mu = 0$$

using a one-sample Z-test (σ known) or T-test (σ unknown), with statistic

$$z = \frac{\overline{x}}{\sigma/\sqrt{n}} \qquad \text{or} \qquad t = \frac{\overline{x}}{s/\sqrt{n}}$$

distributed as Normal$(0, 1)$ or Student$(n-1)$ respectively.

For a **non-parametric** test, we can use the **Wilcoxon Signed Rank** test, which proceeds as follows:

1 Compute the within-individual differences

$$x_i = y_{i1} - y_{i2} \qquad i = 1, \ldots, n$$

If any $x_i = 0$, then that data point is discarded and the sample size adjusted.

2 Sort the **absolute values** $s_1, \ldots, s_n$ of $x_1, x_2, \ldots, x_n$ into **ascending** order, and assign ranks 1 up to n. If there are ties, assign **average** ranks.

3 Form the two rank sums T_+ and T_-, where

$$T_+ = \text{Sum of ranks for those } x_i > 0 \qquad T_- = \text{Sum of ranks for those } x_i < 0$$

The test statistic is a function of these rank sums. Heuristically, if the statistic T_+ is large and T_- is small, this implies that the experimental units where $y_{i1} > y_{i2}$ have a **larger** (in magnitude) difference than those where $y_{i1} < y_{i2}$. This indicates an overall **decrease** between the first and second measurements. Conversely, if the statistic T_- is large and T_+ is small, this implies that the experimental units where $y_{i2} > y_{i1}$ have a **larger** (in magnitude) difference than those where $y_{i2} < y_{i1}$. This indicates an overall **increase** between the first and second measurements.

We test the null hypothesis

$$H_0 \ : \ \text{No change between first and second measurements}$$

against the three alternative hypotheses

$$\begin{aligned} (1)\ H_a &: \text{Significant } \textbf{decrease} \text{ between first and second measurements} \\ (2)\ H_a &: \text{Significant } \textbf{increase} \text{ between first and second measurements} \\ (3)\ H_a &: \text{Significant } \textbf{change} \text{ between first and second measurements} \end{aligned}$$

To test H_0 vs (1), we perform a one-sided test using the statistic T_-; the critical value in the test is denoted T_0, and is determined by the table in McClave and Sincich:

$$\text{If } T_- \leq T_0, \text{ we } \textbf{reject } H_0 \text{ in favour of } H_a \text{ (1)}$$

To test H_0 vs (2), we perform a one-sided test using the statistic T_+; the critical value is T_0 and

$$\text{If } T_+ \leq T_0, \text{ we } \textbf{reject } H_0 \text{ in favour of } H_a \text{ (2)}$$

To test H_0 vs (3), we perform a two-sided test using the statistic $T = \min\{T_-, T_+\}$; the critical value is T_0

$$\text{If } T \leq T_0, \text{ we } \textbf{reject } H_0 \text{ in favour of } H_a \text{ (3)}$$

Notes :

1 The only assumption behind the test is that the difference data x_i are drawn independently from a continuous distribution.

2 **Large Sample Test:** For $n \geq 25$, we can use a large sample version of the test based on T_+, and the Z statistic

$$Z = \frac{T_+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

If H_0 is **true**, then $Z \dot{\sim} \text{Normal}(0, 1)$, so that the test at $\alpha = 0.05$ uses the following critical values

$$\begin{array}{lll} \text{For } H_a \text{ (1)} & \text{use} & C_R = 1.645 \\ \text{For } H_a \text{ (2)} & \text{use} & C_R = -1.645 \\ \text{For } H_a \text{ (3)} & \text{use} & C_R = \pm 1.960 \end{array}$$

EXAMPLE Haemodialysis Data
The following data are measurements of the heparin cofactor II (HCII) to plasma protein ratios in a group of patients at baseline and five months after haemodialysis.

Reference: Toulon, P *et al*. (1987) Antithrombin III and heparin cofactor II in patients with chronic renal failure undergoing regular hemodialysis, *Thrombosis and Haemostasis*, **3**;57(3): pp263-8.

Patient	Before	After					
	y_{i1}	y_{i2}	x_i	s_i	Rank	Ave. Rank	
1	2.11	2.15	-0.04	0.04	3	3.5	
2	1.85	2.11	-0.26	0.26	10	10.0	
3	1.82	1.93	-0.11	0.11	8	8.0	
4	1.75	1.83	-0.08	0.08	6	6.0	
5	1.54	1.90	-0.36	0.36	11	11.0	
6	1.52	1.56	-0.04	0.04	3	3.5	
7	1.49	1.44	0.05	0.05	5	5.0	
8	1.44	1.43	0.01	0.01	1	1.5	
9	1.38	1.28	0.10	0.10	7	7.0	
10	**1.30**	**1.30**	**0.00**	**0.00**	**-**	**-**	**OMIT**
11	1.20	1.21	-0.01	0.01	1	1.5	
12	1.19	1.30	-0.11	0.11	9	9.0	
						$T_+ = 13.5$	
						$T_- = 52.5$	

From the table on p 839, for $n = 12-1 = 11$, we find that the $\alpha = 0.025/0.05$ (one/two-sided) significance level critical value is $T_0 = 11$. Thus using T_+, we **cannot reject** either of the null hypotheses (2) and (3), as $T_+ > T_0$. Note that $Z = -1.734$, so if the approximation was valid, we would be able to reject (2) at $\alpha = 0.05$.

Listing 52: `Haemo` data: Wilcoxon signed rank test

```
> #Haemodialysis data
> wilcox.test(Haemo$after,Haemo$before,paired=T)

        Wilcoxon signed rank test with continuity correction

data:  Haemo$after and Haemo$before
V = 52.5, p-value = 0.09084
alternative hypothesis: true location shift is not equal to 0

Warning messages:
1: In wilcox.test.default(Haemo$after, Haemo$before, paired = T) :
  cannot compute exact p-value with ties
2: In wilcox.test.default(Haemo$after, Haemo$before, paired = T) :
  cannot compute exact p-value with zeroes
> wilcox.test(Haemo$after,Haemo$before,paired=T,alternative="greater")

        Wilcoxon signed rank test with continuity correction

data:  Haemo$after and Haemo$before
V = 52.5, p-value = 0.04542
alternative hypothesis: true location shift is greater than 0

Warning messages:
1: In wilcox.test.default(Haemo$after, Haemo$before, paired = T,
        alternative = "greater") :
  cannot compute exact p-value with ties
2: In wilcox.test.default(Haemo$after, Haemo$before, paired = T,
        alternative = "greater") :
  cannot compute exact p-value with zeroes

>wilcox.test(Haemo$after,Haemo$before,paired=T,alternative="less")

        Wilcoxon signed rank test with continuity correction

data:  Haemo$after and Haemo$before
V = 52.5, p-value = 0.9625
alternative hypothesis: true location shift is less than 0


> #Direct calculation
> xi<-Haemo$after-Haemo$before
> xi<-xi[xi!=0]
> si<-abs(xi)
> n<-length(si)
>
> Tplus<-sum(rank(si)[xi > 0])
> Tminus<-sum(rank(si)[xi < 0])
>
> (Z<-(Tplus-0.25*n*(n+1))/sqrt(n*(n + 1)*(2*n + 1)/24))
[1] 1.733761
> pnorm(Z)
[1] 0.9585198
```

Listing 53: Asthma data: Wilcoxon signed rank test

```
> #PEFRAsthma
> wilcox.test(PEFRAsthma$after,PEFRAsthma$before,paired=T)

        Wilcoxon signed rank test with continuity correction

data:  PEFRAsthma$after and PEFRAsthma$before
V = 1, p-value = 0.01277
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(PEFRAsthma$after, PEFRAsthma$before, paired = T) :
  cannot compute exact p-value with ties

>wilcox.test(PEFRAsthma$after,PEFRAsthma$before,paired=T,alternative="greater")

        Wilcoxon signed rank test with continuity correction

data:  PEFRAsthma$after and PEFRAsthma$before
V = 1, p-value = 0.9955
alternative hypothesis: true location shift is greater than 0

Warning message:
In wilcox.test.default(PEFRAsthma$after, PEFRAsthma$before, paired = T,  :
  cannot compute exact p-value with ties

>wilcox.test(PEFRAsthma$after,PEFRAsthma$before,paired=T,alternative="less")

        Wilcoxon signed rank test with continuity correction

data:  PEFRAsthma$after and PEFRAsthma$before
V = 1, p-value = 0.006386
alternative hypothesis: true location shift is less than 0

Warning message:
In wilcox.test.default(PEFRAsthma$after, PEFRAsthma$before, paired = T,  :
  cannot compute exact p-value with ties


> #Direct calculation
> xi<-PEFRAsthma$after-PEFRAsthma$before
> xi<-xi[xi!=0]
> si<-abs(xi)
> n<-length(si)
>
> Tplus<-sum(rank(si)[xi > 0])
> Tminus<-sum(rank(si)[xi < 0])
>
> (Z<-(Tplus-0.25*n*(n+1))/sqrt(n*(n + 1)*(2*n + 1)/24))
[1] -2.5471
> pnorm(Z)
[1] 0.005431112
```

THREE OR MORE INDEPENDENT SAMPLES

We now seek non-parametric tests that can be used for multiple independent samples, such as those found in the Completely Randomized Design (CRD) and Randomized Block Design (RBD) described in the ANOVA section. The non-parametric equivalents of the Fisher-F tests for these two designs are

- The **Kruskal-Wallis *H* test** for a Completely Randomized Design
- **Friedman's test** for a Randomized Block Design

KRUSKAL-WALLIS TEST

In a CRD, we have k independent groups, corresponding to k different treatments, with sample sizes $n_1, \ldots, n_k$. Let $n = n_1 + \cdots + n_k$. To compute the test statistic, H, we

1 Pool the data, sort them into ascending order, and assign ranks. If there are ties in the data, then average ranks are used.

2 For $j = 1, \ldots, k$, compute the rank sum R_j

$$R_j = \text{Sum of ranks for data from sample } j.$$

To test the hypothesis

H_0 : No difference between the population distributions of the k groups
H_a : At least two population distributions different

the test statistic is

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

If H_0 is **true**, then for large n,

$$H \dot{\sim} \text{Chisquared}(k-1).$$

Notes :

1 The test assumes that the k samples are independently drawn from continuous populations.

2 For the approximation to be valid, there should be at least **five** observations in each sample, and the number of ties should be small.

EXAMPLE Mucociliary efficiency data
The data are measures of mucociliary efficiency from the rate of removal of dust in normal subjects (Group 1), subjects with obstructive airway disease (Group 2), and subjects with asbestosis (Group 3).

Reference: Myles Hollander, M and Douglas A. Wolfe (1973), *Nonparametric statistical inference*, New York: John Wiley & Sons. pp115-120.

Group	1	1	1	1	1	2	2	2	2	3	3	3	3	3
y	2.9	3.0	2.5	2.6	3.2	3.8	2.7	4.0	2.4	2.8	3.4	3.7	2.2	2.0
Rank	8	9	4	5	10	13	6	14	3	7	11	12	2	1

Hence $R_1 = 36$, $R_2 = 36$ and $R_3 = 33$, and the test statistic $H = 0.7714$. To complete the test, we compare with the $\alpha = 0.05$ quantile of the Chisquared$(k-1)$ = Chisquared(2) distribution. We have

$$\text{Chisq}_{0.05}(2) = 5.99 > H \qquad \therefore \qquad \textbf{No evidence to reject } H_0$$

and a p-value of $p = 0.680$.

