MATH 204: PRINCIPLES OF STATISTICS 2

WINTER 2009

MATH 204: PRINCIPLES OF STATISTICS 2

WINTER 2009

Instructor :	David A. Stephens (Burnside 1235)
Email :	d.stephens@math.mcgill.ca
Office Hours :	Monday 15:00-17:00
Tutorial :	TBA
Teaching Assistant :	Ashkan Ertefaie
Textbook :	*Statistics* (10th or 11th Edition) by J. T. McClave and T. Sincich.
Web Site :	http://www.math.mcgill.ca/∼dstephens/204/

TARGET 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	From Friday 16th January 2009
Mid-Term	Week of 2nd February - 9th February 2009 Take Home
Final	Closed book (with formula sheet)

Final mark for course: the larger of

$$15\ \%\ \text{Coursework} + 25\ \%\ \text{Mid-Term} + 60\ \%\ \text{Final}$$

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.

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).

David A. Stephens.
December 11, 2008

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

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

and

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

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

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

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 on page 901 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

$$\text{SST} = 1259.199 \qquad \text{SSE} = 14.539 \qquad \text{SS} = 1273.738$$

so that

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

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
$$\Pr[\, F > \textbf{5.96} \,] = 0.05$$
- if $\nu_1 = 6$ and $\nu_2 = 18$, we know from the table that
$$\Pr[\, F > \textbf{2.66} \,] = 0.05$$
- if $\nu_1 = 20$ and $\nu_2 = 20$, we know from the table that
$$\Pr[\, F > \textbf{2.12} \,] = 0.05$$

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
$$\Pr[\, F \leq 1/\textbf{3.48} \,] = \Pr[\, F \leq \textbf{0.29} \,] = 0.05$$
giving the $\alpha = 0.05$ (left-tail) probability point of the Fisher-F$(10, 4)$ distribution as 0.29.

Table of the Fisher-F distribution

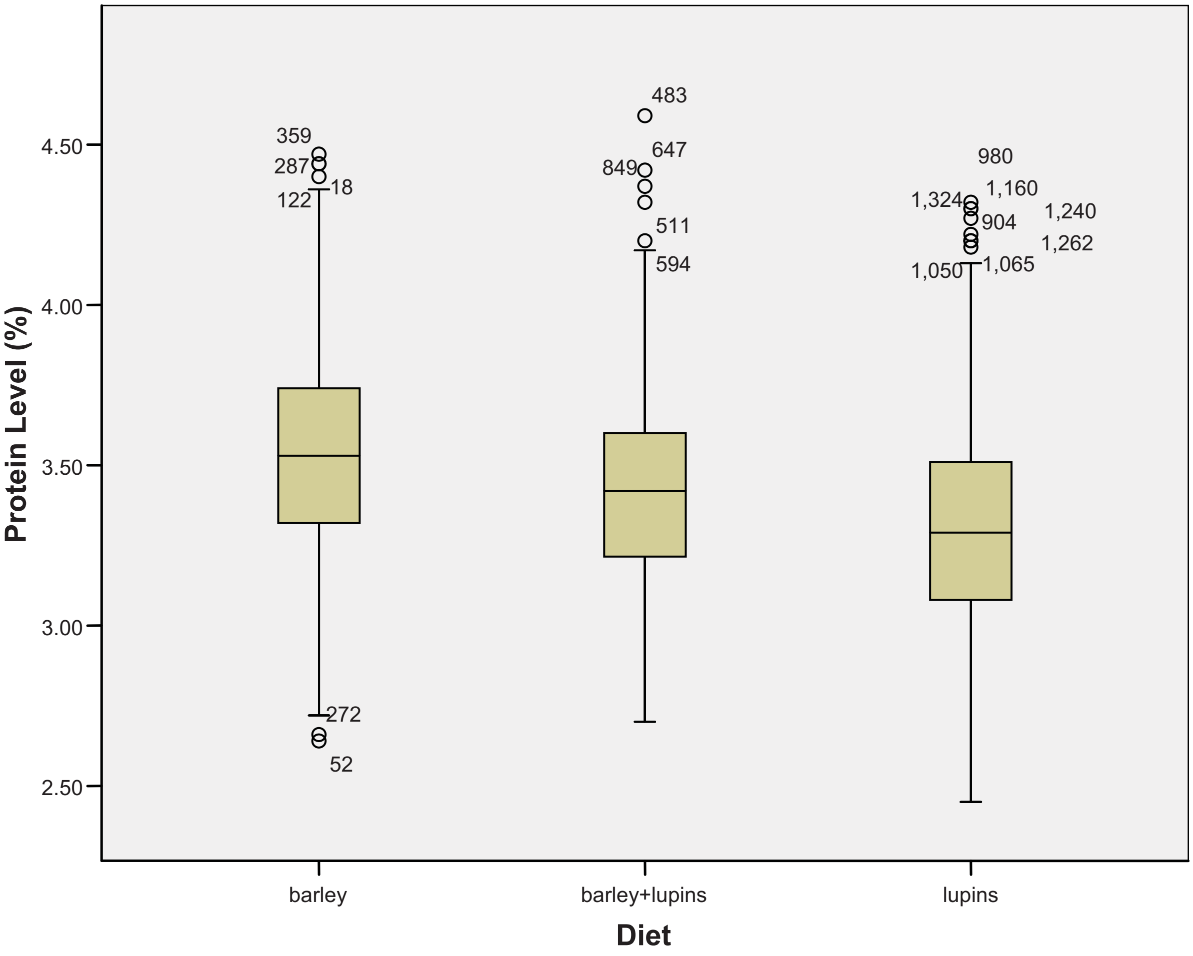
Entries in table are the $\alpha = 0.05$ tail quantile of Fisher-F(ν_1, ν_2) distribution
ν_1 given in columns, ν_2 given in rows.

$\nu_2 \backslash \nu_1$	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

ANOVA F-TEST : EXAMPLES

Diet

Protein Level (%)



Oneway

Descriptives

Protein Level (%)

					95% Confidence Interval for Mean			
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
barley	425	3.5319	.31921	.01548	3.5015	3.5624	2.64	4.47
barley+lupins	459	3.4297	.30234	.01411	3.4020	3.4574	2.70	4.59
lupins	453	3.3124	.33709	.01584	3.2813	3.3435	2.45	4.32
Total	1337	3.4224	.33175	.00907	3.4046	3.4402	2.45	4.59

Test of Homogeneity of Variances

Protein Level (%)

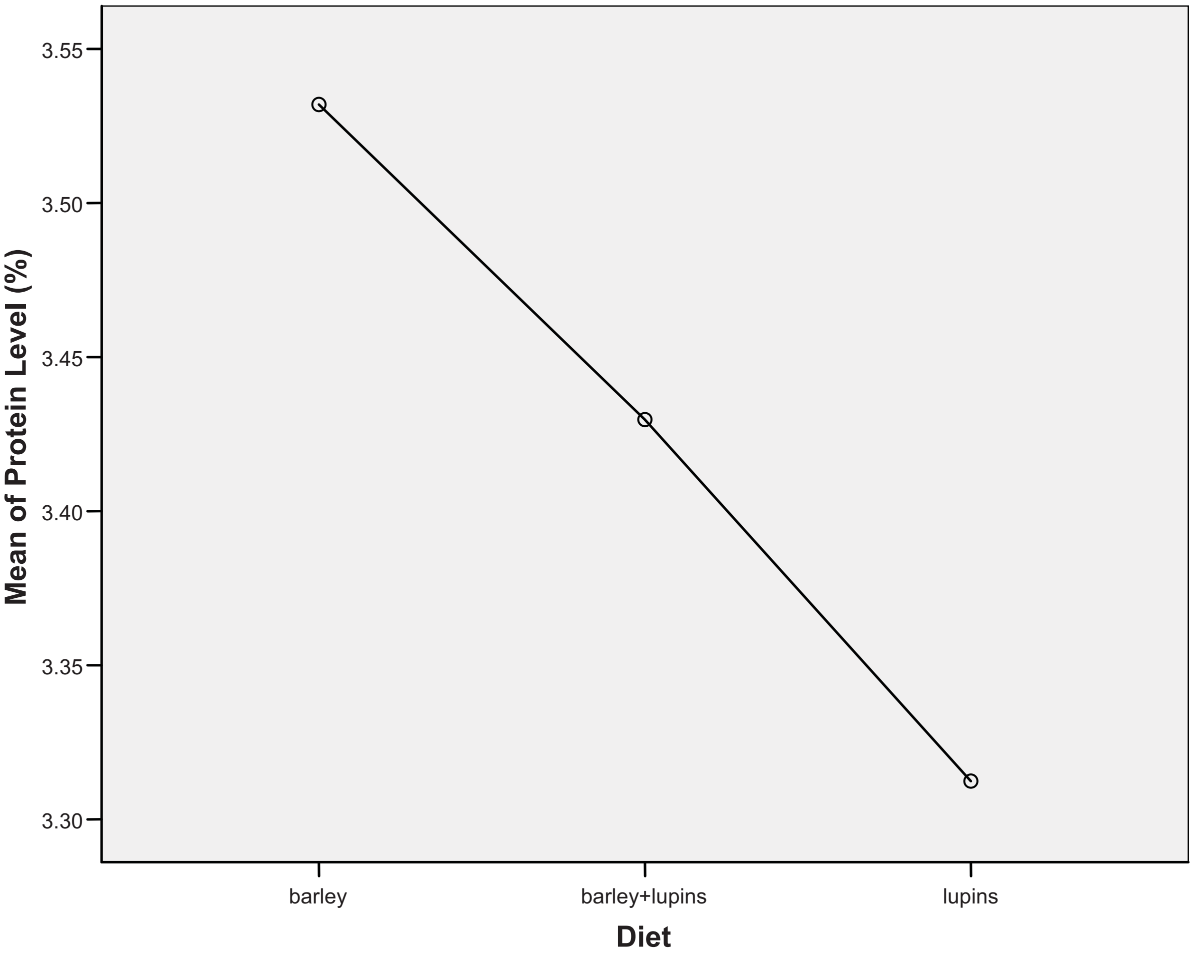
Levene Statistic	df1	df2	Sig.
1.838	2	1334	.160

ANOVA

Protein Level (%)

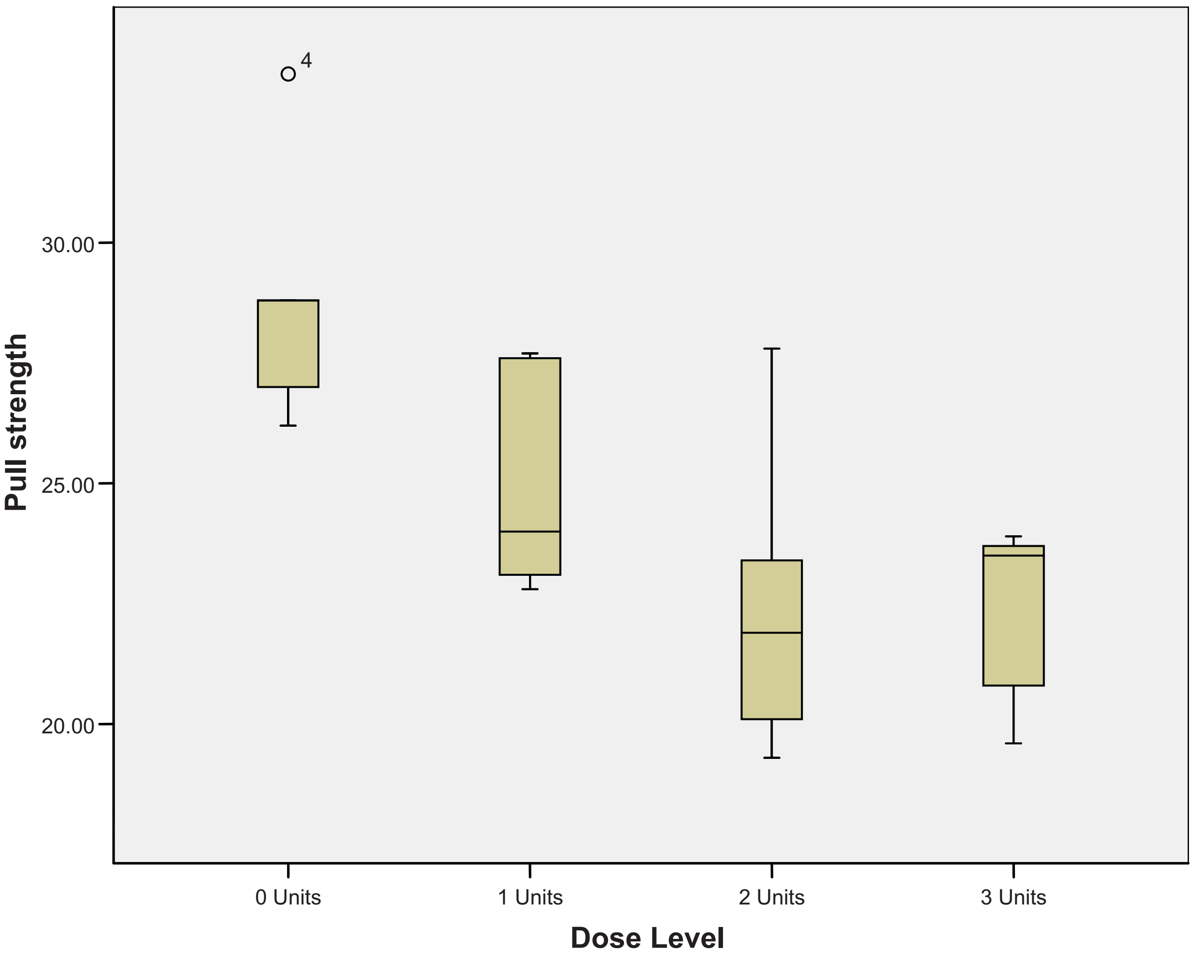
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.606	2	5.303	51.851	.000
Within Groups	136.432	1334	.102		
Total	147.038	1336			

Means Plots



Dose Level

Pull strength



Oneway

Descriptives

Pull strength

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
0 Units	5	28.8600	2.83161	1.26633	25.3441	32.3759	26.2	33.5
1 Units	5	25.0400	2.42343	1.08379	22.0309	28.0491	22.8	27.7
2 Units	5	22.5000	3.36378	1.50433	18.3233	26.6767	19.3	27.8
3 Units	5	22.3000	1.96850	.88034	19.8558	24.7442	19.6	23.9
Total	20	24.6750	3.67364	.82145	22.9557	26.3943	19.3	33.5

Test of Homogeneity of Variances

Pull strength

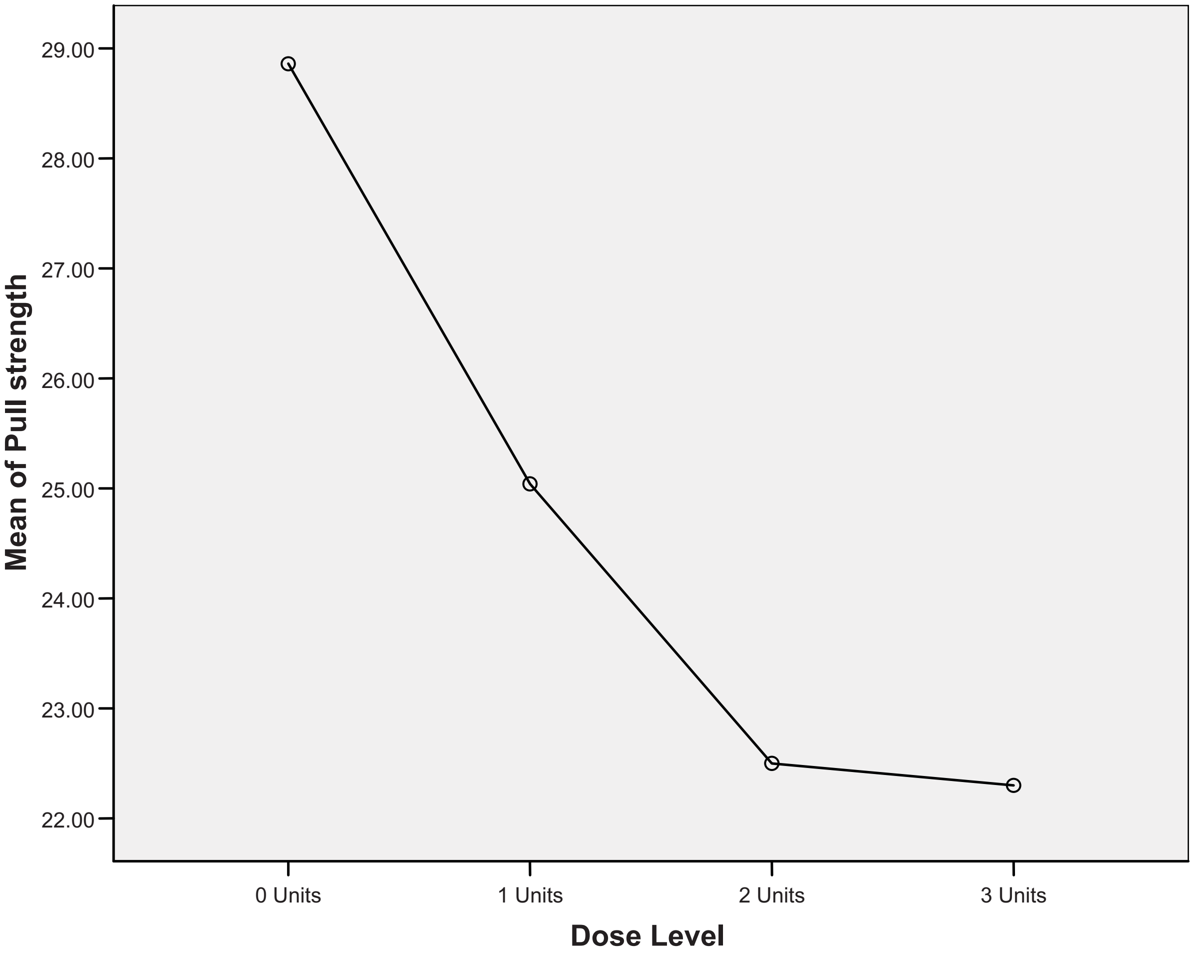
Levene Statistic	df1	df2	Sig.
.295	3	16	.829

ANOVA

Pull strength

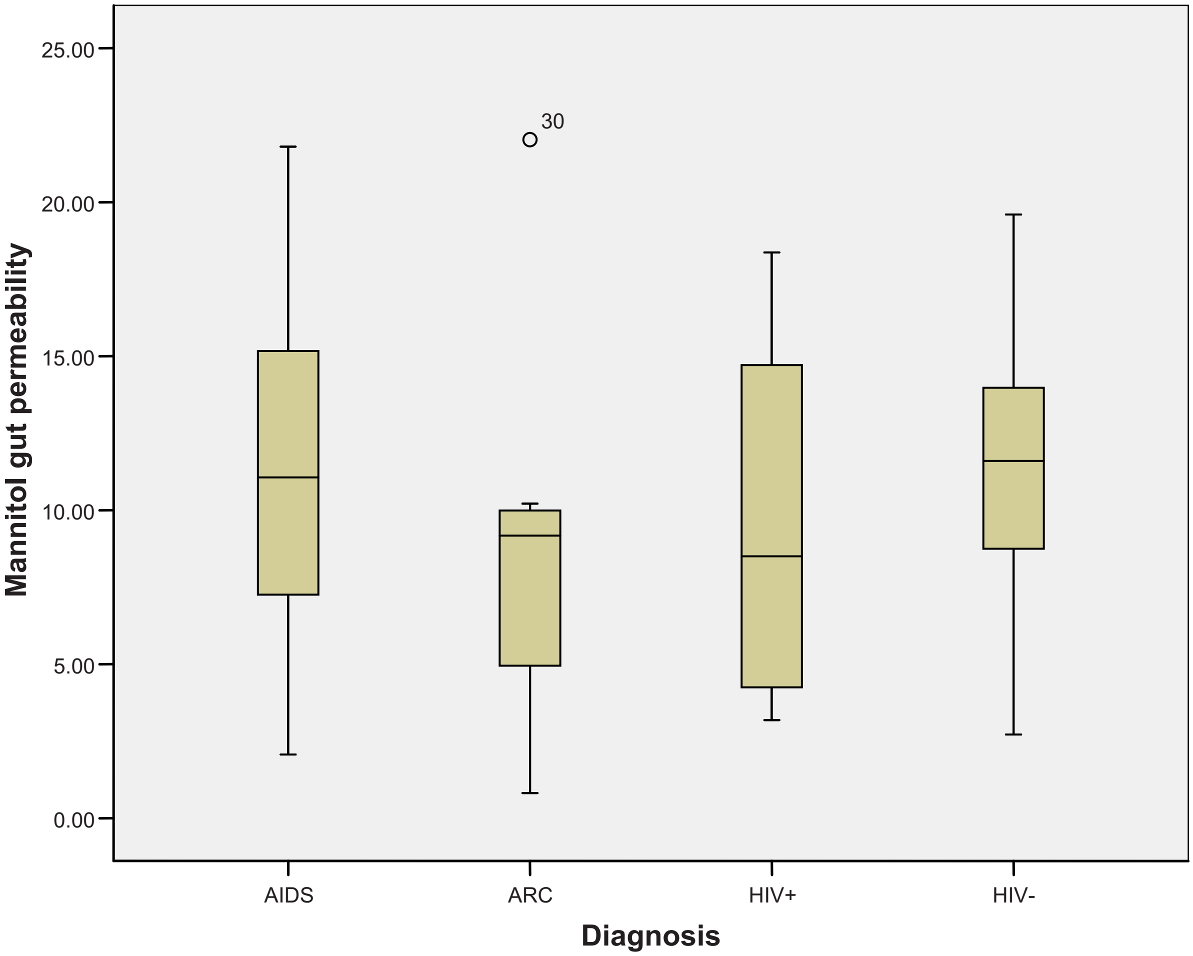
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	140.094	3	46.698	6.423	.005
Within Groups	116.324	16	7.270		
Total	256.418	19			

Means Plots



Diagnosis

Mannitol gut permeability



Oneway

Descriptives

Mannitol gut permeability

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
AIDS	26	11.3312	5.17639	1.01517	9.2404	13.4220	2.07	21.80
ARC	7	8.8419	6.87762	2.59950	2.4811	15.2026	.81	22.03
HIV+	7	9.7104	6.18770	2.33873	3.9878	15.4331	3.18	18.37
HIV-	19	11.3970	4.25972	.97725	9.3439	13.4501	2.72	19.60
Total	59	10.8647	5.18458	.67498	9.5136	12.2159	.81	22.03

