DATA vs DATA

One Input Variable

Compare Variability
- F-ratio test (two levels)
- Bartlett's test (multiple levels)
- Cochran's test (multiple levels)

Compare Means
- Student's T Test (two levels)
- ANOVA (multiple levels)
- Nested ANOVA (multiple levels)

Compare Medians
- Mann-Whitney (two levels)
- Kruskal-Wallis (multiple levels)

Study Source of Variation
- Y vs X plot
- Correlation Coefficient
- Linear Regression

Compare Proportions
- Proportion Test
- Chi-Square Test

Multiple Input Variables

Compare Proportions
- Chi-Square Test

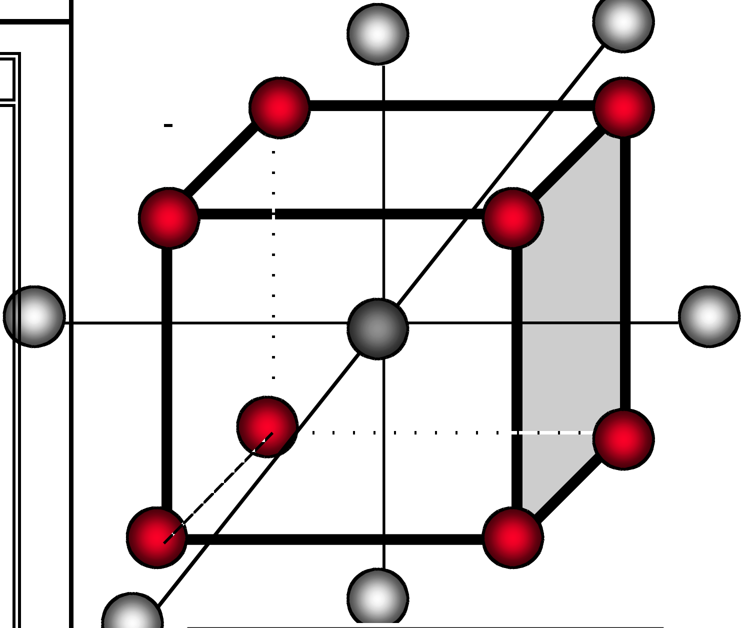
Screening Experiments
- Full Factorial
- Fractional Factorial

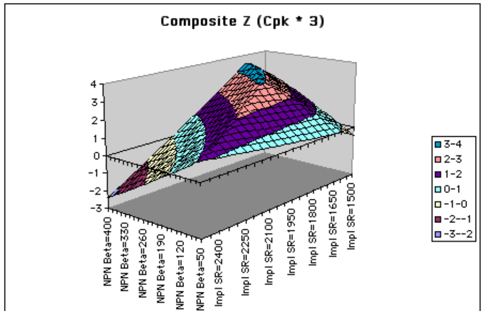
Analysis of Experiments
- ANOVA
- Multiple Linear Regression

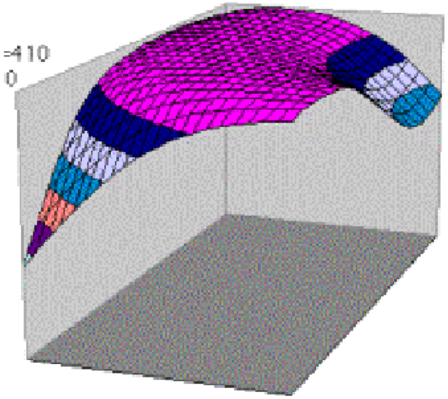
Response Surface Modeling
- Box-Behnken Designs
- Central Composite Designs
- Multiple Linear Regression
- Stepwise Regression
- Contour Plots
- 3 D Mesh Plots

Model Response Distribution
- Monte Carlo Simulation
- Generation of System Moments

Optimization
- Optimization of Expected Value:
- Linear Programming
- Non Linear Programming
- Yield Surface Modeling™







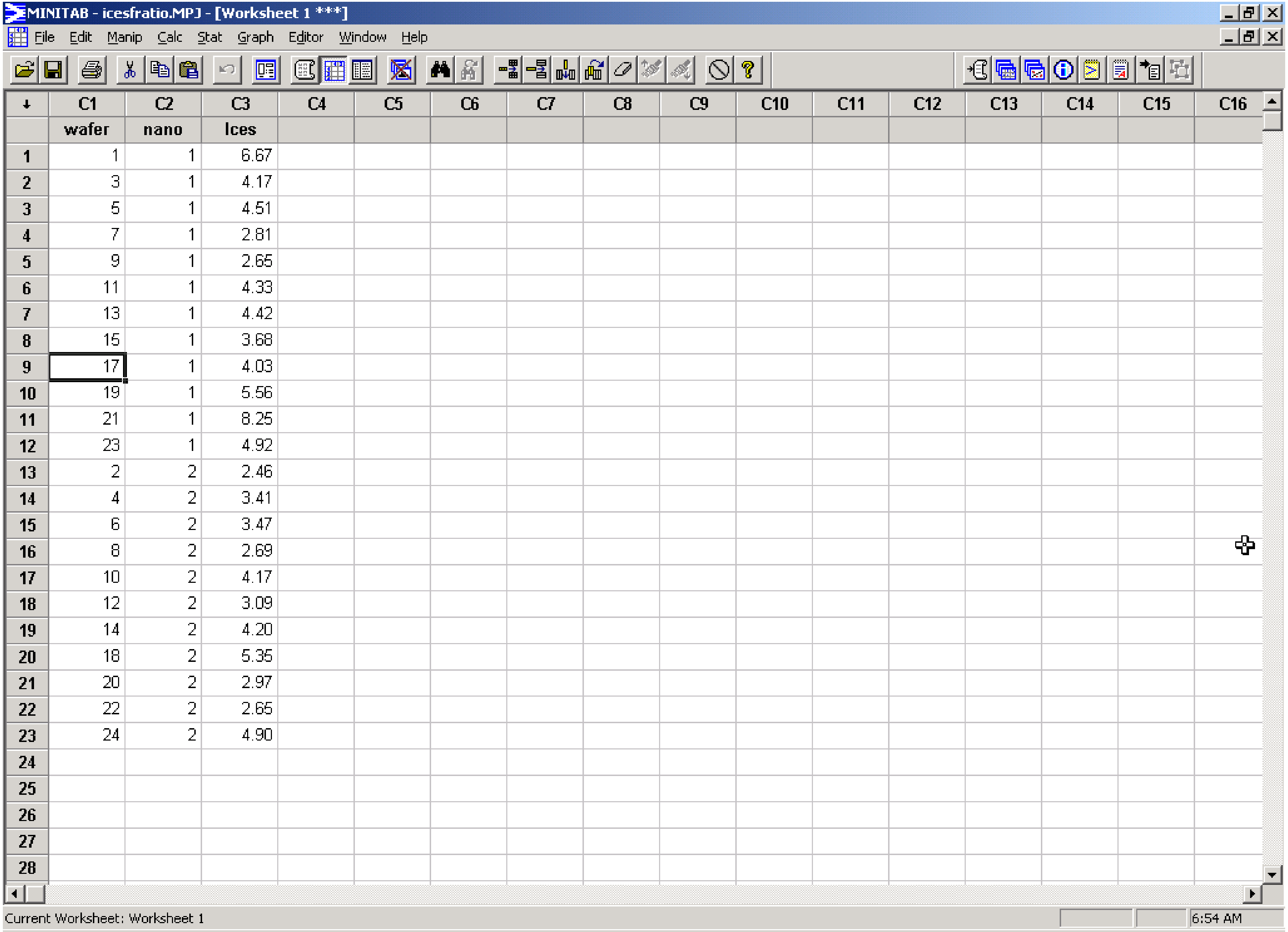
F RATIO -
USED TO TEST IF TWO VARIANCES ARE EQUAL.

$$F = \frac{S_1^2}{S_2^2} \text{ or } \frac{S_2^2}{S_1^2}$$

(put larger sample variance in numerator)