FRIEDMAN TEST

In a RBD, we have k treatment groups, and a blocking factor. For example, we might have k repeated measurements on the same b experimental units, and $n = bk$ observations in total. To compute the test statistic, F_r, we proceed as follows.

1 **Within each block separately**, sort the k data values into ascending order, and assign ranks. If there are ties in the data, then average ranks are used.

2 For $j = 1, \ldots, k$, compute the rank sum R_j

$$R_j = \text{Sum of ranks for data from } \textbf{treatment } j.$$

To test the hypothesis

H_0 : No difference between the population distributions of the k treatment groups

H_a : At least two population distributions different

the test statistic is

$$F_r = \frac{12}{bk(k+1)} \sum_{j=1}^{k} R_j^2 - 3b(k+1)$$

If H_0 is **true**, then for large n,

$$F_r \dot{\sim} \text{Chisq}(k-1)$$

Notes :

1 The test assumes that the data are drawn independently from continuous populations, with random assignment of treatments within blocks.

2 For the approximation to be valid, it is recommended that b or k is at least five, and the number of ties should be small.

EXAMPLE Skin potential under hypnosis
A study was conducted to investigate whether hypnosis has the same effect on skin potential for four different emotions. Eight subjects were asked to display fear, joy, sadness and calmness under hypnosis, and the resulting skin potential (measured in millivolts) was recorded for each emotion. Thus in this experiment, $b = 8$ and $k = 4$.

	Fear		Joy		Sadness		Calmness	
Subject	y	Rank	y	Rank	y	Rank	y	Rank
1	23.1	4	22.7	3	22.5	1	22.6	2
2	57.6	4	53.2	2	53.7	3	53.1	1
3	10.5	3	9.7	2	10.8	4	8.3	1
4	23.6	4	19.6	3	21.1	2	21.6	1
5	11.9	1	13.8	4	13.7	3	13.3	2
6	54.6	4	47.1	3	39.2	2	37.0	1
7	21.0	4	13.6	1	13.7	2	14.8	3
8	20.3	3	23.6	4	16.3	2	14.8	1
Rank Sum		27		20		19		14

Thus the within-treatment rank sums are $R_1 = 27, R_2 = 20, R_3 = 19$ and $R_4 = 14$ and thus $F_r = 6.45$. To complete the test, we compare with the $\alpha = 0.05$ quantile of the

$$\text{Chisquared}(k-1) = \text{Chisquared}(3)$$

distribution. We have

$$\text{Chisq}_{0.05}(3) = 7.81 > F_r \qquad \therefore \qquad \textbf{No evidence to reject } H_0$$

and a p-value of $p = 0.092$.

Listing 54: Kruskal-Wallis test

```
> #Mucociliary data
> #Mean ranks
> xtabs(rank(y) ~ group,data=Muco)/xtabs( ~ group,data=Muco)
group
                    Healthy Obstructive airway disease
                        7.2                         9.0
                 Asbestosis
                        6.6
> kruskal.test(y ~ group,data=Muco)

        Kruskal-Wallis rank sum test

data:  y by group
Kruskal-Wallis chi-squared = 0.7714, df = 2, p-value = 0.68

>
> #MemoryTask
> #Mean ranks
> xtabs(rank(Memory) ~ Task,data=Memory)/xtabs( ~ Task,data=Memory)
Task
   Counting     Rhyming   Adjective     Imagery Intentional
      12.95       13.10       31.50       36.60       33.35
>
> kruskal.test(Memory ~ Task,data=Memory)

        Kruskal-Wallis rank sum test

data:  Memory by Task
Kruskal-Wallis chi-squared = 25.3761, df = 4, p-value = 4.227e-05
```

Listing 55: Friedman test

```
> #Skin potential data (Hypnosis)
>
> friedman.test(SkinResponse ~ EmoCode | Subject,data=Hypnosis)

        Friedman rank sum test

data:  SkinResponse and EmoCode and Subject
Friedman chi-squared = 6.45, df = 3, p-value = 0.09166


>
> #Soil
> friedman.test(sulphur ~ solvent | soil,data=Soil)

        Friedman rank sum test

data:  sulphur and solvent and soil
Friedman chi-squared = 1.08, df = 3, p-value = 0.7819
```

RANK CORRELATION

SPEARMAN'S RANK CORRELATION

A measure of association for two samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ is the **Pearson Product Moment Correlation Coefficient**, r, where

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}\, SS_{yy}}}$$

where

$$SS_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad SS_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2 \qquad SS_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

This quantity measures the **linear** association between the X and Y variables. A measure of the potentially non-linear association between the samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ is the **Spearman Rank Correlation Coefficient**, r_S, which computes the correlation between the **ranks** of the data.

The Spearman Rank Correlation Coefficient is computed as follows:

1. Assign ranks $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ to the data $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ separately by sorting each sample into ascending order and assigning the ranks in order.

2. Compute r_S as

$$r_S = \frac{SS_{uv}}{\sqrt{SS_{uu}\, SS_{vv}}}$$

   where

$$SS_{uu} = \sum_{i=1}^{n}(u_i - \overline{u})^2 \qquad SS_{vv} = \sum_{i=1}^{n}(v_i - \overline{v})^2 \qquad SS_{uv} = \sum_{i=1}^{n}(u_i - \overline{u})(v_i - \overline{v})$$

If there are no ties in the data, then

$$r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad \text{where} \qquad d_i = u_i - v_i \qquad i = 1, \ldots, n$$

Tests for r_S : If the population correlation is ρ, then we may test the hypothesis

$$H_0 \ : \ \rho = 0$$

against the hypotheses

$$\begin{aligned} (1)\ H_a &: \ \rho > 0 \\ (2)\ H_a &: \ \rho < 0 \\ (3)\ H_a &: \ \rho \neq 0 \end{aligned}$$

using the table of the null distribution in McClave and Sincich. If Spearman_α is the α tail quantile of the null distribution, we have the following rejection regions:

$$\begin{aligned} (1) &: \text{Reject } H_0 \text{ if } r_S > \text{Spearman}_\alpha \\ (2) &: \text{Reject } H_0 \text{ if } r_S < -\text{Spearman}_\alpha \\ (3) &: \text{Reject } H_0 \text{ if } |r_S| > \text{Spearman}_{\alpha/2} \end{aligned}$$

EXAMPLE Latitude and dizygotic twinning rates
The relationship between the geographical latitude of a country and its dizygotic twinning (DZT) rate is to be investigated. The data are presented and plotted below.

Reference: James, W.H. (1985) Dizygotic twinning, birth weight and latitude, *Annals of Human Biology*, **12**, 5, pp. 441-447.

Country	Latitude x	Rank u	DZT Rate y	Rank v
Portugal	40	1.5	6.5	2.0
Greece	40	1.5	8.8	13.0
Spain	41	3.0	5.9	1.0
Bulgaria	42	4.0	7.0	3.0
Italy	44	5.0	8.6	11.5
France	47	6.5	7.1	4.0
Switzerland	47	6.5	8.1	7.5
Austria	48	8.0	7.5	6.0
Belgium	51	9.5	7.3	5.0
FR Germany	51	9.5	8.2	9.0
Holland	52	11.5	8.1	7.5
GDR	52	11.5	9.1	16.0
England & Wales	53	13.5	8.9	14.5
Ireland	53	13.5	11.0	18.0
Scotland	56	15.5	8.9	14.5
Denmark	56	15.5	9.6	17.0
Sweden	60	17.0	8.6	11.5
Norway	61	18.0	8.3	10.0
Finland	62	19.0	12.1	19.0

For these data

$$r_S = \frac{SS_{uv}}{\sqrt{SS_{uu}\,SS_{vv}}} = \frac{384.5}{\sqrt{567 \times 568.5}} = 0.677 \qquad r = \frac{SS_{xy}}{\sqrt{SS_{xx}\,SS_{yy}}} = \frac{118.4}{\sqrt{866.105 \times 38.88}} = 0.645$$

indicating a strong positive association.

Listing 56: `DZT` data: correlation coefficients

```
> cor(DZT$Lat,DZT$DZTRate,method="spearman")
[1] 0.6772353
> cor(DZT$Lat,DZT$DZTRate,method="pearson")
[1] 0.6452132
```

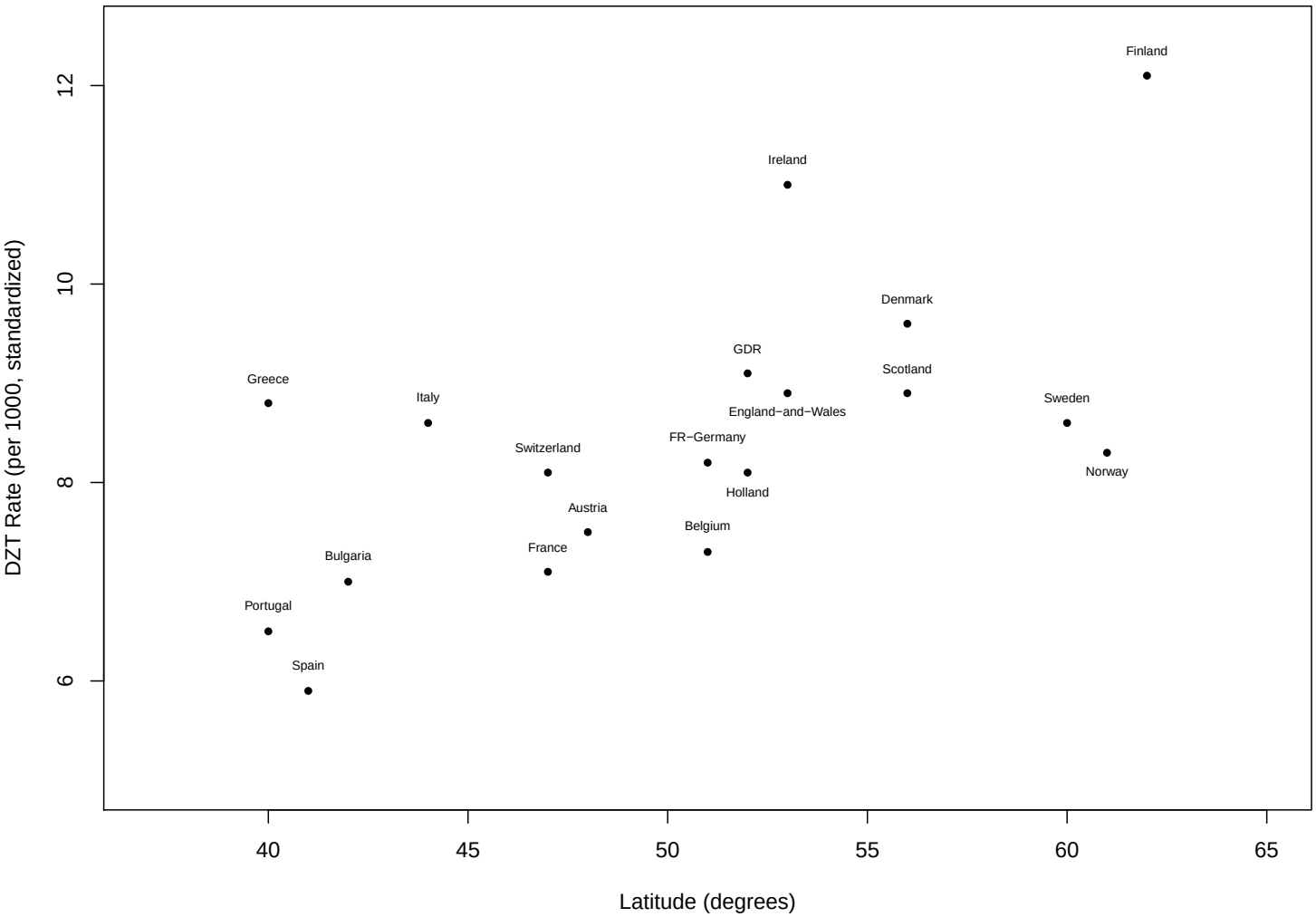


Figure 25: Latitude and dizygotic twinning rates: raw data

RANDOMIZATION AND PERMUTATION PROCEDURES

Randomization or **Permutation** procedures are useful for computing **exact** null distributions for non-parametric test statistics when sample sizes are small.