Test of Homogeneity of Variances

Mannitol gut permeability

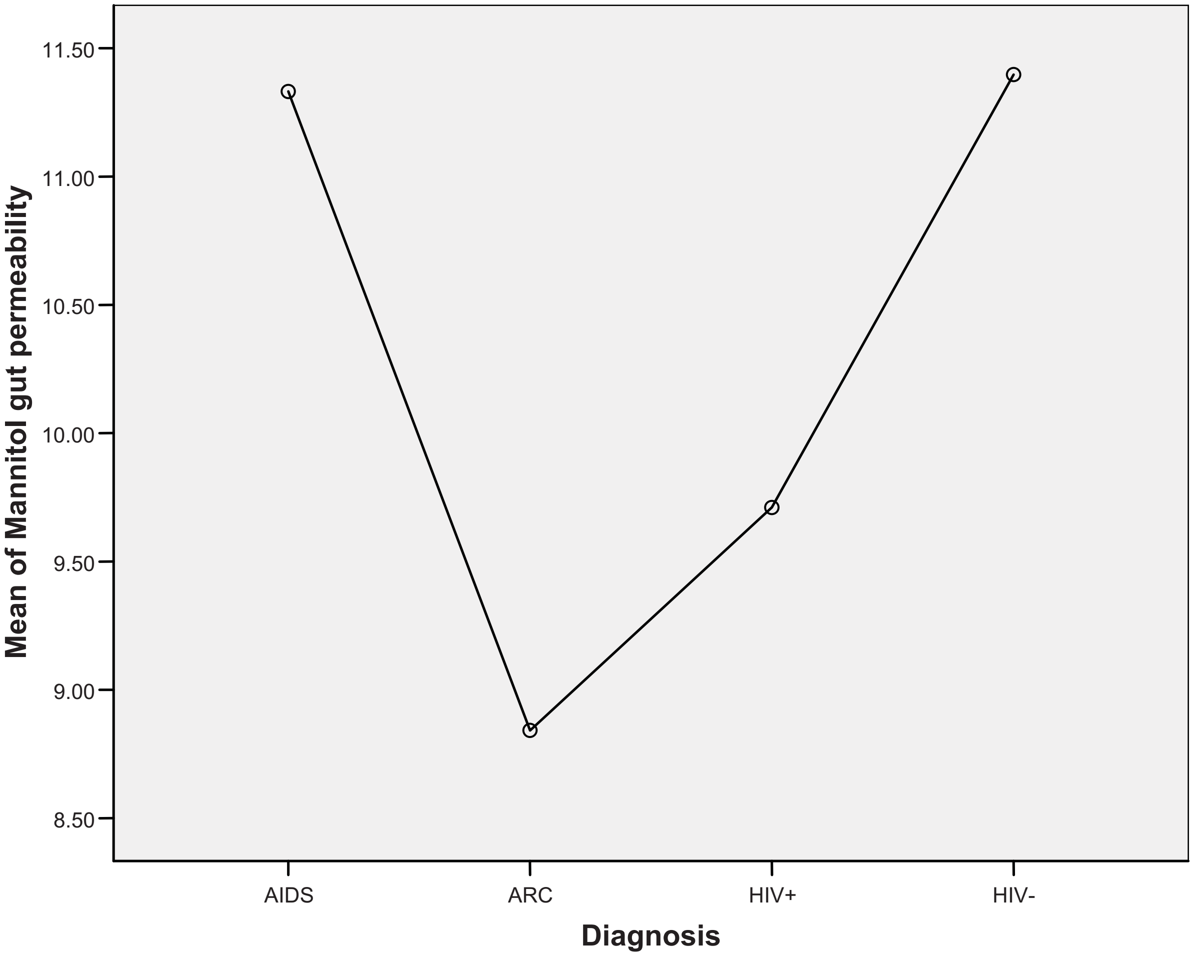
Levene Statistic	df1	df2	Sig.
.866	3	55	.464

ANOVA

Mannitol gut permeability

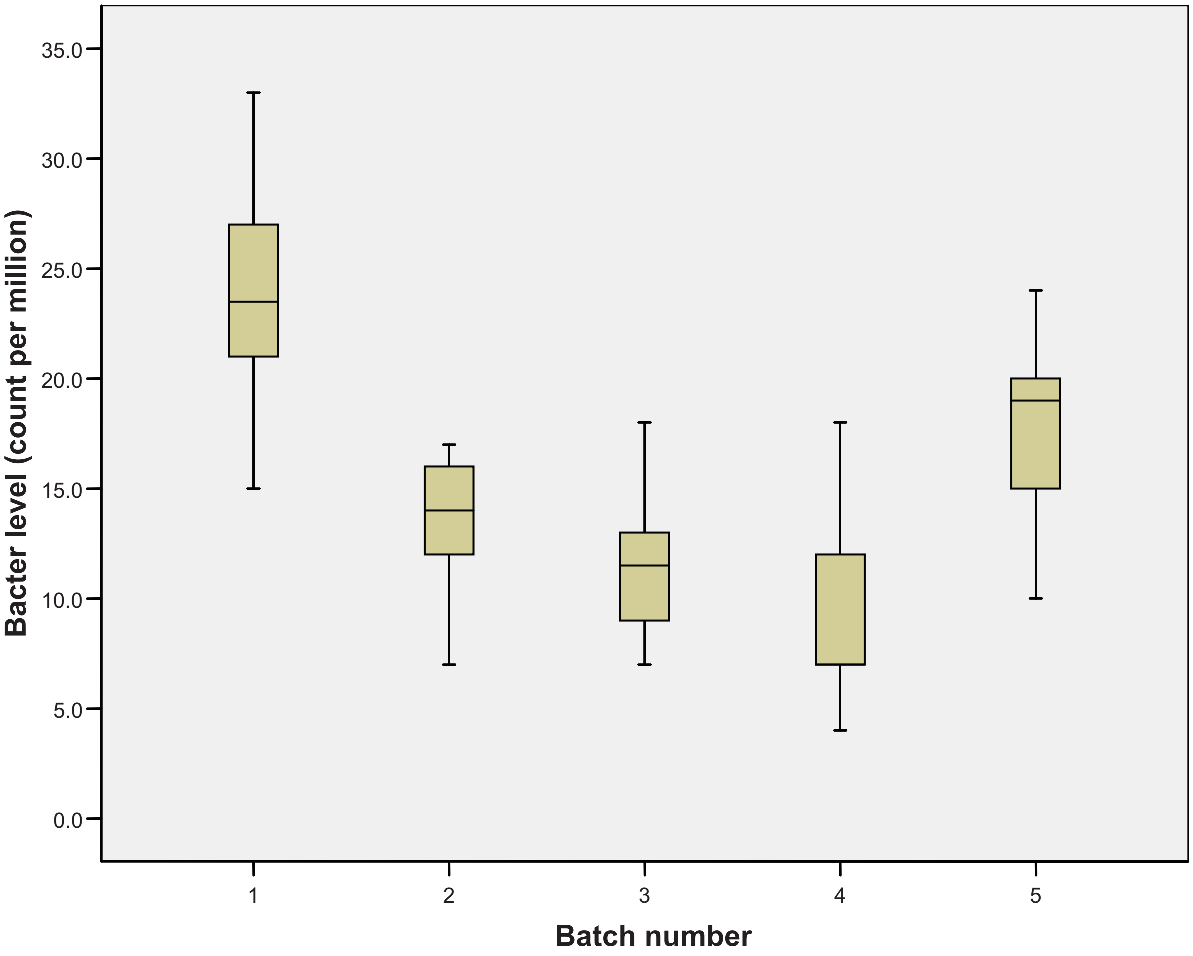
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	49.011	3	16.337	.595	.621
Within Groups	1510.024	55	27.455		
Total	1559.035	58			

Means Plots



Batch number

Bacteria level (count per million)



Oneway

Descriptives

Bacter level (count per million)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	6	23.833	6.0139	2.4552	17.522	30.145	15	33
2	6	13.333	3.5590	1.4530	9.598	17.068	7	17
3	6	11.667	3.7771	1.5420	7.703	15.631	7	18
4	6	9.167	5.0365	2.0562	3.881	14.452	4	18
5	6	17.833	4.7924	1.9565	12.804	22.863	10	24
Total	30	15.167	6.8485	1.2504	12.609	17.724	4	33

Test of Homogeneity of Variances

Bacter level (count per million)

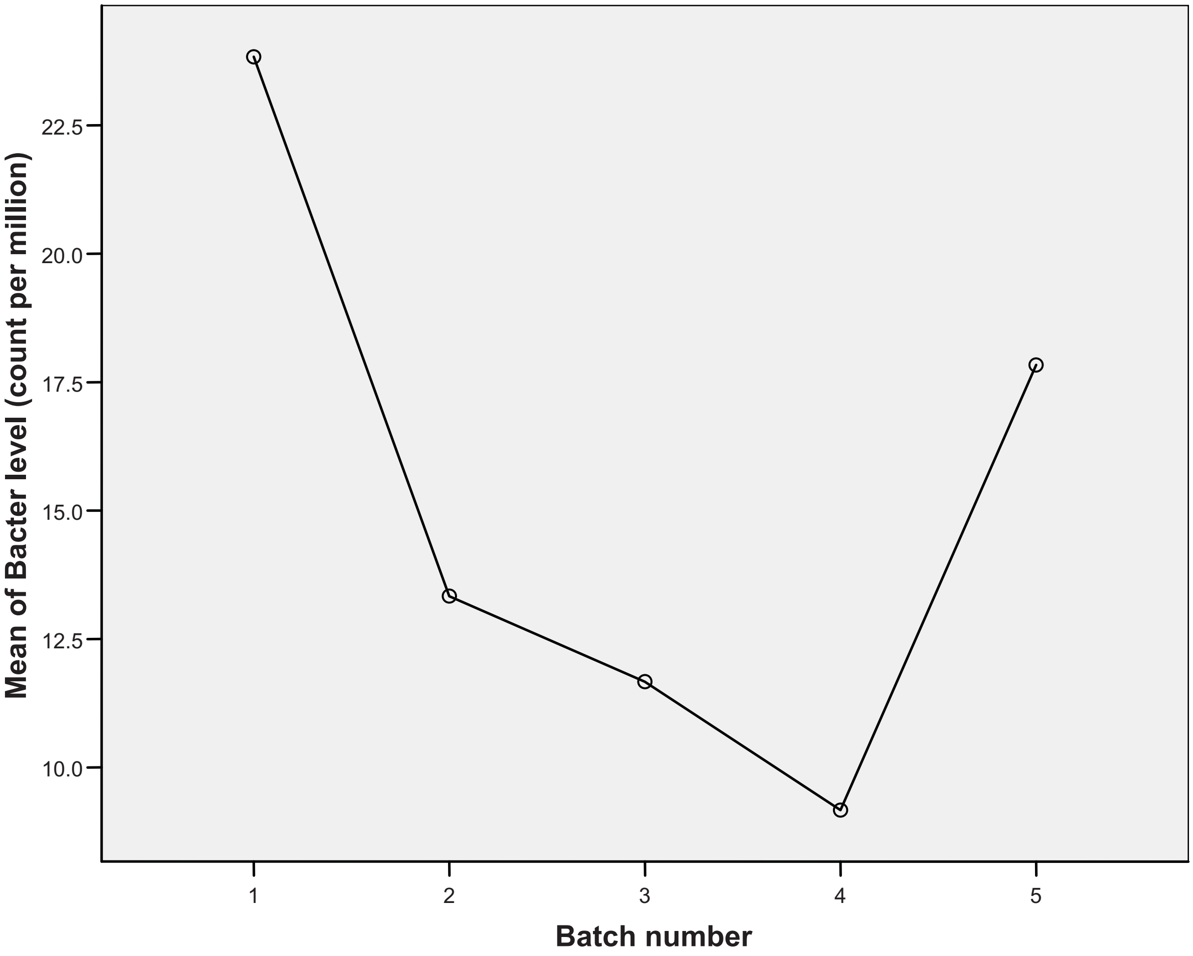
Levene Statistic	df1	df2	Sig.
.384	4	25	.818

ANOVA

Bacter level (count per million)

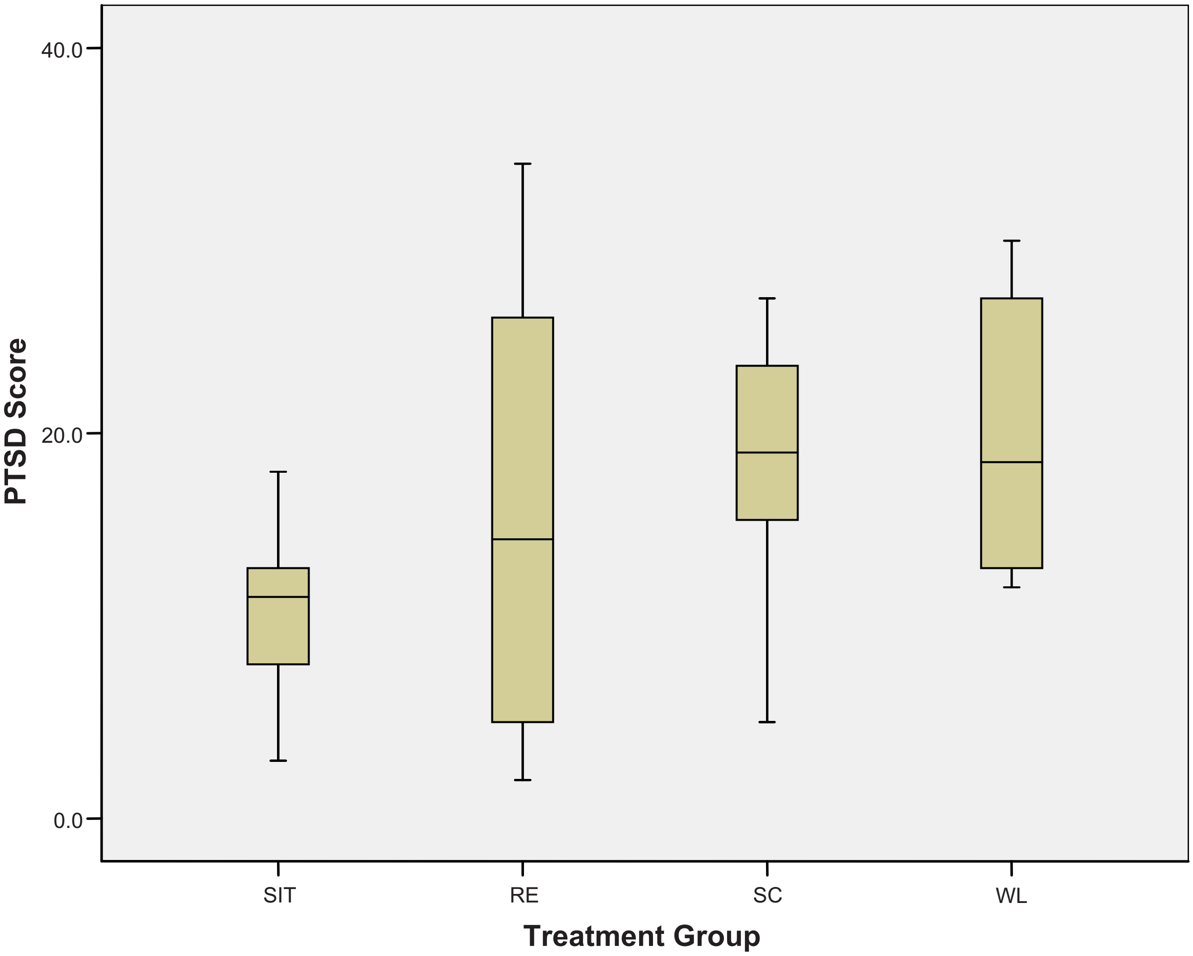
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	803.000	4	200.750	9.008	.000
Within Groups	557.167	25	22.287		
Total	1360.167	29			

Means Plots



Treatment Group

PTSD Score



Oneway

Descriptives

SystBlood

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	19	22.789	13.1596	3.0190	16.447	29.132	-2	44
2	19	18.211	13.5547	3.1097	11.677	24.744	-6	36
3	20	15.800	11.3025	2.5273	10.510	21.090	-3	32
Total	58	18.879	12.8009	1.6808	15.513	22.245	-6	44

Test of Homogeneity of Variances

SystBlood

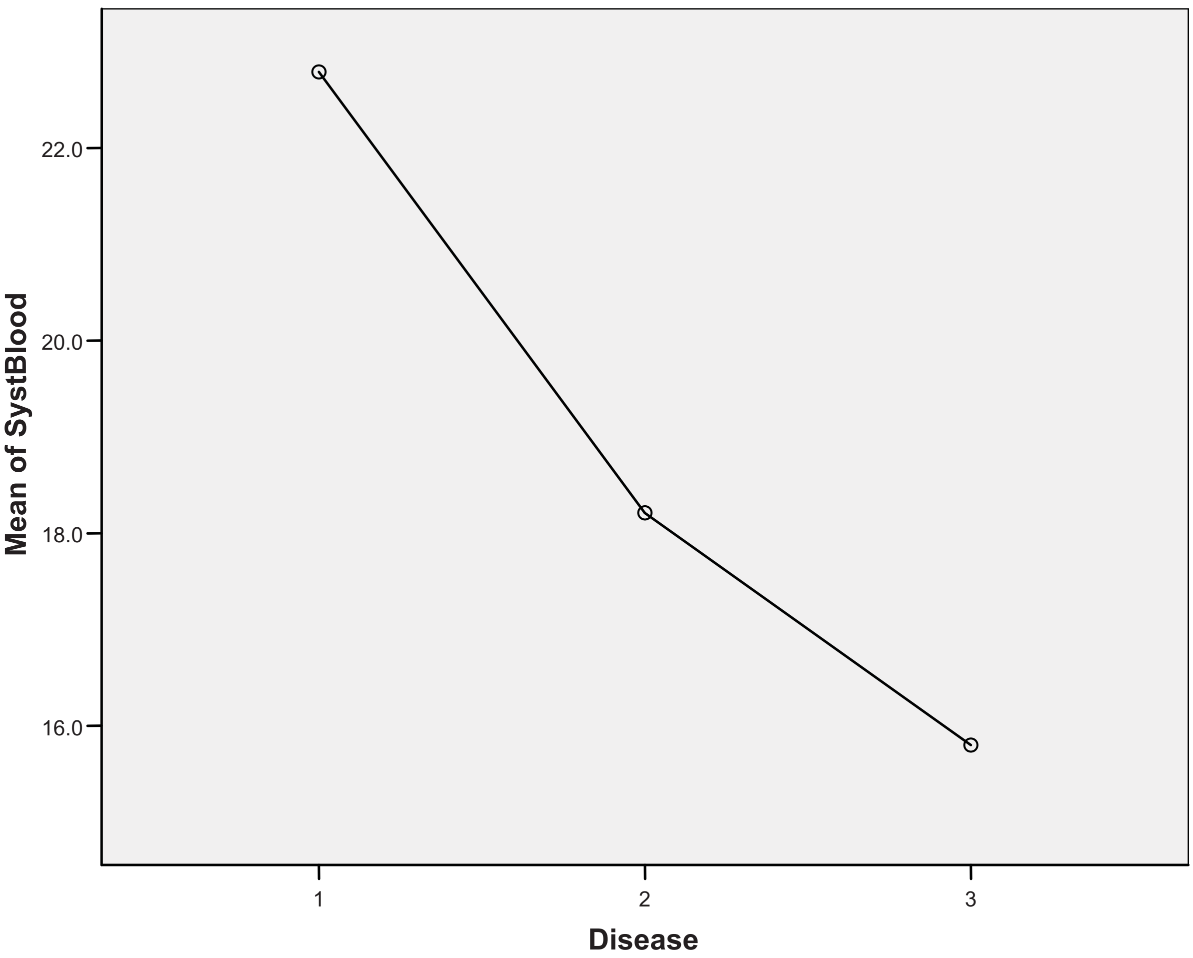
Levene Statistic	df1	df2	Sig.
.317	2	55	.730

ANOVA

SystBlood

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	488.639	2	244.320	1.518	.228
Within Groups	8851.516	55	160.937		
Total	9340.155	57			

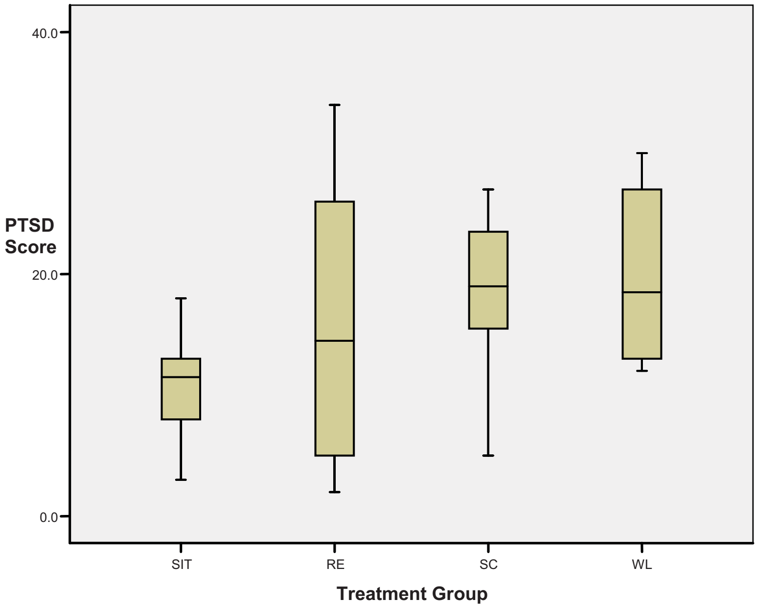
Means Plots



PTSD Analysis

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.

RESPONSE: PTSD Score **FACTOR: Treatment Group (k=4 levels)**



Summary Statistics

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
SIT	14	11.071	3.9509	1.0559	8.790	13.353	3	18
RE	10	15.400	11.1176	3.5157	7.447	23.353	2	34
SC	11	18.091	7.1338	2.1509	13.298	22.883	5	27
WL	10	19.500	7.1063	2.2472	14.416	24.584	12	30
Total	45	15.622	7.9581	1.1863	13.231	18.013	2	34

Test of Homogeneity of Variances

PTSD Score

Levene Statistic	df1	df2	Sig.
6.633	3	41	.001

P-value = 0.001 < 0.05, so the Levene test of equality of variances between the treatment groups REJECTS the hypothesis of equal variances.

ANOVA TABLE

PTSD Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	507.840	3	169.280	3.046	.039
Within Groups	2278.738	41	55.579		
Total	2786.578	44			

P-value = 0.039 < 0.05 so the ANOVA-F test REJECTS the hypothesis of equal treatment means.

IS THE CONCLUSION OF THE ANOVA F-TEST CORRECT IF THE EQUAL VARIANCE ASSUMPTION IS NOT MET ?

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

These data may be downloaded

- in plain text format from

  http://www.math.mcgill.ca/∼dstephens/204/Data/MemoryTask.txt

- in SPSS format from

  http://www.math.mcgill.ca/∼dstephens/204/Data/MemoryTask.sav

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

Test a hypothesis to answer this question using an ANOVA F-test. Specifically

(a) Form the ANOVA table, and report the result of the ANOVA F-test.
(b) Discuss whether the assumptions of behind the ANOVA F-test hold for this example.

MATH 204 - One Way ANOVA Worked Example
Solution

Memory Task Data Set: Response is Number of Words remembered, Factor is Memory Training method.

(a) ANOVA TABLE (from SPSS)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	351.520	4	87.880	9.085	.000
Within Groups	435.300	45	9.673		
Total	786.820	49			

ANOVA F-test statistic F=9.085

ANOVA F-test p-value = 0.000.
(to three decimal places)

Thus the result of the ANOVA F-test implies that we can

REJECT H_0

at significance levels $\alpha = 0.05/0.01$, and conclude that there is a significant difference between the treatment means.