TEST FOR EQUAL VARIANCES - ICES EXAMPLE

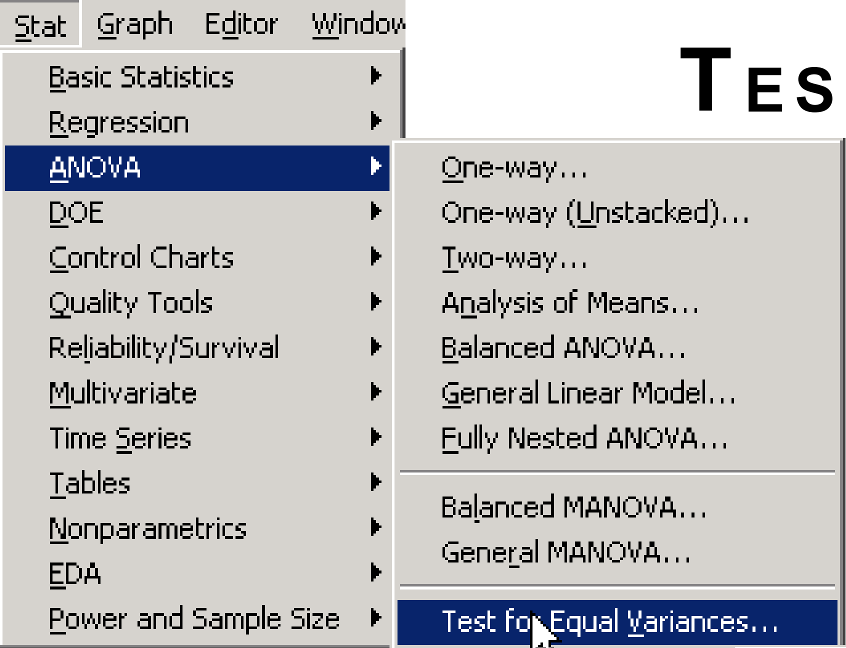
open icesfratio.mpj

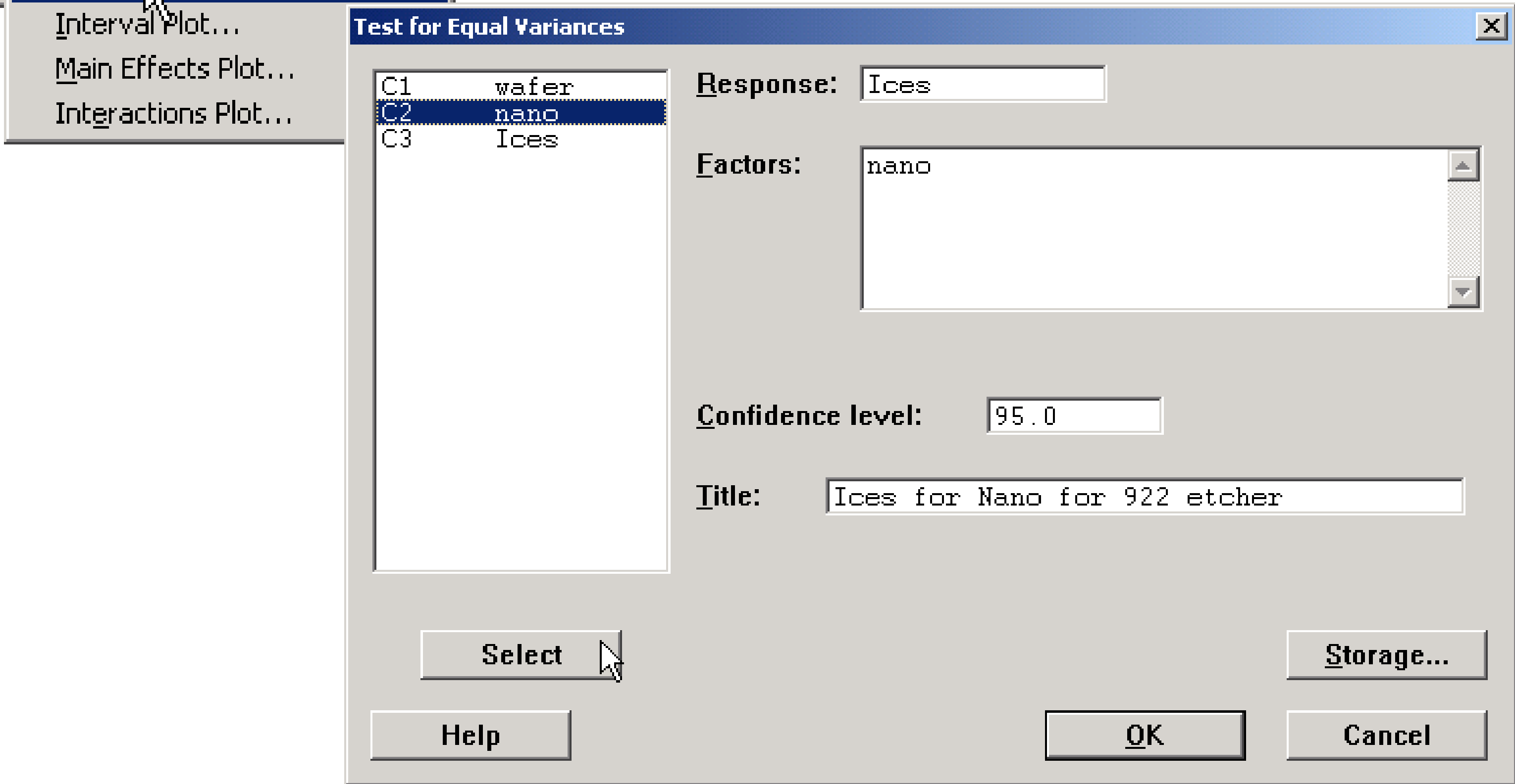


↓	C1	C2	C3
	wafer	nano	Ices
1	1	1	6.67
2	3	1	4.17
3	5	1	4.51
4	7	1	2.81
5	9	1	2.65
6	11	1	4.33
7	13	1	4.42
8	15	1	3.68
9	17	1	4.03
10	19	1	5.56
11	21	1	8.25
12	23	1	4.92
13	2	2	2.46
14	4	2	3.41
15	6	2	3.47
16	8	2	2.69
17	10	2	4.17
18	12	2	3.09
19	14	2	4.20
20	18	2	5.35
21	20	2	2.97
22	22	2	2.65
23	24	2	4.90
24			
25			
26			
27			
28			

TEST FOR EQUAL VARIANCES

Ices vs (Nano or 922 Etcher)