Suppose that two data samples $x_1 \ldots, x_{n_1}$ and $y_1 \ldots, y_{n_2}$ (where $n_1 \geq n_2$) have been obtained, and we wish to carry out a comparison of the two populations from which the samples are drawn. The Wilcoxon test statistic, W, is the sum of the ranks for the second sample. The permutation test proceeds as follows:

1 Let $n = n_1 + n_2$. Assuming that there are no ties, the pooled and ranked samples will have ranks

$$1 \quad 2 \quad 3 \quad \ldots \quad n$$

2 The test statistic is $W = R_2$, the rank sum for sample two items. For the observed data, W will be the sum of n_2 of the ranks given in the list above.

3 If the null hypothesis

$$H_0 \ : \text{ No difference between population 1 and population2}$$

were **true**, then there should be **no pattern** in the group labels when sorted into ascending order; the sorted data would give rise a **random** assortment of group 1 and group 2 labels.

4 To obtain the exact distribution of W under H_0 (for the assessment of statistical significance), we could compute W for all possible permutations of the group labels, and then form the probability distribution of the values of W. We call this the **permutation null distribution**.

5 But W is a rank sum, so we can compute the permutation null distribution simply by tabulating **all possible subsets** of size n_2 of the set of ranks $\{1, 2, 3, \ldots, n\}$.

6 There are

$$\binom{n}{n_2} = \frac{n!}{n_1!\, n_2!} = N$$

say possible subsets of size n_2; for $n = 6$ and $n_2 = 2$, the number of subsets of size n_2 is

$$\binom{8}{2} = \frac{8!}{6!\, 2!} = 28$$

However, the number of subsets increases dramatically as n increases; for $n_1 = n_2 = 10$, so that $n = 20$, the number of subsets of size n_2 is

$$\binom{20}{10} = \frac{20!}{10!\, 10!} = 184756$$

7 The exact rejection region and p-value are computed from the permutation null distribution. Let $W_i, i = 1, \ldots, N$ denote the value of the Wilcoxon statistic for the N possible subsets of the ranks of size n_2. The probability that the test statistic, W, is less than or equal to w is

$$\Pr[W \leq w] = \frac{\text{Number of } W_i \leq w}{N}$$

We seek the values of w that give the appropriate rejection region, $\mathcal{R}$, so that

$$\Pr[W \in \mathcal{R}] = \frac{\text{Number of } W_i \in \mathcal{R}}{N} = \alpha$$

It may not be possible to find critical values, and define $\mathcal{R}$, so that this probability is **exactly** α as the distribution of W is **discrete**.

EXAMPLE Simple Example
Suppose $n_1 = 7$ and $n_2 = 3$. There are

$$\binom{10}{3} = \frac{10!}{7!\,3!} = 120$$

subsets of the ranks $\{1, 2, 3, \ldots, 10\}$ of size 3. The subsets are listed below, together with the rank sums.

Ranks			W	Ranks			W	Ranks			W	Ranks			W
1	2	3	6	1	7	8	16	2	7	10	19	4	6	7	17
1	2	4	7	1	7	9	17	2	8	9	19	4	6	8	18
1	2	5	8	1	7	10	18	2	8	10	20	4	6	9	19
1	2	6	9	1	8	9	18	2	9	10	21	4	6	10	20
1	2	7	10	1	8	10	19	3	4	5	12	4	7	8	19
1	2	8	11	1	9	10	20	3	4	6	13	4	7	9	20
1	2	9	12	2	3	4	9	3	4	7	14	4	7	10	21
1	2	10	13	2	3	5	10	3	4	8	15	4	8	9	21
1	3	4	8	2	3	6	11	3	4	9	16	4	8	10	22
1	3	5	9	2	3	7	12	3	4	10	17	4	9	10	23
1	3	6	10	2	3	8	13	3	5	6	14	5	6	7	18
1	3	7	11	2	3	9	14	3	5	7	15	5	6	8	19
1	3	8	12	2	3	10	15	3	5	8	16	5	6	9	20
1	3	9	13	2	4	5	11	3	5	9	17	5	6	10	21
1	3	10	14	2	4	6	12	3	5	10	18	5	7	8	20
1	4	5	10	2	4	7	13	3	6	7	16	5	7	9	21
1	4	6	11	2	4	8	14	3	6	8	17	5	7	10	22
1	4	7	12	2	4	9	15	3	6	9	18	5	8	9	22
1	4	8	13	2	4	10	16	3	6	10	19	5	8	10	23
1	4	9	14	2	5	6	13	3	7	8	18	5	9	10	24
1	4	10	15	2	5	7	14	3	7	9	19	6	7	8	21
1	5	6	12	2	5	8	15	3	7	10	20	6	7	9	22
1	5	7	13	2	5	9	16	3	8	9	20	6	7	10	23
1	5	8	14	2	5	10	17	3	8	10	21	6	8	9	23
1	5	9	15	2	6	7	15	3	9	10	22	6	8	10	24
1	5	10	16	2	6	8	16	4	5	6	15	6	9	10	25
1	6	7	14	2	6	9	17	4	5	7	16	7	8	9	24
1	6	8	15	2	6	10	18	4	5	8	17	7	8	10	25
1	6	9	16	2	7	8	17	4	5	9	18	7	9	10	26
1	6	10	17	2	7	9	18	4	5	10	19	8	9	10	27

There are 22 possible rank sums, $\{6, 7, 8, \ldots, 25, 26, 27\}$; the number of times each is observed is displayed in the table below, with the corresponding probabilities and cumulative probabilities.

W	6	7	8	9	10	11	12	13	14	15	16
Frequency	1	1	2	3	4	5	7	8	9	10	10
Prob.	0.008	0.008	0.017	0.025	0.033	0.042	0.058	0.067	0.075	0.083	0.083
Cumulative Prob.	0.008	0.017	0.033	0.058	0.092	0.133	0.192	0.258	0.333	0.417	0.500

W	17	18	**19**	20	21	22	23	24	25	26	27
Frequency	10	10	**9**	8	7	5	4	3	2	1	1
Prob.	0.083	0.083	**0.075**	0.067	0.058	0.042	0.033	0.025	0.017	0.008	0.008
Cumulative Prob.	0.583	0.667	0.742	0.808	0.867	0.908	0.942	0.967	0.983	0.992	1.000

Thus, for example, the probability that $W = 19$ is 0.075, with a frequency of 9 out of 120. From this table:

$$\Pr[8 \leq W \leq 25] = 0.983 - 0.017 = 0.966$$

implying that the two-sided rejection region for $\alpha = 0.05$ is the set $\mathcal{R} = \{6, 7, 26, 27\}$.