For completeness: the exact *p*-value is 1.815e-05. Critical values are

- $\alpha = 0.05$, $C_R = F_\alpha(4,45) = 2.579$ (textbook gives $F_\alpha(4,40) = 2.61, F_\alpha(4,60) = 2.53$)
- $\alpha = 0.01$, $CR = F_\alpha(4,45) = 3.767$ (textbook gives $F_\alpha(4,40) = 3.83, F_\alpha(4,60) = 3.65$)

Therefore we reject the hypothesis of equal treatment means at the 5% significance level (and, indeed, at every significance level greater than 0.1%).

(b) Checking the Assumptions:

(i) Independent samples: this is apparently a completely randomized design, so this assumption is met.

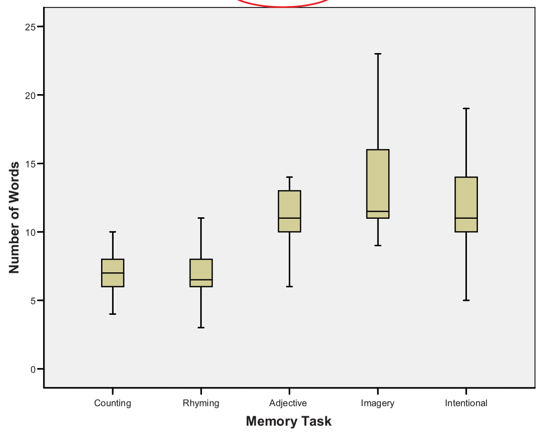
(ii) Normality of the populations: visual inspection of the boxplot below provides no categorical evidence that the normality assumption is violated. This could be tested more formally.

(iii) Equal Variances: Levene's test (below) implies that the equality of variances is not rejected at the 5% level (p=0.054)

SPSS Output

Boxplot indicates that the assumption of normality may be valid, although this is perhaps questionable. A formal test would probably be needed.

Boxplot



Descriptives

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Counting	10	7.00	1.826	.577	5.69	8.31	4	10
Rhyming	10	6.90	2.132	.674	5.38	8.42	3	11
Adjective	10	11.00	2.494	.789	9.22	12.78	6	14
Imagery	10	13.40	4.502	1.424	10.18	16.62	9	23
Intentional	10	12.00	3.742	1.183	9.32	14.68	5	19
Total	50	10.06	4.007	.567	8.92	11.20	3	23

Levene's Test of Homogeneity of Variances

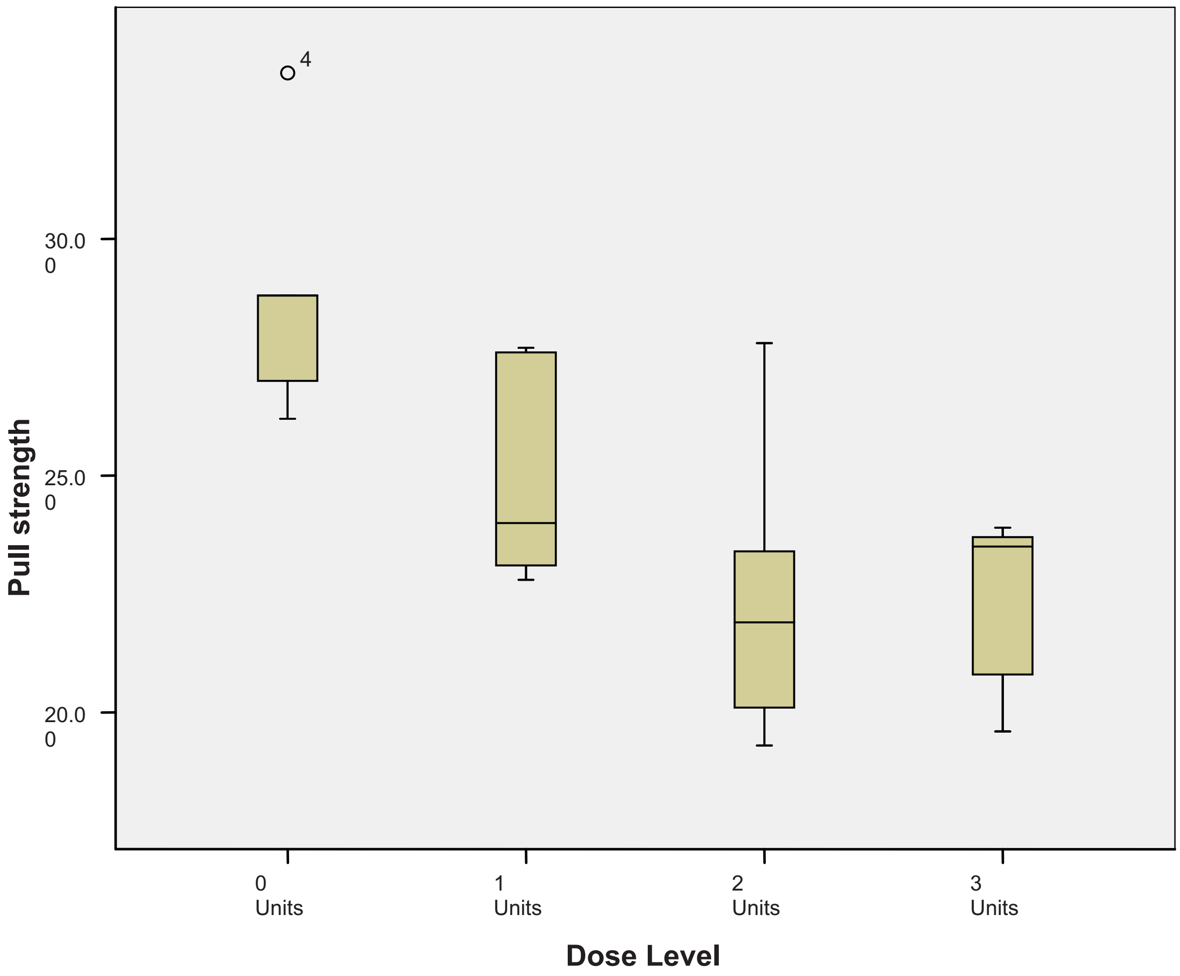
Number of Words

Levene Statistic	df1	df2	Sig.
2.529	4	45	.054

Levene's Test p-value = 0.054.

Therefore no reason to reject the hypothesis of equal variances at the 5% significance level.

Alzheimer’s Study: Dose Level v Pull strength



Levene Statistic	df1	df2	Sig.
.295	3	16	.829

ANOVA F-test suggests the REJECTION of the null hypothesis

H0: No significant difference between means

ANOVA

Pull strength

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	140.094	3	46.698	6.423	.005
Within Groups	116.324	16	7.270		
Total	256.418	19			

Alzheimer’s Study: Dose Level v Pull strength

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Pull strength

	(I) Dose Level	(J) Dose Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	0 Units	1 Units	3.82000	1.70532	.155	-1.0589	8.6989
		2 Units	6.36000(*)	1.70532	.009	1.4811	11.2389
		3 Units	6.56000(*)	1.70532	.007	1.6811	11.4389
	1 Units	0 Units	-3.82000	1.70532	.155	-8.6989	1.0589
		2 Units	2.54000	1.70532	.466	-2.3389	7.4189
		3 Units	2.74000	1.70532	.403	-2.1389	7.6189
	2 Units	0 Units	-6.36000(*)	1.70532	.009	-11.2389	-1.4811
		1 Units	-2.54000	1.70532	.466	-7.4189	2.3389
		3 Units	.20000	1.70532	.999	-4.6789	5.0789
	3 Units	0 Units	-6.56000(*)	1.70532	.007	-11.4389	-1.6811
		1 Units	-2.74000	1.70532	.403	-7.6189	2.1389
		2 Units	-.20000	1.70532	.999	-5.0789	4.6789
Scheffe	0 Units	1 Units	3.82000	1.70532	.213	-1.4957	9.1357
		2 Units	6.36000(*)	1.70532	.016	1.0443	11.6757
		3 Units	6.56000(*)	1.70532	.013	1.2443	11.8757
	1 Units	0 Units	-3.82000	1.70532	.213	-9.1357	1.4957
		2 Units	2.54000	1.70532	.544	-2.7757	7.8557
		3 Units	2.74000	1.70532	.482	-2.5757	8.0557
	2 Units	0 Units	-6.36000(*)	1.70532	.016	-11.6757	-1.0443
		1 Units	-2.54000	1.70532	.544	-7.8557	2.7757
		3 Units	.20000	1.70532	1.000	-5.1157	5.5157
	3 Units	0 Units	-6.56000(*)	1.70532	.013	-11.8757	-1.2443
		1 Units	-2.74000	1.70532	.482	-8.0557	2.5757
		2 Units	-.20000	1.70532	1.000	-5.5157	5.1157
Bonferroni	0 Units	1 Units	3.82000	1.70532	.238	-1.3102	8.9502
		2 Units	6.36000(*)	1.70532	.011	1.2298	11.4902
		3 Units	6.56000(*)	1.70532	.009	1.4298	11.6902
	1 Units	0 Units	-3.82000	1.70532	.238	-8.9502	1.3102
		2 Units	2.54000	1.70532	.935	-2.5902	7.6702
		3 Units	2.74000	1.70532	.766	-2.3902	7.8702
	2 Units	0 Units	-6.36000(*)	1.70532	.011	-11.4902	-1.2298
		1 Units	-2.54000	1.70532	.935	-7.6702	2.5902
		3 Units	.20000	1.70532	1.000	-4.9302	5.3302
	3 Units	0 Units	-6.56000(*)	1.70532	.009	-11.6902	-1.4298
		1 Units	-2.74000	1.70532	.766	-7.8702	2.3902
		2 Units	-.20000	1.70532	1.000	-5.3302	4.9302

* The mean difference is significant at the .05 level.

Starred results indicate significantly different means: in this analysis, we conclude that

-"0 Units" yields a significantly different mean from "2 Units" and "3 Units"

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)

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

- Sum of Squares for Blocks (SSB)

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

- Overall Sum of Squares (SS)

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

The following decomposition holds

$$\mathrm{SS} = \mathrm{SST} + \mathrm{SSB} + \mathrm{SSE} \qquad \therefore \qquad \mathrm{SSE} = \mathrm{SS} - \mathrm{SST} - \mathrm{SSB}$$

For testing

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

in an RBD, the test statistic is

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

where

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

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**

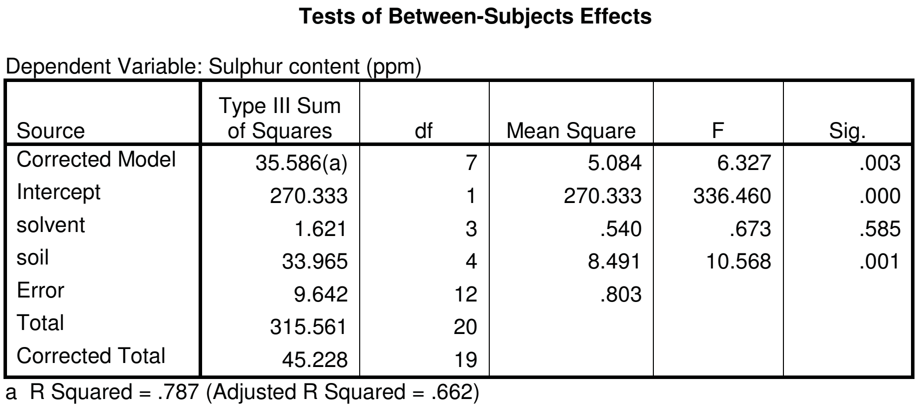
1. Troup, Jackson Co. (*Paleudults* soil)
2. Lakeland, Walton Co. (*Quartzipsamments* soil)
3. Leon, Duval Co. (*Haplaquads* soil)
4. Chipley, Jackson Co. (*Quartzipsamments* soil)
5. Norfolk, Alachua Co. (*Paleudults* soil)

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

Using SPSS, the following ANOVA table was obtained; see the related SPSS screens at

`www.math.mcgill.ca/~dstephens/204/Handouts/Math204-SPSS-RBDANOVA-Screens.pdf`



Tests of Between-Subjects Effects

Dependent Variable: Sulphur content (ppm)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	35.586(a)	7	5.084	6.327	.003
Intercept	270.333	1	270.333	336.460	.000
solvent	1.621	3	.540	.673	.585
soil	33.965	4	8.491	10.568	.001
Error	9.642	12	.803		
Total	315.561	20			
Corrected Total	45.228	19			

a R Squared = .787 (Adjusted R Squared = .662)

This table contains a much information not needed for the ANOVA F-test; the rows headed

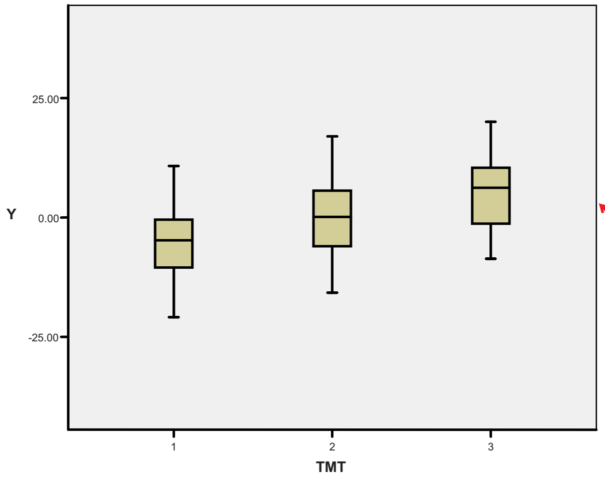
- Corrected Model (row 1)
- Intercept (row 2)
- Total (row 6)

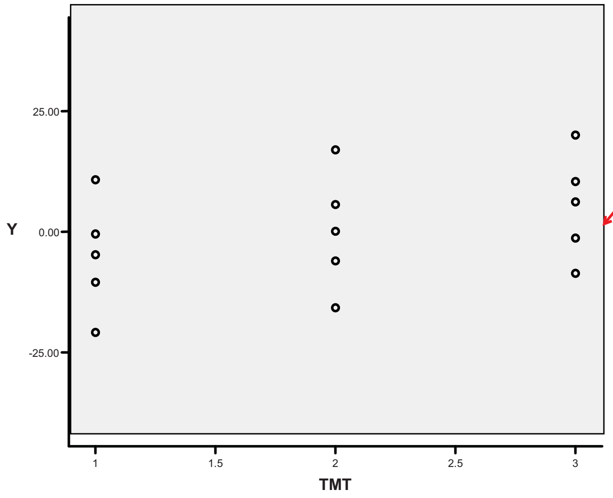
can be ignored. The remaining rows are the standard ANOVA table for the randomized block design. As expected, there is a significant difference between **blocks** (row 4, $F = 10.568$, p-value=0.001), but **no significant difference** between **treatments** (row 3, $F = 0.673$, p-value=0.585).

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**





Boxplot and pointplot display raw data; it appears that the assumptions of normality and equal variances are valid here.

Y

Levene Statistic	df1	df2	Sig.
.012	2	12	.988

Levene's Test indicates that the equal variances assumption is met.

ANOVA for a CRD: Y

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	276.179	2	138.090	1.012	.393
Within Groups	1637.801	12	136.483		
Total	1913.980	14			

Thus the CRD analysis and ANOVA-F test imply that there is **NO DIFFERENCE** between TREATMENTS.

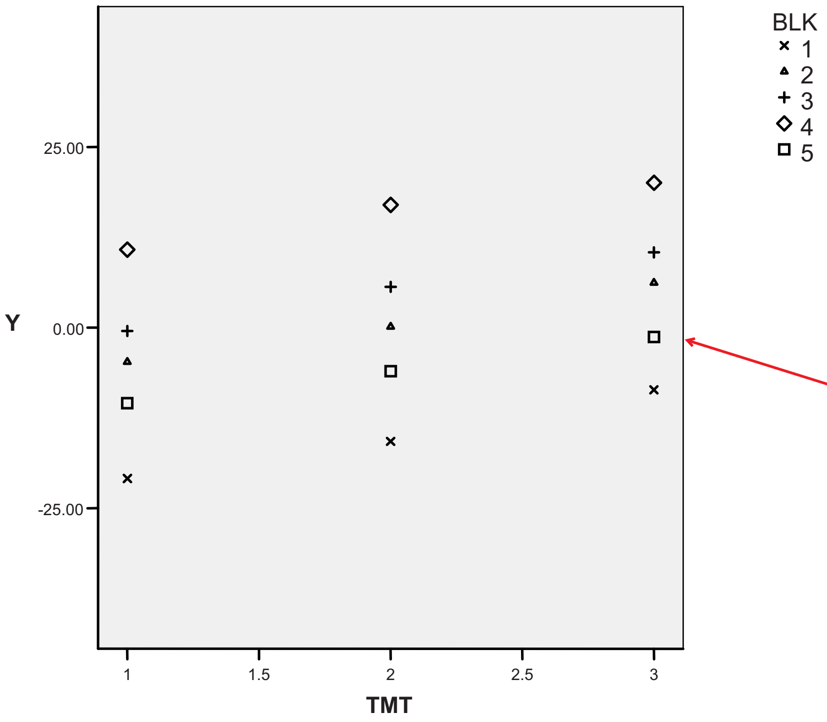
Analysis using RBD with columns taken as blocks:

Dependent Variable: Y

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1907.802	6	317.967	411.720	.000
Intercept	.246	1	.246	.318	.588
TMT	276.179	2	138.090	178.805	.000
BLK	1631.623	4	407.906	528.177	.000
Error	6.178	8	.772		
Total	1914.226	15			
Corrected Total	1913.980	14			

Significant difference between treatments !

Significant difference between blocks defined by columns of data table.



Pointplot reveals hidden structure due to blocking factor; there is systematic variation due to block which is not recognized by the CRD analysis.

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)

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

- Sum of Squares for Blocks (SSB)

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

- Sum of Squares for Interaction (SSI)

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

- Overall Sum of Squares (SS)

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

The following decomposition holds

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

Define

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

and

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

HYPOTHESIS TESTING

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

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

  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

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

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

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

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

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

  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

$$n = 3 \times 3 \times 4 = 36$$

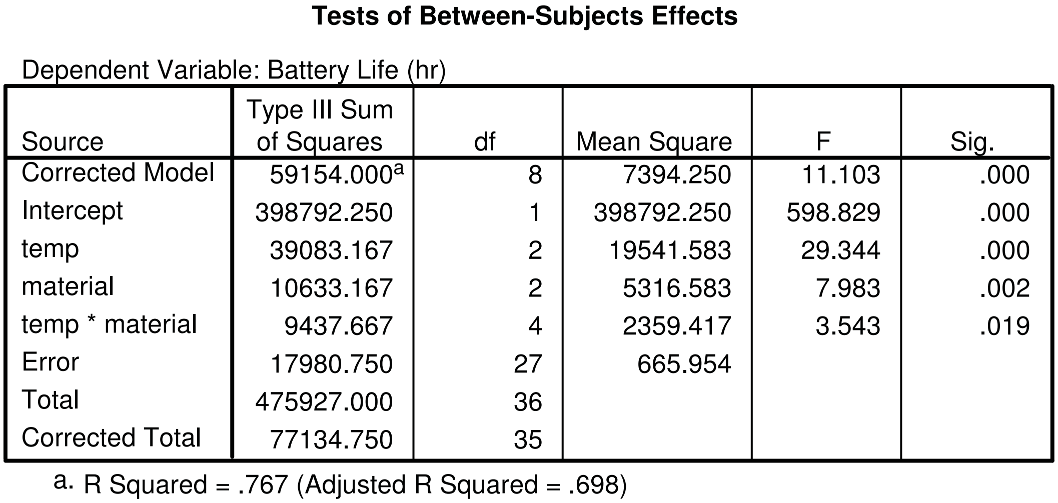
data were obtained in total.

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

Using SPSS, the following ANOVA table was obtained; see the related SPSS screens at

`www.math.mcgill.ca/~dstephens/204/Handouts/Math204-SPSS-RBDANOVAREP-Screens.pdf`



Tests of Between-Subjects Effects

Dependent Variable: Battery Life (hr)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	59154.000[a]	8	7394.250	11.103	.000
Intercept	398792.250	1	398792.250	598.829	.000
temp	39083.167	2	19541.583	29.344	.000
material	10633.167	2	5316.583	7.983	.002
temp * material	9437.667	4	2359.417	3.543	.019
Error	17980.750	27	665.954		
Total	475927.000	36			
Corrected Total	77134.750	35			

a. R Squared = .767 (Adjusted R Squared = .698)

There is a **significant difference** between **blocks** (row 4, material, $F = 7.983$, p-value=0.002), a **significant difference** between **treatments** (row 3, temp, $F = 29.344$, p-value< 0.001), and also a significant interaction (row 5, temp*material, $F = 3.543$, p-value=0.019), Levene’s test reveals that there is no evidence to suspect that the population variances are different:

Levene's Test of Equality of Error Variances[a]

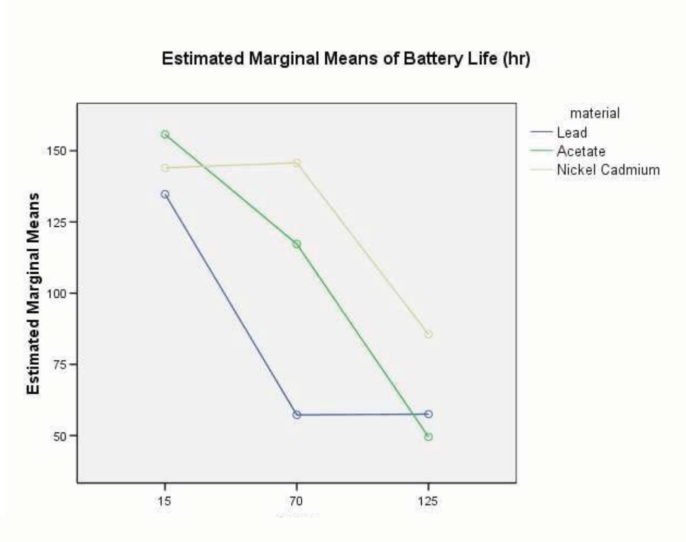
Dependent Variable: Battery Life (hr)

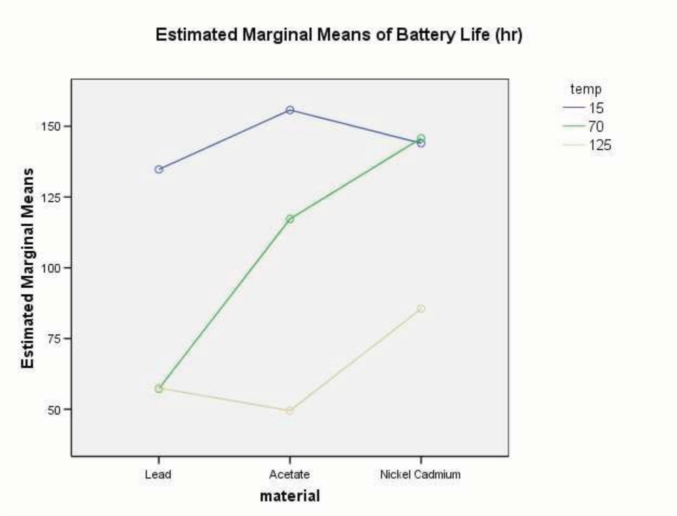
F	df1	df2	Sig.
1.059	8	27	.420

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+temp+material+temp * material

The means plots also indicate some significant interaction.