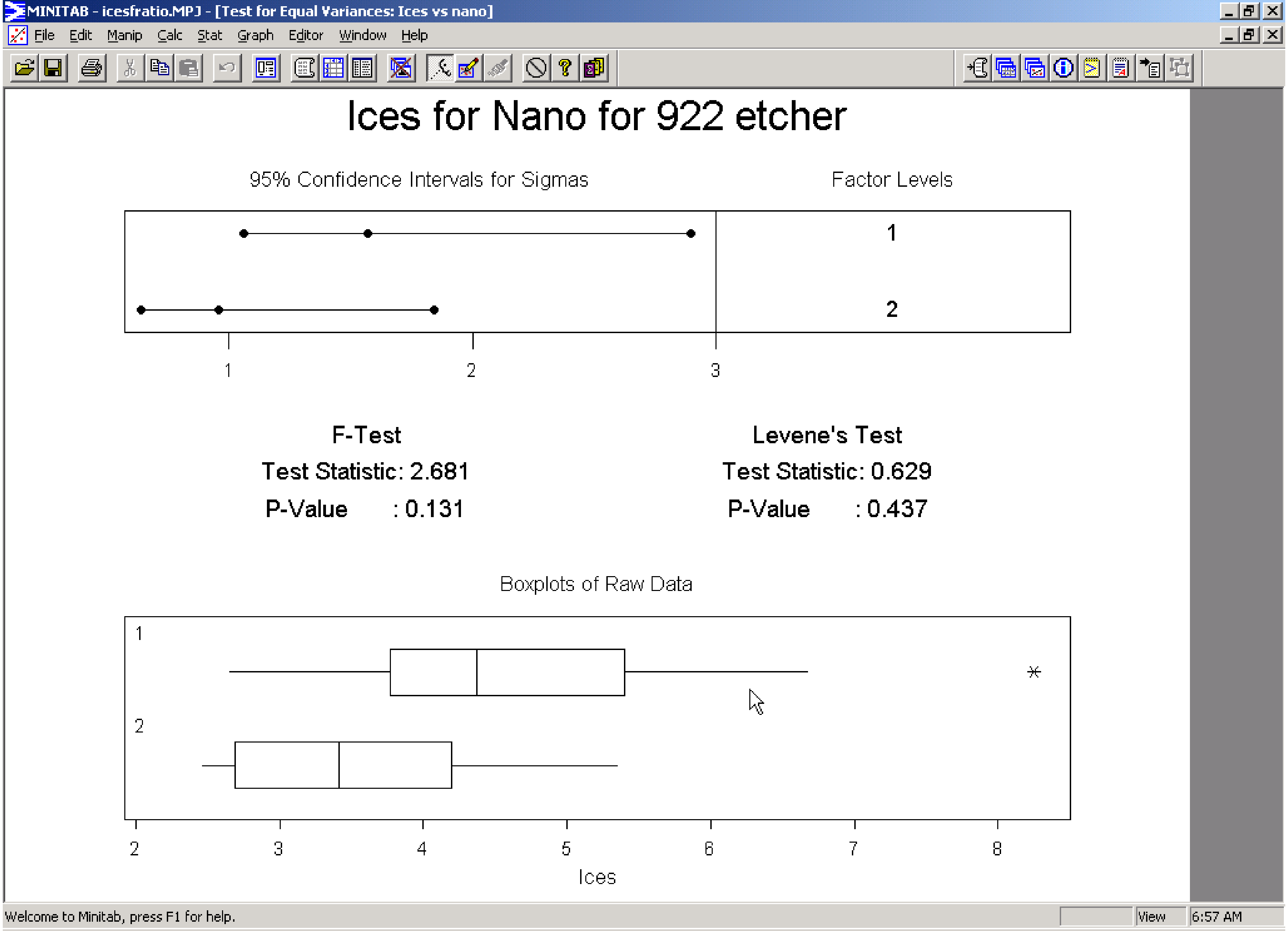
C1 wafer
C2 nano
C3 Ices

Response: Ices

Factors: nano

Confidence level: 95.0

Title: Ices for Nano for 922 etcher



Ices for Nano for 922 etcher

F-Test

Test Statistic: 2.681

P-Value : 0.131

Levene's Test

Test Statistic: 0.629

P-Value : 0.437

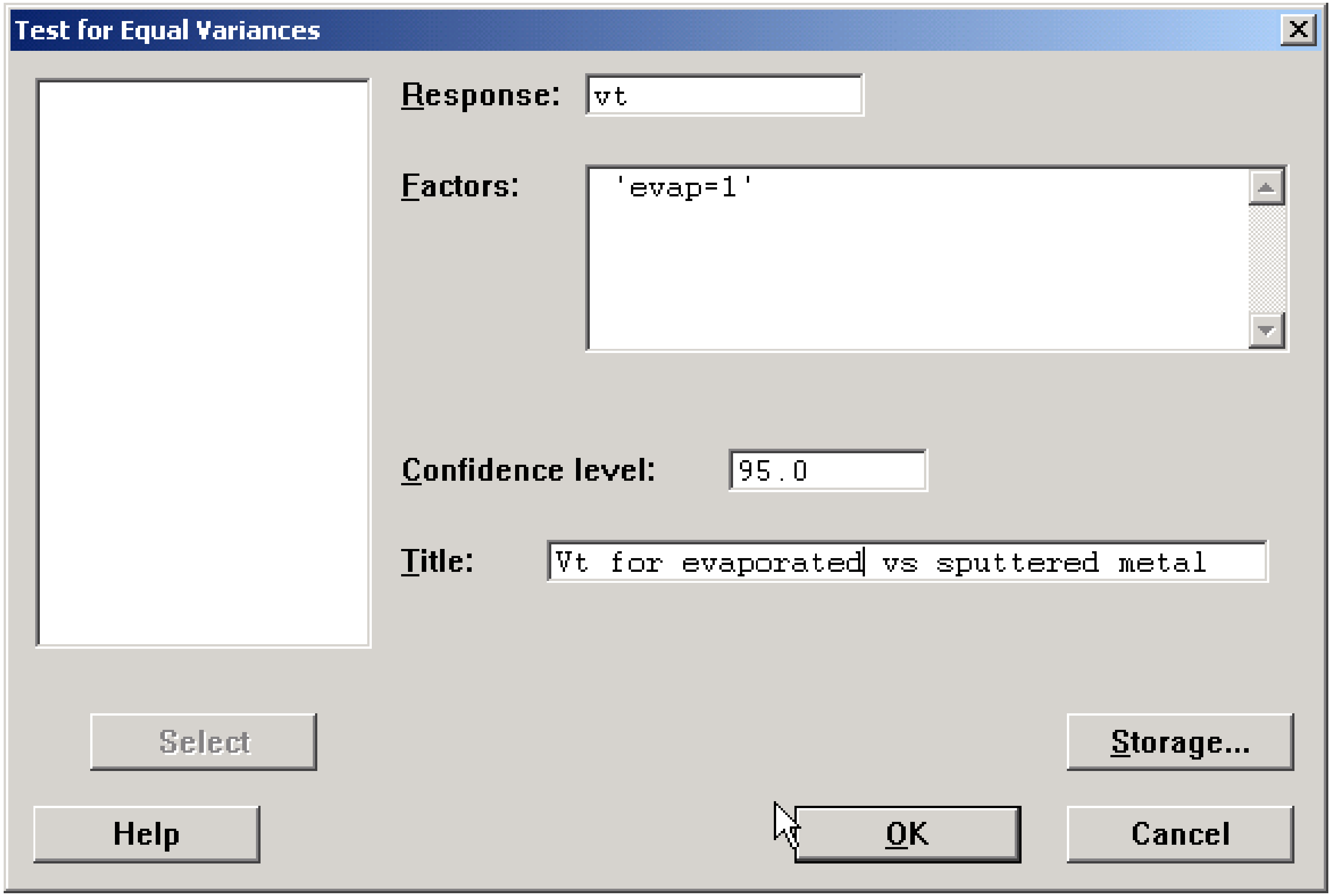
Welcome to Minitab, press F1 for help.

THRESHOLD VOLTAGE
N–CHANNEL

Evaporated Metal	Sputtered Metal
.707	.791
.645	.764
.682	.782
.692	.788

TEST FOR EQUAL VARIANCES

VT VERSUS METAL DEPOSITION EXAMPLE



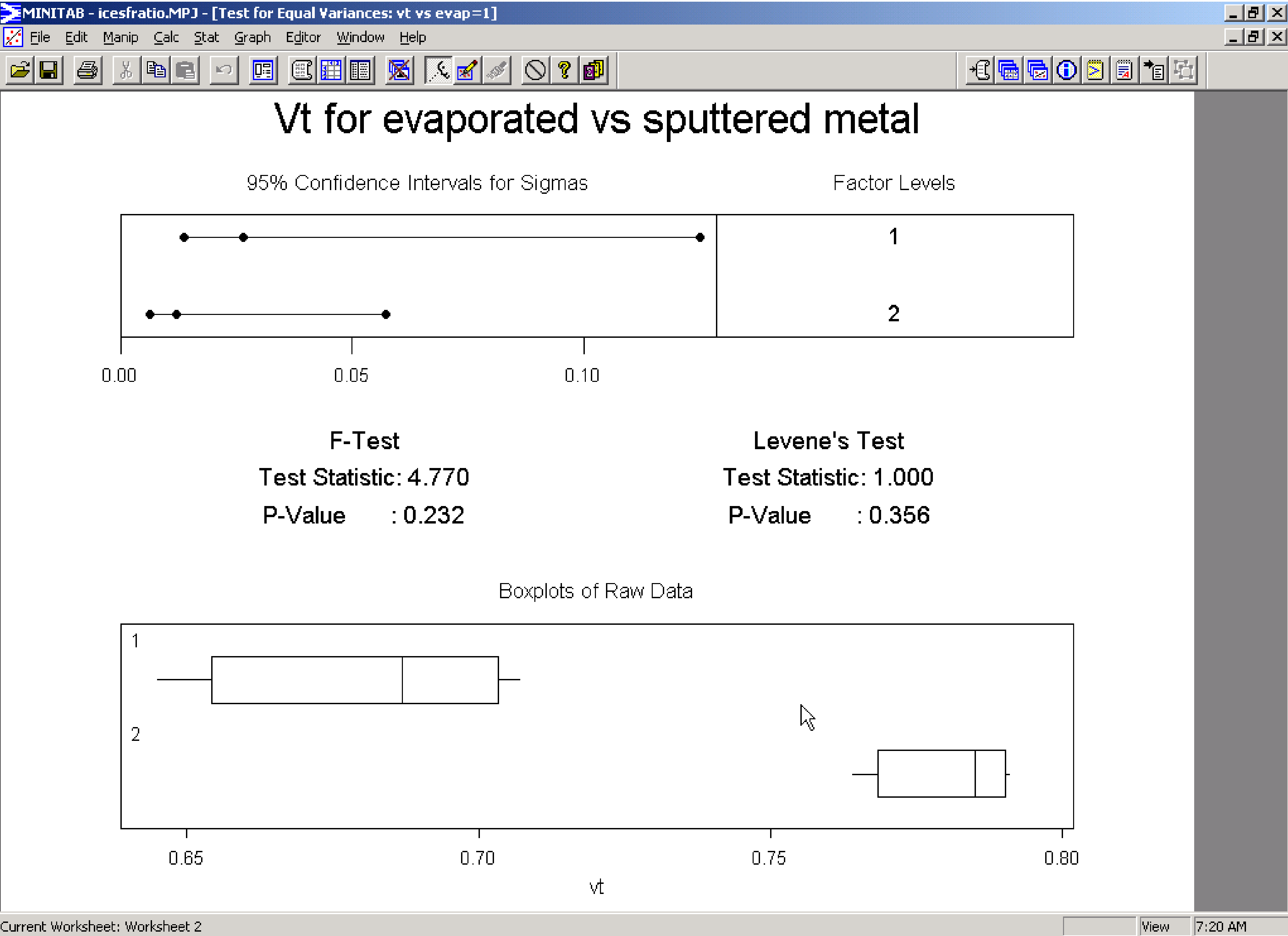
Response: vt

Factors: 'evap=1'

Confidence level: 95.0

Title: Vt for evaporated vs sputtered metal

TEST FOR EQUAL VARIANCES - VT VS METAL DEPOSITION EXAMPLE



Vt for evaporated vs sputtered metal

F-Test

Test Statistic: 4.770

P-Value : 0.232

Levene's Test

Test Statistic: 1.000

P-Value : 0.356

Student's t-test

… was developed

by W.S. Gosset (aka "Student"),

as an approach for testing the

quality of beer at a

GUINNESS

brewery.

