Analysis of Variance
Using Statgraphics Centurion

Dr. Neil W. Polhemus

Analysis of Variance

Analysis of variance (ANOVA) models partition the variability of a response variable into components attributable to one or more explanatory factors.

Factors may be:

- Categorical or quantitative
- Crossed or nested
- Fixed or random
- Fully or partially randomized

Procedures

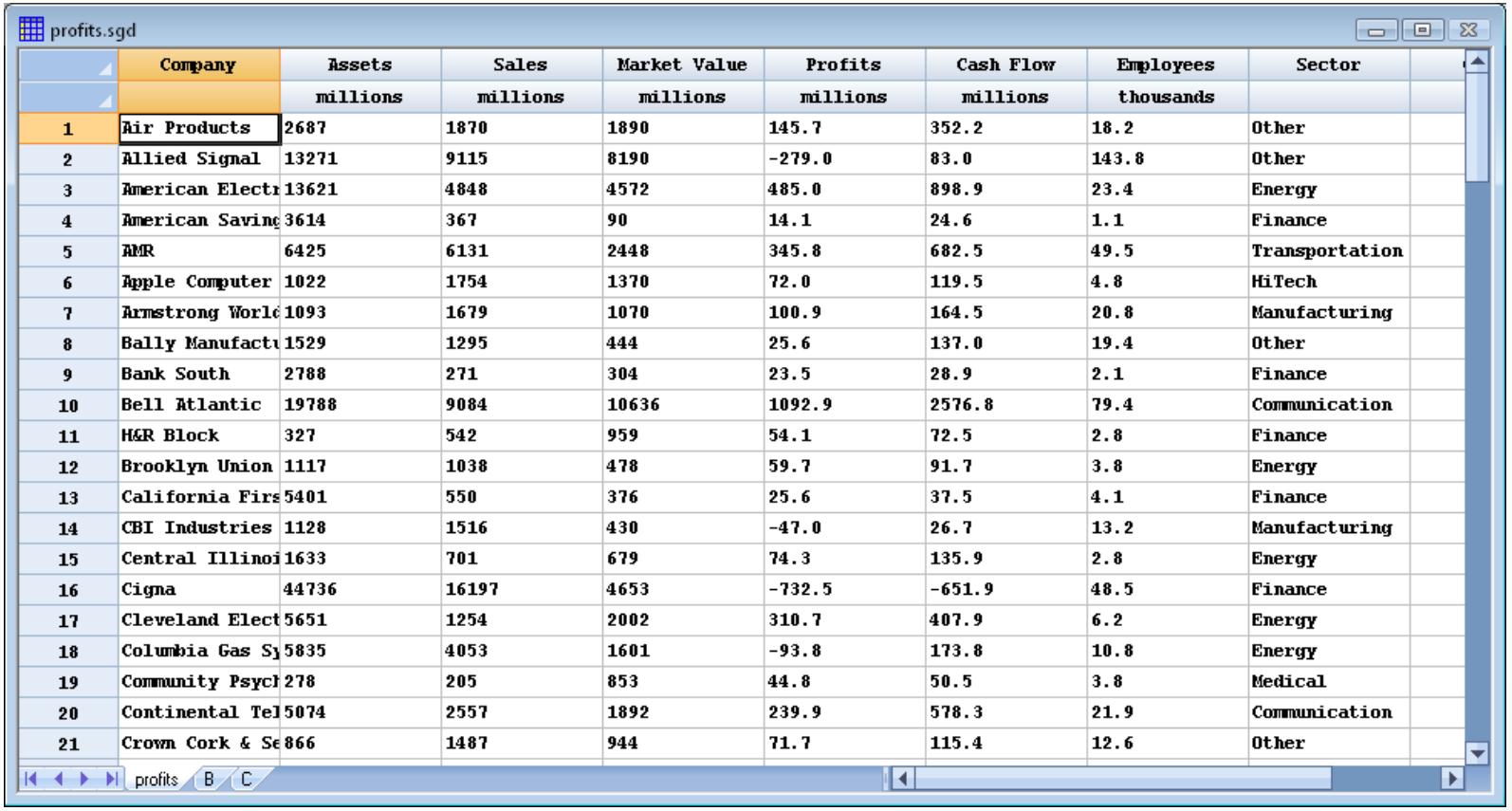
STATGRAPHICS has procedures for:

- Oneway ANOVA – single categorical factor.
- Multifactor ANOVA – multiple crossed categorical factors.
- Variance Components Analysis – multiple nested categorical factors.
- General Linear Models – any combination of factors.

Example #1 – Oneway ANOVA

- Data: 1/10-th systematic sample of companies in the Fortune 500 list for 1986 (Source: DASL – Data and Story Library)
- Response: Y = profit per employee
- Factor: X = sector of economy

Data file: profits.sgd

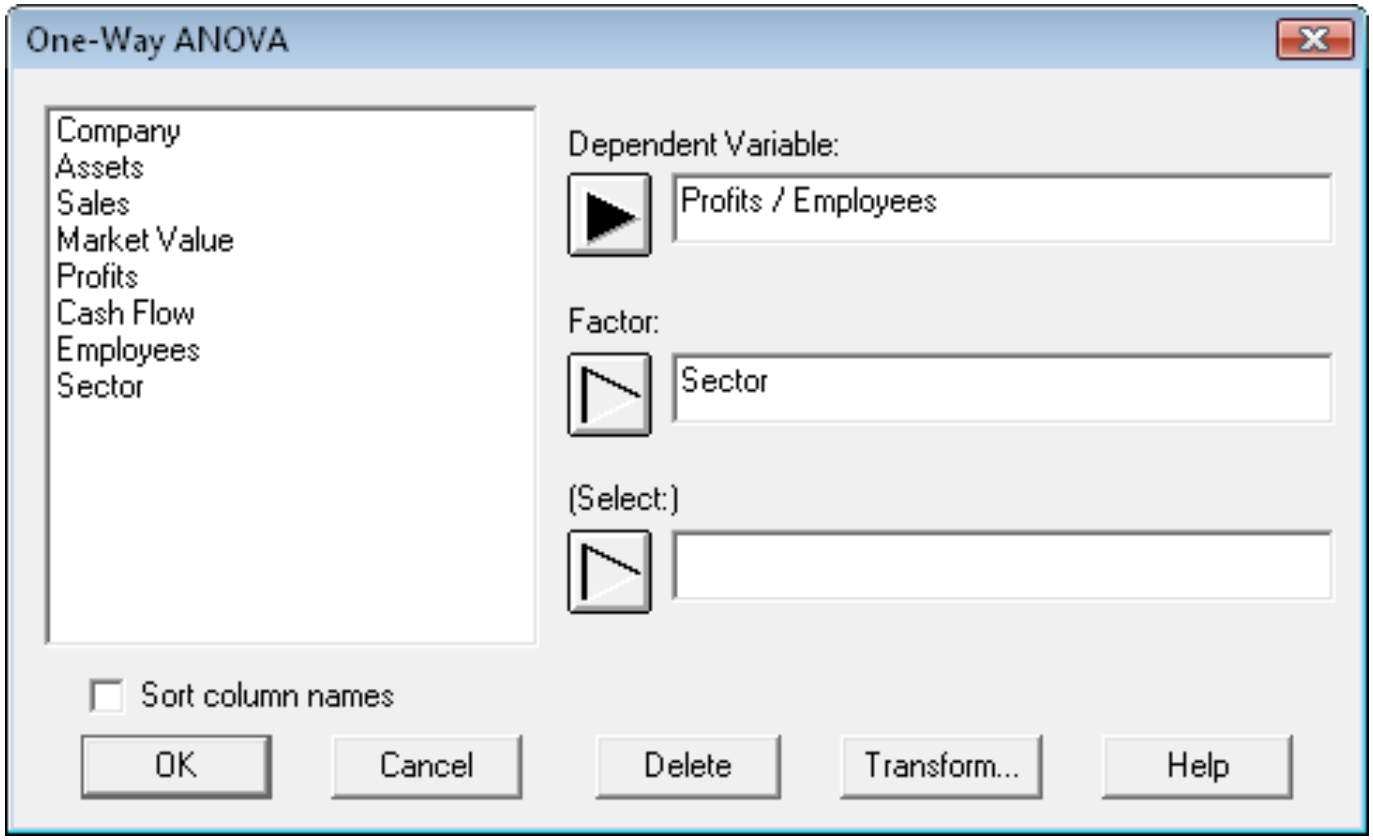


profits.sgd

	Company	Assets	Sales	Market Value	Profits	Cash Flow	Employees	Sector
		millions	millions	millions	millions	millions	thousands	
1	Air Products	2687	1870	1890	145.7	352.2	18.2	Other
2	Allied Signal	13271	9115	8190	-279.0	83.0	143.8	Other
3	American Electr	13621	4848	4572	485.0	898.9	23.4	Energy
4	American Savinç	3614	367	90	14.1	24.6	1.1	Finance
5	AMR	6425	6131	2448	345.8	682.5	49.5	Transportation
6	Apple Computer	1022	1754	1370	72.0	119.5	4.8	HiTech
7	Armstrong Worlc	1093	1679	1070	100.9	164.5	20.8	Manufacturing
8	Bally Manufactu	1529	1295	444	25.6	137.0	19.4	Other
9	Bank South	2788	271	304	23.5	28.9	2.1	Finance
10	Bell Atlantic	19788	9084	10636	1092.9	2576.8	79.4	Communication
11	H&R Block	327	542	959	54.1	72.5	2.8	Finance
12	Brooklyn Union	1117	1038	478	59.7	91.7	3.8	Energy
13	California Firs	5401	550	376	25.6	37.5	4.1	Finance
14	CBI Industries	1128	1516	430	-47.0	26.7	13.2	Manufacturing
15	Central Illinoi	1633	701	679	74.3	135.9	2.8	Energy
16	Cigna	44736	16197	4653	-732.5	-651.9	48.5	Finance
17	Cleveland Elect	5651	1254	2002	310.7	407.9	6.2	Energy
18	Columbia Gas Sy	5835	4053	1601	-93.8	173.8	10.8	Energy
19	Community Psycl	278	205	853	44.8	50.5	3.8	Medical
20	Continental Tel	5074	2557	1892	239.9	578.3	21.9	Communication
21	Crown Cork & Se	866	1487	944	71.7	115.4	12.6	Other