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)

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

- Sum of Squares for Treatments due to factor B (SST_B)

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

- Sum of Squares for Interaction (SSI_AB)

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

- Overall Sum of Squares (SS)

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

The following decomposition holds

$$\text{SS} = \text{SST}_A + \text{SST}_B + \text{SSI}_{AB} + \text{SSE} \qquad \therefore \qquad \text{SSE} = \text{SS} - \text{SST}_A - \text{SST}_B - \text{SSI}_{AB}$$

Define

$$\text{MST}_A = \frac{\text{SST}_A}{a-1} \qquad \text{MST}_B = \frac{\text{SST}_B}{b-1} \qquad \text{MSI}_{AB} = \frac{\text{SSI}_{AB}}{(a-1)(b-1)}$$

and

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

HYPOTHESIS TESTING

- For testing for a **FACTOR A** effect, use
  $$F = \frac{\text{MST}_A}{\text{MSE}}$$
  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
  $$F = \frac{\text{MST}_B}{\text{MSE}}$$
  Under the assumption of **NO FACTOR B EFFECT**, then
  $$F \sim \text{Fisher-F}(b-1, n-ab)$$

- For testing for an **INTERACTION**, use
  $$F = \frac{\text{MSI}_{AB}}{\text{MSE}}$$
  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.

BALANCED COMPLETE FACTORIAL DESIGNS: EXAMPLES

EXAMPLE 1: 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 years 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. The data are available from the course website as **Butterfat.sav**

Results:

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 refit the model, we confirm the strong effect of Factor B ($F = 49.802, p < 0.000$), and then can estimate the Factor B treatment means. Note how the error degrees of freedom changes when terms in the model are omitted.

EXAMPLE 2: 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. The data are available from the course website as **Lyrics.sav**

Results:

1. **Interaction model**: First note that the Levene test **DOES NOT REJECT** the null hypothesis of equal group variances ($p = 0.804$)

   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, F=0.579, $p = 0.450$), and no significant interaction (song*pool, $F = 1.563$, $p = 0.216$.

2. Fits of the main-effects model (Factor A and Factor B but no interaction), and the Factor A only model confirm the results.

EXAMPLE 3: 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.

The data are available from the course website as **Gravel.sav**

Results:

1. **Interaction model**: First note that the Levene test **DOES NOT REJECT** the null hypothesis of equal group variances ($p = 0.969$).

   There is a **significant difference** due to Factor A (*shift*, $F = 13.667$, $p = 0.001$), and due to Factor B (*quarry*, F=19.996, $p < 0.001$), but **no significant interaction** (*shift*quarry*, $F = 1.099$, $p = 0.364$.

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* $F = 13.552, p = 0.001$, *quarry* $F = 19.829, p < 0.001$. 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.

BUTTERFAT DATA: INTERACTION MODEL

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Butterfat (%)

F	df1	df2	Sig.
2.711	9	90	.008

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

[a] Design: Intercept+age+breed+age * breed

BUTTERFAT DATA : INTERACTION MODEL

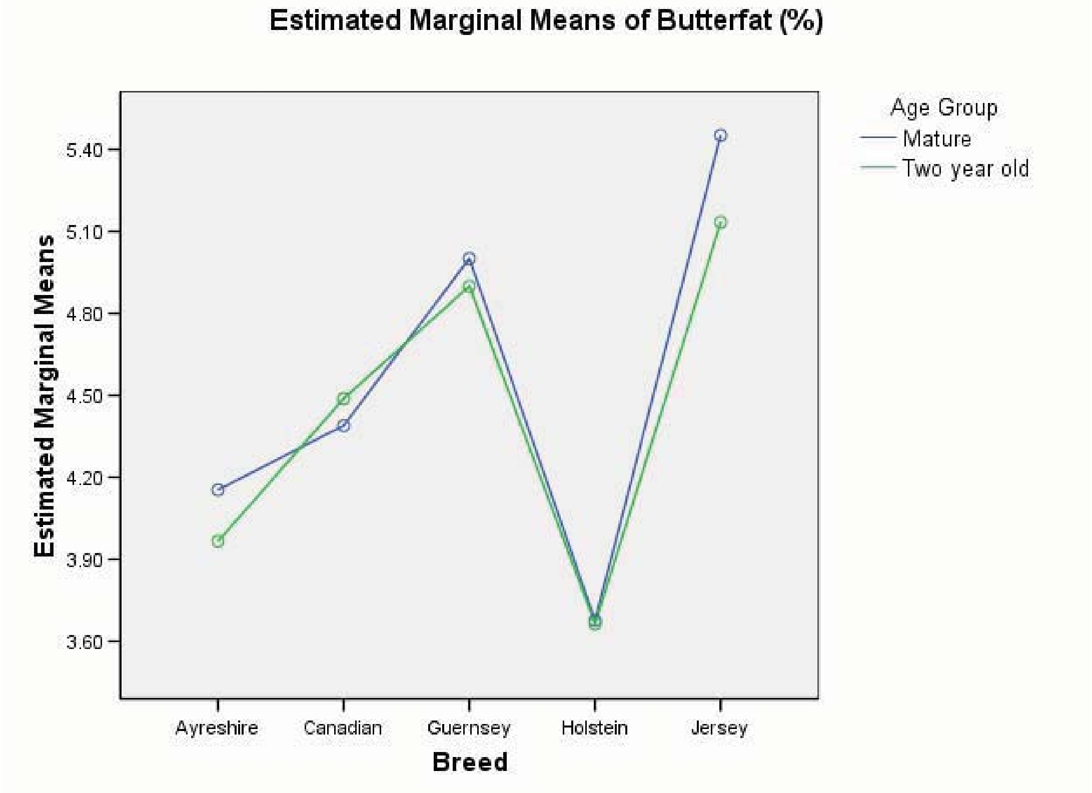
Tests of Between-Subjects Effects

Dependent Variable: Butterfat (%)

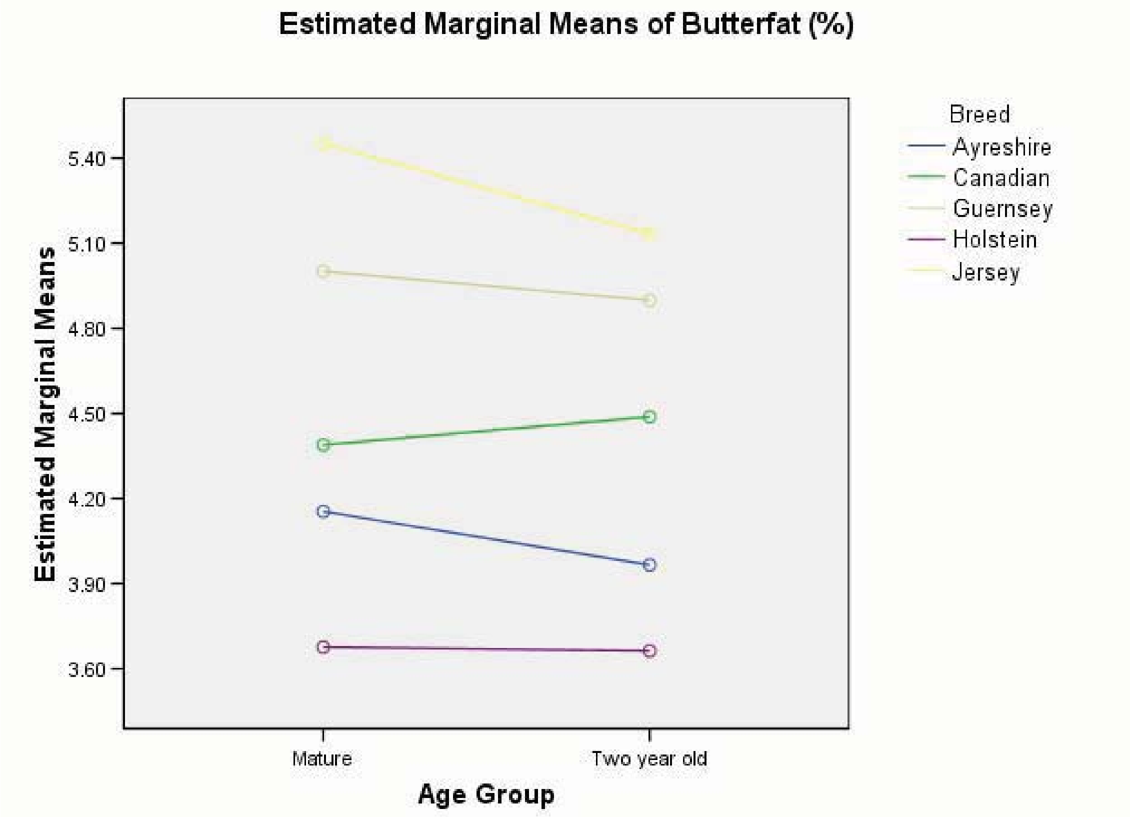
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	35.109[a]	9	3.901	22.534	.000
Intercept	2008.922	1	2008.922	11604.716	.000
age	.274	1	.274	1.580	.212
breed	34.321	4	8.580	49.565	.000
age * breed	.514	4	.128	.742	.566
Error	15.580	90	.173		
Total	2059.611	100			
Corrected Total	50.689	99			

[a] R Squared = .693 (Adjusted R Squared = .662)

BUTTERFAT DATA: INTERACTION MODEL



BUTTERFAT DATA: INTERACTION MODEL



BUTTERFAT DATA: NO INTERACTION, NO AGE MODEL

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Butterfat (%)

F	df1	df2	Sig.
3.766	4	95	.007

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+breed

BUTTERFAT DATA: NO INTERACTION, NO AGE MODEL

Tests of Between-Subjects Effects

Dependent Variable: Butterfat (%)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	34.321[a]	4	8.580	49.802	.000
Intercept	2008.922	1	2008.922	11660.138	.000
breed	34.321	4	8.580	49.802	.000
Error	16.368	95	.172		
Total	2059.611	100			
Corrected Total	50.689	99			

a. R Squared = .677 (Adjusted R Squared = .664)

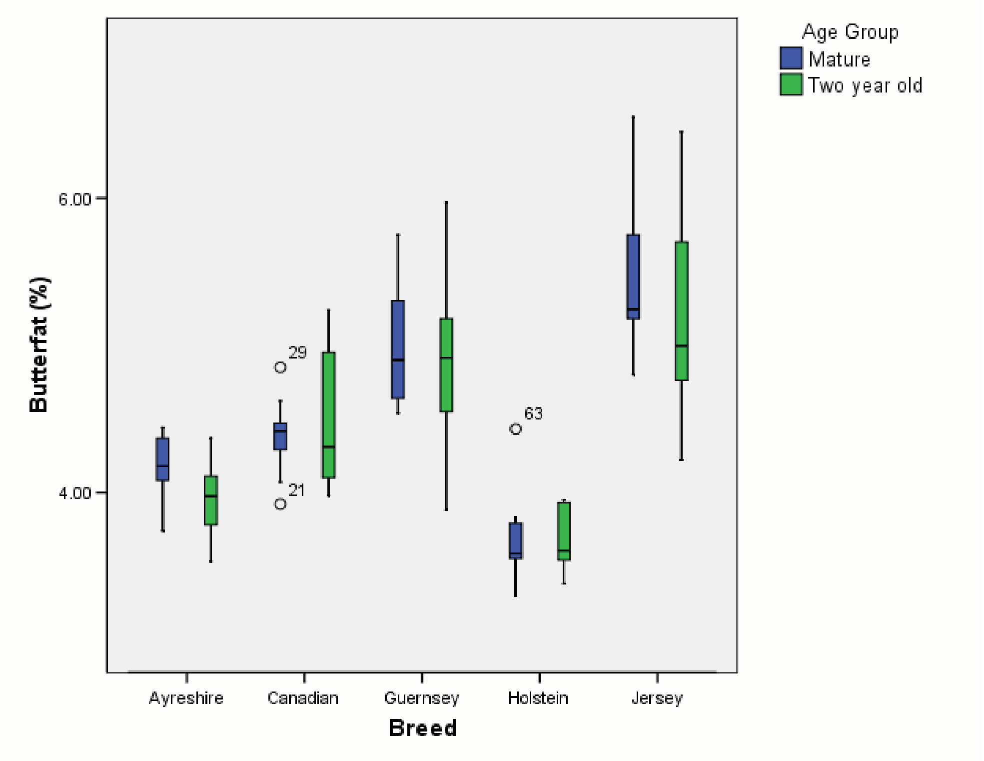
BUTTERFAT DATA: NO INTERACTION, NO AGE MODEL

Breed

Dependent Variable: Butterfat (%)

Breed	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Ayreshire	4.060	.093	3.876	4.244
Canadian	4.439	.093	4.254	4.623
Guernsey	4.950	.093	4.766	5.134
Holstein	3.670	.093	3.485	3.854
Jersey	5.293	.093	5.108	5.477

BUTTERFAT DATA



LYRICS DATA: INTERACTION MODEL

Levene's Test of Equality of Error Variances[a]

Dependent Variable: SCORE

F	df1	df2	Sig.
.329	3	56	.804

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+SONG+POOL+SONG * POOL

LYRICS DATA: INTERACTION MODEL

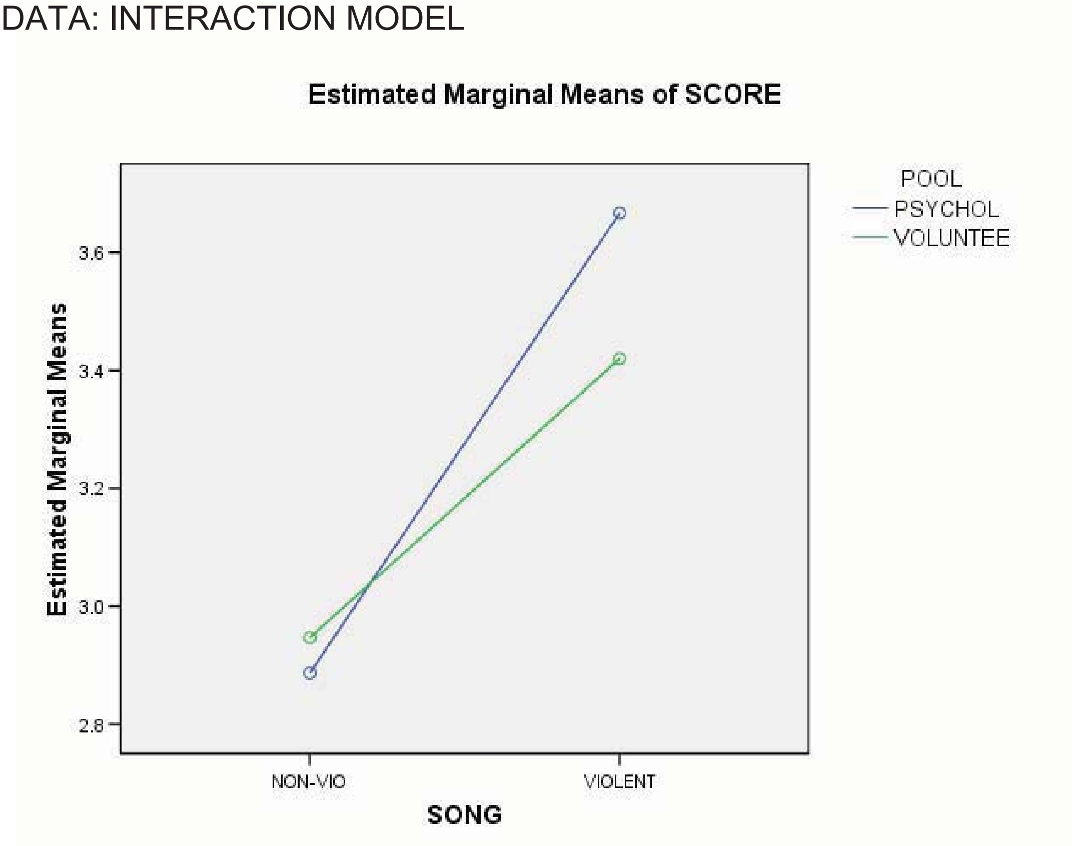
Tests of Between-Subjects Effects

Dependent Variable: SCORE

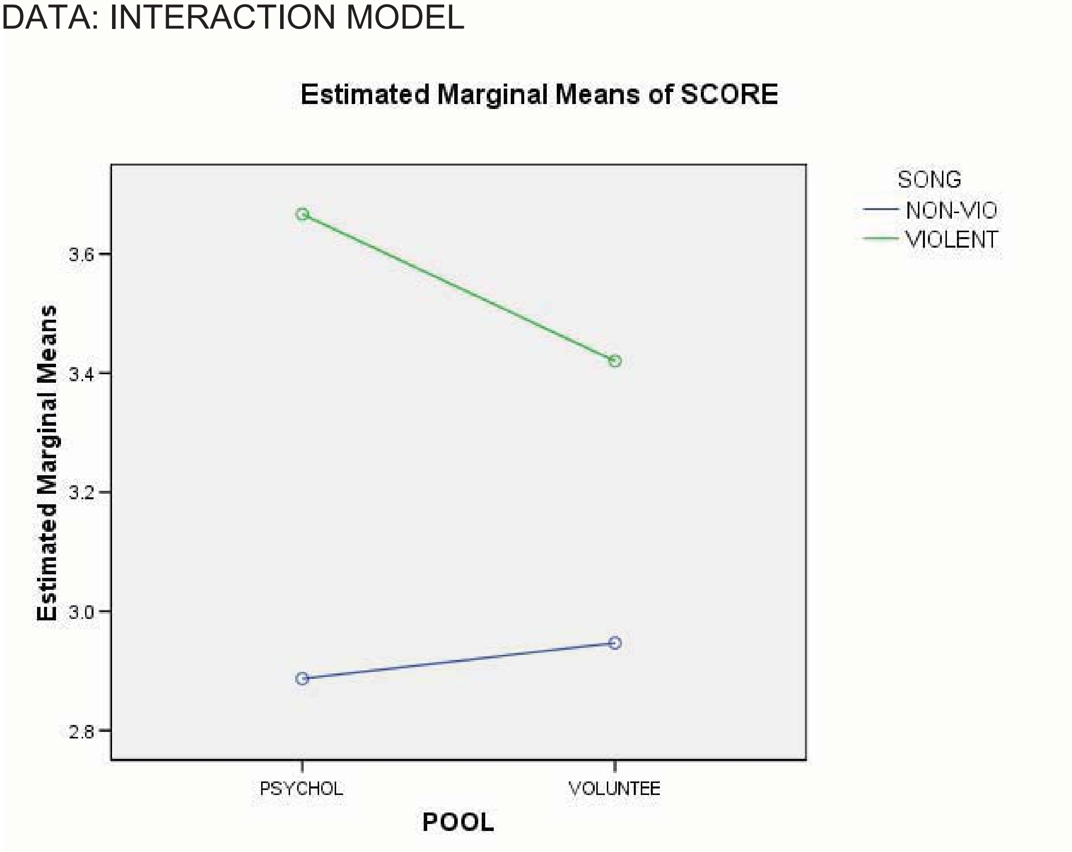
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.374[a]	3	2.125	9.419	.000
Intercept	625.974	1	625.974	2775.059	.000
SONG	5.891	1	5.891	26.114	.000
POOL	.131	1	.131	.579	.450
SONG * POOL	.353	1	.353	1.563	.216
Error	12.632	56	.226		
Total	644.980	60			
Corrected Total	19.006	59			

a. R Squared = .335 (Adjusted R Squared = .300)

LYRICS DATA: INTERACTION MODEL



LYRICS DATA: INTERACTION MODEL



LYRICS DATA: INTERACTION MODEL

Levene's Test of Equality of Error Variances[a]

Dependent Variable: SCORE

F	df1	df2	Sig.
.236	3	56	.871

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

[a] Design: Intercept+SONG+POOL

LYRICS DATA: NO INTERACTION MODEL

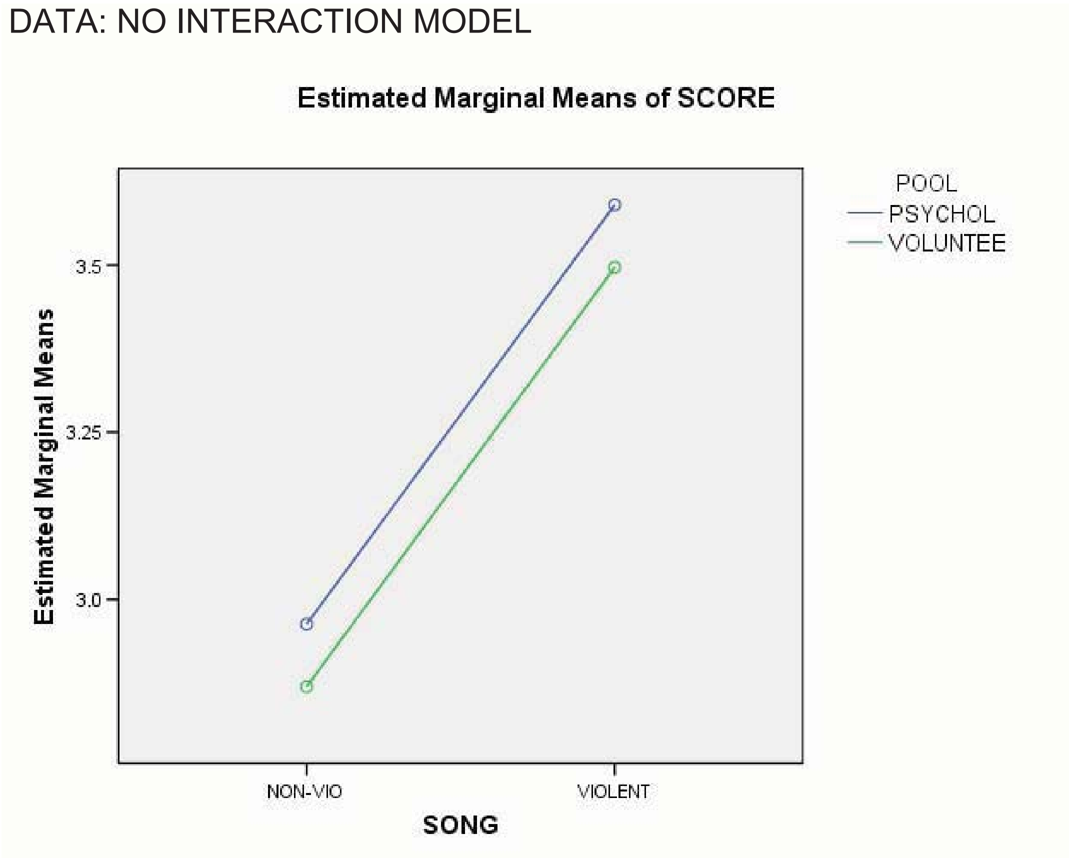
Tests of Between-Subjects Effects

Dependent Variable: SCORE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.021[a]	2	3.011	13.216	.000
Intercept	625.974	1	625.974	2747.896	.000
SONG	5.891	1	5.891	25.859	.000
POOL	.131	1	.131	.574	.452
Error	12.985	57	.228		
Total	644.980	60			
Corrected Total	19.006	59			

[a] R Squared = .317 (Adjusted R Squared = .293)

LYRICS DATA: NO INTERACTION MODEL



LYRICS DATA: NO INTERACTION, NO POOL MODEL

Levene's Test of Equality of Error Variances[a]

Dependent Variable: SCORE

F	df1	df2	Sig.
.017	1	58	.897

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+SONG

LYRICS DATA: NO INTERACTION, NO POOL MODEL

Tests of Between-Subjects Effects

Dependent Variable: SCORE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.891[a]	1	5.891	26.050	.000
Intercept	625.974	1	625.974	2768.248	.000
SONG	5.891	1	5.891	26.050	.000
Error	13.115	58	.226		
Total	644.980	60			
Corrected Total	19.006	59			