DIFFERENCE BETWEEN MEANS OF TWO POPULATIONS

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{S_1^2 + S_2^2}{N}}}$$

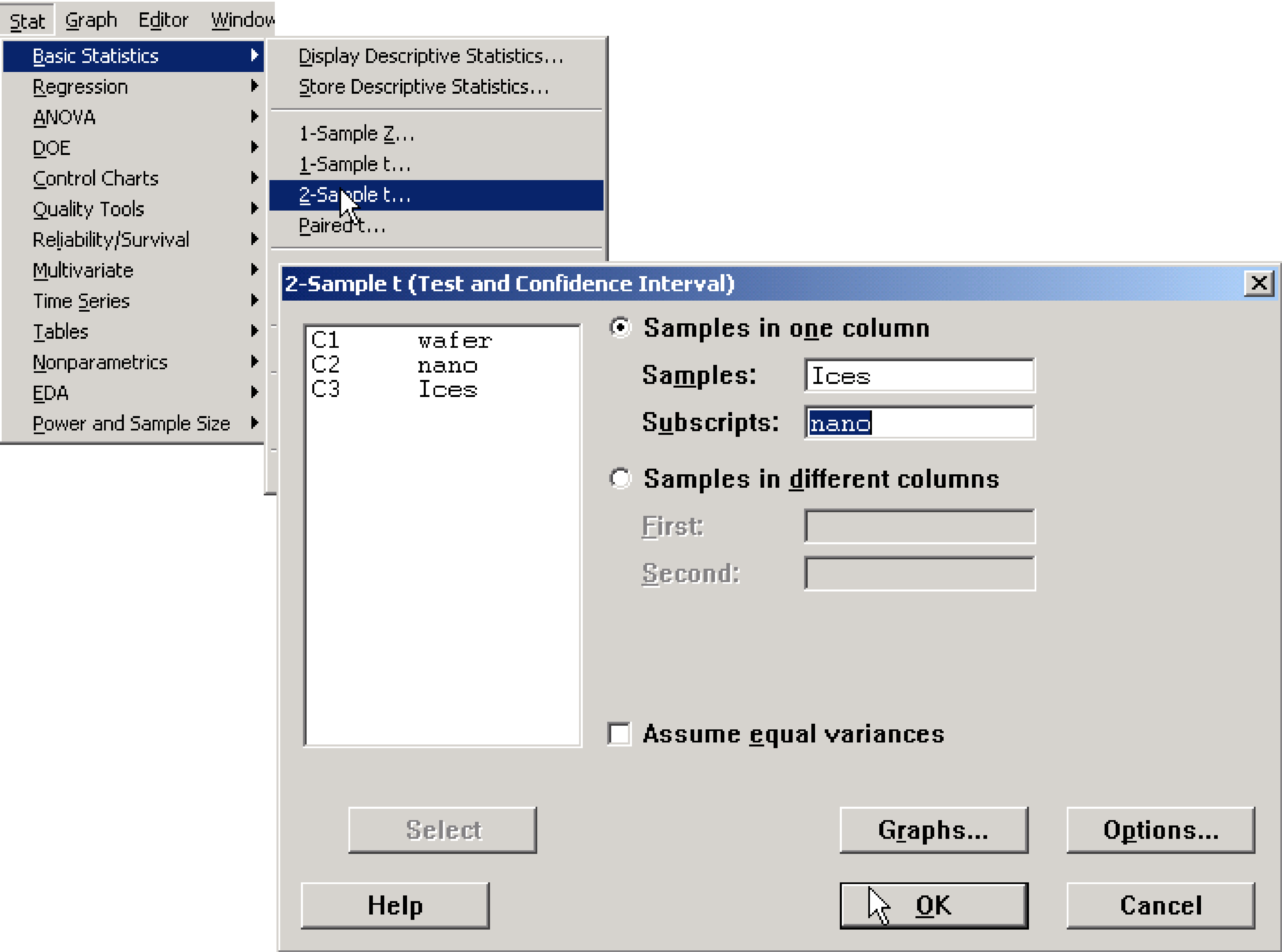
for $N_1 = N_2 = N$

OR

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2}\left(\dfrac{1}{N_1} + \dfrac{1}{N_2}\right)}}$$

$df = N_1 + N_2 - 2$

for N_1 ° N_2



2-Sample t (Test and Confidence Interval)

C1 wafer
C2 nano
C3 Ices

● Samples in one column

Samples: Ices

Subscripts: nano

○ Samples in different columns

First:

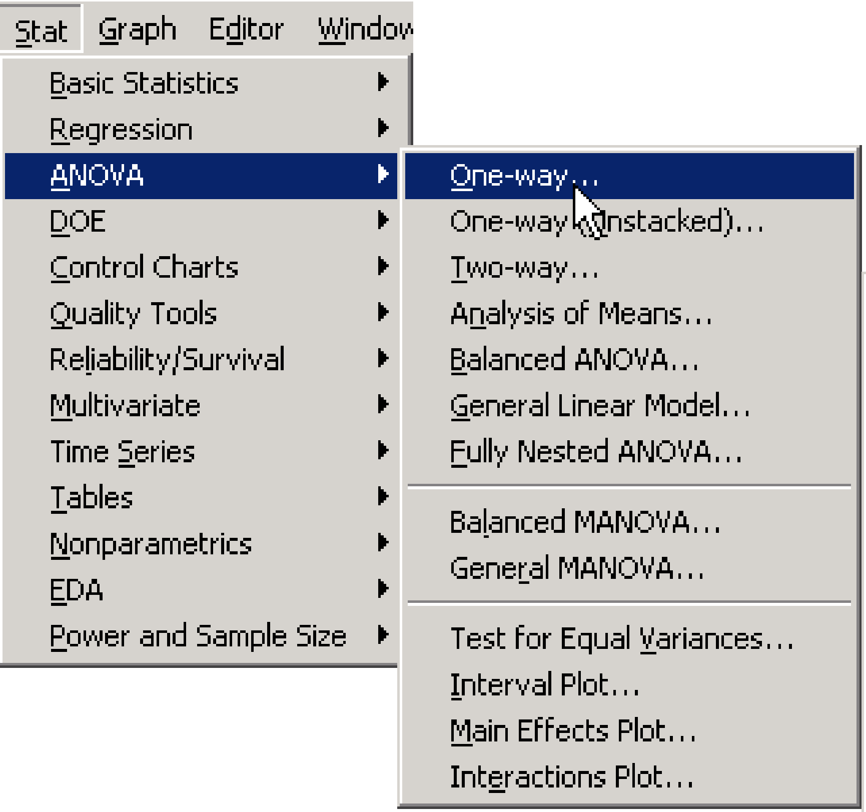
Second:

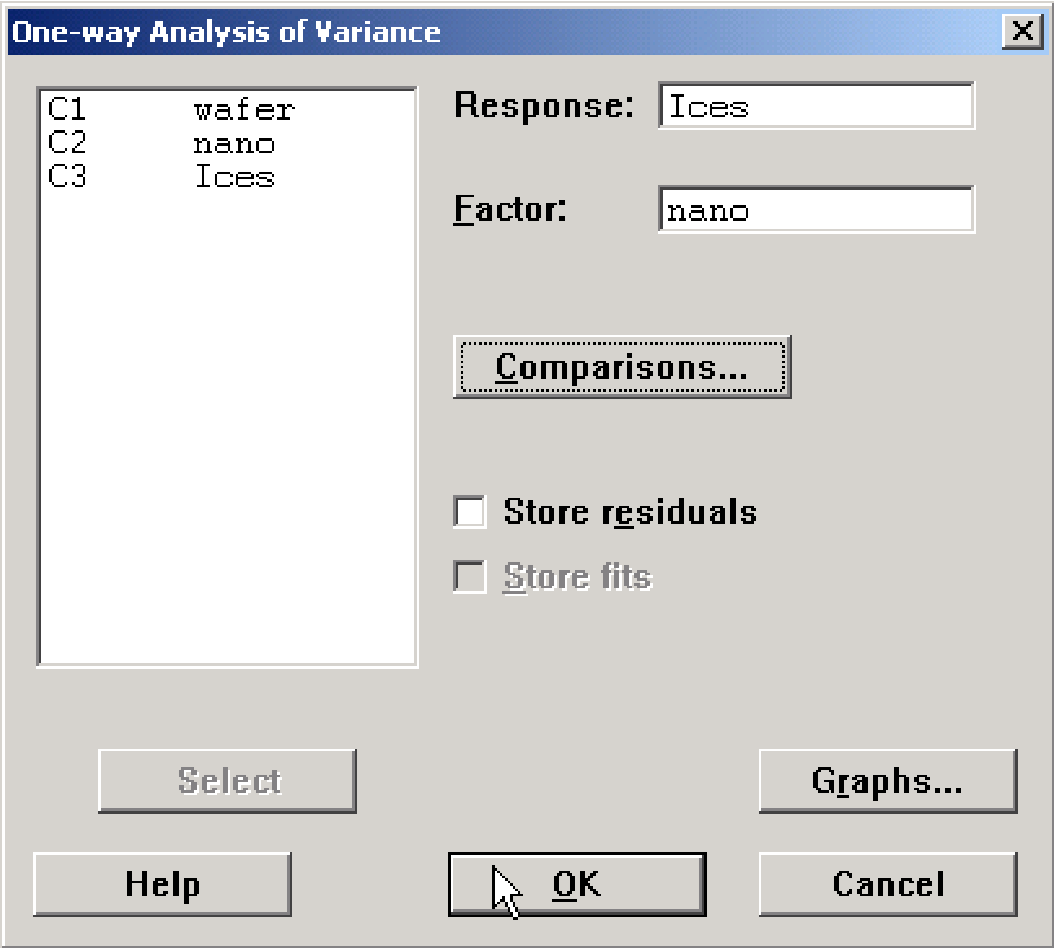
☐ Assume equal variances

Two-Sample T -Test and CI: Ices, nano

```
Two-sample T for Ices

nano         N      Mean     StDev   SE Mean
1           12      4.67      1.57      0.45
2           11     3.578     0.958      0.29

Difference = mu (1) - mu (2)
Estimate for difference:  1.088
95% CI for difference: (-0.040, 2.217)
T-Test of difference = 0 (vs not =): T-Value = 2.03
                                P-Value = 0.058  DF = 18
```