profits B C

Data Input

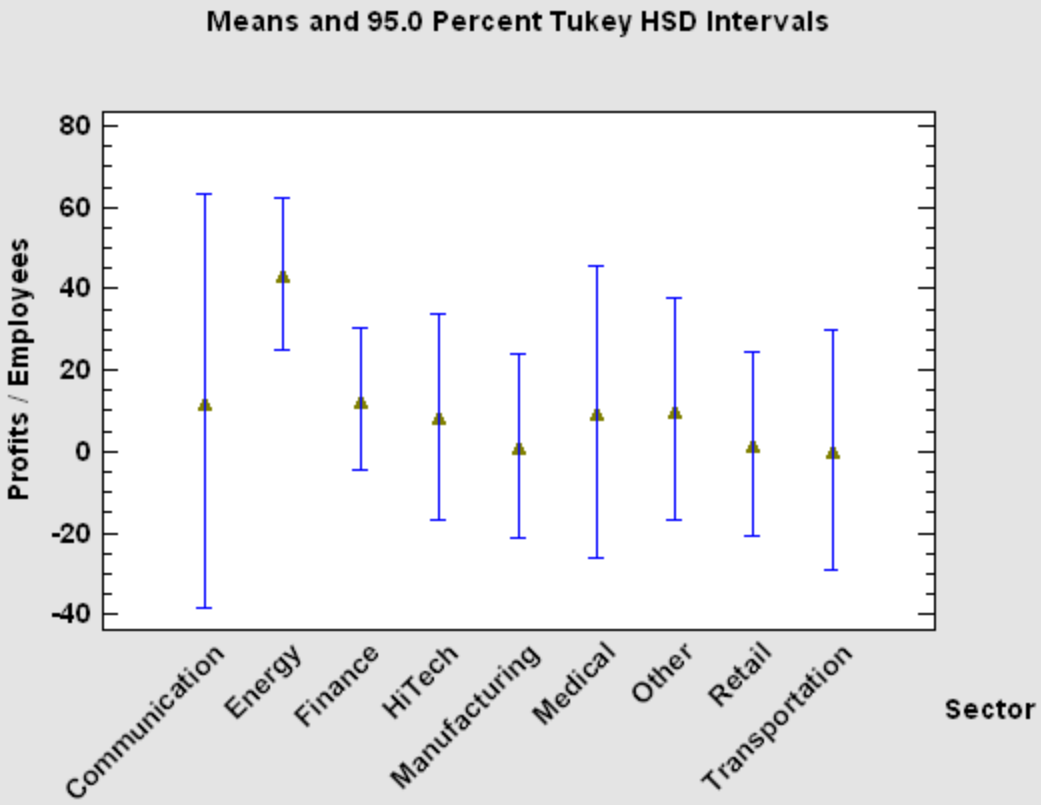

One-Way ANOVA
Company
Assets
Sales
Market Value
Profits
Cash Flow
Employees
Sector
Dependent Variable:
Profits / Employees
Factor:
Sector
(Select:)
Sort column names
OK
Cancel
Delete
Transform...
Help

ANOVA Table

ANOVA Table for Profits / Employees by Sector

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Between groups	17755.7	8	2219.46	2.21	0.0368
Within groups	70331.4	70	1004.73		
Total (Corr.)	88087.1	78			

Means Plot


Means and 95.0 Percent Tukey HSD Intervals
Profits / Employees
80
60
40
20
0
-20
-40
Communication
Energy
Finance
HiTech
Manufacturing
Medical
Other
Retail
Transportation
Sector

Multiple Range Tests

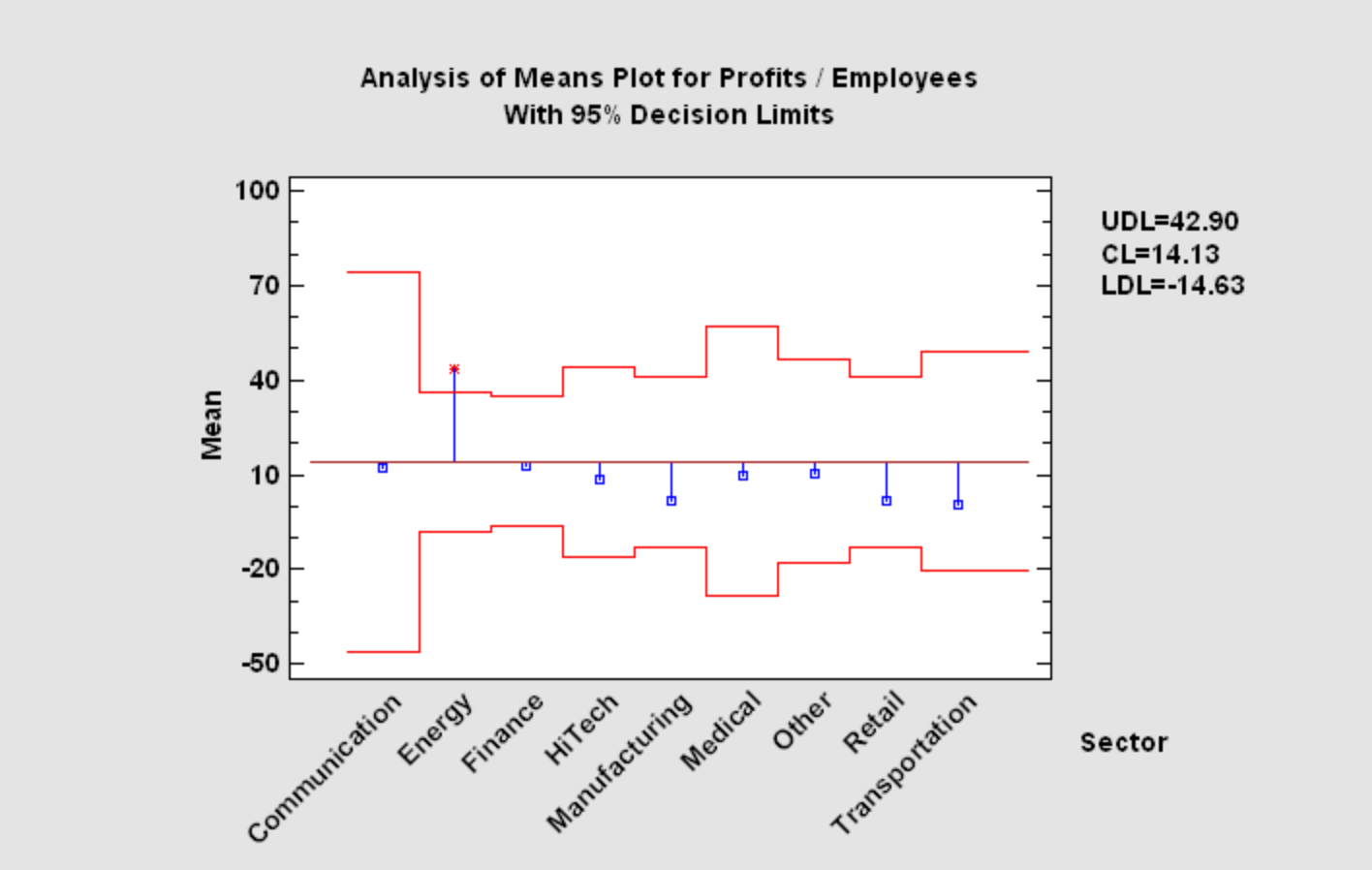
Multiple Range Tests for Profits / Employees by Sector

Method: 95.0 percent Tukey HSD

Sector	*Count*	*Mean*	*Homogeneous Groups*
Transportation	6	0.429218	XX
Manufacturing	10	1.47137	X
Retail	10	1.87889	X
HiTech	8	8.68061	XX
Medical	4	9.7668	XX
Other	7	10.4461	XX
Communication	2	12.3594	XX
Finance	17	12.8905	XX
Energy	15	43.6653	X

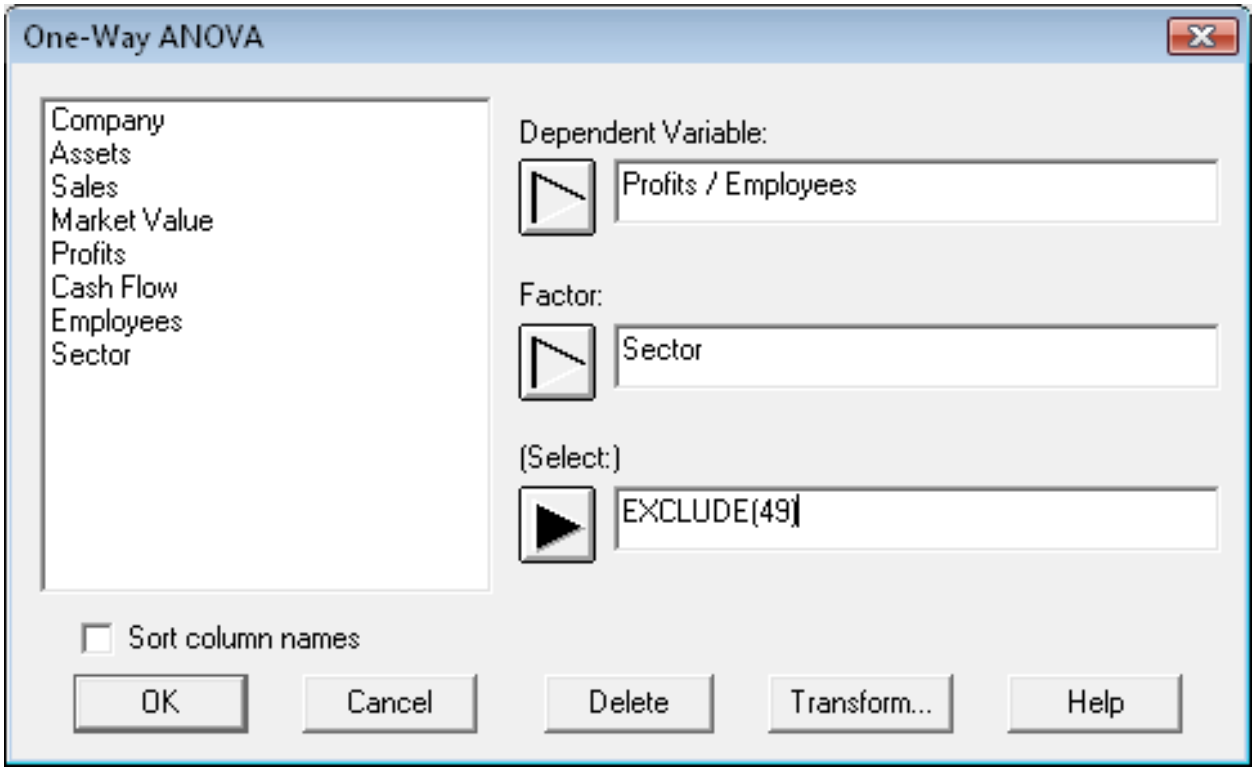
Contrast	*Sig.*	*Difference*	*+/- Limits*
Communication - Energy		-31.3059	76.3733
Communication - Finance		-0.531067	75.8429
Communication - HiTech		3.6788	80.208
Communication - Manufacturing		10.888	78.5875
Communication - Medical		2.59261	87.8635
Communication - Other		1.91329	81.3458
Communication - Retail		10.4805	78.5875
Communication - Transportation		11.9302	82.8385
Energy - Finance		30.7748	35.9404
Energy - HiTech		34.9847	44.4172
Energy - Manufacturing	*	42.1939	41.4192
Energy - Medical		33.8985	57.0925
Energy - Other		33.2192	46.4402
Energy - Retail	*	41.7864	41.4192
Energy - Transportation		43.2361	49.0079