[a] R Squared = .310 (Adjusted R Squared = .298)

GRAVEL DATA: INTERACTION MODEL

Levene's Test of Equality of Error Variances[a]

Dependent Variable: amount

F	df1	df2	Sig.
.248	7	32	.969

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

[a] Design: Intercept+shift+quarry+shift * quarry

GRAVEL DATA: INTERACTION MODEL

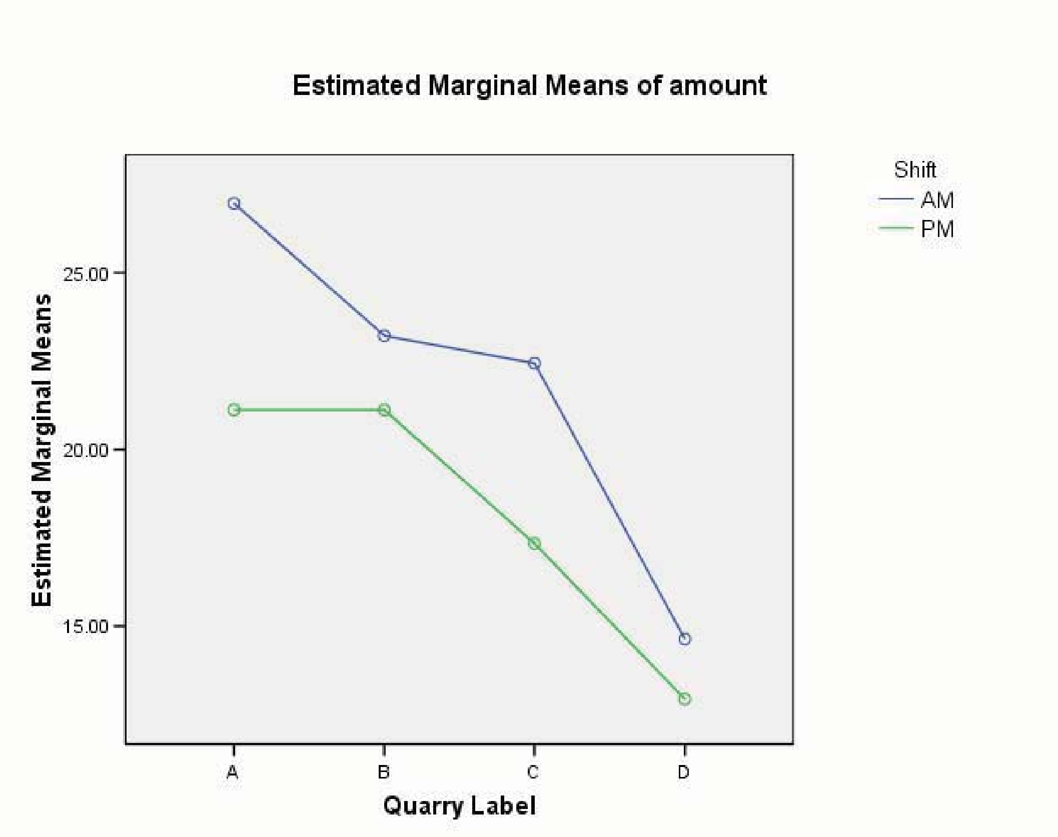
Tests of Between-Subjects Effects

Dependent Variable: amount

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	764.576[a]	7	109.225	10.993	.000
Intercept	15956.030	1	15956.030	1605.921	.000
shift	135.792	1	135.792	13.667	.001
quarry	596.037	3	198.679	19.996	.000
shift * quarry	32.747	3	10.916	1.099	.364
Error	317.944	32	9.936		
Total	17038.550	40			
Corrected Total	1082.520	39			

a. R Squared = .706 (Adjusted R Squared = .642)

GRAVEL DATA: INTERACTION MODEL



GRAVEL DATA: NO INTERACTION MODEL

Levene's Test of Equality of Error Variances[a]

Dependent Variable: amount

F	df1	df2	Sig.
.199	7	32	.983

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

[a] Design: Intercept+shift+quarry

GRAVEL DATA: NO INTERACTION MODEL

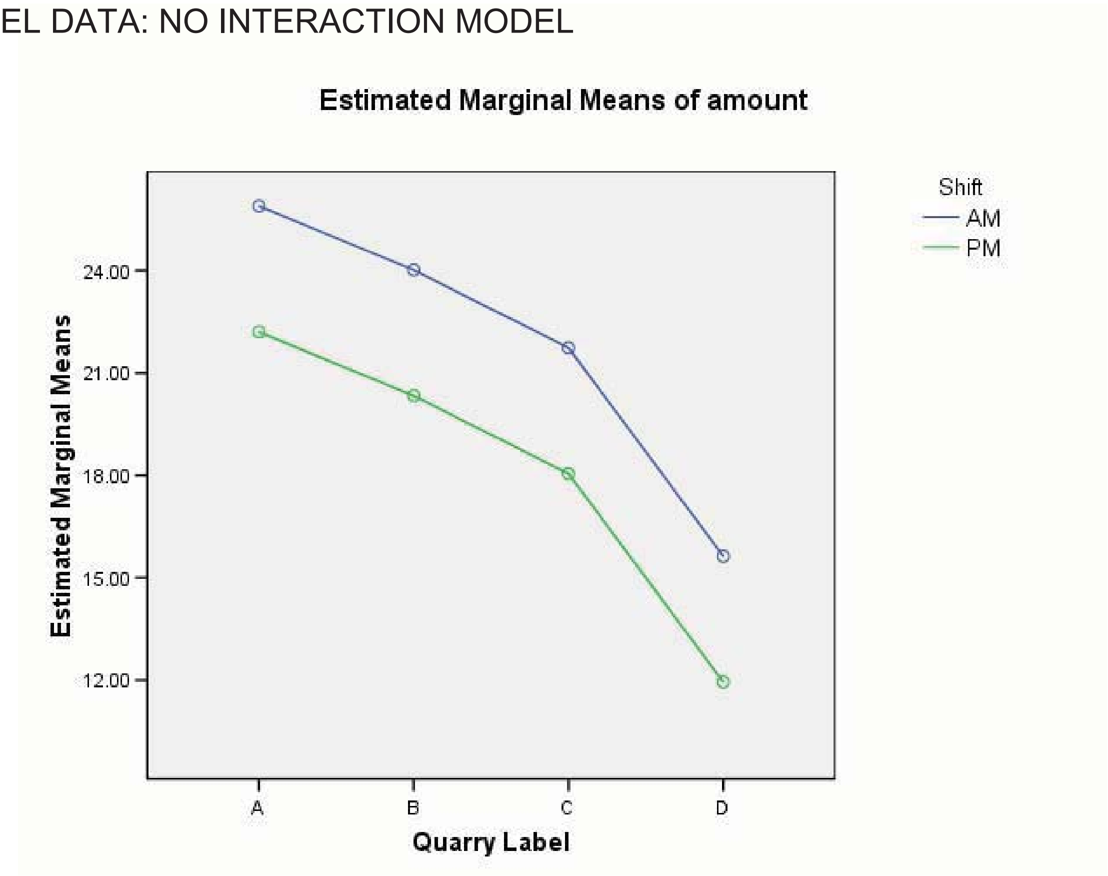
Tests of Between-Subjects Effects

Dependent Variable: amount

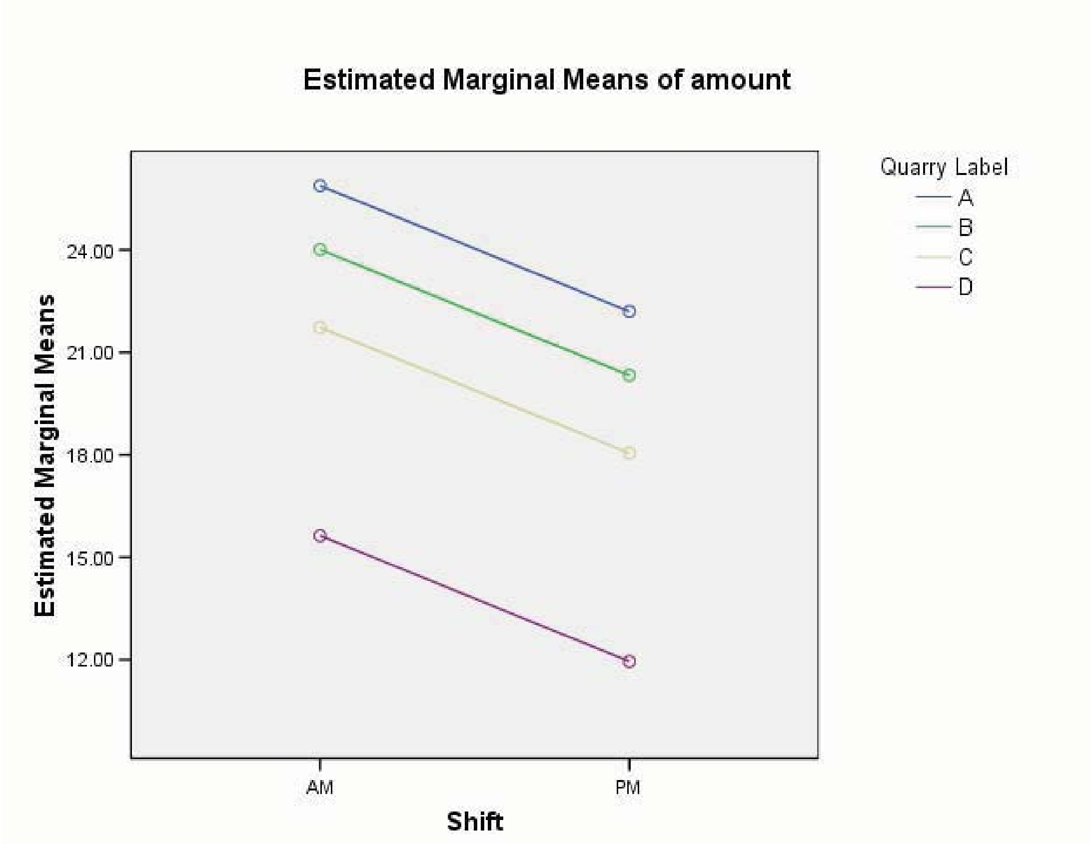
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	731.829[a]	4	182.957	18.260	.000
Intercept	15956.030	1	15956.030	1592.460	.000
shift	135.792	1	135.792	13.552	.001
quarry	596.037	3	198.679	19.829	.000
Error	350.691	35	10.020		
Total	17038.550	40			
Corrected Total	1082.520	39			

[a] R Squared = .676 (Adjusted R Squared = .639)

GRAVEL DATA: NO INTERACTION MODEL



GRAVEL DATA: NO INTERACTION MODEL



Explaining Interaction between Factor Predictors

What the models and the parameters mean

For example: $a = 4, b = 3$.

- Factor A: levels $1, 2, \ldots, a$
- Factor B: levels $1, 2, \ldots, b$

Most complicated model: Main Effects plus Interaction

$$\mathrm{A + B + A.B}$$

that is, we have

- a baseline mean: β_0
- an effect for each level of Factor A: $\beta_i^{(A)}$
- an effect for each level of Factor B: $\beta_j^{(B)}$
- an interaction that modifies the effect of changing levels of Factor *A* at each level of Factor B: $\gamma_{ij}^{(AB)}$

In SPSS, the baseline group is the one where Factor A has level *a* and Factor B has level *b*, but this choice is **arbitrary**; changing this assumption should have no effect on the results we obtain.

Thus we adopt the following modelling strategy:

- Establish a baseline
- Look for changes from baseline introduced by Factor A
- Look for changes from baseline introduced by Factor B
- Look for changes from baseline introduced by Factor A and Factor B **additively**, so that the effect of changing the level of Factor A is **identical** in each level of Factor B, and *vice versa*).
- Look for changes from baseline introduced by Factor A and Factor B **additively with interaction**, so that the effect of changing the level of Factor A is **different** in each level of Factor B, and *vice versa*).

Two-way table: 4 × 3

Factor B

Factor A

	1	2	3
1			
2			
3			
4			

Null Model : Baseline Mean Only

Null Model: cell entries are means for data for each treatment.

Factor B

Factor A

	1	2	3
1	β_0	β_0	β_0
2	β_0	β_0	β_0
3	β_0	β_0	β_0
4	β_0	β_0	β_0

Effect of Factor A only

Main Effect Only: A

Factor B

Factor A

	1	2	3
1	$\beta_0 + \beta_1^{(A)}$	$\beta_0 + \beta_1^{(A)}$	$\beta_0 + \beta_1^{(A)}$
2	$\beta_0 + \beta_2^{(A)}$	$\beta_0 + \beta_2^{(A)}$	$\beta_0 + \beta_2^{(A)}$
3	$\beta_0 + \beta_3^{(A)}$	$\beta_0 + \beta_3^{(A)}$	$\beta_0 + \beta_3^{(A)}$
4	β_0	β_0	β_0

Effect of Factor B only

Main Effect Only: B

Factor B

		1	2	3
Factor A	1	$\beta_0 + \beta_1^{(B)}$	$\beta_0 + \beta_2^{(B)}$	β_0
	2	$\beta_0 + \beta_1^{(B)}$	$\beta_0 + \beta_2^{(B)}$	β_0
	3	$\beta_0 + \beta_1^{(B)}$	$\beta_0 + \beta_2^{(B)}$	β_0
	4	$\beta_0 + \beta_1^{(B)}$	$\beta_0 + \beta_2^{(B)}$	β_0

Effect of Factor A plus Effect of Factor B

Main Effects Only: A + B

Factor B

		1	2	3
Factor A	1	$\beta_0 + \beta_1^{(A)} + \beta_1^{(B)}$	$\beta_0 + \beta_1^{(A)} + \beta_2^{(B)}$	$\beta_0 + \beta_1^{(A)}$
	2	$\beta_0 + \beta_2^{(A)} + \beta_1^{(B)}$	$\beta_0 + \beta_2^{(A)} + \beta_2^{(B)}$	$\beta_0 + \beta_2^{(A)}$
	3	$\beta_0 + \beta_3^{(A)} + \beta_1^{(B)}$	$\beta_0 + \beta_3^{(A)} + \beta_2^{(B)}$	$\beta_0 + \beta_3^{(A)}$
	4	$\beta_0 + \beta_1^{(B)}$	$\beta_0 + \beta_2^{(B)}$	β_0

Main effects plus Interaction between A and B

Main Effects Plus Interaction: A + B + A.B

Factor B

		1	2	3
Factor A	1	$\beta_0 + \beta_1^{(A)} + \beta_1^{(B)} + \gamma_{11}^{(AB)}$	$\beta_0 + \beta_1^{(A)} + \beta_2^{(B)} + \gamma_{12}^{(AB)}$	$\beta_0 + \beta_1^{(A)}$
	2	$\beta_0 + \beta_2^{(A)} + \beta_1^{(B)} + \gamma_{21}^{(AB)}$	$\beta_0 + \beta_2^{(A)} + \beta_2^{(B)} + \gamma_{22}^{(AB)}$	$\beta_0 + \beta_2^{(A)}$
	3	$\beta_0 + \beta_3^{(A)} + \beta_1^{(B)} + \gamma_{31}^{(AB)}$	$\beta_0 + \beta_3^{(A)} + \beta_2^{(B)} + \gamma_{32}^{(AB)}$	$\beta_0 + \beta_3^{(A)}$
	4	$\beta_0 + \beta_1^{(B)}$	$\beta_0 + \beta_2^{(B)}$	β_0

Q. Why are the following models

- A.B
- A + A.B
- B + A.B

not considered ?

A. Because they make specific and perhaps **unrealistic** assumptions about the data, and they imply that the levels of the factors are **not arbitrarily labelled**.

Q. Why are the following models

- A.B
- A + A.B
- B + A.B

not considered ?

A. Because they make specific and perhaps **unrealistic** assumptions about the data, and they imply that the levels of the factors are **not arbitrarily labelled**.

SPSS will not fit such models, although it appears that it does !

Recall the definition of interaction:

- Variation in the effect of changing levels of one factor at the different levels of the other factor.
- For example, the effect on the response mean of moving from level 1 to level 2 for Factor B is **different** at different levels of Factor A.

Consider the model

A.B

this model implies that all parameters apart from the **baseline** and the **interaction** parameters are zero.

Interaction between A and B only

Interaction only: A.B

		Factor B		
		1	2	3
Factor A	1	$\beta_0 + 0 + 0 + \gamma_{11}^{(AB)}$	$\beta_0 + 0 + 0 + \gamma_{12}^{(AB)}$	$\beta_0 + 0$
	2	$\beta_0 + 0 + 0 + \gamma_{21}^{(AB)}$	$\beta_0 + 0 + 0 + \gamma_{22}^{(AB)}$	$\beta_0 + 0$
	3	$\beta_0 + 0 + 0 + \gamma_{31}^{(AB)}$	$\beta_0 + 0 + 0 + \gamma_{32}^{(AB)}$	$\beta_0 + 0$
	4	$\beta_0 \quad + 0$	$\beta_0 \quad + 0$	β_0

In this set-up,

- for Factor A, Level 4: the effect of moving from Level 3 to Level 2 of factor B is **zero**
- for Factor A, Level 3: the effect of moving from Level 3 to Level 2 of factor B is $\gamma_{32}^{(AB)}$.

Therefore, there is a **fundamental difference** between the way that we regard the levels of Factor A.

Main Effect of A plus Interaction between A and B only

Interaction only: A + A.B

		Factor B		
		1	2	3
Factor A	1	$\beta_0 + \beta_1^{(A)} + 0 + \gamma_{11}^{(AB)}$	$\beta_0 + \beta_1^{(A)} + 0 + \gamma_{12}^{(AB)}$	$\beta_0 + \beta_1^{(A)}$
	2	$\beta_0 + \beta_2^{(A)} + 0 + \gamma_{21}^{(AB)}$	$\beta_0 + \beta_2^{(A)} + 0 + \gamma_{22}^{(AB)}$	$\beta_0 + \beta_2^{(A)}$
	3	$\beta_0 + \beta_3^{(A)} + 0 + \gamma_{31}^{(AB)}$	$\beta_0 + \beta_3^{(A)} + 0 + \gamma_{32}^{(AB)}$	$\beta_0 + \beta_3^{(A)}$
	4	$\beta_0 \quad + 0$	$\beta_0 \quad + 0$	β_0

In this set-up,

- for Factor A, Level 4: the effect of moving from Level 3 to Level 2 of factor B is **zero**
- for Factor A, Level 3: the effect of moving from Level 3 to Level 2 of factor B is $\gamma_{32}^{(AB)}$.

Therefore, there is a **fundamental difference** between the way that we regard the levels of Factor A. If we rearrange the labels of the levels of Factor A

we may get a different result.

Therefore, although it is possible **in general** to fit such models, it is no longer possible to talk of the effect of "Factor A".

How does SPSS Handle Such Models ?

It is possible to fit the models

A + A.B B + A.B A.B

in SPSS. For example, for the model A+A.B

- *Analyze* ⟶ *General Linear Model* ⟶ *Univariate*
- Select the *Dependent Variable* and *Fixed Factor(s)*
- Click *Model* to bring up the *Univariate: Model* dialog box.
- Select Factor A as a **Main Effect** using the *Build* pull-down list, click the selection arrow,
- highlight Factor A and Factor B simultaneously, and select **Interaction** from the *Build* pull-down list, and click the selection arrow.
- Click *Continue*, and then *OK*.

How does SPSS Handle Such Models ?

This produces the usual ANOVA table, with terms including

Factor A

and

Factor A * Factor B

However, in fact the model

$$A + B + A.B$$

has been fitted !

- The results are just reported differently
- The terms B and A.B are reported together !

Example: Batteries Data

A - Material
B - Temperature
Model A + B + A.B

Dependent Variable: Battery Life					
Source	Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	59154.000	8	7394.250	11.103	0.000
Intercept	398792.250	1	398792.250	598.829	0.000
material	10633.167	2	5316.583	7.983	0.002
temp	39083.167	2	19541.583	29.344	0.000
material * temp	9437.667	4	2359.417	3.543	0.019
Error	17980.750	27	665.954		
Total	475927.000	36			
Corrected Total	77134.750	35			
R Squared = .767 (Adjusted R Squared = .698)					

$$\mathrm{SS} = \mathrm{SST}_A + \mathrm{SST}_B + \mathrm{SSI}_{AB} + \mathrm{SSE}$$

Example: Batteries Data

A - Material
B - Temperature
Model A + A.B

Dependent Variable: Battery Life					
Source	Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	59154.000	8	7394.250	11.103	0.000
Intercept	398792.250	1	398792.250	598.829	0.000
material	10633.167	2	5316.583	7.983	0.002
material * temp	48520.833	6	8086.806	12.143	0.000
Error	17980.750	27	665.954		
Total	475927.000	36			
Corrected Total	77134.750	35			
R Squared = .767 (Adjusted R Squared = .698)					

$$\mathrm{SS} = \mathrm{SST}_A + \mathrm{SSI}_{B:AB} + \mathrm{SSE}$$

where

$$\mathrm{SSI}_{B:AB} = \mathrm{SST}_B + \mathrm{SSI}_{AB}$$

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**

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

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}:
$$\text{SS}_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2$$
- Sum of Squares SS_{xy}:
$$\text{SS}_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

The **least-squares estimates** are:

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

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

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

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

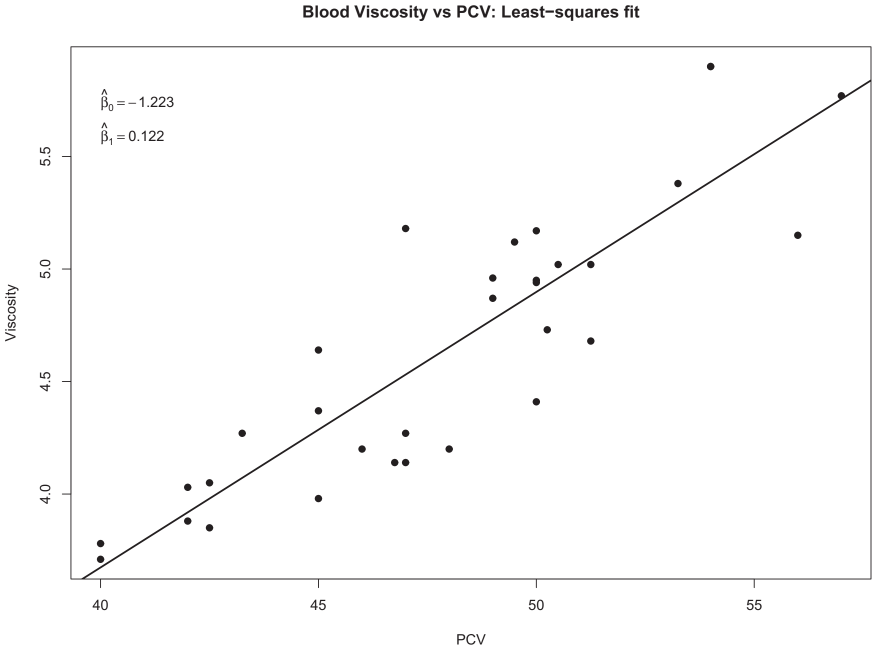
$$\mathrm{SS}_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2 = 615.75 \qquad \mathrm{SS}_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 75.386$$