ANOVA




One-way Analysis of Variance

C1 wafer
C2 nano
C3 Ices

Response: Ices

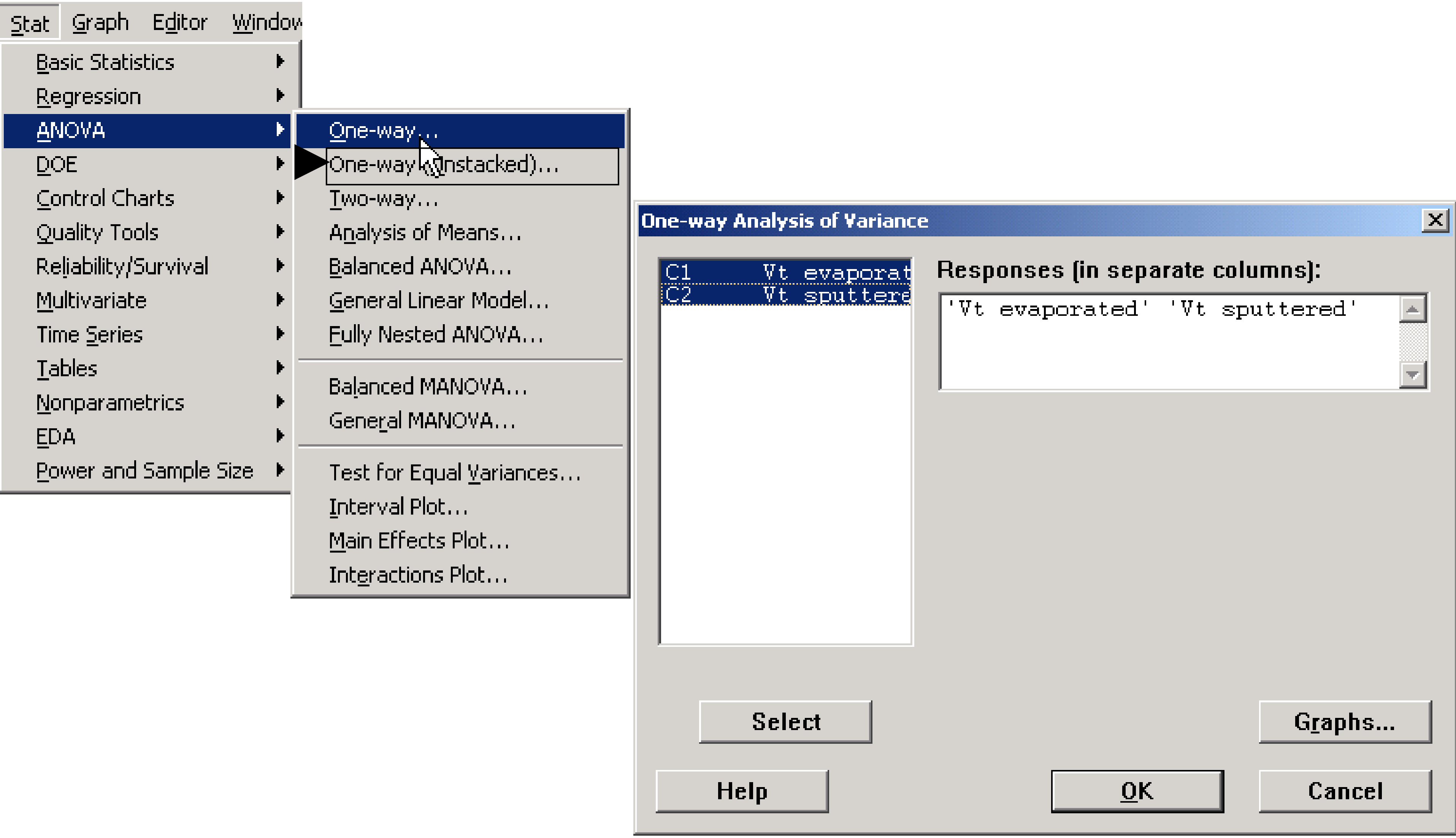
Factor: nano

☐ Store residuals
☐ Store fits

One-way ANOVA: Ices versus nano

```
Analysis of Variance for Ices
Source      DF        SS        MS        F        P
nano         1      6.80      6.80     3.94    0.060
Error       21     36.23      1.73
Total       22     43.03
                                    Individual 95% CIs For Mean
                                    Based on Pooled StDev
Level        N      Mean     StDev  ---+----+----+----+
1           12     4.667     1.568                (----*----)
2           11     3.578     0.958  (-----*----)
                                    ---+----+----+----+
Pooled StDev =     1.313                3.20      4.00      4.80      5.60
```

One Way Unstacked



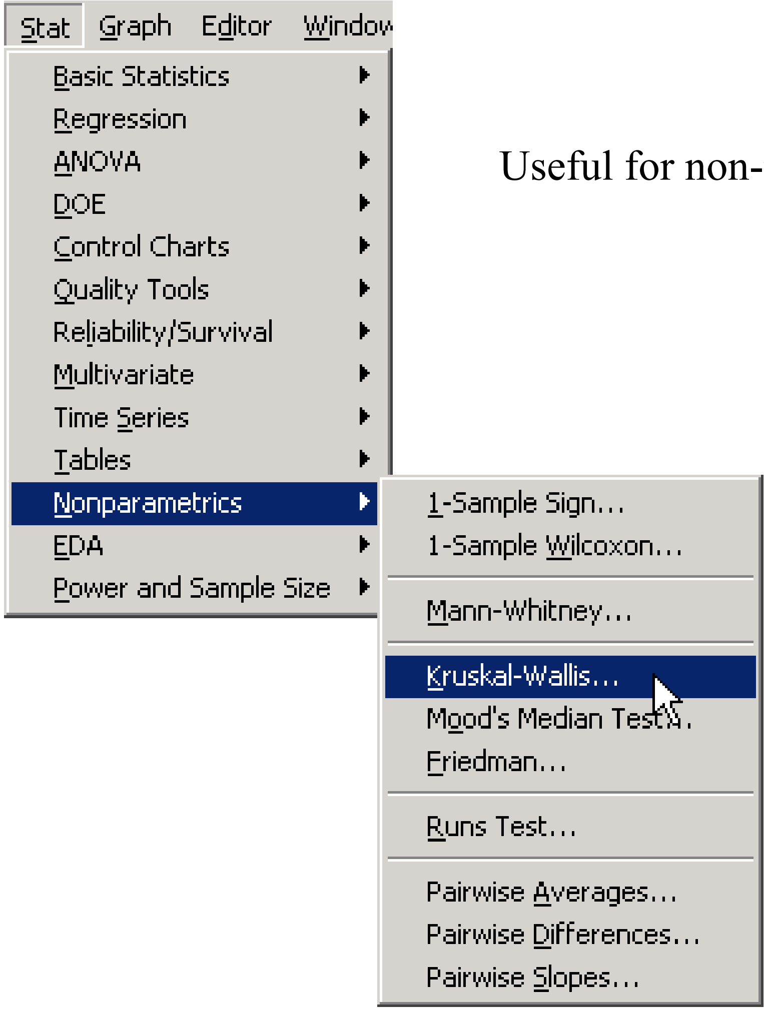
Responses (in separate columns): 'Vt evaporated' 'Vt sputtered'

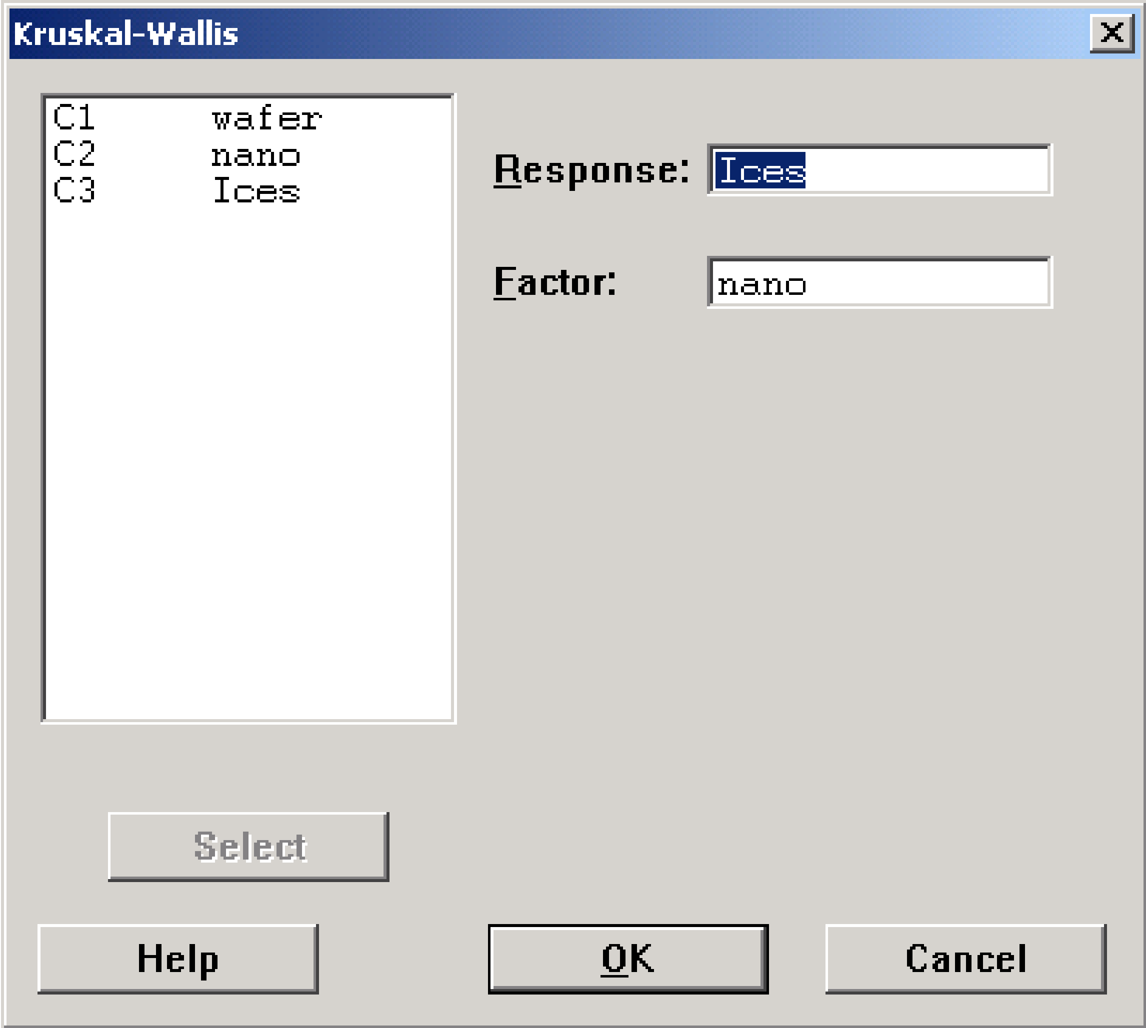
One-way ANOVA: Vt evaporated, Vt sputtered

```
Analysis of Variance
Source     DF        SS        MS        F        P
Factor      1  0.019900  0.019900    47.16    0.000
Error       6  0.002532  0.000422
Total       7  0.022432
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  ---------+---------+---------+-------
Vt evapo    4   0.68150   0.02641  (---*---)
Vt sputt    4   0.78125   0.01209                     (---*---)
                                   ---------+---------+---------+-------
Pooled StDev =  0.02054                  0.700     0.750     0.800
```