Analysis of Means (ANOM)



Residual Plot

Excluding Row #49



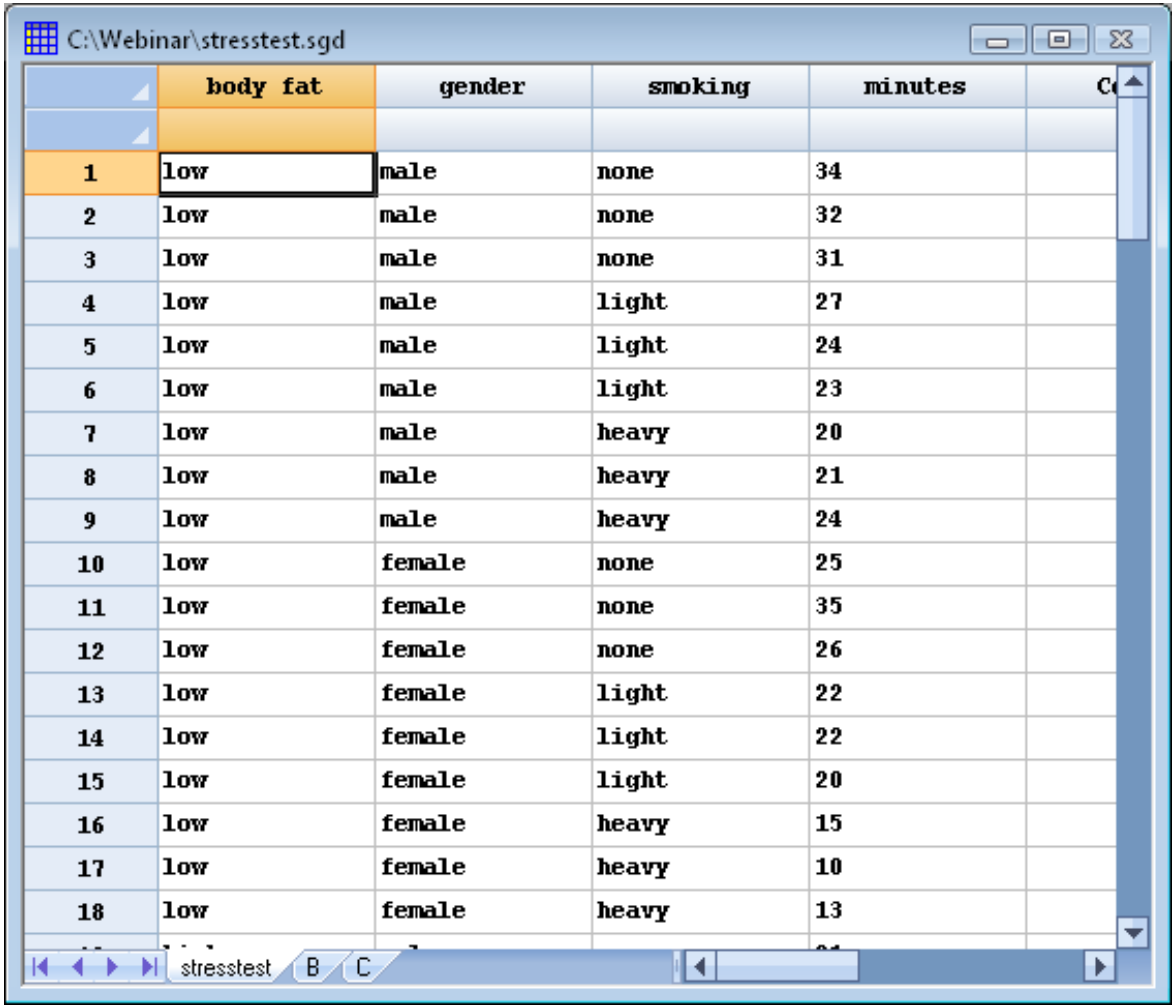
ANOVA Table for Profits / Employees by Sector

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Between groups	6034.98	8	754.372	4.77	0.0001
Within groups	10909.0	69	158.101		
Total (Corr.)	16944.0	77			

Example #2 – Multifactor ANOVA

- Data: Exercise tolerance in a stress test (Applied Linear Statistical Models by Neter et al.)
- Response: Y = minutes until fatigue occurs on a stationary bicycle
- Factors: X1 = gender, X2 = percent body fat, X3 = smoking history
- Experimental design: 2 by 2 by 3 factorial design with 3 subjects at each of the 12 combinations of the factors

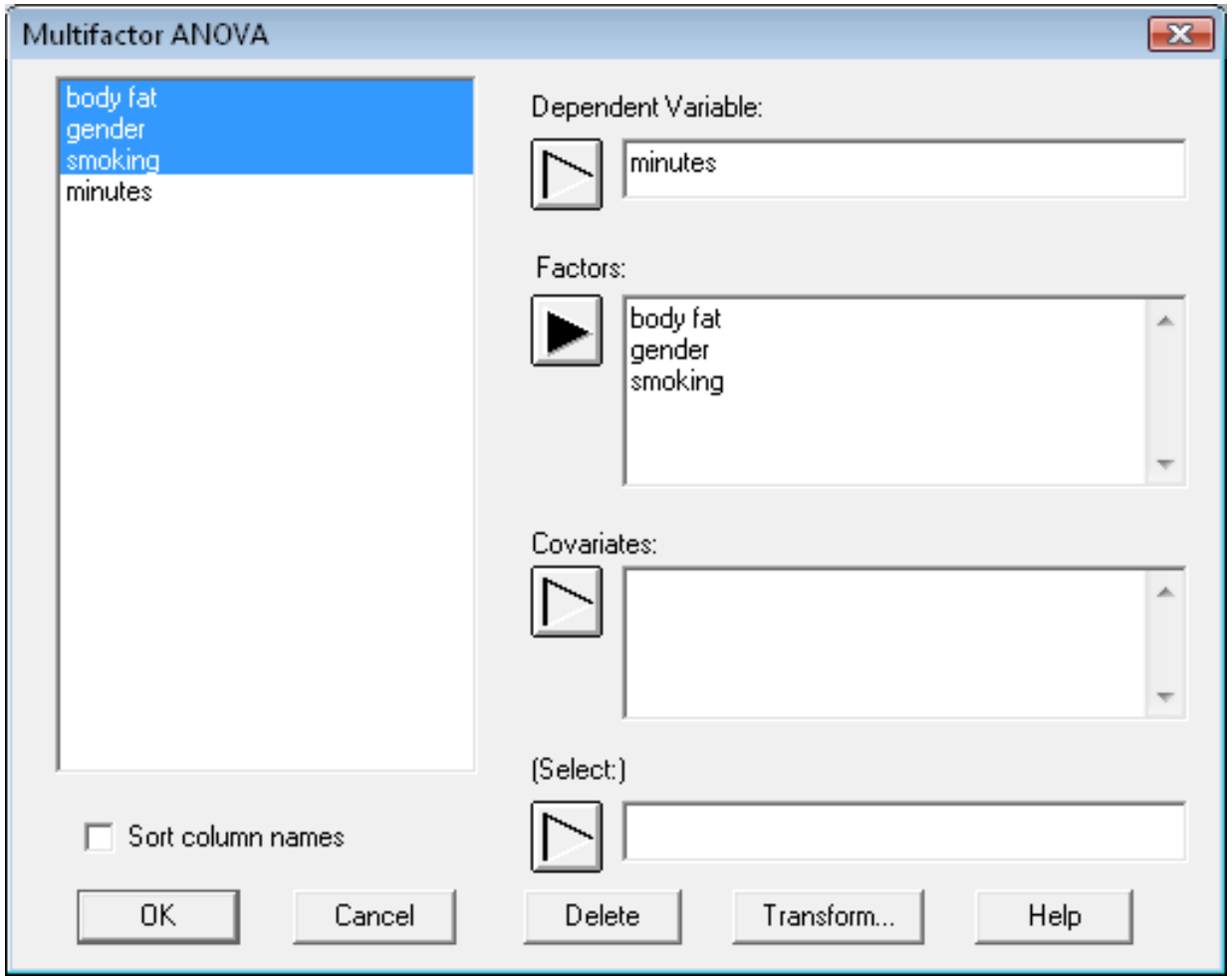
Data file: stresstest.sgd


C:\Webinar\stresstest.sgd

	body fat	gender	smoking	minutes	C
1	low	male	none	34	
2	low	male	none	32	
3	low	male	none	31	
4	low	male	light	27	
5	low	male	light	24	
6	low	male	light	23	
7	low	male	heavy	20	
8	low	male	heavy	21	
9	low	male	heavy	24	
10	low	female	none	25	
11	low	female	none	35	
12	low	female	none	26	
13	low	female	light	22	
14	low	female	light	22	
15	low	female	light	20	
16	low	female	heavy	15	
17	low	female	heavy	10	
18	low	female	heavy	13	

stresstest B C

Data Input


Multifactor ANOVA
body fat
gender
smoking
minutes
Dependent Variable:
minutes
Factors:
body fat
gender
smoking
Covariates:
(Select:)
Sort column names
OK
Cancel
Delete
Transform...
Help