Thus

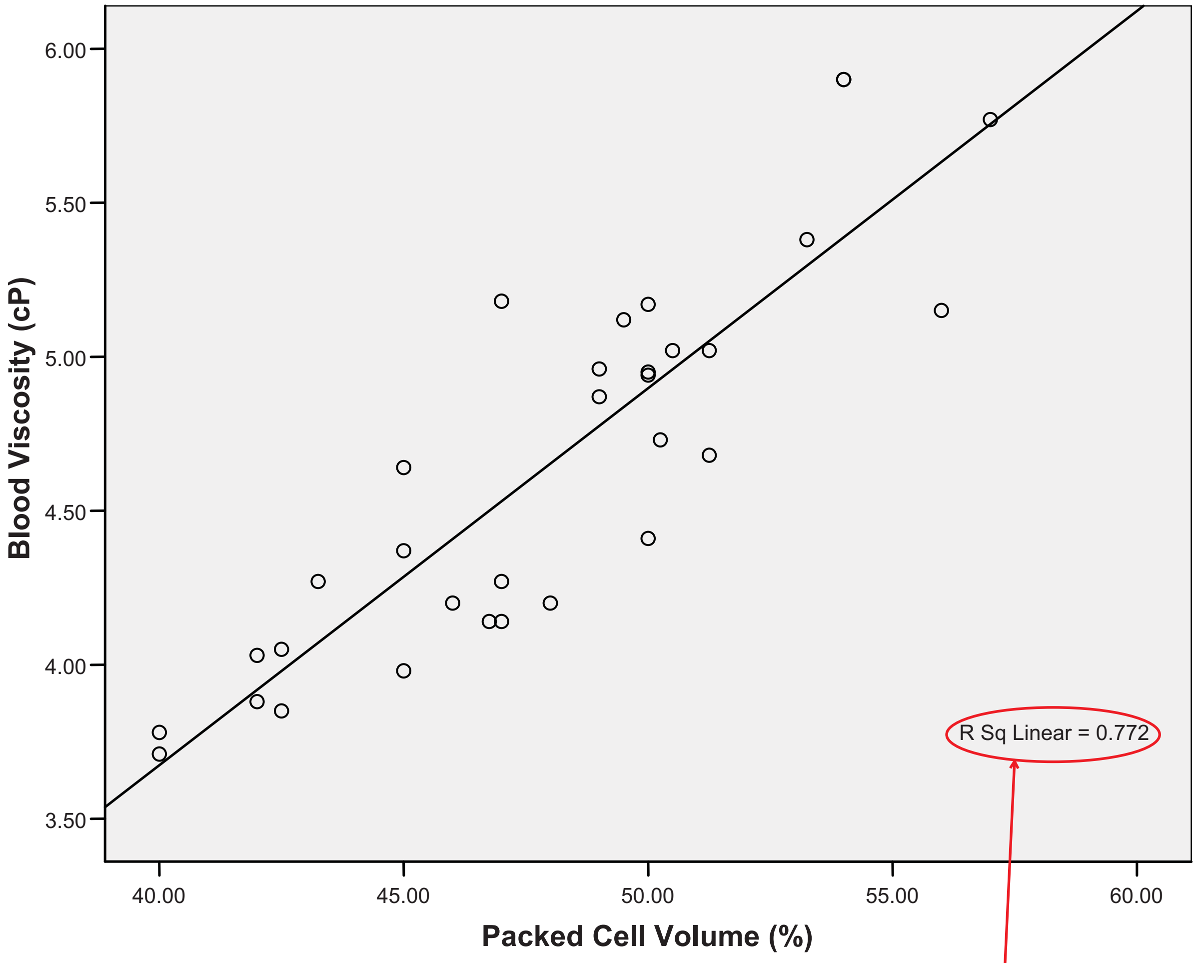
$$\widehat{\beta}_1 = \frac{\mathrm{SS}_{xy}}{\mathrm{SS}_{xx}} = 0.122 \qquad \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = -1.223$$

The estimate of the residual error variance is

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



VISCOSITY vs PCV Regression Analysis



R-squared statistic = 0.772

Descriptive Statistics

	Mean	Std. Deviation	N
Blood Viscosity (cP)	4.6456	.62088	32
Packed Cell Volume (%)	47.9375	4.45678	32

Correlations

		Blood Viscosity (cP)	Packed Cell Volume (%)
Pearson Correlation	Blood Viscosity (cP)	1.000	.879
	Packed Cell Volume (%)	.879	1.000
Sig. (1-tailed)	Blood Viscosity (cP)	.	.000
	Packed Cell Volume (%)	.000	.
N	Blood Viscosity (cP)	32	32
	Packed Cell Volume (%)	32	32

Correlation coefficient r=0.879

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.879	.772	.765	.30116

Summary of Model Fit

ANOVA table testing
H0: No influence of x on y
Ha: x systematically influences y

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9.230	1	9.230	101.764	.000
	Residual	2.721	30	.091		
	Total	11.950	31			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-1.223	.584		-2.094	.045	-2.416	-.030
	PCV (%)	.122	.012	.879	10.088	.000	.098	.147

Parameter estimates and standard errors:
(Constant) corresponds to the estimate of intercept beta0
PCV (%) corresponds to the slope beta1

Test statistics:

t = beta/s.e.(beta)

for estimated beta0 and beta1

P-values in the test of

H0: beta is 0
Ha: beta is not zero

for both beta0 (row 1) and beta1 (row 2).

The ANOVA test is a global test of the regression model; specifically it tests whether the covariate x is an influential variable that is associated with a systematic change in response y.

The F statistic is still of the form

F=MSR/MSE

but now MSR is the Mean Square for Regression. If x not is associated with changing y, then

F ~ Fisher(1,n-2)

which is of precisely the same form as the null distribution in ANOVA - Fisher(k-1,n-k) - where

k = number of parameters estimated = 2

SIMPLE LINEAR REGRESSION: EXAMPLES

EXAMPLE 1: 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.

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

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

SPSS Results:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-50.682	5.193		-9.760	.000	-61.591	-39.772
	status	1.534	.146	.927	10.499	.000	1.227	1.841

a. Dependent Variable: testscore

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.927[a]	.860	.852	3.70509

a. Predictors: (Constant), status

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1513.213	1	1513.213	110.230	.000[a]
	Residual	247.099	18	13.728		
	Total	1760.312	19			

a. Predictors: (Constant), status

b. Dependent Variable: testscore

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

$$t = \frac{r}{\sqrt{(1-r^2)/(n-2)}} = \frac{0.927}{\sqrt{(1-0.927^2)/(20-2)}} = 10.486$$

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**.

EXAMPLE 2: 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.

x	y	x	y	x	y	x	y
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		

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.

SPSS Results:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	146.673	.776		188.911	.000	145.085	148.261
	Pressure	2.253	.038	.996	59.143	.000	2.175	2.330

a. Dependent Variable: Boiling point of Water (C)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.996[a]	.992	.991	.8060

a. Predictors: (Constant), Pressure

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2272.474	1	2272.474	3497.902	.000[a]
	Residual	18.840	29	.650		
	Total	2291.315	30			

a. Predictors: (Constant), Pressure

b. Dependent Variable: Boiling point of Water (C)

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

$$t = \frac{r}{\sqrt{(1-r^2)/(n-2)}} = \frac{0.996}{\sqrt{(1-0.996^2)/(31-2)}} = 60.027$$

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**.

Coleman Data: Regression Analysis

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.927^a	.860	.852	3.70509

a. Predictors: (Constant), status

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1513.213	1	1513.213	110.230	.000^a
	Residual	247.099	18	13.728		
	Total	1760.312	19			

a. Predictors: (Constant), status

b. Dependent Variable: testscore

Coleman Data: Regression Analysis

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-50.682	5.193		-9.760	.000	-61.591	-39.772
	status	1.534	.146	.927	10.499	.000	1.227	1.841

a. Dependent Variable: testscore

Coleman Data: General Linear Model Analysis

Tests of Between-Subjects Effects

Dependent Variable: testscore

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1513.213^a	1	1513.213	110.230	.000
Intercept	1307.621	1	1307.621	95.254	.000
status	1513.213	1	1513.213	110.230	.000
Error	247.099	18	13.728		
Total	1957.567	20			
Corrected Total	1760.312	19			

a. R Squared = .860 (Adjusted R Squared = .852)

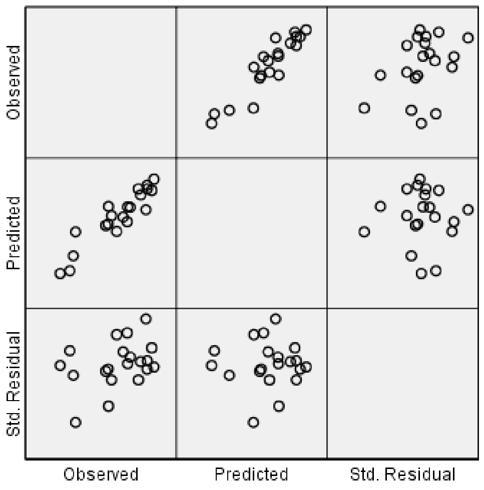
Parameter Estimates

Dependent Variable: testscore

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	-50.682	5.193	-9.760	.000	-61.591	-39.772
status	1.534	.146	10.499	.000	1.227	1.841

Coleman Data: Residuals

Dependent Variable: testscore



Model: Intercept + status

Hooker Data: Regression Analysis

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.996[a]	.992	.991	.8060

a. Predictors: (Constant), Pressure

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2272.474	1	2272.474	3497.902	.000[a]
	Residual	18.840	29	.650		
	Total	2291.315	30			

a. Predictors: (Constant), Pressure

b. Dependent Variable: Boiling point of Water (C)

Hooker Data: Regression Analysis

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	146.673	.776		188.911	.000	145.085	148.261
	Pressure	2.253	.038	.996	59.143	.000	2.175	2.330

a. Dependent Variable: Boiling point of Water (C)

Hooker Data: General Linear Model Analysis

Tests of Between-Subjects Effects

Dependent Variable: Boiling point of Water (C)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2272.474[a]	1	2272.474	3497.902	.000
Intercept	23184.901	1	23184.901	35687.311	.000
pressure	2272.474	1	2272.474	3497.902	.000
Error	18.840	29	.650		
Total	1142542.320	31			
Corrected Total	2291.315	30			

a. R Squared = .992 (Adjusted R Squared = .991)

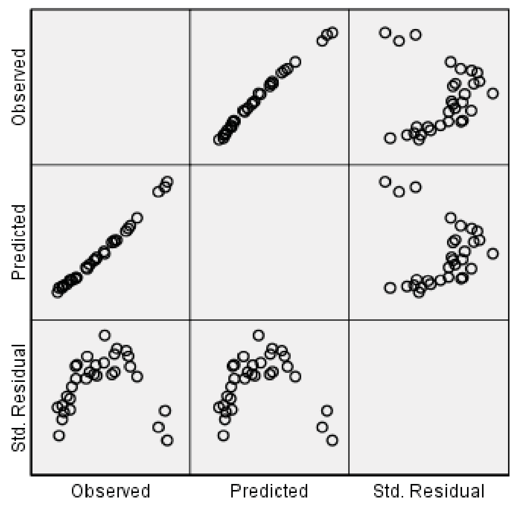
Parameter Estimates

Dependent Variable: Boiling point of Water (C)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	146.673	.776	188.911	.000	145.085	148.261
pressure	2.253	.038	59.143	.000	2.175	2.330

Hooker Data: Residuals

Dependent Variable: Boiling point of Water (C)



Model: Intercept + pressure

Polynomial Regression

Analysis of Hooker data using Quadratic Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.999(a)	.998	.998	.3956

a Predictors: (Constant), Pressure Squared, Pressure

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2286.933	2	1143.467	7306.975	.000(a)
	Residual	4.382	28	.156		
	Total	2291.315	30			

a Predictors: (Constant), Pressure Squared, Pressure
b Dependent Variable: Boiling point of Water (C)

ANOVA Test comparing the two models

H0 : E[Y] = beta0
Ha : E[Y] = beta0 + beta1.x + beta2.x^2

Here the result is highly significant, which implies that the model given by Ha provides a significantly better fit than the model given by H0.

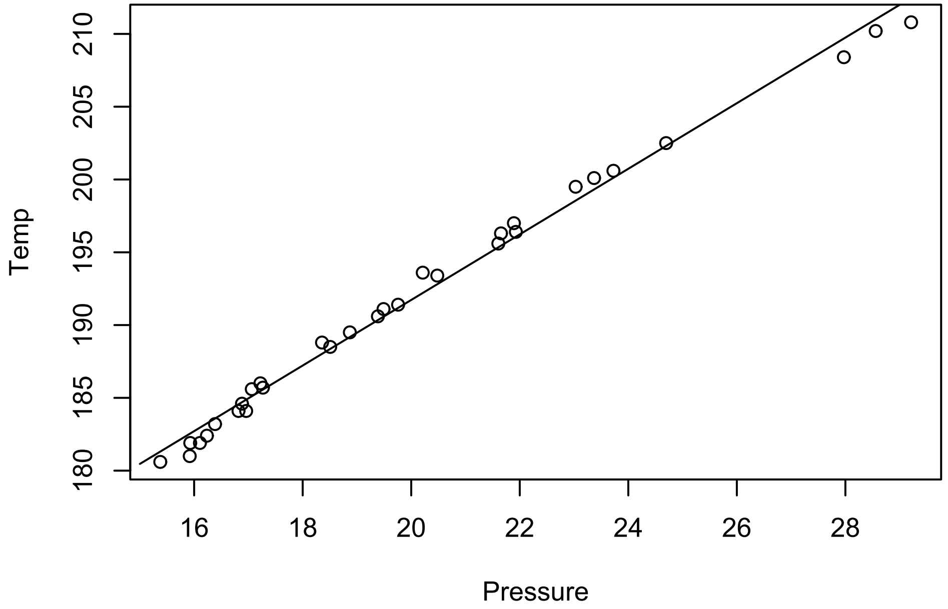
Coefficients(a)

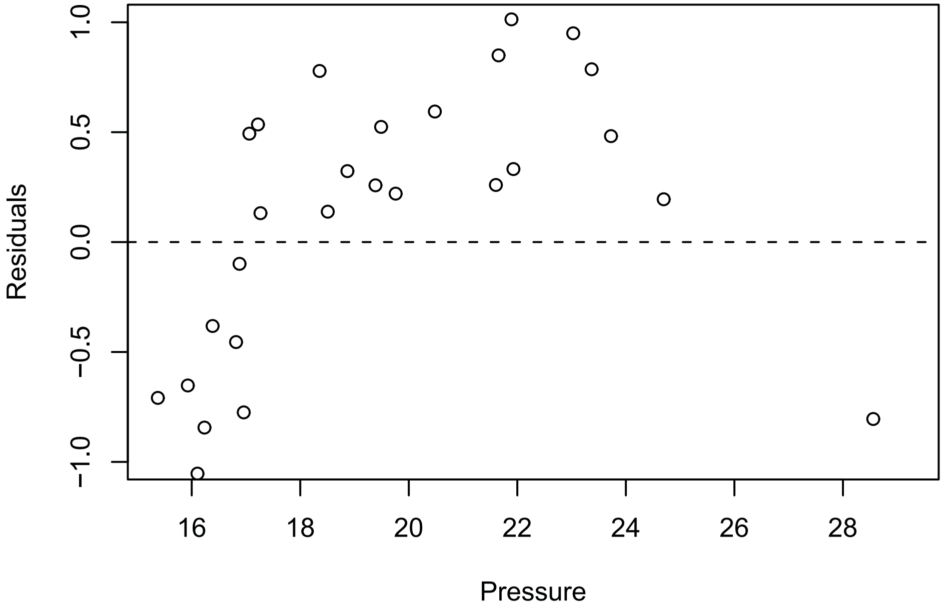
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	126.702	2.112		59.981	.000	122.375	131.029
	Pressure	4.158	.199	1.838	20.885	.000	3.750	4.565
	Pressure Squared	-.044	.005	-.846	-9.612	.000	-.053	-.034

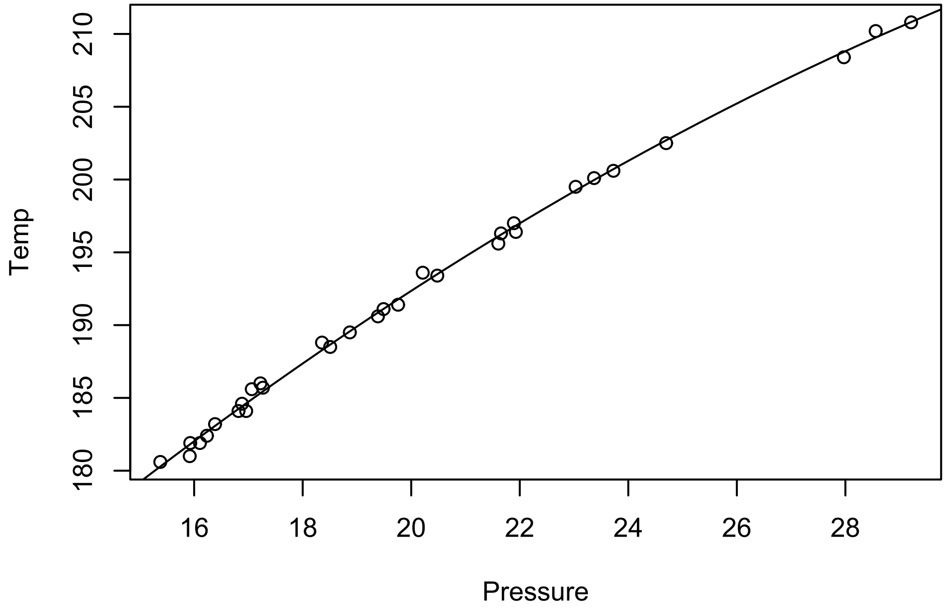
a Dependent Variable: Boiling point of Water (C)

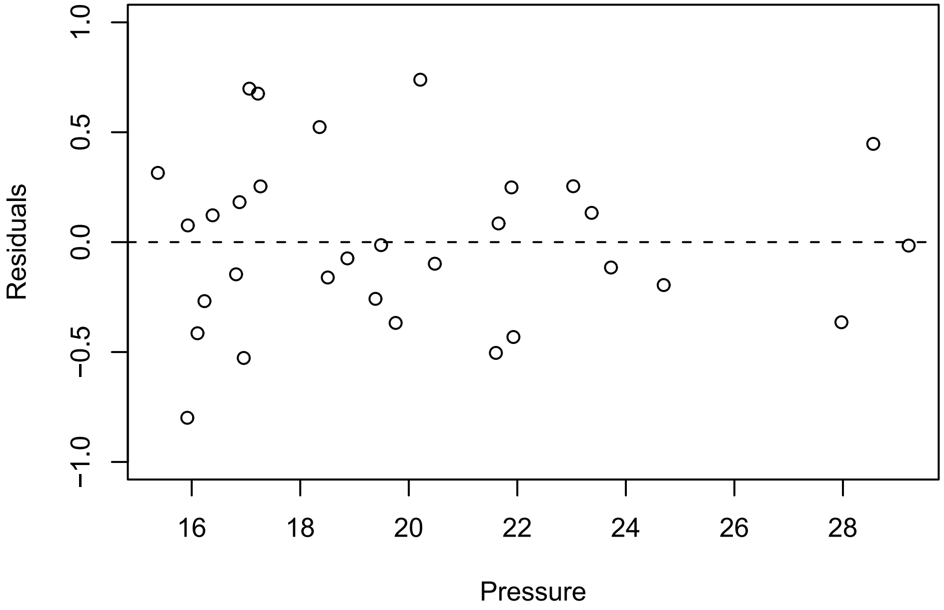
Estimates from Quadratic Regression Model.

The p-values are all < 0.001, so each beta coefficient is significantly different from zero.


Hooker Data: Simple Linear Regression


Hooker Data: Linear Regression Residuals


Hooker Data: Quadratic Regression


Hooker Data: Quadratic Regression Residuals

MATRICES

(MATERIAL NOT EXAMINABLE)

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

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

Some rules for manipulating matrices are given below:

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

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

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

  $$A = \begin{bmatrix} 5 & -4 & 0 & 1 \\ 3 & 5 & -2 & 0 \end{bmatrix} \qquad \Longleftrightarrow \qquad A^\mathsf{T} = \begin{bmatrix} 5 & 3 \\ -4 & 5 \\ 0 & -2 \\ 1 & 0 \end{bmatrix}$$

  A square matrix A is termed **symmetric** if $A = A^\mathsf{T}$.

- **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

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

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

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

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

It can be shown that

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

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

- SSE:

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

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

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

- Variance/Standard Errors of the Parameter estimates:

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

  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 :

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

  say, where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}\boldsymbol{X}^\mathsf{T}$

- 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 $\boldsymbol{x}_p = [1, x_p]^\mathsf{T}$, then the prediction is at the value x_p is

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

  and the prediction error variances are

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

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:

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

Multiple Regression

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.885[a]	.784	.761	.30370

a. Predictors: (Constant), Plasma Protein (g/100ml), Plasma Fibrinogen (mg/100ml), Packed Cell Volume (%)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9.368	3	3.123	33.856	.000[a]
	Residual	2.582	28	.092		
	Total	11.950	31			

a. Predictors: (Constant), Plasma Protein (g/100ml), Plasma Fibrinogen (mg/100ml), Packed Cell Volume (%)

b. Dependent Variable: Blood Viscosity (cP)

Multiple Regression: Parameter Estimates

Tests are of the hypotheses
H0 : beta equal to 0
Ha : beta not equal to zero

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-1.378	.897		-1.537	.136	-3.215	.458
	Packed Cell Volume (%)	.117	.014	.839	8.584	.000	.089	.145
	Plasma Fibrinogen (mg/100ml)	.000	.000	.111	1.147	.261	.000	.001
	Plasma Protein (g/100ml)	.040	.097	.037	.412	.683	-.159	.239

a. Dependent Variable: Blood Viscosity (cP)

Only the packed cell volume coefficient is significantly different from zero ($p < 0.001$)

The other covariates do not seem to be significantly different from zero.