APPENDIX A: R CODE

```
##########################################################
##########################################################
##########                                      ##########
##########  R Scripts for MATH 204 Winter 2013 ##########
##########                                      ##########
##########################################################
##########################################################

#Understanding the ANOVA F-statistic
n<-18;ni<-rep(6,3)
k<-3
X<-c(-0.88,0.24,-0.46,0.78,-0.47,-0.38,-0.75,0.11,0.64,1.98,-1.03,1.84,
1.38,1.20,0.42,0.05,-1.29,-0.04)

tmt<-factor(rep(1:3,each=6),labels=c("TMT1","TMT2","TMT3"))

#Compute treament means and treatment variances
tapply(X,tmt,mean)
tapply(X,tmt,var)

#Compute overall mean
mean(X)

#Compute pooled variance
sum((ni-1)*tapply(X,tmt,var))/(n-k)

#Compute SST
SST<-sum(ni*(tapply(X,tmt,mean)-mean(X))^2)
#Compute SSE
SSE<-sum((ni-1)*tapply(X,tmt,var))
#Compute SS
SS<-sum((n-1)*var(X))
print(c(SST,SSE,SS))

#Compute F
(F<-(SST/(k-1))/(SSE/(n-k)))

#Compute Critical Value
alpha<-0.05;(CR<-qf(1-alpha,k-1,n-k))

#Compute the p-value
(pval<-1-pf(F,k-1,n-k))

#Plot the null distribution
xv<-seq(0,5,by=0.01)
yv<-df(xv,k-1,n-k)
plot(xv,yv,type="l",xlab="x",ylab="f(x)")
abline(v=CR,col="red")
text(F,0,"x")

#Means plot
library(gplots)
plotmeans(X~tmt)

########################################
```

```
#Second Data Set
X<-c(-0.88,0.24,-0.46,0.78,-0.47,-0.38,
9.25,10.11,10.64,11.98,8.97,11.84,
21.38,21.20,20.42,20.05,18.71,19.96)

#Compute treament means
tapply(X,tmt,mean)

#Compute treatment variances
tapply(X,tmt, var)

#Compute overall mean
mean(X)

#Compute pooled variance
sum((ni-1)*tapply(X,tmt,var))/(n-k)

#Compute SST
SST<-sum(ni*(tapply(X,tmt,mean)-mean(X))^2)

#Compute SSE
SSE<-sum((ni-1)*tapply(X,tmt,var))

#Compute SS
SS<-sum((n-1)*var(X))
print(c(SST,SSE,SS))


#Compute F
(F<-(SST/(k-1))/(SSE/(n-k)))

#Compute Critical Value
alpha<-0.05
(CR<-qf(1-alpha,k-1,n-k))

#Compute the p-value
(pval<-1-pf(F,k-1,n-k))


#Plot the null distribution and test statistoc
xv<-seq(0,5,by=0.01)
yv<-df(xv,k-1,n-k)
plot(xv,yv,type="l",xlab="x",ylab="f(x)")
abline(v=CR,col="red")
text(F,0,"x")

#Means plot
require(gplots)
plotmeans(X~tmt)




##################################################################
```

```
#Diet vs Protein Level

boxplot(protein~Diet,data=Milk,xlab="Diet",ylab="Protein Level
(%)")

#A summary function
my.summary<-function(x){

        my.sum<-data.frame(n=length(x),Mean=mean(x),
                           StDev=sd(x),StdErr=sd(x)/sqrt(length(x)),
                           Lower=mean(x)-1.96*sd(x)/sqrt(length(x)),
                           Upper=mean(x)+1.96*sd(x)/sqrt(length(x)),
                           Min=min(x),Max=max(x))
        return(round(my.sum,5))

}

tapply(Milk$protein,Milk$Diet,my.summary)
my.summary(Milk$protein)

#Levene's Test for Equality of variances

library(car)
leveneTest(protein~Diet,center='mean',data=Milk)
leveneTest(protein~Diet,center='median',data=Milk)

#ANOVA

#Total
summary(aov(protein~1,data=Milk))
anova(lm(protein~1,data=Milk))

#Fit and Error (Residual)
summary(aov(protein~Diet,data=Milk))
anova(lm(protein~Diet,data=Milk))
summary(lm(protein~Diet,data=Milk))

#Means plot
require(gplots)
plotmeans(protein~Diet,data=Milk)

##################################################################
#Pull strength

boxplot(Pull~Dose,data=Pull,xlab="Dose level",ylab="Pull
strength")

tapply(Pull$Pull,Pull$Dose,my.summary)
my.summary(Pull$Pull)

#Levene's Test for Equality of variances

library(car)
leveneTest(Pull~Dose,center='mean',data=Pull)
leveneTest(Pull~Dose,center='median',data=Pull)
```

```
166
167 #ANOVA
168
169 #Total
170 summary(aov(Pull ~ 1,data=Pull))
171 anova(lm(Pull ~ 1,data=Pull))
172
173 #Fit and Error (Residual)
174 summary(aov(Pull ~ Dose,data=Pull))
175 anova(lm(Pull ~ Dose,data=Pull))
176 summary(lm(Pull ~ Dose,data=Pull))
177
178 #Means plot
179 require(gplots)
180 plotmeans(protein ~ Diet,data=Milk)
181
182 ###################################################################
183 #Mannitol gut permeability
184
185 boxplot(mann ~ diag,data=AIDS,xlab="Diagnosis",ylab="Permeability")
186
187 tapply(AIDS$mann,AIDS$diag,my.summary)
188 my.summary(AIDS$mann)
189
190 #Levene's Test for Equality of variances
191
192 library(car)
193 leveneTest(mann ~ diag,center='mean',data=AIDS)
194 leveneTest(mann ~ diag,center='median',data=AIDS)
195
196 #ANOVA
197
198 #Total
199 summary(aov(mann ~ 1,data=AIDS))
200 anova(lm(mann ~ 1,data=AIDS))
201
202 #Fit and Error (Residual)
203 summary(aov(mann ~ diag,data=AIDS))
204 anova(lm(mann ~ diag,data=AIDS))
205 summary(lm(mann ~ diag,data=AIDS))
206
207 #Means plot
208 require(gplots)
209 plotmeans(mann ~ diag,data=AIDS)
210
211 ###################################################################
212 #Milk bacteria level
213
214 boxplot(BacLevel ~ Batch,data=MilkBac,xlab="Batch",
215                       ylab="bacteria Level (per million)")
216
217 tapply(MilkBac$BacLevel,MilkBac$Batch,my.summary)
218 my.summary(MilkBac$BacLevel)
219
220 #Levene's Test for Equality of variances
```

```
221
222 library(car)
223 leveneTest(BacLevel ~ Batch,center='mean',data=MilkBac)
224 leveneTest(BacLevel ~ Batch,center='median',data=MilkBac)
225
226 #ANOVA
227 #Total
228 summary(aov(BacLevel ~ 1,data=MilkBac))
229 anova(lm(BacLevel ~ 1,data=MilkBac))
230
231 #Fit and Error (Residual)
232 summary(aov(BacLevel ~ Batch,data=MilkBac))
233 anova(lm(BacLevel ~ Batch,data=MilkBac))
234 summary(lm(BacLevel ~ Batch,data=MilkBac))
235
236 #Means plot
237 require(gplots)
238 plotmeans(BacLevel ~ Batch,data=MilkBac,ylab="bacteria Level (per million)")
239
240 ##################################################################
241 #PTSD Treatment
242
243 boxplot(PTSD ~ GROUP,data=PTSD,xlab="GROUP",ylab="PTSD Score")
244
245 tapply(PTSD$PTSD,PTSD$GROUP,my.summary)
246 my.summary(PTSD$PTSD)
247
248 #Levene's Test for Equality of variances
249
250 library(car)
251 leveneTest(PTSD ~ GROUP,center='mean',data=PTSD)
252 leveneTest(PTSD ~ GROUP,center='median',data=PTSD)
253
254 #ANOVA
255 #Total
256 summary(aov(PTSD ~ 1,data=PTSD))
257 anova(lm(PTSD ~ 1,data=PTSD))
258
259 #Fit and Error (Residual)
260 summary(aov(PTSD ~ GROUP,data=PTSD))
261 anova(lm(PTSD ~ GROUP,data=PTSD))
262 summary(lm(PTSD ~ GROUP,data=PTSD))
263
264 #Means plot
265 require(gplots)
266 plotmeans(PTSD ~ GROUP,data=PTSD,xlab="GROUP",ylab="PTSD Score")
267
268
269 ##################################################################
270 ##################################################################
271 #Memory Task
272
273 boxplot(Memory ~ Task,data=Memory,xlab="Task",ylab="Memory
274 Score")
275
```

```
tapply(Memory$Memory,Memory$Task,my.summary)
my.summary(Memory$Memory)

#Levene's Test for Equality of variances

library(car)
leveneTest(Memory ~ Task,center='mean',data=Memory)
leveneTest(Memory ~ Task,center='median',data=Memory)

#ANOVA

#Total
summary(aov(Memory ~ 1,data=Memory))
anova(lm(Memory ~ 1,data=Memory))

#Fit and Error (Residual)
summary(aov(Memory ~ Task,data=Memory))
anova(lm(Memory ~ Task,data=Memory))
summary(lm(Memory ~ Task,data=Memory))

#Means plot
require(gplots)
plotmeans(Memory ~ Task,data=Memory,xlab="Task",ylab="Memory Score")

#####################################################################
#Post-hoc tests

#Anxiety data
summary(aov(Pull ~ Dose,data=Pull))

pairwise.t.test(Pull$Pull, Pull$Dose, p.adj = "none")
pairwise.t.test(Pull$Pull, Pull$Dose, p.adj = "bonferroni")

TukeyHSD(aov(Pull ~ Dose,data=Pull))

plot(TukeyHSD(aov(Pull ~ Dose,data=Pull)))

#PTSD Data
summary(aov(PTSD ~ GROUP,data=PTSD))

pairwise.t.test(PTSD$PTSD, PTSD$GROUP, p.adj = "none")
pairwise.t.test(PTSD$PTSD, PTSD$GROUP, p.adj = "bonferroni")

TukeyHSD(aov(PTSD ~ GROUP,data=PTSD))

plot(TukeyHSD(aov(PTSD ~ GROUP,data=PTSD)))


#############################################################
#############################################################

#Randomized block design
#Soil data
#Table of data
xtabs(sulphur ~ solvent+soil,data=Soil)
```

```
summary(aov(sulphur ~ 1,data=Soil))
summary(aov(sulphur ~ solvent+soil,data=Soil))

require(car)
#Not run
#leveneTest(sulphur ~ solvent+soil,center='mean',data=Soil)
#leveneTest(sulphur ~ solvent+soil,center='median',data=Soil)

#The need for blocking - example

Y<-c(-20.88,-4.76,-0.46,10.78,-10.47,
-15.75,0.11,5.64,16.98,-6.03,
-8.62,6.20,10.42,20.05,-1.29)

Group<-gl(3,5,labels=c("Group1","Group2","Group3"))
Block<-gl(5,1,15,labels=c(1:5))

library(gplots)
plotmeans(Y ~ Group,ylim=range(-40,40))
stripchart(Y ~ Group,vertical=T,pch=4,cex=0.75,lwd=2,add=T)

require(car)
leveneTest(Y ~ Group,center='mean')
leveneTest(Y ~ Group,center='median')

#Ignoring the blocking in ANOVA
summary(aov(Y ~ 1))
summary(aov(Y ~ Group))

#Acknowledging the blocking in ANOVA
summary(aov(Y ~ Group+Block))

blocklabel<-paste(rep("Block",5),c(1:5))
plotmeans(Y ~ Group,ylim=range(-40,40))
points(rep(0.9,5),Y[Group=="Group1"],pch=c(1:5))
points(rep(1.9,5),Y[Group=="Group2"],pch=c(1:5))
points(rep(2.9,5),Y[Group=="Group3"],pch=c(1:5))
legend(0.5,40,blocklabel,pch=(1:5))

#################################################################
#Randomized block designs with replication
#Batteries data

#Table of data
#Counts of replicates
xtabs( ~ temp+material,data=Batteries)
#Means
xtabs(life ~ temp+material,data=Batteries)/xtabs( ~ temp+material,data=Batteries)

summary(aov(life ~ 1,data=Batteries))
summary(aov(life ~ temp+material+temp:material,data=Batteries))
summary(aov(life ~ temp*material,data=Batteries))
leveneTest(life ~ temp*material,data=Batteries,center='mean')
leveneTest(life ~ temp*material,data=Batteries,center='median')
```

```
386
387
388 ######################################################################
389 ######################################################################
390 #Randomized complete factorial designs
391 #Butterfat data
392
393 m3<-aov(butterfat~breed*age,data=Butterfat)
394 summary(m3)
395 m2<-aov(butterfat~breed+age,data=Butterfat)
396 summary(m2)
397 m1<-aov(butterfat~breed,data=Butterfat)
398 summary(m1)
399 m0<-aov(butterfat~1,data=Butterfat)
400 anova(m0,m1,m2,m3)
401
402 #Set-up means plot
403 plotmeans(0*butterfat~breed,data=Butterfat,ylim=range(3,7),
404                          ylab="Butterfat",pch=19)
405 plotmeans(butterfat~breed,data=Butterfat,subset=(age=="Two year old"),add=TRUE,
406         n.label=FALSE,col="blue",ccol="blue",ylab="Butterfat",pch=19)
407 plotmeans(butterfat~breed,data=Butterfat,subset=(age=="Mature"),add=T,
408         n.label=FALSE,lty=2,pch=15,col="red",barcol="red",ccol="red")
409 legend(0.5,7,c("Two year old","Mature"),pch=c(19,15),lty=c(1,2),
410                          col=c("blue","red"))
411
412
413 #Set-up means plot
414 plotmeans(0*butterfat~age,data=Butterfat,ylim=range(3,7),ylab="Butterfat",pch=19)
415 plotmeans(butterfat~age,data=Butterfat,subset=(breed=="Ayreshire"),add=TRUE,
416     n.label=FALSE,ylim=range(3,6),col="blue",ccol="blue",ylab="Butterfat",pch=19)
417 plotmeans(butterfat~age,data=Butterfat,subset=(breed=="Canadian"),add=T,
418     n.label=FALSE,lty=2,pch=15,col="red",barcol="red",ccol="red")
419 plotmeans(butterfat~age,data=Butterfat,subset=(breed=="Guernsey"),add=T,
420     n.label=FALSE,lty=3,pch=17,col="green",barcol="green",ccol="green")
421 plotmeans(butterfat~age,data=Butterfat,subset=(breed=="Holstein"),add=T,
422     n.label=FALSE,lty=4,pch=4,col="darkgreen",barcol="darkgreen",ccol="darkgreen")
423 plotmeans(butterfat~age,data=Butterfat,subset=(breed=="Jersey"),add=T,
424     n.label=FALSE,lty=5,pch=3,col="purple",barcol="purple",ccol="purple")
425 legend(1.5,7,unique(Butterfat$breed),pch=c(19,15,17,4,3),lty=c(1:5),
426     col=c("blue","red","green","darkgreen","purple"))
427
428 leveneTest(m1,center='mean')
429 leveneTest(m1,center='median')
430 leveneTest(m3,center='mean')
431 leveneTest(m3,center='median')
432
433 ######################################################################
434 ######################################################################
435 #Lyrics data
436
437 m3<-aov(SCORE~SONG*POOL,data=Lyrics)
438 summary(m3)
439 m2<-aov(SCORE~SONG+POOL,data=Lyrics)
440 summary(m2)
```

```
m1<-aov(SCORE ~ SONG,data=Lyrics)
summary(m1)
m0<-aov(SCORE ~ 1,data=Lyrics)
anova(m0,m1,m2,m3)

#Set-up means plot
plotmeans(0*SCORE ~ SONG,data=Lyrics,ylim=range(2.5,4),ylab="SCORE",pch=19)
plotmeans(SCORE ~ SONG,subset=(POOL=="PSYCHOL "),data=Lyrics,
    ylim=range(2.5,4),n.label=FALSE,pch=19,col="blue",ccol="blue",)
plotmeans(SCORE ~ SONG,subset=(POOL=="VOLUNTEE"),data=Lyrics,
    add=T,lty=2,pch=15,n.label=FALSE,col="red",barcol="red",ccol="red")
legend(0.5,4,c("PSYCHOL ","VOLUNTEE"),pch=c(19,15),lty=c(1,2),col=c("blue","red"))

#Set-up means plot
plotmeans(0*SCORE ~ POOL,data=Lyrics,ylim=range(2.5,4),ylab="SCORE",pch=19)
plotmeans(SCORE ~ POOL,subset=(SONG=="NON-VIO     "),data=Lyrics,ylim=range(2.5,4),
    n.label=FALSE,pch=19,col="blue",ccol="blue",)
plotmeans(SCORE ~ POOL,subset=(SONG=="VIOLENT     "),data=Lyrics,
    add=T,lty=2,pch=15,n.label=FALSE,col="red",barcol="red",ccol="red")
legend(1.5,4,c("NON-VIO ","VIOLENT"),pch=c(19,15),lty=c(1,2),col=c("blue","red"))

leveneTest(m1,center='mean')
leveneTest(m1,center='median')
leveneTest(m3,center='mean')
leveneTest(m3,center='median')

plotmeans(Y ~ Group,ylim=range(-40,40))


######################################################################
######################################################################
#Gravel data

m3<-aov(amount ~ shift*quarry,data=Gravel)
summary(m3)
m2<-aov(amount ~ shift+quarry,data=Gravel)
summary(m2)
m1<-aov(amount ~ quarry,data=Gravel)
m0<-aov(amount ~ 1,data=Gravel)
anova(m0,m1,m2,m3)

#Set-up means plot
plotmeans(0*amount ~ quarry,data=Gravel,ylim=range(5,35),ylab="Amount",pch=19)
plotmeans(amount ~ quarry,subset=(shift=="AM"),data=Gravel,
    n.label=FALSE,pch=19,col="blue",ccol="blue",add=T)
plotmeans(amount ~ quarry,subset=(shift=="PM"),data=Gravel,add=T,lty=2,pch=15,
    n.label=FALSE,col="red",barcol="red",ccol="red")
legend(3.5,35,c("AM","PM"),pch=c(19,15),lty=c(1,2),col=c("blue","red"))

plotmeans(0*amount ~ shift,data=Gravel,ylim=range(5,35),ylab="Amount",pch=19)
plotmeans(amount ~ shift,subset=(quarry=="A"),data=Gravel,
    n.label=FALSE,pch=19,col="blue",ccol="blue",add=T)
plotmeans(amount ~ shift,subset=(quarry=="B"),data=Gravel,add=T,lty=2,pch=15,
    n.label=FALSE,col="red",barcol="red",ccol="red")
plotmeans(amount ~ shift,subset=(quarry=="C"),data=Gravel,add=T,
```

```
    n.label=FALSE,lty=3,pch=17,col="green",barcol="green",ccol="green")
plotmeans(amount ~ shift,subset=(quarry=="D"),data=Gravel,add=T,
    n.label=FALSE,lty=4,pch=4,col="darkgreen",barcol="darkgreen",ccol="darkgreen")
legend(1.5,35,c("A","B","C","D"),pch=c(19,15,17,4),lty=c(1:4),
        col=c("blue","red","green","darkgreen"))

leveneTest(m1,center='mean')
leveneTest(m1,center='median')
leveneTest(m3,center='mean')
leveneTest(m3,center='median')

#####################################################################
#####################################################################
#Models: interactions without main effects

summary(aov(life ~ material*temp,data=Batteries))
summary(aov(life ~ material+material:temp,data=Batteries))
summary(aov(life ~ temp+material:temp,data=Batteries))

#############################################################
#############################################################

#Simple Linear Regression
attach(Viscosity)
plot(pcv,viscosity,pch=19)
detach(Viscosity)

plot(Viscosity$pcv,Viscosity$viscosity,pch=19,xlab="pcv",ylab="Viscosity")

summary(Viscosity$pcv)
summary(Viscosity$viscosity)
cor.test(Viscosity$pcv,Viscosity$viscosity)

m0.visc<-lm(viscosity ~ pcv,data=Viscosity)
abline(coef(m0.visc))
be0.hat<-round(coef(m0.visc)[1],3)
be1.hat<-round(coef(m0.visc)[2],3)
text(40,5.9,substitute(hat(beta)[0] == b, list(b = be0.hat)),adj=0)
text(40,5.7,substitute(hat(beta)[1] == b, list(b = be1.hat)),adj=0)
summary(m0.visc)


#############################################################
#############################################################
#Coleman data
attach(Coleman)
plot(status,testscore,pch=19)
detach(Coleman)

m0.cole<-lm(testscore ~ status,data=Coleman)
summary(m0.cole)
abline(coef(m0.cole))

summary(aov(m0.cole))
```

```
551 ############################################################
552 ############################################################
553 #Hooker data
554 attach(Hooker)
555 plot(pressure,temp,pch=19)
556 detach(Hooker)
557
558 m0.hooker<-lm(temp~pressure,data=Hooker)
559 summary(m0.hooker)
560 abline(coef(m0.hooker))
561
562 summary(aov(m0.hooker))
563
564 ############################################################
565 #Polynomial
566 attach(Hooker)
567 plot(pressure,temp,pch=19)
568 detach(Hooker)
569 m1.hooker<-lm(temp~pressure+I(pressure^2),data=Hooker)
570 summary(m1.hooker)
571
572 #Fitted values for quadratic regression
573 xp<-seq(15,30,by=0.01)
574 m1.fit<-cbind(rep(1,length(xp)),xp,xp^2) %*% coef(m1.hooker)
575 lines(xp,m1.fit,col="red",lty=2)
576 abline(coef(m0.hooker))
577 legend(16,210,c("Straight Line","Quadratic"),lty=c(1,2),col=c("black","red"))
578 title("Hooker data")
579
580 par(mfrow=c(2,2))
581 attach(Hooker)
582 plot(pressure,temp,pch=19)
583 abline(coef(m0.hooker))
584 title("Straight-line fit")
585 plot(pressure,residuals(m0.hooker),pch=19,ylab="Residual",ylim=range(-1.5,1.5))
586 title("Residuals from straight-line fit")
587 abline(h=0,lty=2)
588 plot(pressure,temp,pch=19)
589 lines(xp,m1.fit,lty=2)
590 title("Quadratic fit")
591 plot(pressure,residuals(m1.hooker),pch=19,ylab="Residual",ylim=range(-1.5,1.5))
592 title("Residuals from quadratic fit")
593 abline(h=0,lty=2)
594 detach(Hooker)
595 par(mfrow=c(1,1))
596 anova(m0.hooker,m1.hooker)
597
598 ############################################################
599 ############################################################
600 #Multiple regression
601
602 mv.visc<-lm(viscosity~pcv+plasfib+plaspro,data=Viscosity)
603 summary(mv.visc)
604
605 m1.visc<-lm(viscosity~pcv,data=Viscosity)
```

```
summary(m1.visc)
m2.visc<-lm(viscosity~plasfib,data=Viscosity)
summary(m2.visc)
m3.visc<-lm(viscosity~plaspro,data=Viscosity)
summary(m3.visc)

anova(m1.visc,mv.visc)
anova(m2.visc,mv.visc)
anova(m3.visc,mv.visc)

############################################################
############################################################
#Creating dummy variables for factor predictors
#Diabetes data
summary(lm(glutest~group,data=Diabetes))

summary(lm(glutest~-1+group,data=Diabetes))

Diabetes.copy<-Diabetes
Diabetes.copy$group2<-factor(Diabetes.copy$group,
        c("Normal","Chemically Diabetic","Overt Diabetic"))
summary(lm(glutest~group2,data=Diabetes.copy))
summary(lm(glutest~-1+group2,data=Diabetes.copy))

boxplot(glutest~group2,data=Diabetes.copy)
title("Glucose level on test \n (original scale)")

Diabetes.copy$group3 <- relevel(Diabetes.copy$group, ref="Normal")
summary(lm(glutest~group3,data=Diabetes.copy))
summary(lm(glutest~-1+group3,data=Diabetes.copy))

leveneTest(glutest~group2,data=Diabetes.copy)

boxplot(log(glutest)~group2,data=Diabetes.copy)
title("Glucose level on test \n (log scale)")

leveneTest(log(glutest)~group2,data=Diabetes.copy)

Diabetes.copy$X1<-as.numeric(Diabetes.copy$group=="Normal")
Diabetes.copy$X2<-as.numeric(Diabetes.copy$group=="Chemically Diabetic")
Diabetes.copy$X3<-as.numeric(Diabetes.copy$group=="Overt Diabetic")

summary(lm(glutest~-1+X1+X2+X3,data=Diabetes.copy))

############################################################
#Factor predictor and covariate
attach(Diabetes)
plot(log(glufast),log(glutest),pch=19)
plot(log(glufast),log(glutest),type="n")
points(log(glufast[group=="Normal"]),log(glutest[group=="Normal"]),pch=17)
points(log(glufast[group=="Chemically Diabetic"]),
        log(glutest[group=="Chemically Diabetic"]),pch=4)
points(log(glufast[group=="Overt Diabetic"]),
        log(glutest[group=="Overt Diabetic"]),pch=3)
legend(4.2,7.3,c("Normal","Chem. Diabetic","Overt Diabetic"),pch=c(17,4,3))
```

```
detach(Diabetes)

m.Diabetes<-lm(log(glutest)~log(glufast)*group2,data=Diabetes.copy)
summary(m.Diabetes)

attach(Diabetes)
par(mfrow=c(2,2))
plot(log(glufast),log(glutest),type="n")
points(log(glufast[group=="Normal"]),log(glutest[group=="Normal"]),pch=17)
be0<-coef(m.Diabetes)[1]
be1<-coef(m.Diabetes)[2]
abline(be0,be1)
title("Normal")
legend(4.2,7.3,c("Normal"),pch=c(17))

plot(log(glufast),log(glutest),type="n")
points(log(glufast[group=="Chemically Diabetic"]),
        log(glutest[group=="Chemically Diabetic"]),pch=4)
be0<-coef(m.Diabetes)[1]+coef(m.Diabetes)[3]
be1<-coef(m.Diabetes)[2]+coef(m.Diabetes)[5]
abline(be0,be1)
title("Chemically Diabetic")
legend(4.2,7.3,c("Chem. Diabetic"),pch=c(4))

plot(log(glufast),log(glutest),type="n")
points(log(glufast[group=="Overt Diabetic"]),
        log(glutest[group=="Overt Diabetic"]),pch=3)
be0<-coef(m.Diabetes)[1]+coef(m.Diabetes)[4]
be1<-coef(m.Diabetes)[2]+coef(m.Diabetes)[6]
abline(be0,be1)
title("Overt Diabetic")
legend(4.2,7.3,c("Overt Diabetic"),pch=c(3))

plot(log(glufast),log(glutest),type="n")
be0<-coef(m.Diabetes)[1]
be1<-coef(m.Diabetes)[2]
abline(be0,be1)
be0<-coef(m.Diabetes)[1]+coef(m.Diabetes)[3]
be1<-coef(m.Diabetes)[2]+coef(m.Diabetes)[5]
abline(be0,be1,lty=2)
be0<-coef(m.Diabetes)[1]+coef(m.Diabetes)[4]
be1<-coef(m.Diabetes)[2]+coef(m.Diabetes)[6]
abline(be0,be1,lty=3)
legend(4.2,7.3,c("Normal","Chem. Diabetic","Overt Diabetic"),lty=c(1,2,3))
par(mfrow=c(1,1))
detach(Diabetes)

#Results separately for groups
summary(lm(log(glutest)~log(glufast),data=Diabetes,
                            subset=(group=="Normal")))
summary(lm(log(glutest)~log(glufast),data=Diabetes,
                            subset=(group=="Chemically Diabetic")))
summary(lm(log(glutest)~log(glufast),data=Diabetes,
                            subset=(group=="Overt Diabetic")))
```

```
#Model selection
#Try to drop the interaction
drop1(m.Diabetes,test="F")
#Cannot drop !

################################################################
#Unbalanced designs

m.potato<-lm(leak~potato*regime*temp,data=Potato)
summary(m.potato)
summary(aov(m.potato))

drop1(m.potato,test="F")
#Can drop 3-way interaction

m1.potato<-update(m.potato, ~ .-potato:regime:temp)

drop1(m1.potato,test="F")

#Can drop 2-way interaction regime:temp
m2.potato<-update(m1.potato, ~ .-regime:temp)
drop1(m2.potato,test="F")

#Can drop 2-way interaction potato:temp
m3.potato<-update(m2.potato, ~ .-potato:temp)
drop1(m3.potato,test="F")

summary(m3.potato)

summary(aov(m3.potato))

#####################################################################
#Two factors, one covariate

m.Task<-lm(Errors~Group*Task*Distract,data=TaskDistract)
summary(m.Task)

summary(aov(m.Task))
drop1(m.Task,test="F")

#Can drop 3-way interaction

m1.Task<-update(m.Task, ~ .-Group:Task:Distract)
drop1(m1.Task,test="F")

#Maybe drop Group:Distract
m2.Task<-update(m1.Task, ~ .-Group:Distract)
drop1(m2.Task,test="F")

summary(m1.Task)
summary(aov(m1.Task))

m3.Task<-lm(Errors~Group+Task+Distract,data=TaskDistract)
summary(m3.Task)