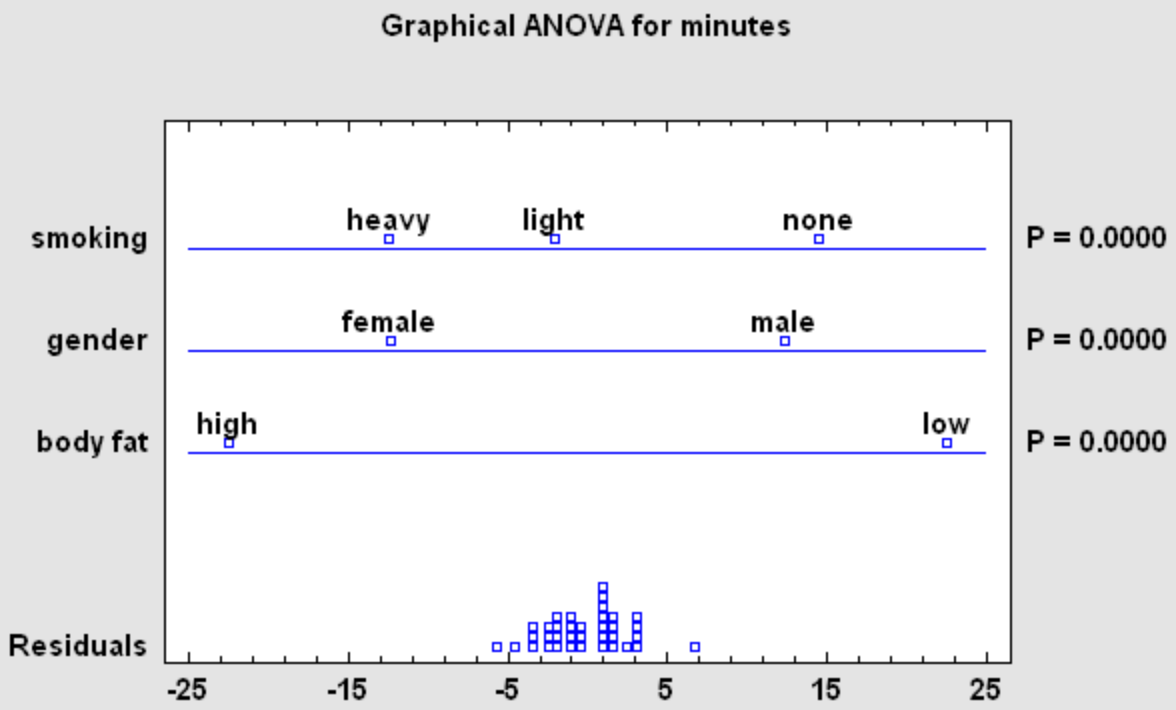
Analysis of Variance Table

Analysis of Variance for minutes - Type III Sums of Squares

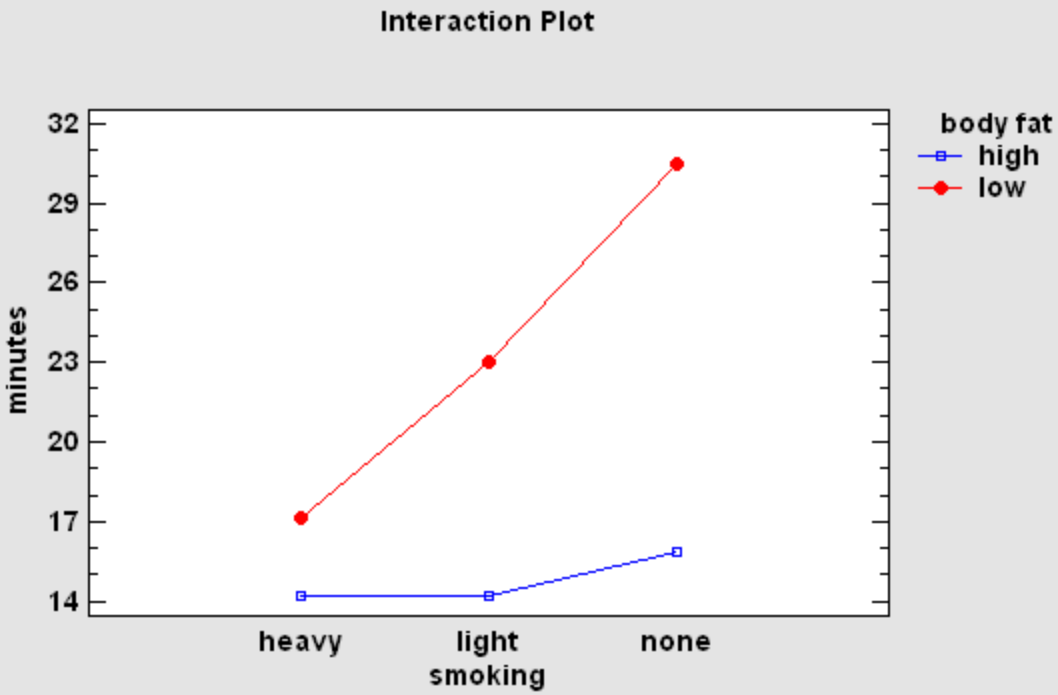
Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
MAIN EFFECTS					
A:body fat	702.25	1	702.25	79.10	0.0000
B:gender	210.25	1	210.25	23.68	0.0000
C:smoking	343.056	2	171.528	19.32	0.0000
INTERACTIONS					
AB	2.25	1	2.25	0.25	0.6189
AC	204.167	2	102.083	11.50	0.0003
BC	21.5	2	10.75	1.21	0.3142
RESIDUAL	230.833	26	8.87821		
TOTAL (CORRECTED)	1714.31	35			

All F-ratios are based on the residual mean square error.

Graphical ANOVA



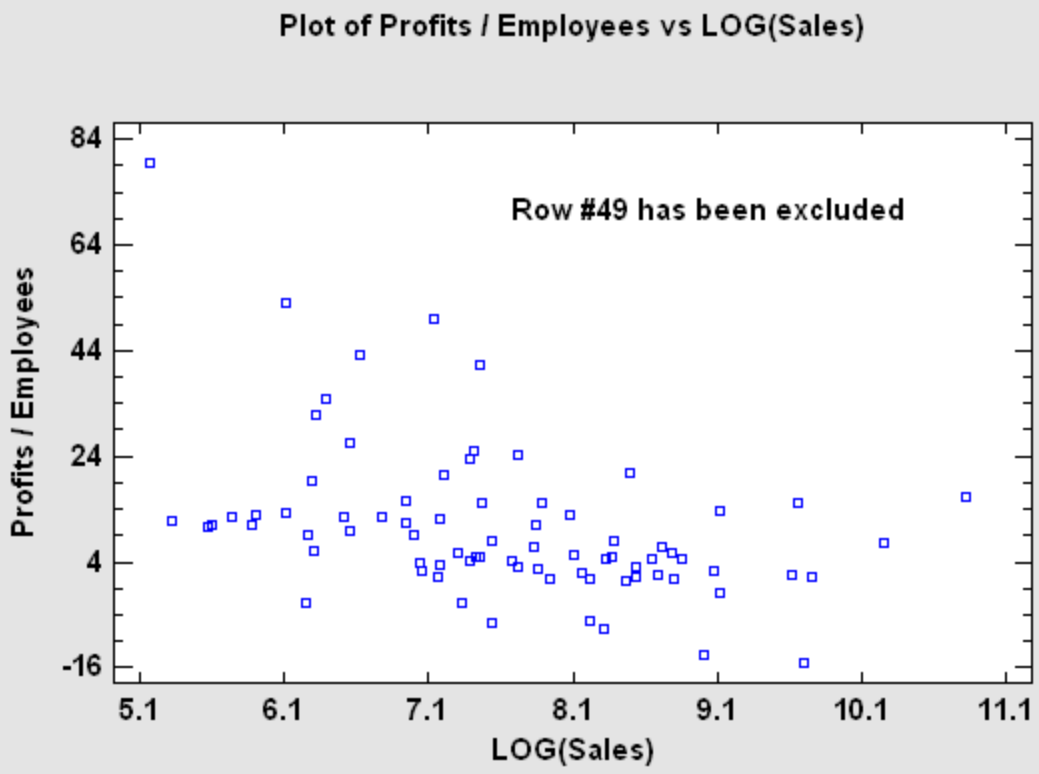
Interaction Plot



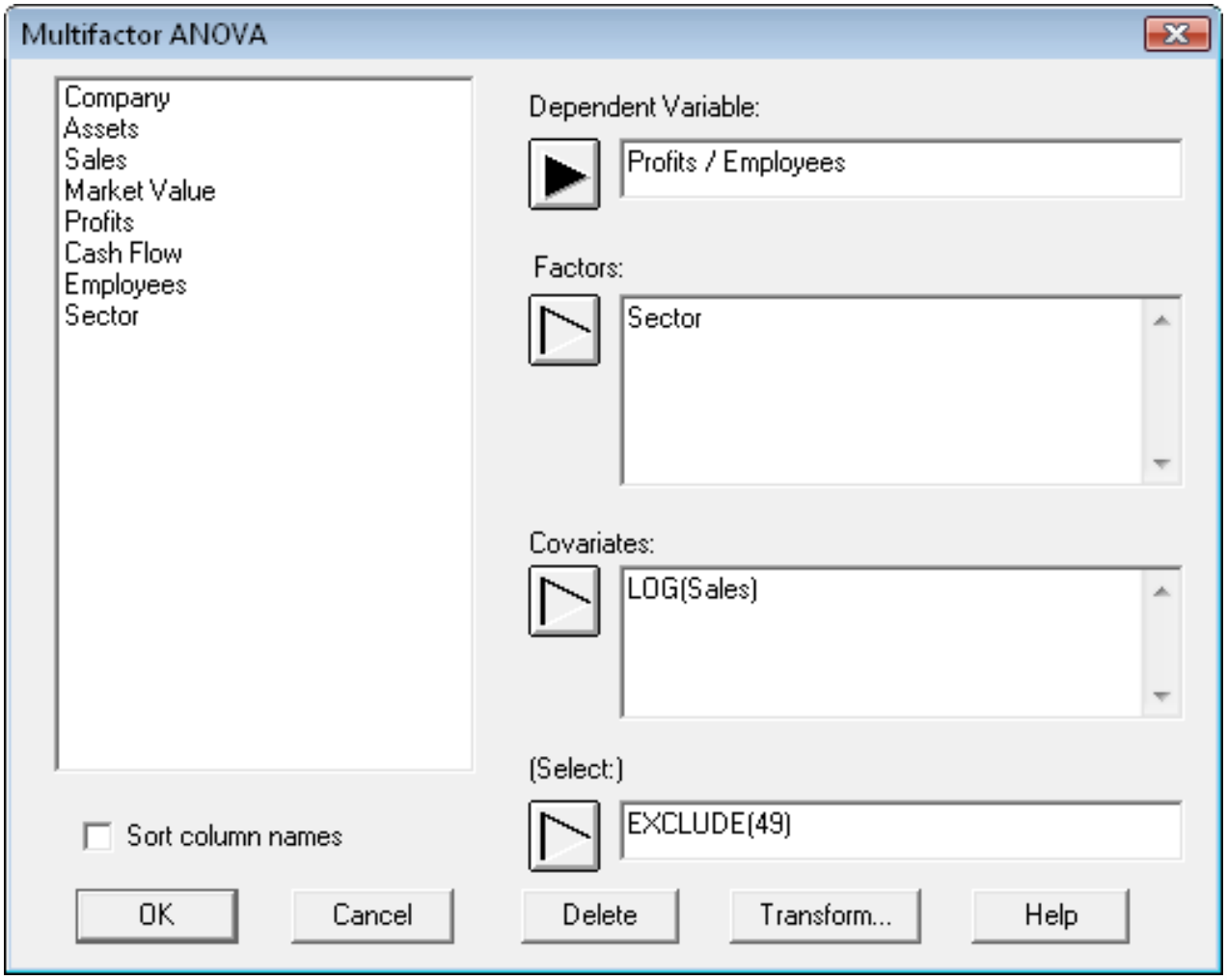
Example #3: Analysis of Covariance

- Tests whether certain factors have an effect on the response after removing the effect of one or more quantitative factors.
- Response: Y = profit per employee
- Factor: X = sector of economy
- Covariate: LOG(sales)