Regression on Packed Cell Volume only

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.879[a]	.772	.765	.30116

a. Predictors: (Constant), Packed Cell Volume (%)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9.230	1	9.230	101.764	.000[a]
	Residual	2.721	30	.091		
	Total	11.950	31			

a. Predictors: (Constant), Packed Cell Volume (%)

b. Dependent Variable: Blood Viscosity (cP)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	-1.223	.584		-2.094	.045	-2.416	-.030
	Packed Cell Volume (%)	.122	.012	.879	10.088	.000	.098	.147

a. Dependent Variable: Blood Viscosity (cP)

PCV is a significant term in the model ($p < 0.001$)

Regression on Plasma Protein only

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.457[a]	.209	.183	.56129

a. Predictors: (Constant), Plasma Fibrinogen (mg/100ml)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.499	1	2.499	7.932	.009[a]
	Residual	9.451	30	.315		
	Total	11.950	31			

a. Predictors: (Constant), Plasma Fibrinogen (mg/100ml)

b. Dependent Variable: Blood Viscosity (cP)

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	3.871	.292		13.236	.000	3.274	4.468
	Plasma Fibrinogen (mg/100ml)	.002	.001	.457	2.816	.009	.000	.003

a. Dependent Variable: Blood Viscosity (cP)

Plasfib is a significant term in the model (p = 0.009)

Regression on Plasma Fibrinogen only

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.101[a]	.010	-.023	.62791

a. Predictors: (Constant), Plasma Protein (g/100ml)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.122	1	.122	.310	.582[a]
	Residual	11.828	30	.394		
	Total	11.950	31			

a. Predictors: (Constant), Plasma Protein (g/100ml)

b. Dependent Variable: Blood Viscosity (cP)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	5.296	1.174		4.510	.000	2.898	7.694
	Plasma Protein (g/100ml)	-.110	.198	-.101	-.556	.582	-.515	.295

a. Dependent Variable: Blood Viscosity (cP)

Plaspro is not a significant term in the model (p =0.582)

FACTOR PREDICTOR REGRESSION USING DUMMY VARIABLES

We can fit a factor predictor using the *Linear Regression* pulldown in SPSS by 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.SAV** 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	group	Dummy 1	Dummy 2	Dummy 3
	y	x	x_1	x_2	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

The analysis below indicates that the estimated coefficients and the ANOVA results are identical whether we use the *General Linear Model* or *Regression* pulldown menus.

Factor Predictor Fitted Using *General Linear Model*

Between-Subjects Factors

		Value Label	N
Clinical Group	1	Overt Diabetic	32
	2	Chemically Diabetic	36
	3	Normal	76

Tests of Between-Subjects Effects

Dependent Variable: Log(GluTest)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
~~Corrected Model~~	~~24.344(a)~~	~~2~~	~~12.172~~	~~412.568~~	~~.000~~
~~Intercept~~	~~4969.483~~	~~1~~	~~4969.483~~	~~168437.466~~	~~.000~~
group	24.344	2	12.172	412.568	.000
Error	4.160	141	.030		
~~Total~~	~~5509.040~~	~~144~~			
Corrected Total	28.504	143			

a R Squared = .854 (Adjusted R Squared = .852)

Parameter Estimates

Dependent Variable: Log(GluTest)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	5.852	.020	297.026	.000	5.813	5.891
[group=1]	1.039	.036	28.704	.000	.967	1.111
[group=2]	.344	.035	9.905	.000	.276	.413
[group=3]	0(a)	.	.	.	.	.

a This parameter is set to zero because it is redundant.

ANOVA results identical

R squared identical

Factor Predictor Fitted Using *Linear Regression*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.924(a)	.854	.852	.17177

a Predictors: (Constant), Group = 2, Group = 1

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	24.344	2	12.172	412.568	.000(a)
	Residual	4.160	141	.030		
	Total	28.504	143			

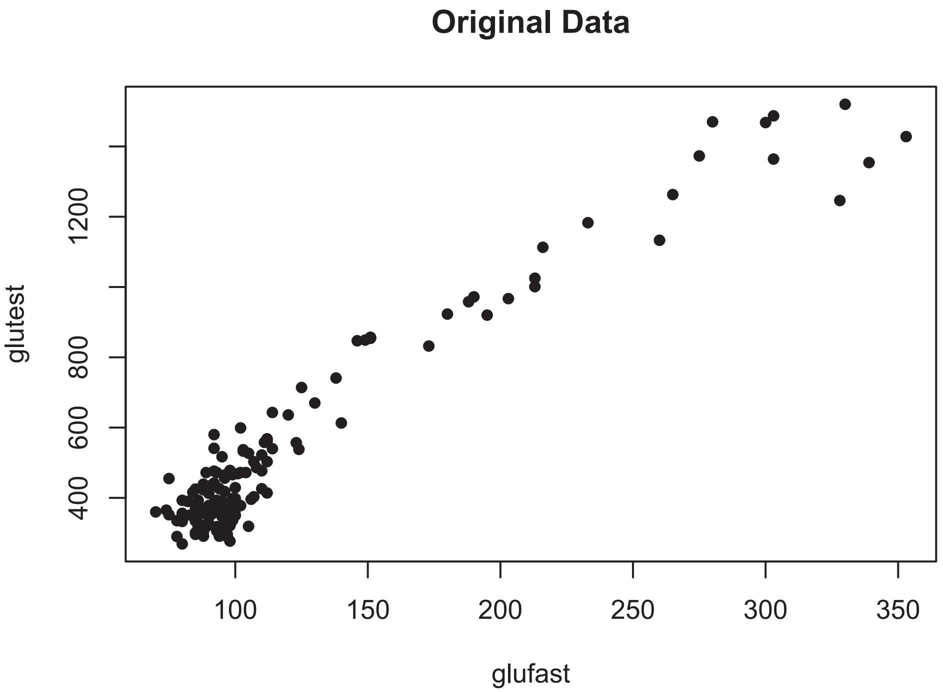
a Predictors: (Constant), Group = 2, Group = 1
b Dependent Variable: Log(GluTest)

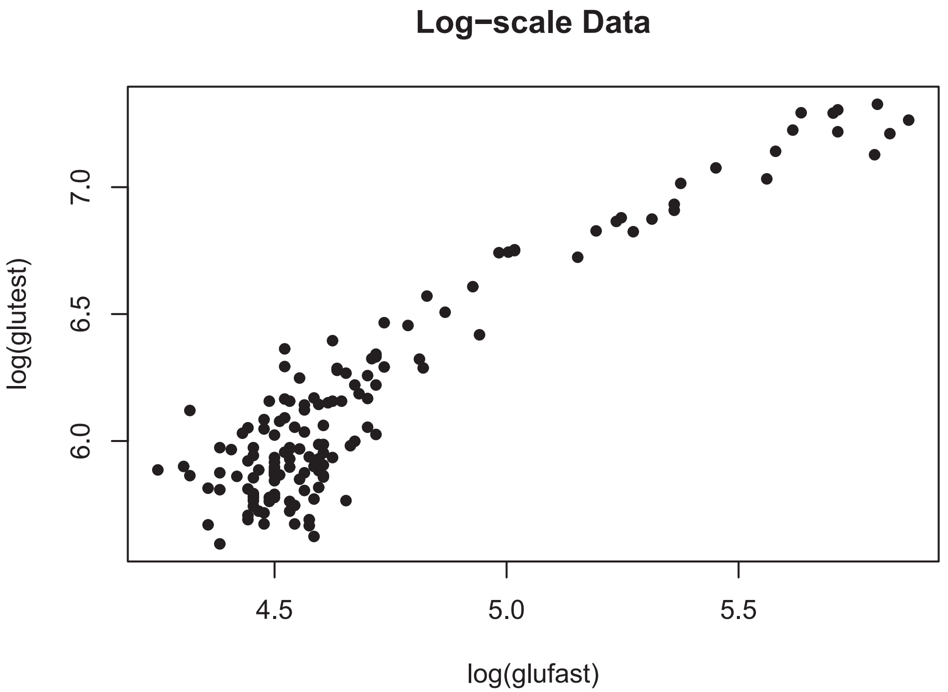
Coefficients(a)

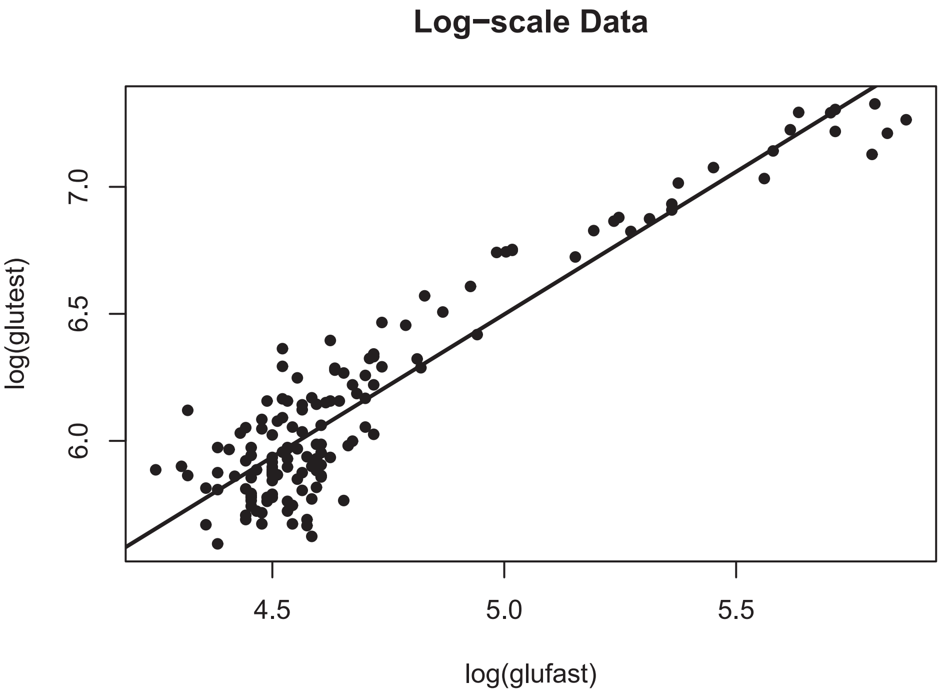
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	5.852	.020		297.026	.000	5.813	5.891
	Group = 1	1.039	.036	.971	28.704	.000	.967	1.111
	Group = 2	.344	.035	.335	9.905	.000	.276	.413

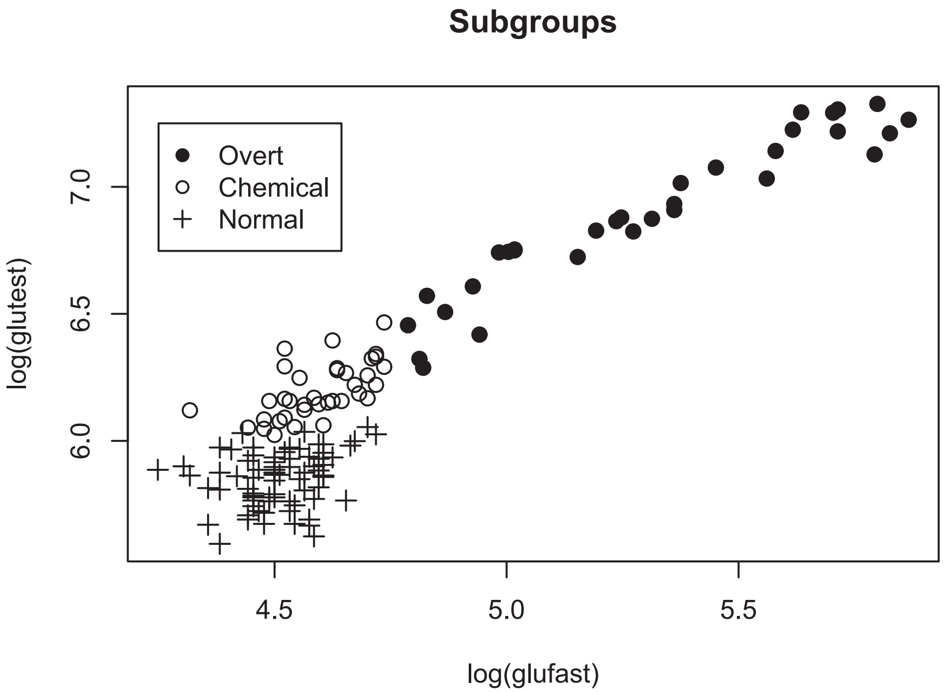
a Dependent Variable: Log(GluTest)

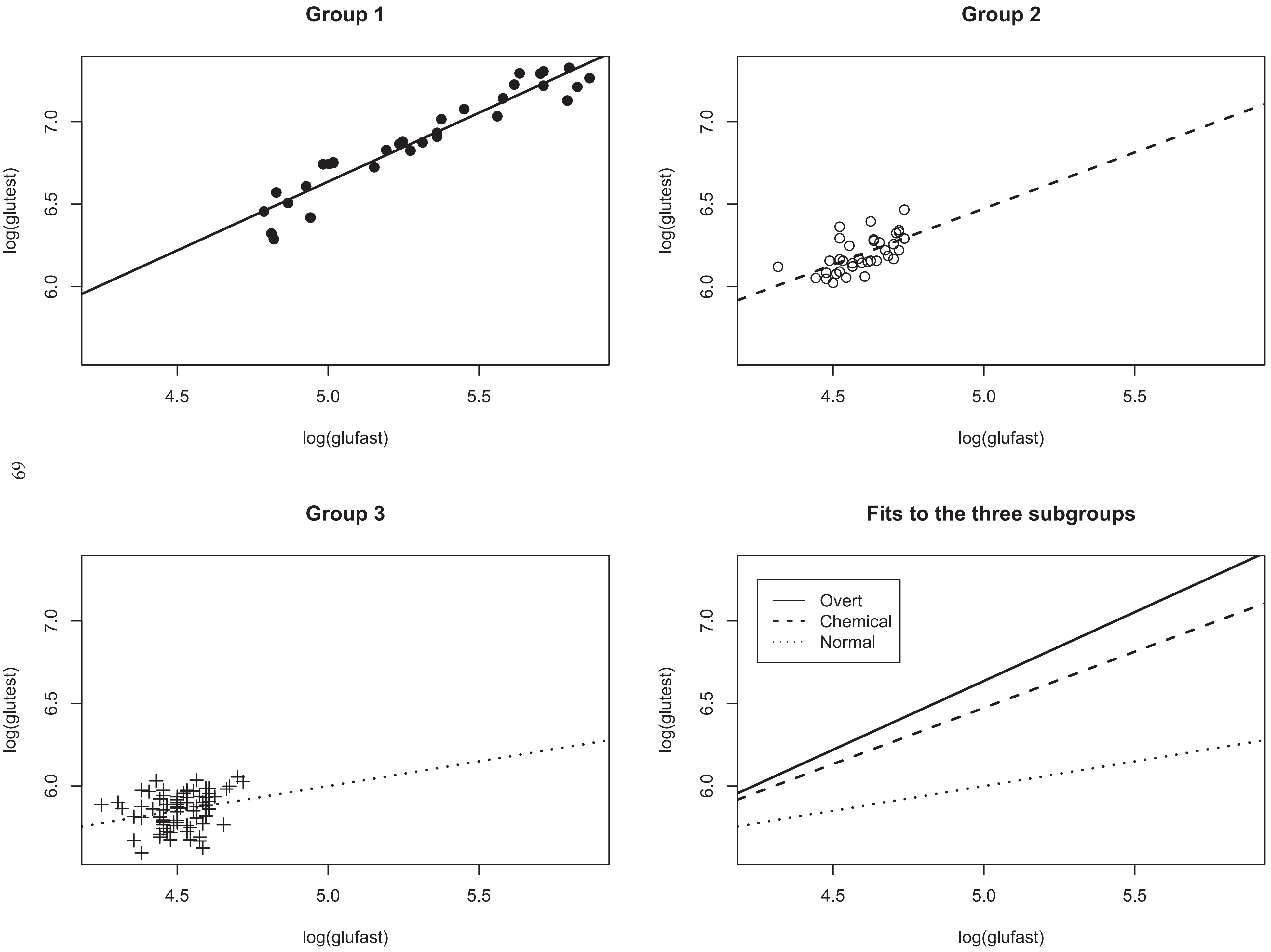
Estimates of coefficients, standard errors etc. are identical


Original Data


Log−scale Data


Log−scale Data


Subgroups


Group 1
Group 2
Group 3
Fits to the three subgroups
log(glutest)
log(glufast)
6.0
6.5
7.0
4.5
5.0
5.5
Overt
Chemical
Normal

HOOKER'S DATA: SPSS COMPARISON OF LINEAR AND QUADRATIC MODELS

Regression with Linear Term

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.996(a)	.992	.991	.8060

a Predictors: (Constant), Pressure

REDUCED MODEL FIT

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2272.474	1	2272.474	3497.902	.000(a)
	Residual	18.840	29	.650		
	Total	2291.315	30			

(a) Predictors: (Constant), Pressure (b) Dependent Variable: Boiling point of Water (C)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	146.673	.776		188.911	.000	145.085	148.261
	Pressure	2.253	.038	.996	59.143	.000	2.175	2.330

a Dependent Variable: Boiling point of Water (C)

Regression with Linear and Quadratic Terms

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.999(a)	.998	.998	.3956

a Predictors: (Constant), Pressure Squared, Pressure

COMPLETE MODEL FIT

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2286.933	2	1143.467	7306.975	.000(a)
	Residual	4.382	28	.156		
	Total	2291.315	30			

(a) Predictors: (Constant), Pressure Squared, Pressure (b) Dependent Variable: Boiling point of Water (C)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	126.702	2.112		59.981	.000	122.375	131.029
	Pressure	4.158	.199	1.838	20.885	.000	3.750	4.565
	Pressure Squared	-.044	.005	-.846	-9.612	.000	-.053	-.034

a Dependent Variable: Boiling point of Water (C)

DIABETES DATA: STEPWISE MODEL COMPARISON

MODEL 4

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
group	.104	2	.052	5.447	.005
loggluf	.675	1	.675	70.702	.000
group * loggluf	.155	2	.077	8.099	.000
Error	1.318	138	.010		
Corrected Total	28.504	143			

a R Squared = .954 (Adjusted R Squared = .952)

MODEL 3

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
group	2.266	2	1.133	107.717	.000
loggluf	2.688	1	2.688	255.565	.000
Error	1.472	140	.011		
Corrected Total	28.504	143			

a R Squared = .948 (Adjusted R Squared = .947)

MODEL 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
group	24.344	2	12.172	412.568	.000
Error	4.160	141	.030		
Corrected Total	28.504	143			

a R Squared = .854 (Adjusted R Squared = .852)

MODEL 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
loggluf	24.766	1	24.766	940.846	.000
Error	3.738	142	.026		
Corrected Total	28.504	143			

a R Squared = .869 (Adjusted R Squared = .868)

Diabetes Data: ANOVA Table for MODEL 4: main effects plus interaction

loggluf + group + loggluf.group

NOTE: C = A + B is the decomposition SS = SSR + SSE.

k = Total number of parameters - 1
= 6 - 1 = 5

Dependent Variable: Log(GluTest)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	A 27.187[a]	5	5.437	569.463	.000
Intercept	.973	1	.973	101.906	.000
group	.104	2	.052	5.447	.005
loggluf	.675	1	.675	70.702	.000
group * loggluf	.155	2	.077	8.099	.000
Error	B 1.318	138	.010		
Total	5509.040	144			
Corrected Total	C 28.504	143			

a. R Squared = .954 (Adjusted R Squared = .952)

Number of additional parameters needed to introduce each of the main effects and the interaction

R squared/Adjusted R squared terms give an indication of whether the regression terms contribute significantly to the model. As a rule of thumb:

R squared > 0.7 implies a good fit
R squared > 0.4 implies that although the fit might not be good, there is some explanatory power in the predictors.

General Linear Model of an Unbalanced Factorial Design

Potato Damage Data

This is an unbalanced design as we have different numbers of replicated in the 2 x 2 x 2 = 8 cells of the table.

Temperature Pre-treatment * Potato variety * Acclimatization Routine Crosstabulation

Count

Acclimatization Routine			Potato variety		Total
			Variety 1	Variety 2	
Room Temp	Temperature Pre-treatment	-4 C	5	13	18
		-8 C	5	13	18
	Total		10	26	36
Cold Room	Temperature Pre-treatment	-4 C	12	7	19
		-8 C	13	7	20
	Total		25	14	39

Number of parameters:
k=7

Three-way Interaction Model (COMPLETE MODEL)

Dependent Variable: Damage Score: Ion Leakage

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8842.339(a)	7	1263.191	17.033	.000
Intercept	8055.406	1	8055.406	108.619	.000
potato	1892.313	1	1892.313	25.516	.000
regime	1493.822	1	1493.822	20.143	.000
temp	803.280	1	803.280	10.831	.002
potato * regime	2087.539	1	2087.539	28.148	.000
potato * temp	48.135	1	48.135	.649	.423
regime * temp	13.891	1	13.891	.187	.667
potato * regime * temp	89.198	1	89.198	1.203	.277
Error	4968.876	67	74.162		
Total	27481.316	75			
Corrected Total	13811.215	74			

a R Squared = .640 (Adjusted R Squared = .603)

It appears that the fit is moderate (R squared = 0.640), but that there is some explanatory power in the variables.

Note that we cannot interpret the quoted F statistics, as this is an unbalanced design, and therefore the stated p-values are not in general exact. However, these results do give an indication of which terms might be omitted.

Number of parameters:
g=4

REDUCED MODEL 1

Dependent Variable: Damage Score: Ion Leakage

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8717.469(a)	4	2179.367	29.950	.000
Intercept	8060.776	1	8060.776	110.774	.000
potato	1890.703	1	1890.703	25.983	.000
regime	1492.360	1	1492.360	20.509	.000
temp	1225.714	1	1225.714	16.844	.000
potato * regime	2089.928	1	2089.928	28.721	.000
Error	5093.746	70	72.768		
Total	27481.316	75			
Corrected Total	13811.215	74			

a R Squared = .631 (Adjusted R Squared = .610)

Interaction terms omitted:

Three-way interaction:
potato*regime*temp

Two-way interactions:
potato*temp
regime*temp

Number of parameters:
g=3

REDUCED MODEL 2

Dependent Variable: Damage Score: Ion Leakage

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6627.541(a)	3	2209.180	21.834	.000
Intercept	13233.292	1	13233.292	130.792	.000
potato	1502.970	1	1502.970	14.855	.000
regime	1977.340	1	1977.340	19.543	.000
temp	1255.583	1	1255.583	12.410	.001
Error	7183.674	71	101.179		
Total	27481.316	75			
Corrected Total	13811.215	74			

a R Squared = .480 (Adjusted R Squared = .458)

Remaining interaction term
potato*regime omitted

Number of parameters:
g=3

REDUCED MODEL 3

Dependent Variable: Damage Score: Ion Leakage

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7491.755(a)	3	2497.252	28.057	.000
Intercept	8119.673	1	8119.673	91.226	.000
potato	1862.829	1	1862.829	20.929	.000
regime	1467.591	1	1467.591	16.489	.000
potato * regime	2119.797	1	2119.797	23.816	.000
Error	6319.460	71	89.006		
Total	27481.316	75			
Corrected Total	13811.215	74			

a R Squared = .542 (Adjusted R Squared = .523)

Interaction replaced,
but temp main effect
removed.