```

```
771 ######################################################################
772 #Prediction
773 #Prediction A: Subject in Group "NonSmoker", doing Task "Driving Simulation",
774 #               with Distraction level 180.
775 #Prediction B: Subject in Group "Delayed Smoker", doing Task "Cognitive Task",
776 #               with Distraction level 180.
777 #Prediction C: Subject in Group "Active Smoker", doing Task "Pattern Recognition",
778 #               with Distraction level 100.
779
780 newTask<-data.frame(Group=c("NonSmoker","Delayed Smoker","Active Smoker"),
781     Task=c("Driving Simulation","Cognitive Task","Pattern Recognition"),
782     Distract=c(180,180,100))
783
784 #Prediction plus confidence interval
785 predict(m.Task,newdata=newTask,interval="confidence")
786 predict(m3.Task,newdata=newTask,interval="confidence")
787
788 #Prediction plus prediction interval
789 predict(m.Task,newdata=newTask,interval="prediction")
790 predict(m3.Task,newdata=newTask,interval="prediction")
791
792
793 ######################################################################
794 #Stepwise selection using AIC
795
796 best1<-step(m.Task,k=2,direction="backward")
797 best2<-step(m.Task,k=2,direction="both")
798 formula(best1)
799 formula(best2)
800
801 #Stepwise selection using BIC
802
803 best3<-step(m.Task,k=log(nrow(TaskDistract)),direction="backward")
804 best4<-step(m.Task,k=log(nrow(TaskDistract)),direction="both")
805 formula(best3)
806 formula(best4)
807
808 ######################################################################
809 ######################################################################
810 #Residuals
811
812 Task.full<-lm(formula(m.Task),data=TaskDistract)
813 Task.best1<-lm(formula(best1),data=TaskDistract)
814 Task.best3<-lm(formula(best3),data=TaskDistract)
815
816 res.Task<-residuals(m.Task)
817 res.best1<-residuals(Task.best1)
818 res.best3<-residuals(Task.best3)
819
820 stres.Task<-rstandard(m.Task)
821 stres.best1<-rstandard(Task.best1)
822 stres.best3<-rstandard(Task.best3)
823
824 fit.Task<-fitted(m.Task)
825 fit.best1<-fitted(Task.best1)
```

```
fit.best3<-fitted(Task.best3)

par(mfrow=c(2,2))
plot(TaskDistract$Distract,res.Task,
        xlab="Distraction Level",ylab="Residuals",
    ylim=range(-25,25),pch=20)
plot(TaskDistract$Distract,stres.Task,
        xlab="Distraction Level",ylab="Standardized Residuals",
    ylim=range(-3,3),pch=20)
plot(fit.Task,stres.Task,
        xlab="Fitted response",ylab="Standardized Residuals",
    ylim=range(-3,3),pch=20)
plot(TaskDistract$Errors,fit.Task,
        xlab="Observed response",ylab="Fitted response",xlim=range(0,70),
    ylim=range(0,70),pch=20)
abline(0,1,lty=2)

par(mfrow=c(2,2))
plot(TaskDistract$Distract,res.best1,
    xlab="Distraction Level",ylab="Residuals",ylim=range(-25,25),pch=20)
plot(TaskDistract$Distract,stres.best1,
    xlab="Distraction Level",ylab="Standardized Residuals",
    ylim=range(-3,3),pch=20)
plot(fit.best1,stres.best1,xlab="Fitted response",ylab="Standardized Residuals",
    ylim=range(-3,3),pch=20)
plot(TaskDistract$Errors,fit.best1,xlab="Observed response",
    ylab="Fitted response", xlim=range(0,70),ylim=range(0,70),pch=20)
abline(0,1,lty=2)

par(mfrow=c(2,2))
plot(TaskDistract$Distract,res.best3,
    xlab="Distraction Level",ylab="Residuals",ylim=range(-25,25),pch=20)
plot(TaskDistract$Distract,stres.best3,
    xlab="Distraction Level",ylab="Standardized Residuals",
    ylim=range(-3,3),pch=20)
plot(fit.best3,stres.best1,
    xlab="Fitted response",ylab="Standardized Residuals",
    ylim=range(-3,3),pch=20)
plot(TaskDistract$Errors,fit.best3,
    xlab="Observed response",ylab="Fitted response",
    xlim=range(0,70),ylim=range(0,70),pch=20)
abline(0,1,lty=2)

#####################################################################
summary(m.Diabetes)
par(mfrow=c(2,2))
plot(log(Diabetes$glufast),residuals(m.Diabetes),
    xlab="log(glufast)",ylab="Residuals",ylim=range(-0.5,0.5),pch=20)
plot(log(Diabetes$glufast),rstandard(m.Diabetes),
    xlab="log(glufast)",ylab="Standardized Residuals",
    ylim=range(-3,3),pch=20)
plot(fitted(m.Diabetes),rstandard(m.Diabetes),
    xlab="Fitted response",ylab="Standardized Residuals",
    ylim=range(-3,3),pch=20)
plot(log(Diabetes$glutest),fitted(m.Diabetes),
```

```
    xlab="Observed response",ylab="Fitted response",xlim=range(5.5,7.3),
    ylim=range(5.5,7.3),pch=20)
abline(0,1,lty=2)
par(mfrow=c(1,1))

#####################################################################
#Simulated example demonstrating the influence of outliers
set.seed(43723)
n<-25
X<-sort(runif(n,0,20))
b0<-2.0
b1<-3.0
Y<-b0+b1*X+rnorm(n)*3
cbind(X,Y)

par(mfrow=c(2,2))
plot(X,Y,pch=20)
abline(coef(lm(Y~X)))
text(0,60,substitute(hat(beta)[0] == b,
    list(b = round(coef(lm(Y~X))[1],3))),adj=0)
text(0,55,substitute(hat(beta)[1] == b,
    list(b = round(coef(lm(Y~X))[2],3))),adj=0)
title("Original data")

newY<-Y
newY[20]<-Y[20]-30
plot(X,newY,pch=20)
abline(coef(lm(newY~X)))
abline(coef(lm(Y~X)),lty=2)
points(X[20],newY[20],cex=2,col="red")

text(0,60,substitute(hat(beta)[0] == b,
    list(b = round(coef(lm(newY~X))[1],3))),adj=0)
text(0,55,substitute(hat(beta)[1] == b,
    list(b = round(coef(lm(newY~X))[2],3))),adj=0)
title("Data with outlier")

plot(X,rstandard(lm(Y~X)),ylim=range(-5,5),pch=20,
        ylab="Standardized residuals")
abline(h=c(-2,0,2),lty=c(2,3,2))

plot(X,rstandard(lm(newY~X)),ylim=range(-5,5),pch=20,
        ylab="Standardized residuals")
abline(h=c(-2,0,2),lty=c(2,3,2))

par(mfrow=c(1,1))

#####################################################################
#####################################################################

#DNA data Chisq Test
DNA.data<-c(38514,24631,25685,38249)
DNA.fitted<-rep(1/4,4)*sum(DNA.data)

(Xsq<-sum((DNA.data-DNA.fitted)^2/DNA.fitted))
```

```
(pval<-1-pchisq(Xsq,df=3))

#DNA data Chisq Test: alternative null
DNA.data<-c(38514,24631,25685,38249)
DNA.fitted<-c(0.3,0.2,0.2,0.3)*sum(DNA.data)

(Xsq<-sum((DNA.data-DNA.fitted)^2/DNA.fitted))
(pval<-1-pchisq(Xsq,df=3))


############################################################
#Hair & Eye colour

xtabs(count~eye+hair,data=HairEyes)

chisq.test(xtabs(count~eye+hair,data=HairEyes))

#Direct calculation
HairEyes.table<-xtabs(count~eye+hair,data=HairEyes)
HE.rowtots<-apply(HairEyes.table,1,sum)
HE.coltots<-apply(HairEyes.table,2,sum)

#Form the fitted values from the row and column totals
HE.fitted<-outer(HE.rowtots,HE.coltots,"*")/sum(HairEyes.table)

Xsq<-sum((HairEyes.table-HE.fitted)^2/HE.fitted)

(pval<-1-pchisq(Xsq,df=9))

############################################################
############################################################
#Non-parametric tests

#Mann-Whitney-Wilcoxon test

Sample1<-c(0.31,0.48,1.02,3.11)
Sample2<-c(0.16,0.20,1.97)
Y<-c(Sample1,Sample2)
Samp<-rep(c(1,2),c(4,3))

rY<-rank(Y)
(R1<-sum(rY[Samp==1]))
(R2<-sum(rY[Samp==2]))

wilcox.test(Y~Samp)
wilcox.test(Y~Samp,alternative="less")
wilcox.test(Y~Samp,alternative="greater")

library(coin)
wilcox_test(Y~as.factor(Samp)) #Asymptotic approximation
wilcox_test(Y~as.factor(Samp),distribution="exact") #Exact null
wilcox_test(Y~as.factor(Samp),distribution="approximate") #Approximate null

#################################################################
#Sign test
```

```
991
992 eta<-9
993 binom.test(sum(Sign1$h2o>eta),nrow(Sign1),p=0.5,alternative="greater")
994 binom.test(sum(Sign1$h2o>eta),nrow(Sign1),p=0.5,alternative="less")
995 binom.test(sum(Sign1$h2o>eta),nrow(Sign1),p=0.5,alternative="two.sided")
996
997 #Test statistic
998 S<-sum(Sign1$h2o>9)
999 #First p-value
1000 1-pbinom(S-1,nrow(Sign1),0.5)
1001 #Second p-value
1002 pbinom(S,nrow(Sign1),0.5)
1003 #Third p-value
1004 2*(1-pbinom(max(S,nrow(Sign1)-S)-1,nrow(Sign1),0.5))
1005
1006 ##########################################################################
1007 eta<-9
1008 binom.test(sum(Sign2$h2o>eta),nrow(Sign2),p=0.5,alternative="greater")
1009 binom.test(sum(Sign2$h2o>eta),nrow(Sign2),p=0.5,alternative="less")
1010 binom.test(sum(Sign2$h2o>eta),nrow(Sign2),p=0.5,alternative="two.sided")
1011
1012 #Test statistic
1013 S<-sum(Sign2$h2o>9)
1014 #First p-value
1015 1-pbinom(S-1,nrow(Sign2),0.5)
1016 #Second p-value
1017 pbinom(S,nrow(Sign2),0.5)
1018 #Third p-value
1019 2*(1-pbinom(max(S,nrow(Sign2)-S)-1,nrow(Sign2),0.5))
1020
1021 ##########################################################################
1022 #Two sample Wilcoxon:Birthweight
1023
1024 attach(Birthweight)
1025 (R1<-sum(rank(BW)[gp==1]))
1026 (R2<-sum(rank(BW)[gp==2]))
1027 n1<-sum(gp==1)
1028 n2<-sum(gp==2)
1029 detach(Birthweight)
1030
1031 W<-R2
1032 U<-W-0.5*n2*(n2+1)
1033 Z<-(U-0.5*n1*n2)/sqrt(n1*n2*(n1+n2+1)/12)
1034
1035 wilcox.test(BW~Group,data=Birthweight)
1036 wilcox.test(BW~Group,data=Birthweight,alternative="greater")
1037 wilcox.test(BW~Group,data=Birthweight,alternative="less")
1038
1039 wilcox_test(BW~Group,data=Birthweight,dist="exact",ties.method="mid-ranks")
1040 wilcox_test(BW~Group,data=Birthweight,dist="exact",alternative="greater")
1041 wilcox_test(BW~Group,data=Birthweight,dist="exact",alternative="less")
1042
1043 ##########################################################################
1044
1045 #Two sample Wilcoxon:Treadmill
```

```
attach(Treadmill)
(R1<-sum(rank(Time)[Group==1]))
(R2<-sum(rank(Time)[Group==2]))
n1<-sum(Group==1);n2<-sum(Group==2)
detach(Treadmill)

W<-R2
U<-W-0.5*n2*(n2+1)
Z<-(U-0.5*n1*n2)/sqrt(n1*n2*(n1+n2+1)/12)

wilcox.test(Time~Group,data=Treadmill)
wilcox.test(Time~Group,data=Treadmill,alternative="greater")
wilcox.test(Time~Group,data=Treadmill,alternative="less")

wilcox_test(Time~as.factor(Group),data=Treadmill,
    dist="exact",ties.method="mid-ranks")
wilcox_test(Time~as.factor(Group),data=Treadmill,
    dist="exact",alternative="greater")
wilcox_test(Time~as.factor(Group),data=Treadmill,
    dist="exact",alternative="less")


#####################################################################
#Paired test
#Haemodialysis
wilcox.test(Haemo$after,Haemo$before,paired=T)
wilcox.test(Haemo$after,Haemo$before,paired=T,alternative="greater")
wilcox.test(Haemo$after,Haemo$before,paired=T,alternative="less")

#Direct calculation
xi<-Haemo$after-Haemo$before
xi<-xi[xi!=0]
si<-abs(xi)
n<-length(si)

Tplus<-sum(rank(si)[xi > 0])
Tminus<-sum(rank(si)[xi < 0])

(Z<-(Tplus-0.25*n*(n+1))/sqrt(n*(n + 1)*(2*n + 1)/24))
pnorm(Z)

#PEFRAsthma
wilcox.test(PEFRAsthma$after,PEFRAsthma$before,paired=T)
wilcox.test(PEFRAsthma$after,PEFRAsthma$before,paired=T,alternative="greater")
wilcox.test(PEFRAsthma$after,PEFRAsthma$before,paired=T,alternative="less")

#Direct calculation
xi<-PEFRAsthma$after-PEFRAsthma$before
xi<-xi[xi!=0]
si<-abs(xi)
n<-length(si)

Tplus<-sum(rank(si)[xi > 0]);Tminus<-sum(rank(si)[xi < 0])