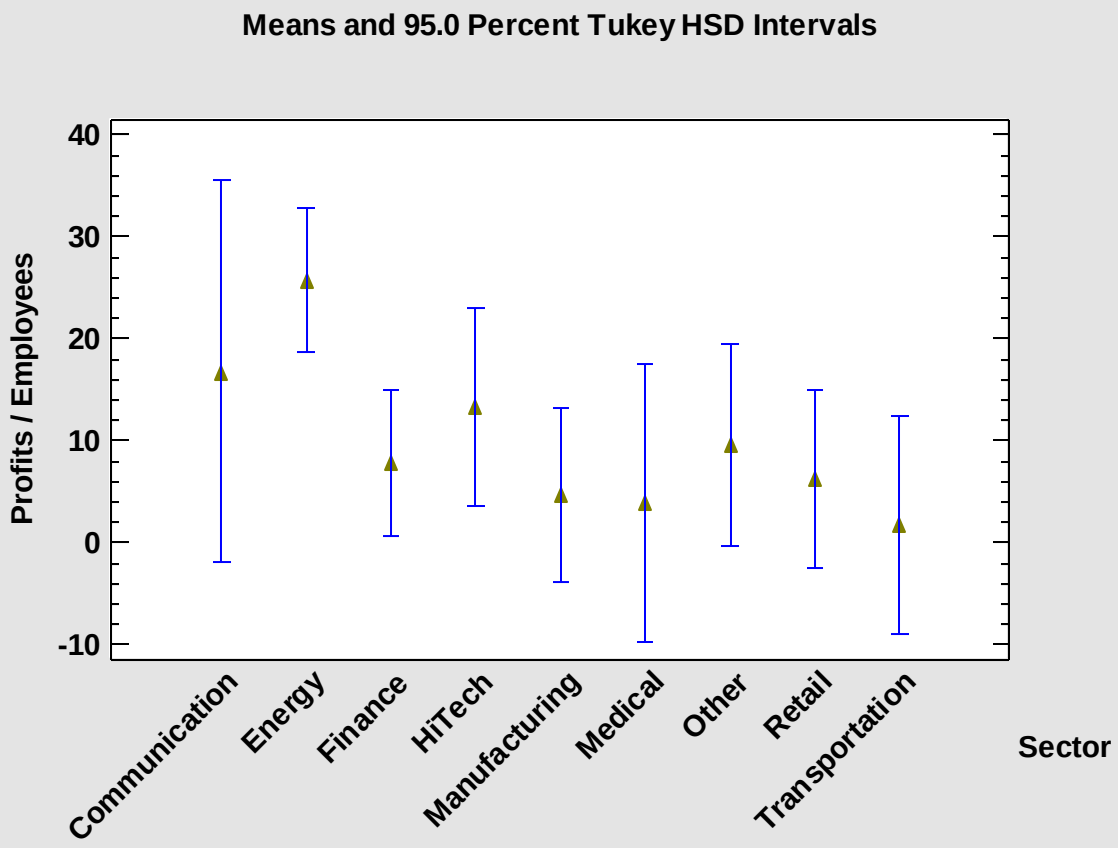
X-Y Plot


Plot of Profits / Employees vs LOG(Sales)
Row #49 has been excluded
Profits / Employees
84
64
44
24
4
-16
5.1
6.1
7.1
8.1
9.1
10.1
11.1
LOG(Sales)

Data Input


Multifactor ANOVA
Company
Assets
Sales
Market Value
Profits
Cash Flow
Employees
Sector
Dependent Variable:
Profits / Employees
Factors:
Sector
Covariates:
LOG(Sales)
(Select:)
EXCLUDE(49)
Sort column names
OK
Cancel
Delete
Transform...
Help

Means Plot


Means and 95.0 Percent Tukey HSD Intervals
Profits / Employees
40
30
20
10
0
-10
Communication
Energy
Finance
HiTech
Manufacturing
Medical
Other
Retail
Transportation
Sector

Example #4 – Variance Components Study

- Data: Pigment paste example (Statistics for Experimenters by Box, Hunter and Hunter)
- Response: Y = moisture contents
- Factors: X1 = batch, X2 = sample within batch, X3 = test within sample
- Experimental design: 15 by 2 by 2 hierarchical design

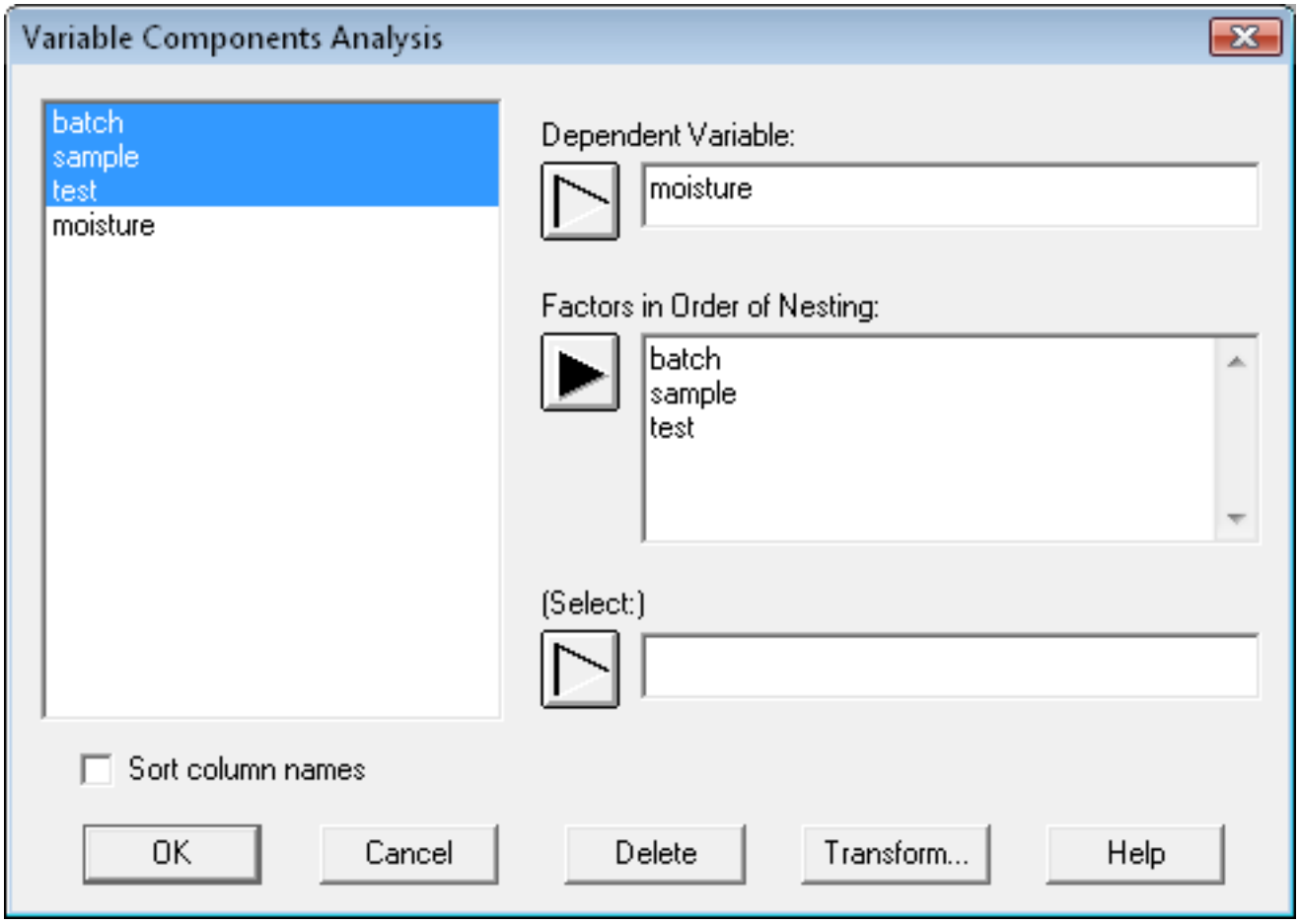
Data file: pigment.sgd

C:\DocData16\pigment.sgd

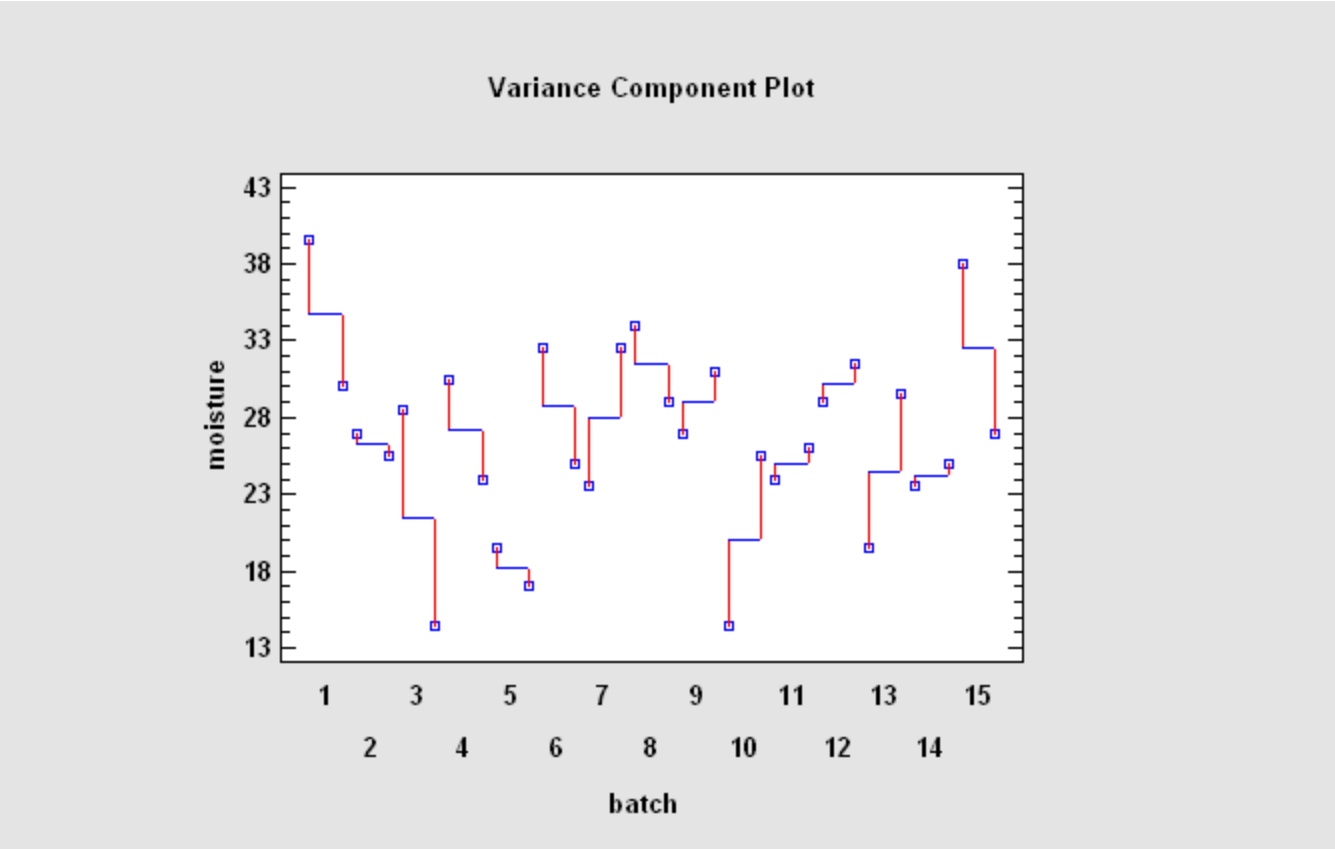
	batch	sample	test	moisture
1	1	1	1	40
2	1	1	2	39
3	1	2	1	30
4	1	2	2	30
5	2	3	1	26
6	2	3	2	28
7	2	4	1	25
8	2	4	2	26
9	3	5	1	29
10	3	5	2	28
11	3	6	1	14
12	3	6	2	15
13	4	7	1	30
14	4	7	2	31
15	4	8	1	24
16	4	8	2	24
17	5	9	1	19
18	5	9	2	20