Balanced two-factor predictor, one covariate linear model

Task Distraction Data

This is a balanced design, with 15 replicates in each of the 3 x 3 = 9 cells of the table. For a balanced design, the quoted p-values are more reliable as indications of significance

COMPLETE MODEL

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39671.916(a)	17	2333.642	48.240	.000
Intercept	309.723	1	309.723	6.402	.013
Group	252.027	2	126.014	2.605	.078
Task	450.584	2	225.292	4.657	.011
Distract	2790.513	1	2790.513	57.684	.000
Group * Task	172.095	4	43.024	.889	.473
Group * Distract	335.100	2	167.550	3.463	.035
Task * Distract	2535.238	2	1267.619	26.203	.000
Group * Task * Distract	142.924	4	35.731	.739	.567
Error	5660.010	117	48.376		
Total	90341.000	135			
Corrected Total	45331.926	134			

a R Squared = .875 (Adjusted R Squared = .857)

k=17

REDUCED MODEL 1

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	37704.447(a)	9	4189.383	68.656	.000
Intercept	537.895	1	537.895	8.815	.004
Group	228.483	2	114.242	1.872	.158
Task	494.293	2	247.147	4.050	.020
Distract	3575.111	1	3575.111	58.589	.000
Group * Distract	343.795	2	171.898	2.817	.064
Task * Distract	2540.469	2	1270.235	20.817	.000
Error	7627.479	125	61.020		
Total	90341.000	135			
Corrected Total	45331.926	134			

a R Squared = .832 (Adjusted R Squared = .820)

g=9

REDUCED MODEL 2

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	37360.652(a)	7	5337.236	85.034	.000
Intercept	619.535	1	619.535	9.871	.002
Group	433.379	2	216.690	3.452	.035
Task	500.794	2	250.397	3.989	.021
Distract	3796.748	1	3796.748	60.491	.000
Task * Distract	2597.561	2	1298.780	20.692	.000
Error	7971.274	127	62.766		
Total	90341.000	135			
Corrected Total	45331.926	134			

a R Squared = .824 (Adjusted R Squared = .814)

g=7

REDUCED MODEL 3

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	36927.272(a)	5	7385.454	113.357	.000
Intercept	522.634	1	522.634	8.022	.005
Task	513.356	2	256.678	3.940	.022
Distract	3565.647	1	3565.647	54.728	.000
Task * Distract	2750.062	2	1375.031	21.105	.000
Error	8404.654	129	65.152		
Total	90341.000	135			
Corrected Total	45331.926	134			

a R Squared = .815 (Adjusted R Squared = .807)

g=5

REDUCED MODEL 4

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	34177.211(a)	3	11392.404	133.791	.000
Intercept	1192.389	1	1192.389	14.003	.000
Task	23726.782	2	11863.391	139.323	.000
Distract	5515.685	1	5515.685	64.776	.000
Error	11154.715	131	85.150		
Total	90341.000	135			
Corrected Total	45331.926	134			

a R Squared = .754 (Adjusted R Squared = .748)

g=3

Task Distraction Data: Follow-Up Analysis

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39671.916(a)	17	2333.642	48.240	.000
Intercept	309.723	1	309.723	6.402	.013
Group	252.027	2	126.014	2.605	.078
Task	450.584	2	225.292	4.657	.011
Distract	2790.513	1	2790.513	57.684	.000
Group * Task	172.095	4	43.024	.889	.473
Group * Distract	335.100	2	167.550	3.463	.035
Task * Distract	2535.238	2	1267.619	26.203	.000
Group * Task * Distract	142.924	4	35.731	.739	.567
Error	5660.010	117	48.376		
Total	90341.000	135			
Corrected Total	45331.926	134			

a R Squared = .875 (Adjusted R Squared = .857)

Now we omit the three-way interaction only

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39528.992(a)	13	3040.692	63.403	.000
Intercept	305.861	1	305.861	6.378	.013
Group	316.691	2	158.345	3.302	.040
Task	481.392	2	240.696	5.019	.008
Distract	2802.472	1	2802.472	58.436	.000
Group * Task	1824.545	4	456.136	9.511	.000
Group * Distract	414.362	2	207.181	4.320	.015
Task * Distract	2643.278	2	1321.639	27.558	.000
Error	5802.934	121	47.958		
Total	90341.000	135			
Corrected Total	45331.926	134			

a R Squared = .872 (Adjusted R Squared = .858)

Here, to compare these models,

$$F = \frac{(5802.934 - 5660.010)/(17-13)}{5660.010/(135-17-1)} = 0.739$$

We compare this with the Fisher-F(17-13,135-17-1) = Fisher-F(4,117) distribution:
the 0.05 tail quantile Critical Value is 2.45.

Therefore we do not reject the simpler model as an adequate simplification: we CAN drop the three-way interaction.

Now we try to drop the least significant two-way interaction: group*distract

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39114.630(a)	11	3555.875	70.348	.000
Intercept	332.570	1	332.570	6.579	.012
Group	364.584	2	182.292	3.606	.030
Task	521.312	2	260.656	5.157	.007
Distract	2871.665	1	2871.665	56.812	.000
Group * Task	1753.978	4	438.495	8.675	.000
Task * Distract	2725.029	2	1362.514	26.955	.000
Error	6217.296	123	50.547		
Total	90341.000	135			
Corrected Total	45331.926	134			

a R Squared = .863 (Adjusted R Squared = .851)

Here, to compare these models,

```
F = (6217.296 - 5802.934)/(13-11)   =  4.320
    ---------------------------------------
         5802.934/(135-13-1)
```

We compare this with the Fisher-F(13-11,135-17-1) = Fisher-F(2,121) distribution:
the approximate 0.05 tail quantile Critical Value is 3.07.

Therefore we reject the simpler model as an adequate simplification.

The conclusion is that the most appropriate model in terms of ANOVA F-test selection is

group + task + distract + group.task + group.distract + task.distract

Note that there is very little difference between the R squared statistics for the models.

SUMMARY OF ISSUES IN ANOVA, REGRESSION AND GENERAL 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

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

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

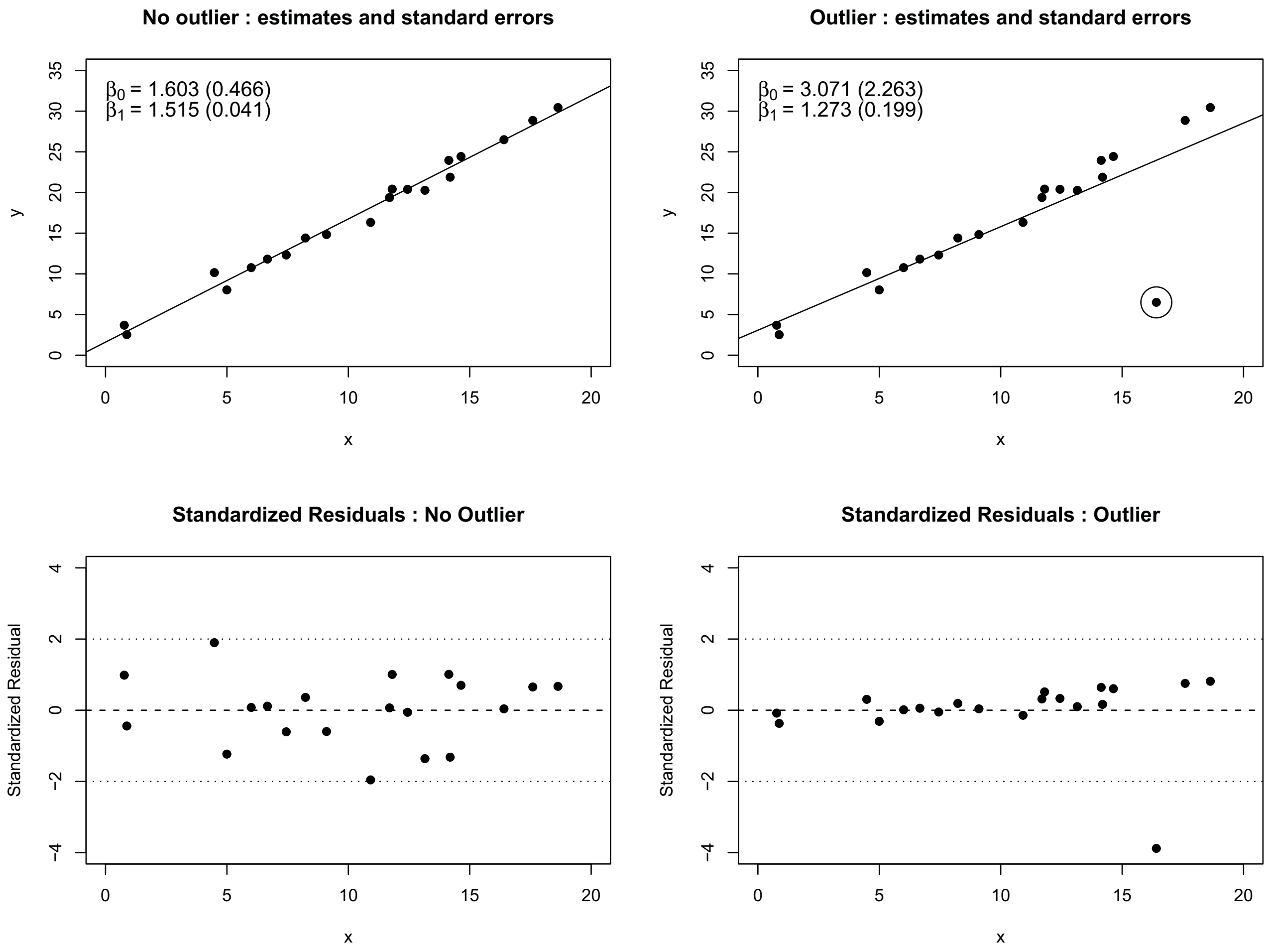
   (i) *Normality*

   (ii) *Dependence* on the covariates

   (iii) *Constant variance*

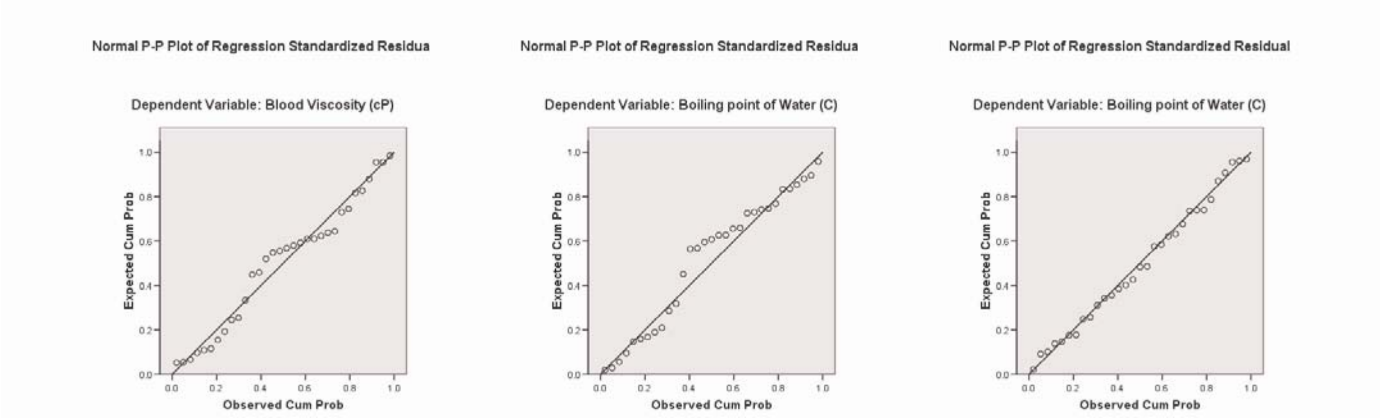
   (iv) *Outliers*: An *outlier* is an response value that gives rise to a residual which is large in magnitude, indicating that the fit of the model is poor for that data point.

   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).



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. These plots are available on the *Linear Regression* menu, after clicking the *Plots* button.

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.



(a) Viscosity vs PCV (b) Hooker Data: Linear Model (c) Hooker Data: Quadratic Model

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, $X^2 \overset{\cdot}{\sim}$ 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

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

EXAMPLE 1: 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, p. 898,

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

so H_0 is **rejected**.

EXAMPLE 2: 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, p. 898,

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

so H_0 is **rejected**

Chi-Squared test for the nucleotide count data

Use

Analyze → Nonparametric Tests → Chi-Square

pulldown menus.

For the test of

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

First null hypothesis

Nucleotide

	Observed N	Expected N	Residual
A	38514	31769.8	6744.3
C	24631	31769.8	-7138.8
G	25685	31769.8	-6084.8
T	38249	31769.8	6479.3
Total	127079		

Chi-squared Statistic = 5522.597

Test Statistics

	Nucleotide
Chi-Square(a)	5522.597
df	3
Asymp. Sig.	.000

p-value < 0.001

a 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 31769.8.

For the test of

$$H_0 : p_1 = p_4 =0.3 \quad p_2 = p_3 = 0.2$$

Second null hypothesis

Nucleotide

	Observed N	Expected N	Residual
A	38514	38123.7	390.3
C	24631	25415.8	-784.8
G	25685	25415.8	269.2
T	38249	38123.7	125.3
Total	127079		

Test Statistics

Chi-squared Statistic = 31.492

	Nucleotide
Chi-Square(a)	31.492
df	3
Asymp. Sig.	.000

p-value < 0.001

a 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 25415.8.

Chi-Squared test for the Hair and Eye colour count data

Use

Analyze → Descriptive Statistics → Crosstabs

pulldown menus.

For the test of

H_0 : Hair and Eye colour are assigned independently

Eye Colour * Hair Colour Crosstabulation

Count

		Hair Colour				Total
		Black	Brown	Red	Blond	
Eye Colour	Brown	68	119	26	7	220
	Blue	20	84	17	94	215
	Hazel	15	54	14	10	93
	Green	5	29	14	16	64
Total		108	286	71	127	592

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	138.290(a)	9	.000
Likelihood Ratio	146.444	9	.000
Linear-by-Linear Association	28.292	1	.000
N of Valid Cases	592		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 7.68.

p-value < 0.001

Chi-square statistic = 138.290

Note the comment returned by SPSS: The chi-squared test is not appropriate if any of the cells in the table have expected count less than 5 under the null hypothesis.

In this case, there is no problem as the cell counts are large enough.

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

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

1 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 $X \sim$ 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 $X \sim$ 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 $X \sim$ Binomial$(n, 1/2)$, then it can be shown that

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

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

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

   and note that if H_0 is true,

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

   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

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

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

2 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

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

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

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

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

EXAMPLES

EXAMPLE 1: 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.

This test can be implemented using SPSS, using the

Analyze → *Nonparametric Tests* → *Binomial*

pulldown menus. The test can be carried out by

(a) Selecting the *test variable* from the variables list
(b) Set the *Cut Point* equal to $\eta_0 = 9$.

A **two-sided** test is carried out at the $\alpha = 0.05$ level. The SPSS output is presented below for the two experiments in turn:

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
% Water content	Group 1	<= 9	8	.80	.50	.109
	Group 2	> 9	2	.20		
	Total		10	1.00		

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
% Water content	Group 1	<= 9	17	.85	.50	.003
	Group 2	> 9	3	.15		
	Total		20	1.00		

EXAMPLE 2: 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), we find that

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	

$$R_1 = 48 \qquad R_2 = 105$$

so that the two statistics are

$$\begin{aligned} \text{Wilcoxon} \quad W &= R_2 = 105 \\ \text{Mann-Whitney} \quad U &= R_2 - \frac{n_2(n_2+1)}{2} = 105 - 36 = 69 \end{aligned}$$

- 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

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

  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 $\alpha = 0.05$.

This test can be implemented using SPSS, using the

Analyze → Nonparametric Tests → Two Independent Samples

pulldown menus. Note, however, that SPSS uses different rules for defining the test statistics, although it yields the same conclusions for a two-sided test.

EXAMPLE 3: 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:

$$\begin{array}{lllllllllll} \text{Disease :} & n = 10 & 864 & 636 & 638 & 708 & 786 & 600 & 1320 & 750 & 594 & 750 \\ \text{Healthy :} & n = 8 & 1014 & 684 & 810 & 990 & 840 & 978 & 1002 & 1110 \end{array}$$

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

$$\begin{aligned} \text{Wilcoxon} \quad W &= R_2 = 101 \\ \text{Mann-Whitney} \quad U &= R_2 - \frac{n_2(n_2+1)}{2} = 101 - 36 = 65 \end{aligned}$$

- 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

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

  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 $\alpha = 0.05$.

3 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

   $$\begin{aligned} T_+ &= \text{Sum of ranks for those } x_i > 0 \\ T_- &= \text{Sum of ranks for those } x_i < 0 \end{aligned}$$

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

(1) H_a : Significant **decrease** between first and second measurements
(2) H_a : Significant **increase** between first and second measurements
(3) H_a : Significant **change** between first and second measurements

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 and

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

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

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

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

For H_a (1)	use	$C_R = 1.645$
For H_a (2)	use	$C_R = -1.645$
For H_a (3)	use	$C_R = \pm 1.960$

EXAMPLE 1: 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 $\alpha = 0.05$.

4 THREE OR MORE INDEPENDENT SAMPLES: THE KRUSKAL-WALLIS AND FRIEDMAN TESTS

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

4.1 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

$$\begin{aligned} H_0 &: \text{No difference between the population distributions of the } k \text{ groups} \\ H_a &: \text{At least two population distributions different} \end{aligned}$$

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,

$$H \stackrel{\cdot}{\sim} \text{Chisquared}(k-1).$$

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 2: 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 $p = 0.680$.

4.2 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

$$\begin{aligned} H_0 &: \text{No difference between the population distributions of the } k \text{ treatment groups} \\ H_a &: \text{At least two population distributions different} \end{aligned}$$

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,

$$F_r \stackrel{\cdot}{\sim} \text{Chisq}(k-1)$$

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 3: 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 $p = 0.092$.

RANK CORRELATION

5 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

$$r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

where

$$d_i = u_i - v_i \qquad i = 1, \ldots, n$$

Tests for r_S : If the population correlation is ρ, then we may test the hypothesis

$$H_0 \;:\; \rho = 0$$

against the hypotheses

$$\begin{aligned} (1)\, H_a &\;:\; \rho > 0 \\ (2)\, H_a &\;:\; \rho < 0 \\ (3)\, H_a &\;:\; \rho \neq 0 \end{aligned}$$

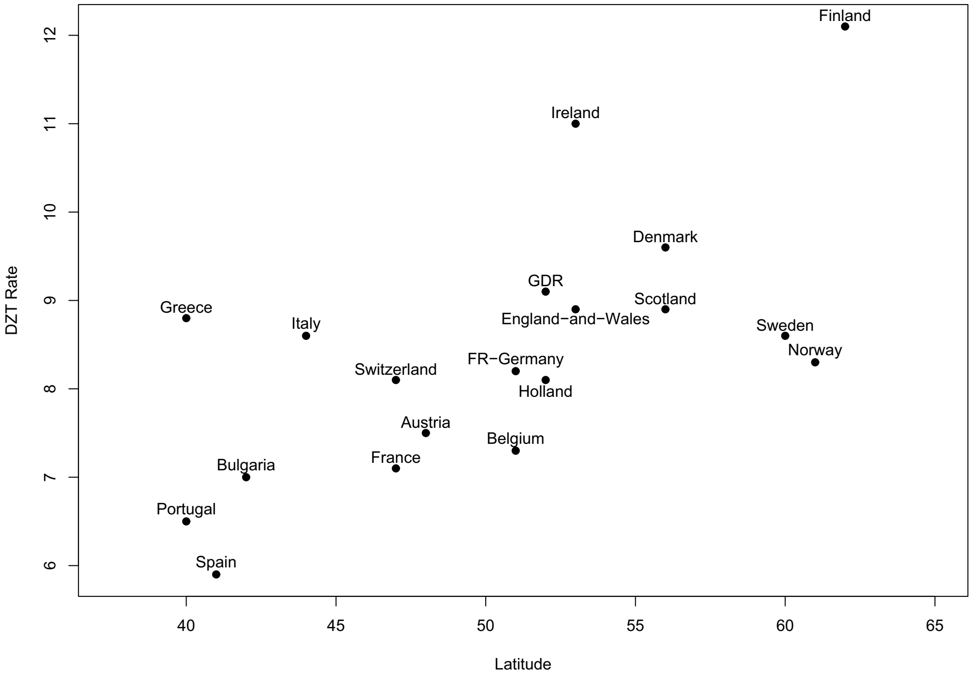
using the table of the null distribution in McClave and Sincich. If Spearman_α is the α tail quantile of the null distribution, we have the following rejection regions:

$$\begin{aligned} (1) &\;:\; \text{Reject } H_0 \text{ if } r_S > \text{Spearman}_\alpha \\ (2) &\;:\; \text{Reject } H_0 \text{ if } r_S < -\text{Spearman}_\alpha \\ (3) &\;:\; \text{Reject } H_0 \text{ if } |r_S| > \text{Spearman}_{\alpha/2} \end{aligned}$$

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.

6 THE ROLE OF RANDOMIZATION/PERMUTATION TESTS

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 $\mathcal{R} = \{6, 7, 26, 27\}$.

Tests for Two Independent Samples

1. Birthweight Data

Mann-Whitney Test

Ranks

	gp	N	Mean Rank	Sum of Ranks
BW	A	9	5.33	48.00
	B	8	13.13	105.00
	Total	17		

Test Statistics[b]

	BW
Mann-Whitney U	3.000
Wilcoxon W	48.000
Z	-3.187
Asymp. Sig. (2-tailed)	.001
Exact Sig. [2*(1-tailed Sig.)]	.001[a]

a. Not corrected for ties.

b. Grouping Variable: gp

2. Treadmill test Data

Mann-Whitney Test

Ranks

	Group	N	Mean Rank	Sum of Ranks
Time	1	8	12.63	101.00
	2	10	7.00	70.00
	Total	18		

Test Statistics[b]

	Time
Mann-Whitney U	15.000
Wilcoxon W	70.000
Z	-2.222
Asymp. Sig. (2-tailed)	.026
Exact Sig. [2*(1-tailed Sig.)]	.027[a]

a. Not corrected for ties.

b. Grouping Variable: Group

Two Dependent Samples (Paired Data)

1. Haemodialysis Data

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
after - before	Negative Ranks	3[a]	4.50	13.50
	Positive Ranks	8[b]	6.56	52.50
	Ties	1[c]		
	Total	12		

a. after < before

b. after > before

c. after = before

Test Statistics[b]

	after - before
Z	-1.736[a]
Asymp. Sig. (2-tailed)	.083

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

2. PEFR/Asthma Data

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
PERF after - PEFR before	Negative Ranks	8[a]	5.50	44.00
	Positive Ranks	1[b]	1.00	1.00
	Ties	0[c]		
	Total	9		

a. PERF after < PEFR before

b. PERF after > PEFR before

c. PERF after = PEFR before

Test Statistics[b]

	PERF after - PEFR before
Z	-2.549[a]
Asymp. Sig. (2-tailed)	.011

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

K Independent Samples

1. Mucociliary efficiency data

Kruskal-Wallis Test

Ranks

	group	N	Mean Rank
y	Healthy	5	7.20
	Obstructive airway disease	4	9.00
	Asbestosis	5	6.60
	Total	14	

Test Statistics[a,b]

	y
Chi-Square	.771
df	2
Asymp. Sig.	.680

a. Kruskal Wallis Test

b. Grouping Variable: group

2. Memory Task Data

Kruskal-Wallis Test

Ranks

	Memory Task	N	Mean Rank
Number of Words	Counting	10	12.95
	Rhyming	10	13.10
	Adjective	10	31.50
	Imagery	10	36.60
	Intentional	10	33.35
	Total	50	

Test Statistics[a,b]

	Number of Words
Chi-Square	25.376
df	4
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: Memory Task

K Dependent Samples

1. Hypnosis Data

Friedman Test

Ranks

	Mean Rank
Fear	3.38
Joy	2.50
Sadness	2.38
Calmness	1.75

Test Statistics[a]

N	8
Chi-Square	6.450
df	3
Asymp. Sig.	.092

a. Friedman Test

2. Soil sulphur content

Friedman Test

Ranks

	Mean Rank
CaCl	2.60
NH4OAc	2.00
Ca(H2P04)2	2.80
Water	2.60

Test Statistics[a]

N	5
Chi-Square	1.080
df	3
Asymp. Sig.	.782

a. Friedman Test