```

```
(Z<-(Tplus-0.25*n*(n+1))/sqrt(n*(n + 1)*(2*n + 1)/24))
pnorm(Z)

#########################################################################
#Mucociliary data
#Mean ranks
xtabs(rank(y) ~ group,data=Muco)/xtabs( ~ group,data=Muco)
kruskal.test(y ~ group,data=Muco)

#MemoryTask
#Mean ranks
xtabs(rank(Memory) ~ Task,data=Memory)/xtabs( ~ Task,data=Memory)

kruskal.test(Memory ~ Task,data=Memory)

#########################################################################
#Skin potential data (Hypnosis)
friedman.test(SkinResponse ~ EmoCode | Subject,data=Hypnosis)

#Soil
friedman.test(sulphur ~ solvent | soil,data=Soil)

#########################################################################
#Rank correlation:DZT data

plot(DZT$Lat,DZT$DZT,pch=20,xlab="Latitude (degrees)",
    ylab="DZT Rate (per 1000, standardized)",
    xlim=range(37,65),ylim=range(5,12.5))
DZT.ylab<-DZT$DZT+0.25
DZT.ylab[c(11,13,18)]<-DZT.ylab[c(11,13,18)]-0.45
text(DZT$Lat,DZT.ylab,DZT$Country,cex=0.6)

cor(DZT$Lat,DZT$DZTRate,method="spearman")
cor(DZT$Lat,DZT$DZTRate,method="pearson")

#########################################################################
#########################################################################
#Permutation/small sample distribution of the Wilcoxon statistic

n1<-7;n2<-3;n<-n1+n2

Rvec<-c(1:n)
require(combinat)
Cmat<-combn(n,n2)
Scores<-apply(Cmat,2,sum)
Prob<-as.numeric(table(Scores))/length(Scores)
CProb<-cumsum(Prob)
par(mfrow=c(1,2))
plot(unique(Scores),Prob,pch=20,xlab="Score",ylab="Prob.")
plot(unique(Scores),CProb,pch=20,xlab="Score",ylab="Cumulative Prob.")
par(mfrow=c(1,1))
```