pigment B C

Data Input


Variable Components Analysis
batch
sample
test
moisture
Dependent Variable:
moisture
Factors in Order of Nesting:
batch
sample
test
(Select:)
Sort column names
OK
Cancel
Delete
Transform...
Help

Variance Components Plot



Analysis of Variance

Analysis of Variance for moisture

Source	*Sum of Squares*	*Df*	*Mean Square*	*Var. Comp.*	*Percent*
TOTAL (CORRECTED)	2108.18	59			
batch	1210.93	14	86.4952	7.12798	19.49
sample	869.75	15	57.9833	28.5333	78.01
test	27.5	30	0.916667	0.916667	2.51

Example #5 – Split-Plot Design

- Data: Corrosion resistance example (Statistics for Experimenters by Box, Hunter and Hunter)
- Response: Y = corrosion resistance of steel bars
- Factors: X1 = furnace temperature, X2 = coating
- Experimental design: Since furnace temperature is hard to change, it will be randomized over a larger experimental unit than coating.

Experimental Design

Run of furnace	Temperature setting °C	Position 1	Position 2	Position 3	Position 4
1	360	C2	C3	C1	C4
2	370	C1	C3	C4	C2
3	380	C3	C1	C2	C4
4	380	C4	C3	C2	C1
5	370	C4	C1	C3	C2
6	360	C1	C4	C2	C3

Each of the 6 runs of the furnace is a “whole plot”.
Temperature is a “whole plot factor” and is randomized across the runs.

Each position in the furnace is a “subplot”.
Coating is a subplot factor that is randomized across the positions.

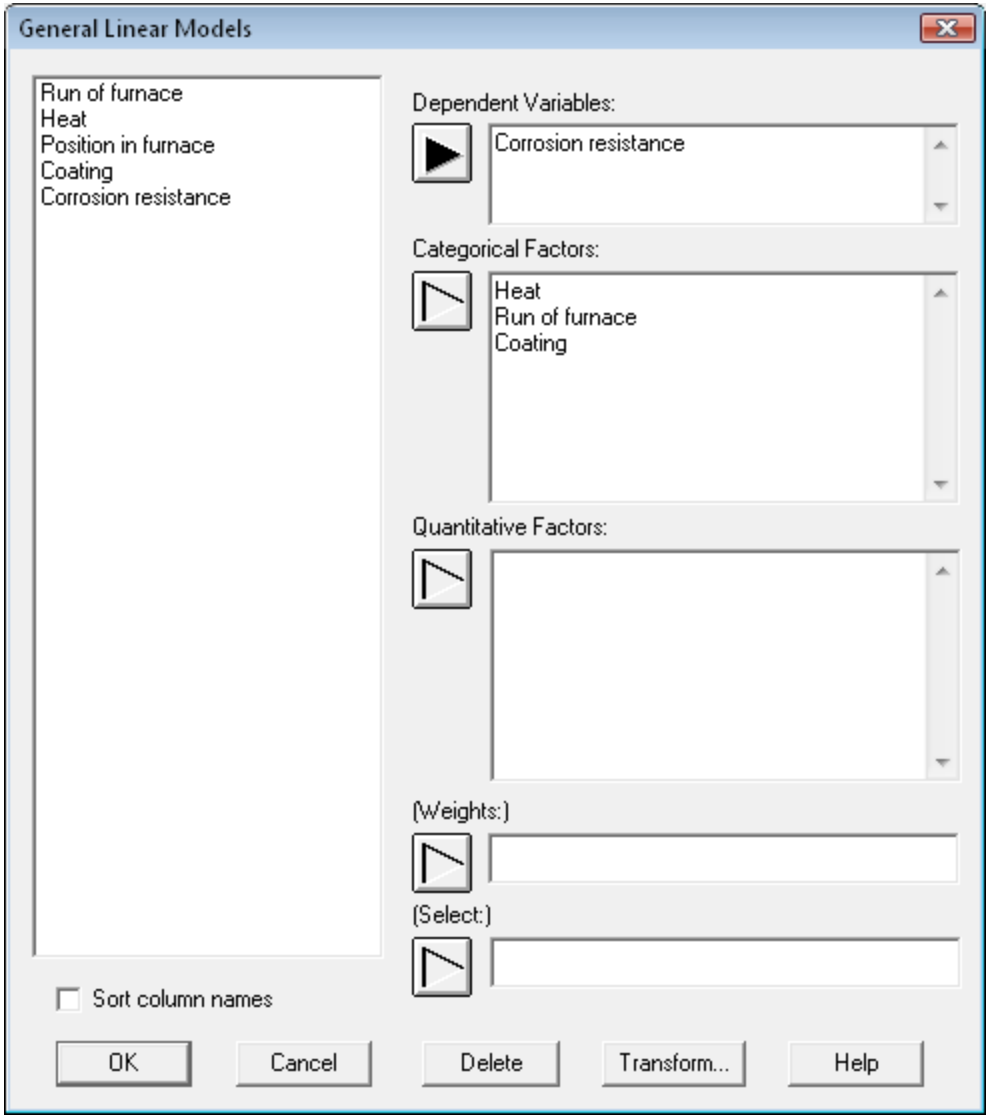
Data file: steelbars.sgd

steelbars.sgd

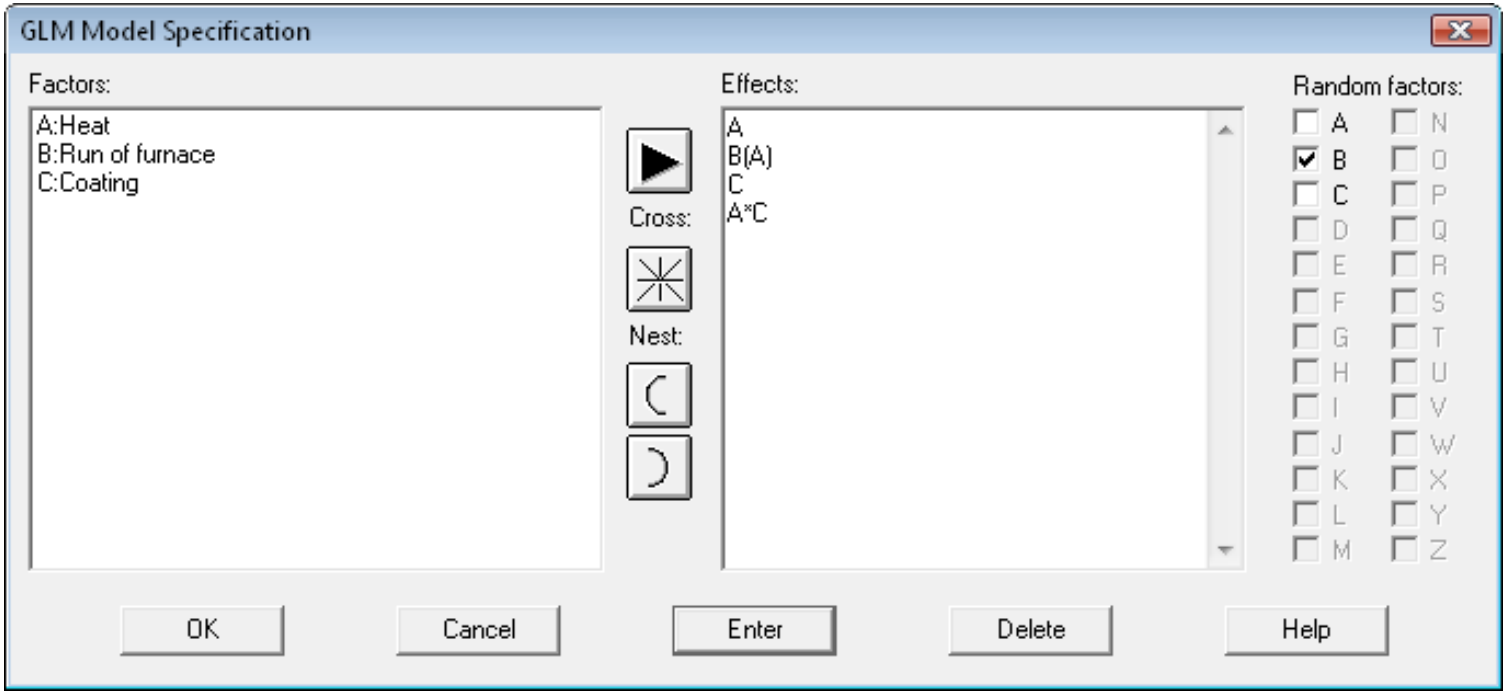
	Run of furnace	Heat	Position in furnace	Coating	Corrosion resistance
	whole plot		subplot		
1	1	360	1	2	73
2	1	360	2	3	83
3	1	360	3	1	67
4	1	360	4	4	89
5	2	370	1	1	65
6	2	370	2	3	87
7	2	370	3	4	86
8	2	370	4	2	91
9	3	380	1	3	147
10	3	380	2	1	155
11	3	380	3	2	127
12	3	380	4	4	212
13	4	380	1	4	153
14	4	380	2	3	90
15	4	380	3	2	100
16	4	380	4	1	108
17	5	370	1	4	150
18	5	370	2	1	140
19	5	370	3	3	121
20	5	370	4	2	142
21	6	360	1	1	33
22	6	360	2	4	54
23	6	360	3	2	8
24	6	360	4	3	46

steelbars B C

Data Input - GLM



Model Specification



Analysis of Variance

Analysis of Variance for Corrosion resistance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	48517.8	14	3465.55	27.83	0.0000
Residual	1120.88	9	124.542		
Total (Corr.)	49638.6	23			