Alternative to ANOVA - Kruskal-Wallis uses Medians rather than Means

Useful for non-normal distributions (although ANOVA is rather robust)





C1 wafer
C2 nano
C3 Ices

Response: Ices

Factor: nano

Kruskal-Wallis Test: Ices versus nano

```
Kruskal-Wallis Test on Ices

nano          N    Median    Ave Rank         Z
1            12     4.375        14.5      1.85
2            11     3.410         9.3     -1.85
Overall      23                  12.0

H = 3.41  DF = 1  P = 0.065
H = 3.41  DF = 1  P = 0.065 (adjusted for ties)
```

SIGNIFICANCE TEST FOR COMPARING TWO PROPORTIONS

$H_O : P_1 = P_2$

$H_A : P_1 \neq P_2$

$$Z = \frac{P_1 - P_2}{\sqrt{P_T (1 - P_T) \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

P_1 = Proportion bad in group 1

P_2 = Proportion bad in group 2

P_T = Proportion bad overall : $\frac{\text{Total bad}}{\text{Total tested}}$

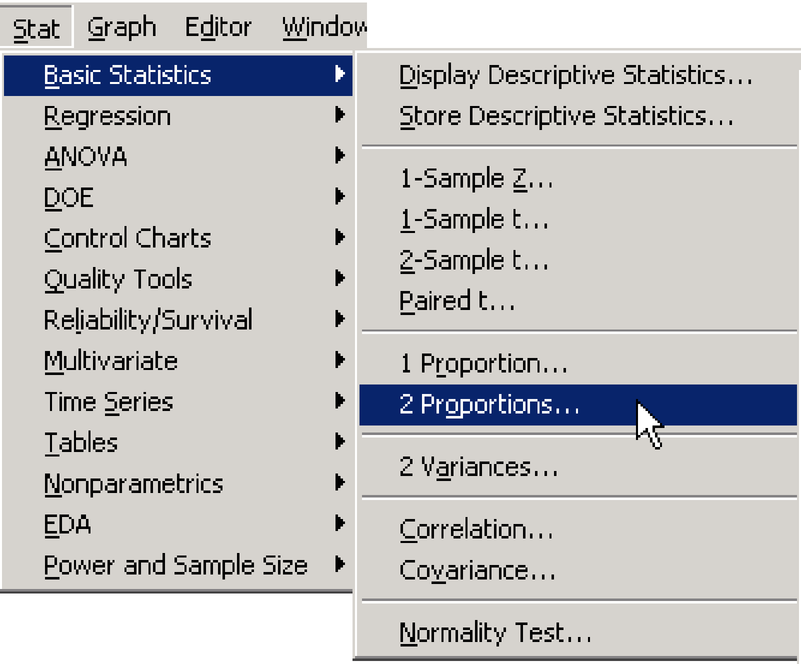
N_1 = Total tested in group 1

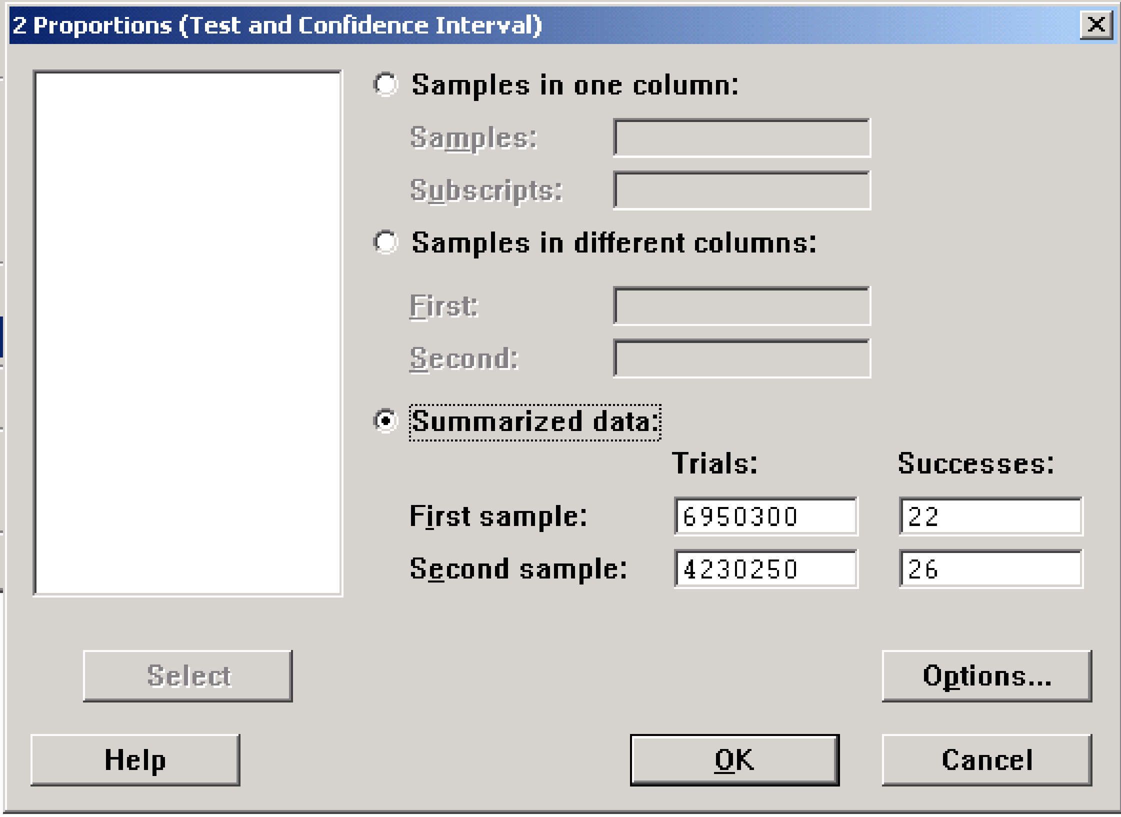
N_2 = Total tested in group 2

Requirement : Total bad $\geq$ 5, total good $\geq$ 5

(Use Z table for alpha risk)

Is this year's 3 ppm failure rate (22 failures out of 6,950,300) better than last year's 6 ppm failure rate? (26 failures out of 4,230,250)