APPENDIX B: STATISTICAL TABLES

Table of the Fisher-F distribution

Entries in table are the $\alpha = 0.05$ tail quantile of Fisher-F(ν_1, ν_2) distribution
ν_1 given in columns, ν_2 given in rows.

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.91	244.69	245.36	245.95	246.46	246.92	247.32	247.69	248.01	248.31	248.58	248.83	249.05	249.26
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.42	19.43	19.43	19.44	19.44	19.44	19.45	19.45	19.45	19.45	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70	8.69	8.68	8.67	8.67	8.66	8.65	8.65	8.64	8.64	8.63
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86	5.84	5.83	5.82	5.81	5.80	5.79	5.79	5.78	5.77	5.77
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62	4.60	4.59	4.58	4.57	4.56	4.55	4.54	4.53	4.53	4.52
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94	3.92	3.91	3.90	3.88	3.87	3.86	3.86	3.85	3.84	3.83
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51	3.49	3.48	3.47	3.46	3.44	3.43	3.43	3.42	3.41	3.40
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22	3.20	3.19	3.17	3.16	3.15	3.14	3.13	3.12	3.12	3.11
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96	2.95	2.94	2.93	2.92	2.91	2.90	2.89
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80	2.79	2.77	2.76	2.75	2.75	2.74	2.73
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67	2.66	2.65	2.64	2.63	2.62	2.61	2.60
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57	2.56	2.54	2.53	2.52	2.51	2.51	2.50
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48	2.47	2.46	2.45	2.44	2.43	2.42	2.41
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41	2.40	2.39	2.38	2.37	2.36	2.35	2.34
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35	2.34	2.33	2.32	2.31	2.30	2.29	2.28
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.30	2.29	2.28	2.26	2.25	2.24	2.24	2.23
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.26	2.24	2.23	2.22	2.21	2.20	2.19	2.18
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.22	2.20	2.19	2.18	2.17	2.16	2.15	2.14
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.18	2.17	2.16	2.14	2.13	2.12	2.11	2.11
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20	2.18	2.17	2.15	2.14	2.12	2.11	2.10	2.09	2.08	2.07
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12	2.11	2.10	2.08	2.07	2.06	2.05	2.05
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17	2.15	2.13	2.11	2.10	2.08	2.07	2.06	2.05	2.04	2.03	2.02
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.18	2.15	2.13	2.11	2.09	2.08	2.06	2.05	2.04	2.02	2.01	2.01	2.00
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.15	2.13	2.11	2.09	2.07	2.05	2.04	2.03	2.01	2.00	1.99	1.98	1.97
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11	2.09	2.07	2.05	2.04	2.02	2.01	2.00	1.98	1.97	1.96	1.96
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.12	2.09	2.07	2.05	2.03	2.02	2.00	1.99	1.98	1.97	1.96	1.95	1.94
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13	2.10	2.08	2.06	2.04	2.02	2.00	1.99	1.97	1.96	1.95	1.94	1.93	1.92
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.09	2.06	2.04	2.02	2.00	1.99	1.97	1.96	1.95	1.93	1.92	1.91	1.91
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.08	2.05	2.03	2.01	1.99	1.97	1.96	1.94	1.93	1.92	1.91	1.90	1.89
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01	1.99	1.98	1.96	1.95	1.93	1.92	1.91	1.90	1.89	1.88
31	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15	2.11	2.08	2.05	2.03	2.00	1.98	1.96	1.95	1.93	1.92	1.91	1.90	1.88	1.88	1.87
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.10	2.07	2.04	2.01	1.99	1.97	1.95	1.94	1.92	1.91	1.90	1.88	1.87	1.86	1.85

Table of the Fisher-F distribution

Entries in table are the $\alpha = 0.05$ tail quantile of Fisher-F(ν_1, ν_2) distribution
ν_1 given in columns, ν_2 given in rows.