Type III Sums of Squares

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Heat	26519.3	2	13259.6	2.75	0.2093
Run of furnace(Heat)	14439.6	3	4813.21	38.65	0.0000
Coating	4289.13	3	1429.71	11.48	0.0020
Heat*Coating	3269.75	6	544.958	4.38	0.0241
Residual	1120.88	9	124.542		
Total (corrected)	49638.6	23			

F-tests and Error Components

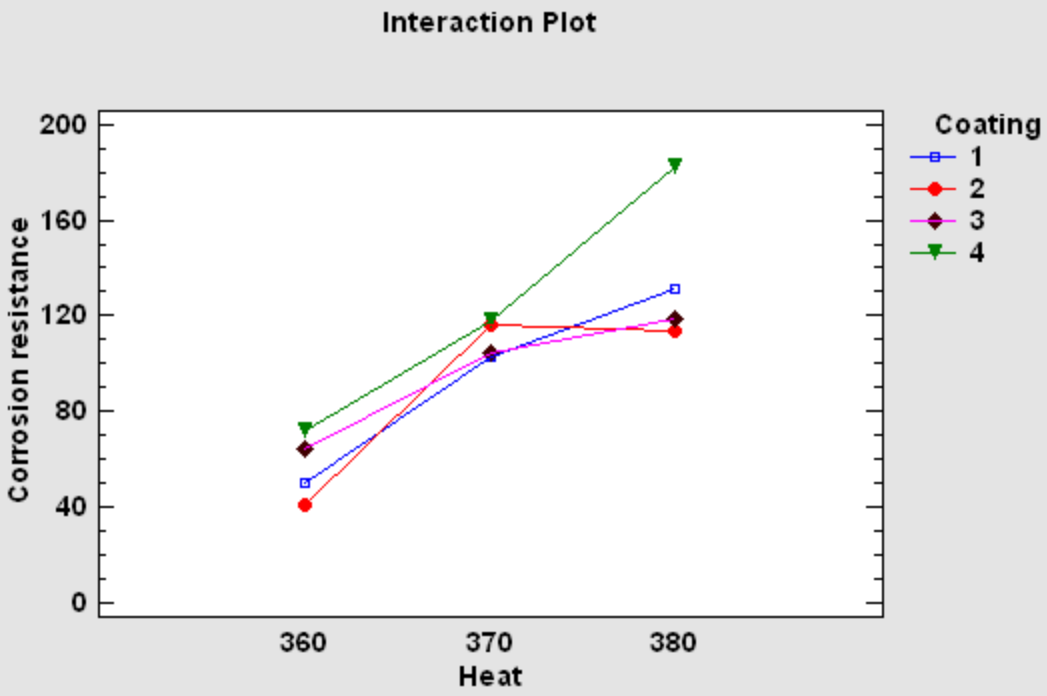
F-Test Denominators

Source	*Df*	*Mean Square*	*Denominator*
Heat	3.00	4813.21	(2)
Run of furnace(Heat)	9.00	124.542	(5)
Coating	9.00	124.542	(5)
Heat*Coating	9.00	124.542	(5)

Variance Components

Source	*Estimate*
Run of furnace(Heat)	1172.17
Residual	124.542

Interaction Plot



Example #6 – Repeated Measures Design

Subject	Drug	T1	T2	T3	T4
1	AX23	72	86	81	77
2	AX23	78	83	88	81
…					
9	BWW9	85	86	80	84
10	BWW9	82	86	80	84
…					
17	CONTROL	69	73	72	74
18	CONTROL	66	62	67	73
…					

Each of 3 drugs was given to 8 different patients. Their heart rate was measured at 4 distinct times.

There are 2 experimental units: “subject” to which a particular drug is assigned, and “time period” in which measurements are taken.

Source: Analysis of Messy Data by Milliken and Johnson.

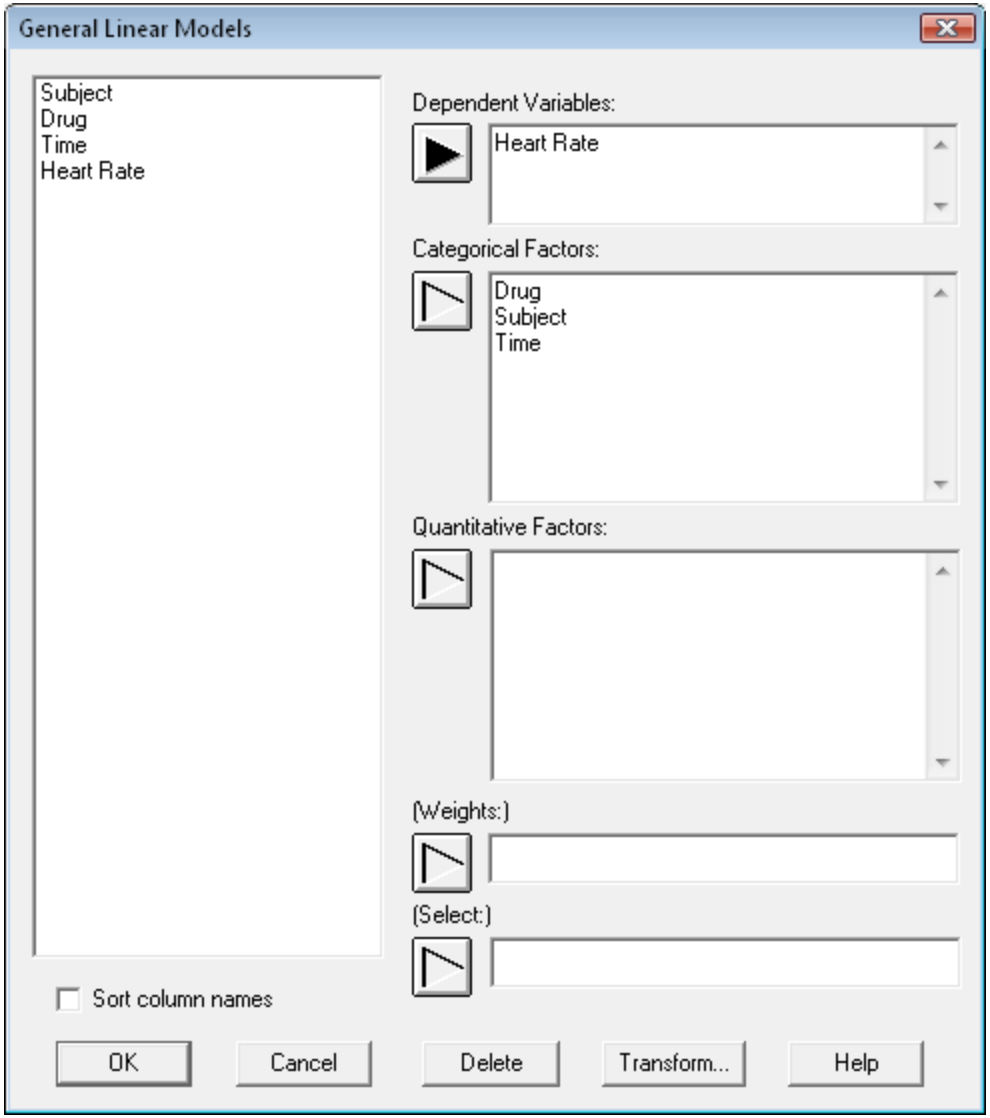
Data file: heartrate.sgd

heartrate.sgd

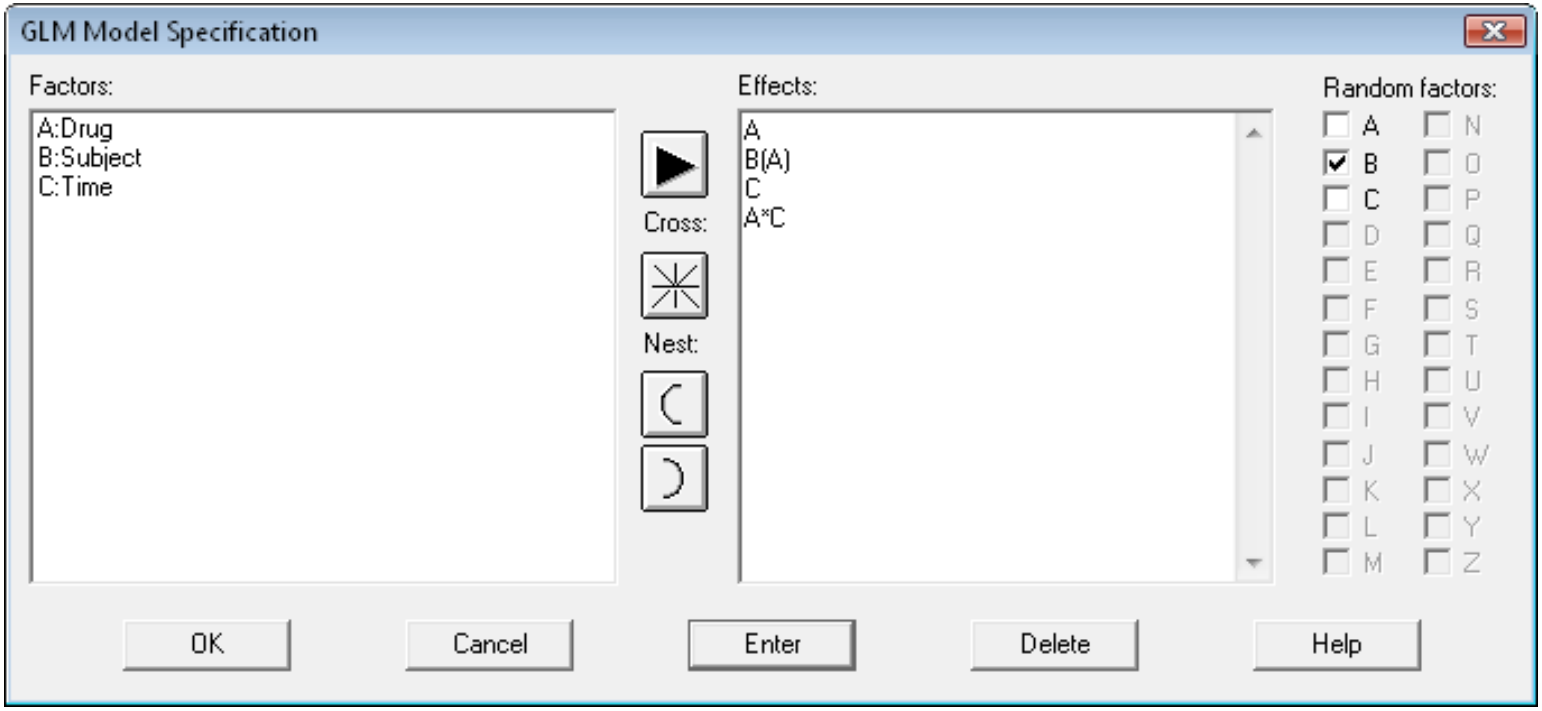
	Subject	Drug	Time	Heart Rate
1	1	AX23	T1	72
2	1	AX23	T2	86
3	1	AX23	T3	81
4	1	AX23	T4	77
5	2	BWW9	T1	85
6	2	BWW9	T2	86
7	2	BWW9	T3	83
8	2	BWW9	T4	80
9	3	CONTROL	T1	69
10	3	CONTROL	T2	73
11	3	CONTROL	T3	72
12	3	CONTROL	T4	74
13	4	AX23	T1	78
14	4	AX23	T2	83
15	4	AX23	T3	88
16	4	AX23	T4	81
17	5	BWW9	T1	82
18	5	BWW9	T2	86

heartrate | B | C

Data input


General Linear Models
Subject
Drug
Time
Heart Rate
Dependent Variables:
Heart Rate
Categorical Factors:
Drug
Subject
Time
Quantitative Factors:
(Weights:)
(Select:)
Sort column names
OK
Cancel
Delete
Transform...
Help