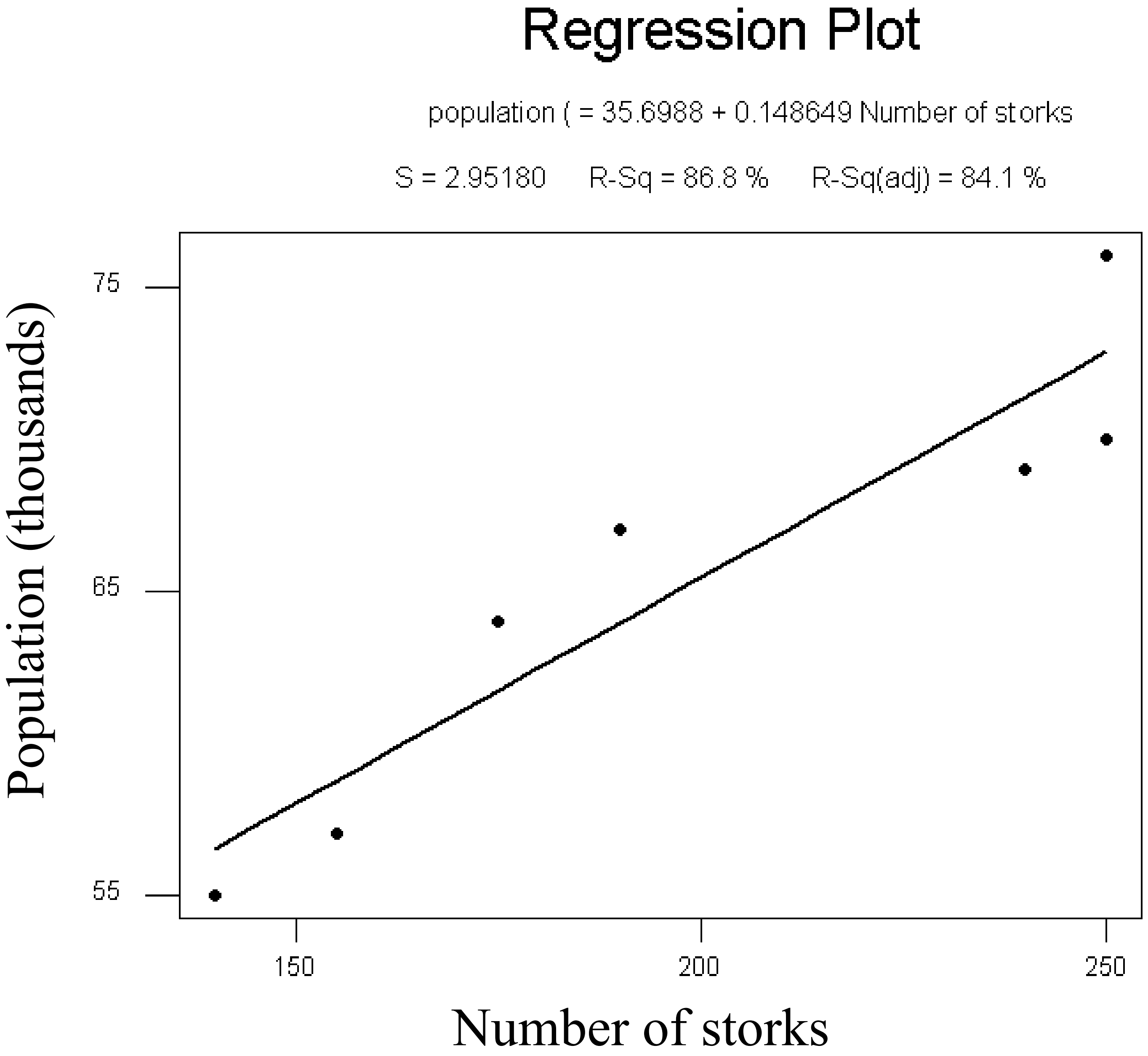


● Summarized data:

	Trials:	Successes:
First sample:	6950300	22
Second sample:	4230250	26

Test and CI for Two Proportions

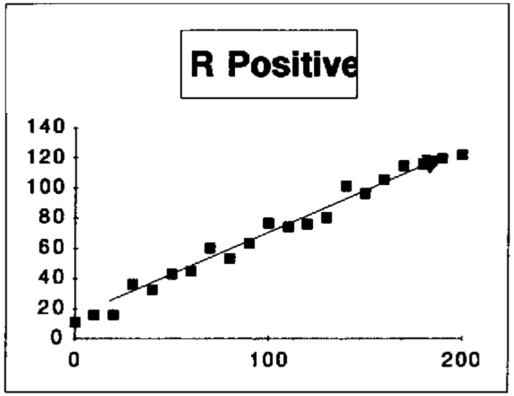
```
Sample        X       N  Sample p
1            22   7E+06  0.000003
2            26   4E+06  0.000006

Estimate for p(1) - p(2):  -0.00000298088
95% CI for p(1) - p(2):  (-0.00000568842, -0.000000273337)
Test for p(1) - p(2) = 0 (vs not = 0):  Z = -2.16  P-Value = 0.031
```

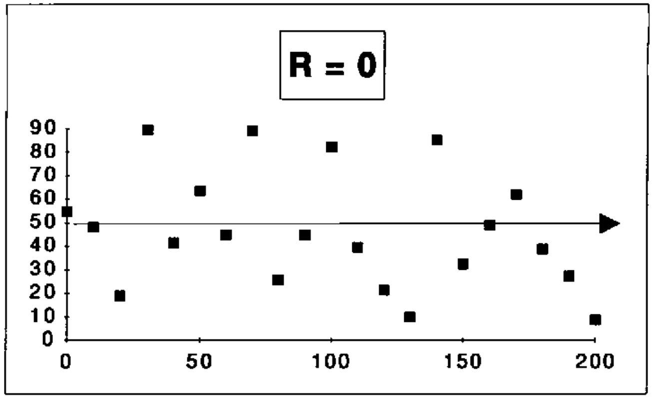

Regression Plot
population (= 35.6988 + 0.148649 Number of storks
S = 2.95180
R-Sq = 86.8 %
R-Sq(adj) = 84.1 %
75
65
55
150
200
250
Population (thousands)
Number of storks

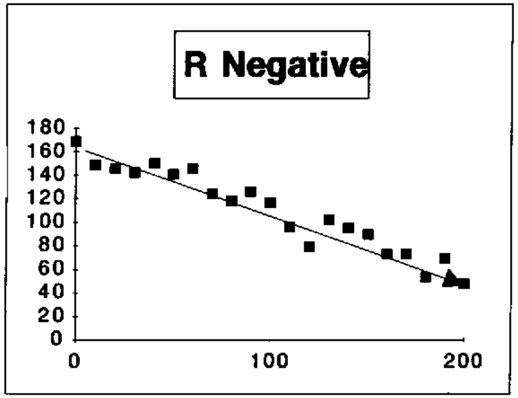
CORRELATION AND CAUSATION

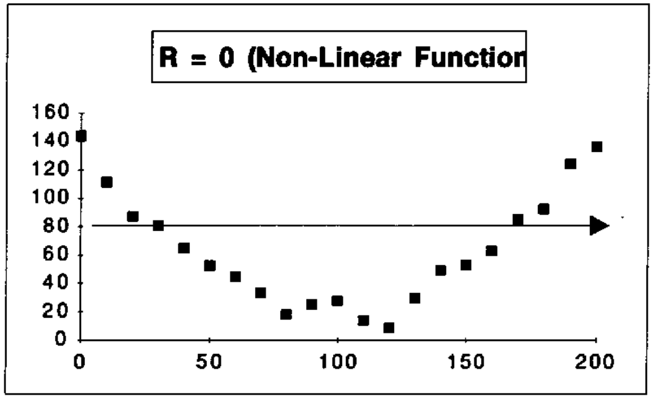
Some cautions must be noted at this point. First, you cannot determine the cause of the relationship from the correlation coefficient. Two variables may be highly correlated for one of three reasons: (1) X causes Y, (2) Y causes X, or (3) both X and Y are caused by some third variable. A well known story that illustrates the danger of inferring causation from a correlation coefficient between the number of storks and the number of births in European cities (that is, the more storks, the more births). Instead of issuing a dramatic announcement supporting the mythical powers of storks, further investigation was carried out. It was found that storks nest in chimneys, which in turn led to the conclusion that a third variable was responsible for the relationship between storks and births - size of city. Large cities had more people, and hence more births; and more houses, and hence more chimneys, and hence more storks. Thus, attribution causality is a logical or scientific problem, not a statistical one.

CORRELATION COEFFICIENTS









"THE MEANING OF R²"

Assume deterministic model:

$$Y = f\left(x_1, x_2, \ldots., x_n\right)$$

Then the variance of Y is: (making some assumptions)

$$S_Y^2 = \left(\frac{\partial Y}{\partial X_1} \cdot S_{x_1}\right)^2 + \left(\frac{\partial Y}{\partial X_2} \cdot S_{x_2}\right)^2 + \ldots. + \left(\frac{\partial Y}{\partial X_n} \cdot S_{x_n}\right)^2$$

The % of variance of Y due to X_1 is:

$$\frac{\left(\frac{\partial Y}{\partial X_1} \cdot S_{x_1}\right)^2}{S_Y^2} \text{ X } 100\%$$

In regression, we assume $Y = bx_1 + a + \text{error}$

$$R^2 = \left(\frac{b\,S_{x_1}}{S_Y}\right)^2 \Rightarrow$$ **The proportion of the variance Y**

due to X_1, assuming Y is a linear function of X_1.