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	11	12
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.91
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
31	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15	2.11	2.08
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.10	2.07

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	11	12
33	4.14	3.28	2.89	2.66	2.50	2.39	2.30	2.23	2.18	2.13	2.09	2.06
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	2.08	2.05
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.07	2.04
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	2.07	2.03
37	4.11	3.25	2.86	2.63	2.47	2.36	2.27	2.20	2.14	2.10	2.06	2.02
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02
39	4.09	3.24	2.85	2.61	2.46	2.34	2.26	2.19	2.13	2.08	2.04	2.01
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00
41	4.08	3.23	2.83	2.60	2.44	2.33	2.24	2.17	2.12	2.07	2.03	2.00
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.03	1.99
43	4.07	3.21	2.82	2.59	2.43	2.32	2.23	2.16	2.11	2.06	2.02	1.99
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98
45	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	2.01	1.97
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04	2.00	1.97
47	4.05	3.20	2.80	2.57	2.41	2.30	2.21	2.14	2.09	2.04	2.00	1.96
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03	1.99	1.96
49	4.04	3.19	2.79	2.56	2.40	2.29	2.20	2.13	2.08	2.03	1.99	1.96
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95
51	4.03	3.18	2.79	2.55	2.40	2.28	2.20	2.13	2.07	2.02	1.98	1.95
52	4.03	3.18	2.78	2.55	2.39	2.28	2.19	2.12	2.07	2.02	1.98	1.94
53	4.02	3.17	2.78	2.55	2.39	2.28	2.19	2.12	2.06	2.01	1.97	1.94
54	4.02	3.17	2.78	2.54	2.39	2.27	2.18	2.12	2.06	2.01	1.97	1.94
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01	1.97	1.93
56	4.01	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.96	1.93
57	4.01	3.16	2.77	2.53	2.38	2.26	2.18	2.11	2.05	2.00	1.96	1.93
58	4.01	3.16	2.76	2.53	2.37	2.26	2.17	2.10	2.05	2.00	1.96	1.92
59	4.00	3.15	2.76	2.53	2.37	2.26	2.17	2.10	2.04	2.00	1.96	1.92
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
61	4.00	3.15	2.76	2.52	2.37	2.25	2.16	2.09	2.04	1.99	1.95	1.91
62	4.00	3.15	2.75	2.52	2.36	2.25	2.16	2.09	2.03	1.99	1.95	1.91
63	3.99	3.14	2.75	2.52	2.36	2.25	2.16	2.09	2.03	1.98	1.94	1.91
64	3.99	3.14	2.75	2.52	2.36	2.24	2.16	2.09	2.03	1.98	1.94	1.91

Table of the Chi-squared distribution

Entries in table are $\chi^2_\alpha(\nu)$: the α tail quantile of Chi-squared(ν) distribution
α given in columns, ν given in rows.

	Left-tail					Right-tail				
ν	0.99500	0.99000	0.97500	0.95000	0.90000	0.10000	0.05000	0.02500	0.01000	0.00500
1	0.00004	0.00016	0.00098	0.00393	0.01579	2.70554	3.84146	5.02389	6.63490	7.87944
2	0.01003	0.02010	0.05064	0.10259	0.21072	4.60517	5.99146	7.37776	9.21034	10.59663
3	0.07172	0.11483	0.21580	0.35185	0.58437	6.25139	7.81473	9.34840	11.34487	12.83816
4	0.20699	0.29711	0.48442	0.71072	1.06362	7.77944	9.48773	11.14329	13.27670	14.86026
5	0.41174	0.55430	0.83121	1.14548	1.61031	9.23636	11.07050	12.83250	15.08627	16.74960
6	0.67573	0.87209	1.23734	1.63538	2.20413	10.64464	12.59159	14.44938	16.81189	18.54758
7	0.98926	1.23904	1.68987	2.16735	2.83311	12.01704	14.06714	16.01276	18.47531	20.27774
8	1.34441	1.64650	2.17973	2.73264	3.48954	13.36157	15.50731	17.53455	20.09024	21.95495
9	1.73493	2.08790	2.70039	3.32511	4.16816	14.68366	16.91898	19.02277	21.66599	23.58935
10	2.15586	2.55821	3.24697	3.94030	4.86518	15.98718	18.30704	20.48318	23.20925	25.18818
11	2.60322	3.05348	3.81575	4.57481	5.57778	17.27501	19.67514	21.92005	24.72497	26.75685
12	3.07382	3.57057	4.40379	5.22603	6.30380	18.54935	21.02607	23.33666	26.21697	28.29952
13	3.56503	4.10692	5.00875	5.89186	7.04150	19.81193	22.36203	24.73560	27.68825	29.81947
14	4.07467	4.66043	5.62873	6.57063	7.78953	21.06414	23.68479	26.11895	29.14124	31.31935
15	4.60092	5.22935	6.26214	7.26094	8.54676	22.30713	24.99579	27.48839	30.57791	32.80132
16	5.14221	5.81221	6.90766	7.96165	9.31224	23.54183	26.29623	28.84535	31.99993	34.26719
17	5.69722	6.40776	7.56419	8.67176	10.08519	24.76904	27.58711	30.19101	33.40866	35.71847
18	6.26480	7.01491	8.23075	9.39046	10.86494	25.98942	28.86930	31.52638	34.80531	37.15645
19	6.84397	7.63273	8.90652	10.11701	11.65091	27.20357	30.14353	32.85233	36.19087	38.58226
20	7.43384	8.26040	9.59078	10.85081	12.44261	28.41198	31.41043	34.16961	37.56623	39.99685
21	8.03365	8.89720	10.28290	11.59131	13.23960	29.61509	32.67057	35.47888	38.93217	41.40106
22	8.64272	9.54249	10.98232	12.33801	14.04149	30.81328	33.92444	36.78071	40.28936	42.79565
23	9.26042	10.19572	11.68855	13.09051	14.84796	32.00690	35.17246	38.07563	41.63840	44.18128
24	9.88623	10.85636	12.40115	13.84843	15.65868	33.19624	36.41503	39.36408	42.97982	45.55851
25	10.51965	11.52398	13.11972	14.61141	16.47341	34.38159	37.65248	40.64647	44.31410	46.92789
26	11.16024	12.19815	13.84390	15.37916	17.29188	35.56317	38.88514	41.92317	45.64168	48.28988
27	11.80759	12.87850	14.57338	16.15140	18.11390	36.74122	40.11327	43.19451	46.96294	49.64492
28	12.46134	13.56471	15.30786	16.92788	18.93924	37.91592	41.33714	44.46079	48.27824	50.99338
29	13.12115	14.25645	16.04707	17.70837	19.76774	39.08747	42.55697	45.72229	49.58788	52.33562
30	13.78672	14.95346	16.79077	18.49266	20.59923	40.25602	43.77297	46.97924	50.89218	53.67196
40	20.70654	22.16426	24.43304	26.50930	29.05052	51.80506	55.75848	59.34171	63.69074	66.76596
50	27.99075	29.70668	32.35736	34.76425	37.68865	63.16712	67.50481	71.42020	76.15389	79.48998
60	35.53449	37.48485	40.48175	43.18796	46.45889	74.39701	79.08194	83.29767	88.37942	91.95170
70	43.27518	45.44172	48.75756	51.73928	55.32894	85.52704	90.53123	95.02318	100.42518	104.21490
80	51.17193	53.54008	57.15317	60.39148	64.27784	96.57820	101.87947	106.62857	112.32879	116.32106
90	59.19630	61.75408	65.64662	69.12603	73.29109	107.56501	113.14527	118.13589	124.11632	128.29894
100	67.32756	70.06489	74.22193	77.92947	82.35814	118.49800	124.34211	129.56120	135.80672	140.16949

Wilcoxon Rank Sum Test: Independent Samples.

Critical Values T_L and T_U

Test statistic is the rank sum associated with the **smaller sample** (if equal sample sizes, either rank sum can be used)

$\alpha = .025$ one-tailed, $\alpha = .05$ two-tailed.

	n_1															
	3		4		5		6		7		8		9		10	
n_2	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	5	16	6	18	6	21	7	23	7	26	8	28	8	31	9	33
4	6	18	11	25	12	28	12	32	13	35	14	38	15	41	16	44
5	6	21	12	28	18	37	19	41	20	45	21	49	22	53	24	56
6	7	23	12	32	19	41	26	52	28	56	29	61	31	65	32	70
7	7	26	13	35	20	45	28	56	37	68	39	73	41	78	43	83
8	8	28	14	38	21	49	29	61	39	73	49	87	51	93	54	98
9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	66	114
10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131

$\alpha = .05$ one-tailed, $\alpha = .10$ two-tailed.

	n_1															
	3		4		5		6		7		8		9		10	
n_2	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	6	15	7	17	7	20	8	22	9	24	9	27	10	29	11	31
4	7	17	12	24	13	27	14	30	15	33	16	36	17	39	18	42
5	7	20	13	27	19	36	20	40	22	43	24	46	25	50	26	54
6	8	22	14	30	20	40	28	50	30	54	32	58	33	63	35	67
7	9	24	15	33	22	43	30	54	39	66	41	71	43	76	46	80
8	9	27	16	36	24	46	32	58	41	71	52	84	54	90	57	95
9	10	29	17	39	25	50	33	63	43	76	54	90	66	105	69	111
10	11	31	18	42	26	54	35	67	46	80	57	95	69	111	83	127

Wilcoxon Matched-Pairs Signed Rank Test.

Critical Values T_0

One Tailed	Two Tailed	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$\alpha = 0.05$	$\alpha = 0.10$	1	2	4	6	8	11
$\alpha = 0.025$	$\alpha = 0.05$		1	2	4	6	8
$\alpha = 0.010$	$\alpha = 0.020$			0	2	3	5
$\alpha = 0.005$	$\alpha = 0.010$				0	2	3
		$n = 11$	$n = 12$	$n = 13$	$n = 14$	$n = 15$	$n = 16$
$\alpha = 0.05$	$\alpha = 0.10$	14	17	21	26	30	36
$\alpha = 0.025$	$\alpha = 0.05$	11	14	17	21	25	30
$\alpha = 0.010$	$\alpha = 0.020$	7	10	13	16	20	24
$\alpha = 0.005$	$\alpha = 0.010$	5	7	10	13	16	19
		$n = 17$	$n = 18$	$n = 19$	$n = 20$	$n = 21$	$n = 22$
$\alpha = 0.05$	$\alpha = 0.10$	41	47	54	60	68	75
$\alpha = 0.025$	$\alpha = 0.05$	35	40	46	52	59	66
$\alpha = 0.010$	$\alpha = 0.020$	28	33	38	43	49	56
$\alpha = 0.005$	$\alpha = 0.010$	23	28	32	37	43	49
		$n = 23$	$n = 24$	$n = 25$	$n = 26$	$n = 27$	$n = 28$
$\alpha = 0.05$	$\alpha = 0.10$	83	92	101	110	120	130
$\alpha = 0.025$	$\alpha = 0.05$	73	81	90	98	107	117
$\alpha = 0.010$	$\alpha = 0.020$	62	69	77	85	93	102
$\alpha = 0.005$	$\alpha = 0.010$	55	61	68	76	84	92
		$n = 29$	$n = 30$	$n = 31$	$n = 32$	$n = 33$	$n = 34$
$\alpha = 0.05$	$\alpha = 0.10$	141	152	163	175	188	201
$\alpha = 0.025$	$\alpha = 0.05$	127	137	148	159	171	183
$\alpha = 0.010$	$\alpha = 0.020$	111	120	130	141	151	162
$\alpha = 0.005$	$\alpha = 0.010$	100	109	118	128	138	149
		$n = 35$	$n = 36$	$n = 37$	$n = 38$	$n = 39$	$n = 40$
$\alpha = 0.05$	$\alpha = 0.10$	214	228	242	256	271	287
$\alpha = 0.025$	$\alpha = 0.05$	195	208	222	235	250	264
$\alpha = 0.010$	$\alpha = 0.020$	174	186	198	211	224	238
$\alpha = 0.005$	$\alpha = 0.010$	160	171	183	195	208	221
		$n = 41$	$n = 42$	$n = 43$	$n = 44$	$n = 45$	$n = 46$
$\alpha = 0.05$	$\alpha = 0.10$	303	319	336	353	371	389
$\alpha = 0.025$	$\alpha = 0.05$	279	295	311	327	344	361
$\alpha = 0.010$	$\alpha = 0.020$	252	267	281	297	313	329
$\alpha = 0.005$	$\alpha = 0.010$	234	248	262	277	292	307
		$n = 47$	$n = 48$	$n = 49$	$n = 50$		
$\alpha = 0.05$	$\alpha = 0.10$	408	427	446	466		
$\alpha = 0.025$	$\alpha = 0.05$	379	397	415	434		
$\alpha = 0.010$	$\alpha = 0.020$	345	362	380	398		
$\alpha = 0.005$	$\alpha = 0.010$	323	339	356	373		