Model Specification



Analysis of Variance

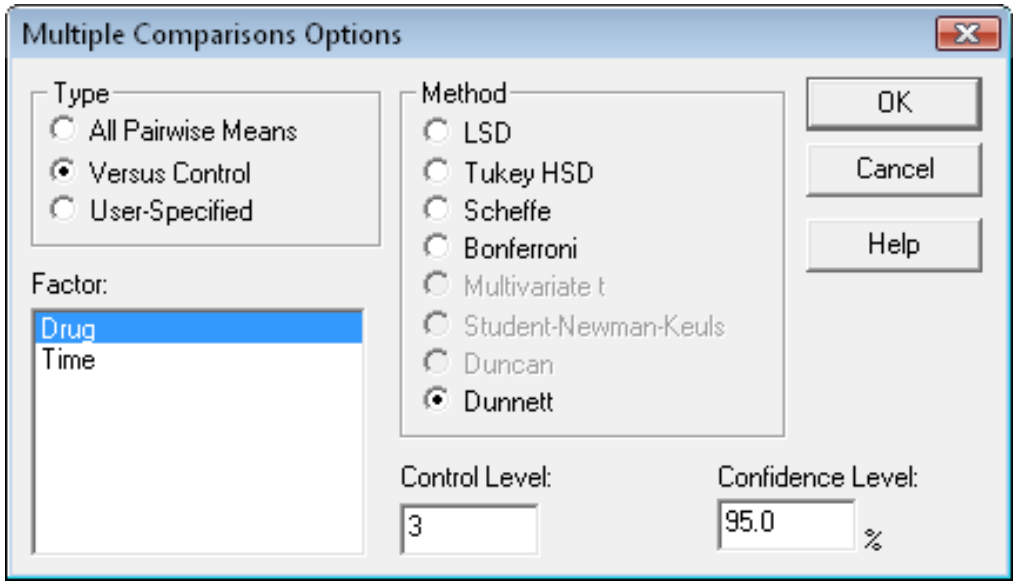
Analysis of Variance for Heart Rate

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	4487.94	32	140.248	18.83	0.0000
Residual	469.219	63	7.44792		
Total (Corr.)	4957.16	95			

Type III Sums of Squares

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Drug	1333.0	2	666.5	5.99	0.0088
Subject(Drug)	2337.91	21	111.329	14.95	0.0000
Time	289.615	3	96.5382	12.96	0.0000
Drug*Time	527.417	6	87.9028	11.80	0.0000
Residual	469.219	63	7.44792		
Total (corrected)	4957.16	95			

Comparison to Control



Multiple Comparisons for Heart Rate by Drug

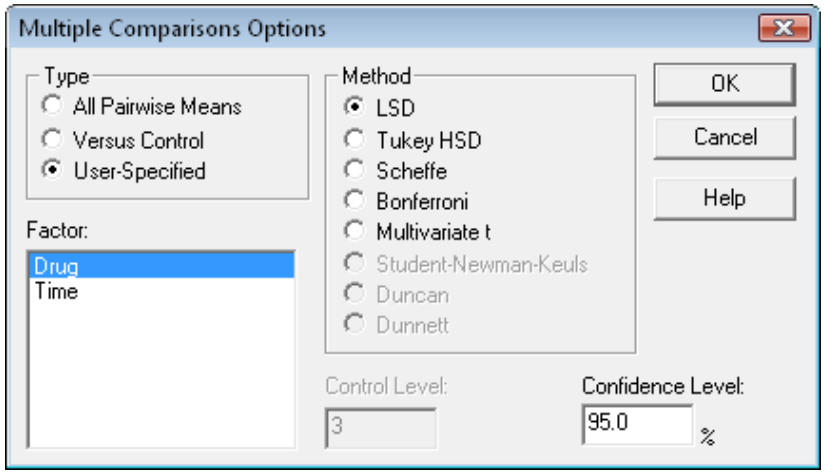
Method: 95.0 percent Dunnett

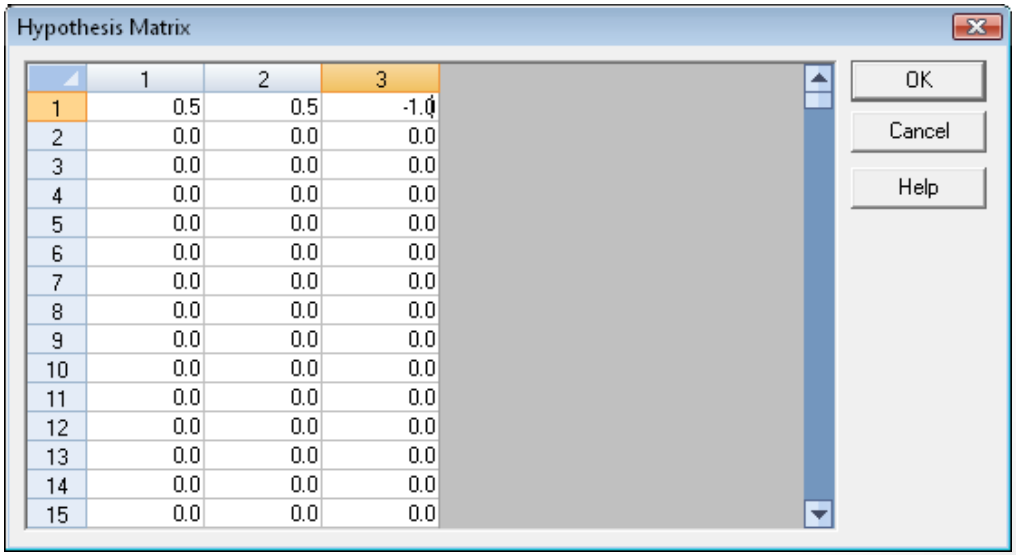
Contrast	*Sig.*	*Difference*	*+/- Limits*
AX23 - CONTROL		4.375	6.34025
BWW9 - CONTROL	*	9.125	6.34025

* denotes a statistically significant difference.

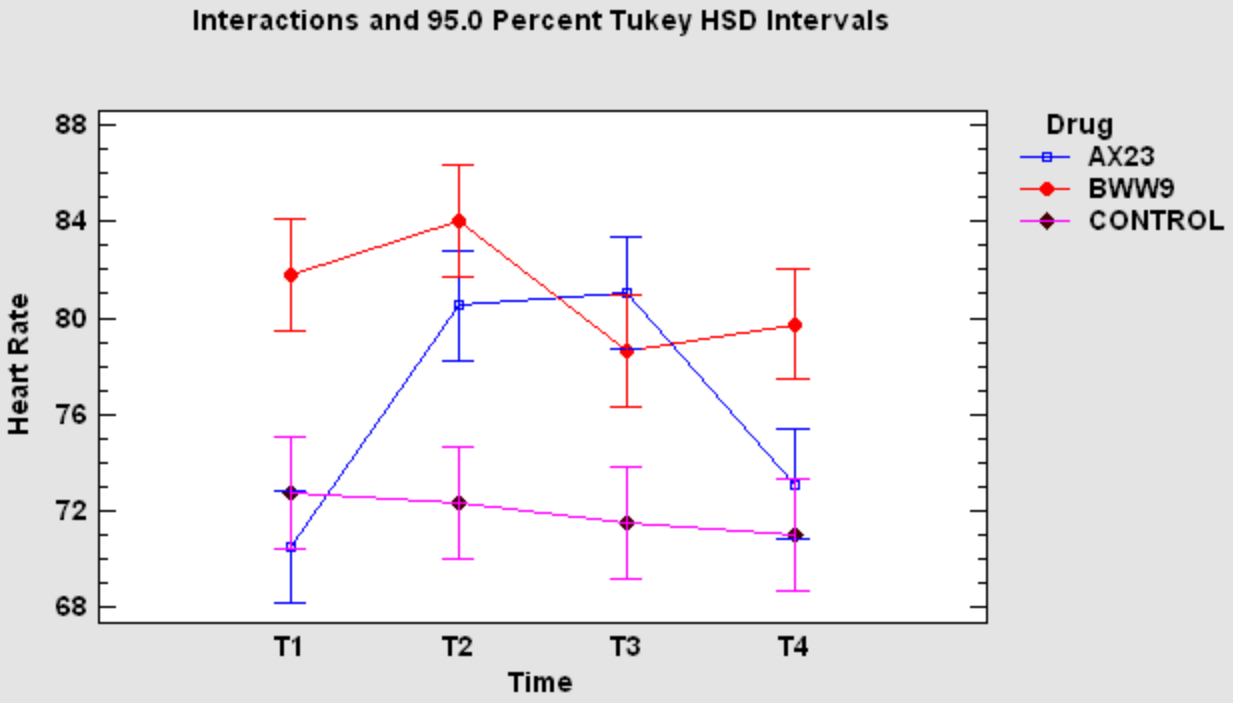
User-Defined Contrast

$L = 0.5*\mu_{AX23} + 0.5*\mu_{BWW9} - \mu_{CONTROL}$





Interaction Plot



References

- Box, G. E. P., Hunter, W. G. and Hunter, J. S. (2005). Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, 2nd edition. New York: John Wiley and Sons.
- DASL – Data and Story Library. (lib.stat.cmu.edu/DASL)
- Milliken, G. A. and Johnson, D. E. (1992). Analysis of Messy Data - Volume 1: Designed Experiments, reprint edition. New York: Van Nostrand Reinhold.
- Neter, J., Kutner, M.H., Wasserman, W., and Nachtscheim, C.J. (1996). Applied Linear Statistical Models, 4th edition. *Homewood, Illinois: Irwin.*