CRITICAL VALUES OF THE PEARSON

df (= N - 2; N =number of pairs)	Level of significance for one-tailed test			
	.05	.025	.01	.005
	Level of significance for two-tailed test			
	.10	.05	.02	.01
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4	.729	.811	.882	.917
5	.669	.754	.833	.874
6	.622	.707	.789	.834
7	.582	.666	.750	.798
8	.549	.632	.716	.765
9	.521	.602	.685	.735
10	.497	.576	.658	.708
11	.476	.553	.634	.684
12	.458	.532	.612	.661
13	.441	.514	.592	.641
14	.426	.497	.574	.623
15	.412	.482	.558	.606

CRITICAL VALUES OF THE PEARSON (Cont’d)

	Level of significance for one-tailed test			
df (= N - 2; N =number of pairs)	.05	.025	.01	.005
	Level of significance for two-tailed test			
	.10	.05	.02	.01
16	.400	.468	.542	.590
17	.389	.456	.528	.575
18	.378	.444	.516	.561
19	.369	.433	.503	.549
20	.360	.423	.492	.537
21	.352	.413	.482	.526
22	.344	.404	.472	.515
23	.337	.396	.462	.505
24	.330	.388	.453	.496
25	.323	.381	.445	.487
26	.317	.374	.437	.479
27	.311	.367	.430	.471
28	.306	.361	.423	.463
29	.301	.355	.416	.456
30	.296	.349	.409	.449

CRITICAL VALUES OF THE PEARSON (Cont'd)

	Level of significance for one-tailed test			
df	.05	.025	.01	.005
(= N - 2;	Level of significance for two-tailed test			
N =number of pairs)	.10	.05	.02	.01
35	.275	.325	.381	.418
40	.257	.304	.358	.393
45	.243	.288	.338	.372
50	.231	.273	.322	.354
60	.211	.250	.295	.325
70	.195	.232	.274	.302
80	.183	.217	.256	.283
90	.173	.205	.242	.267
100	.164	.195	.230	.254

TESTING THE SIGNIFICANCE OF THE CORRELATION COEFFICIENT

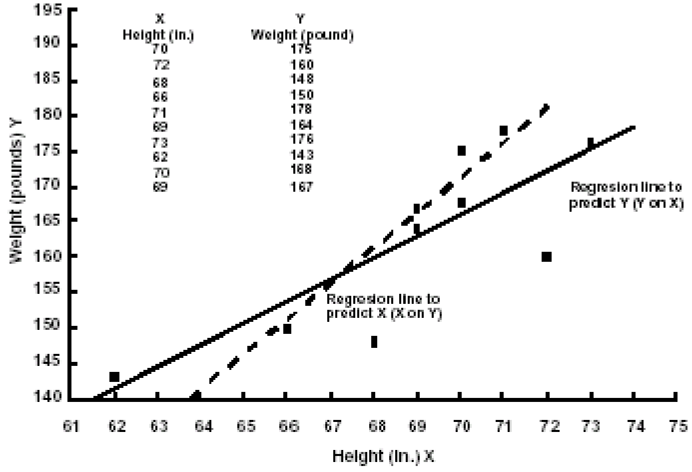
$$t = \frac{r\sqrt{N - 2}}{\sqrt{1 - r^2}}$$

$$F = t^2 = (N - 2)\left(\frac{r^2}{1 - r^2}\right)$$

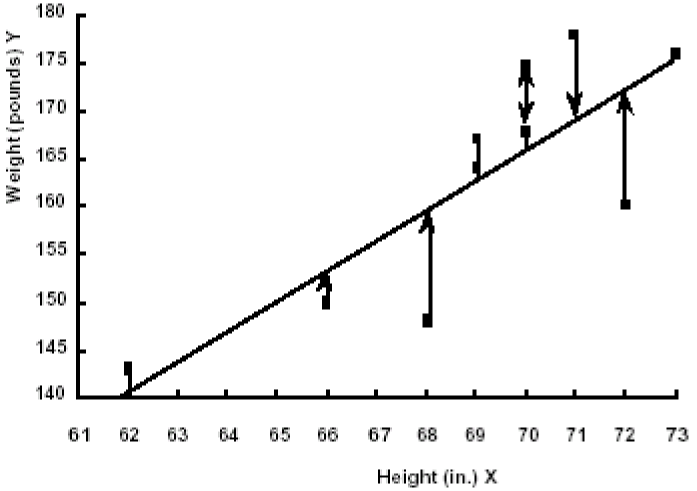
Where N = number of pairs of scores

Linear Regression and Prediction

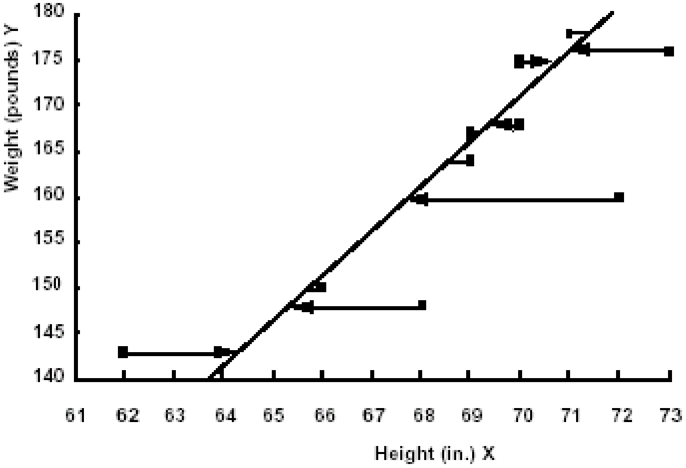
Regression lines when height scores (X) are plotted against weight scores (Y) for ten adult males.



X Height (In.)	Y Weight (pound)
70	175
72	160
68	148
66	150
71	178
69	164
73	176
62	143
70	168
69	167



Regression line of Y on X showing extent of error (difference between actual weight score and predicted weight score).



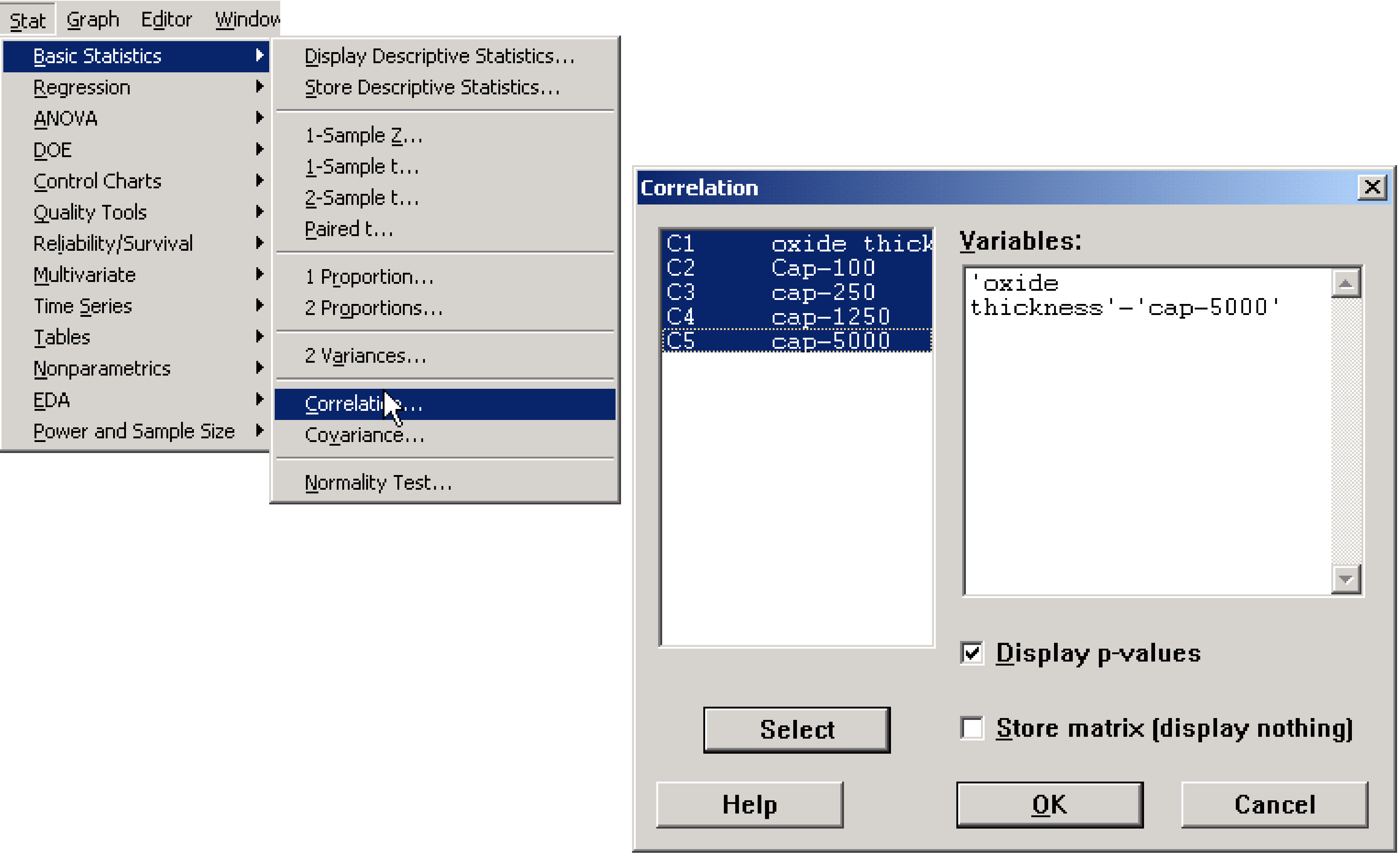
Regression line of X on Y showing extent of error (difference between actual height score and predicted height score).

“LINEAR” CURVE FITS
(BIVARIATE)

	Transform
$y = Ae^{BX}$	$\ln y = \ln (A) + BX$
$y = AX^{B}$	$\ln y = \ln(A) + B \text{ in}(X)$
$y = A + (B/X)$	$y = A + B\ (1/X)$
$y = 1/(A + BX)$	$1/y = A + BX$
$y = X/(A + BX)$	$1/y = B + A\ (1/X)$

	OXIDE THICKNESS X_1	CAP-100 Y_1	CAP-250 Y_2	CAP-1250 Y_3	CAP-5000
1	2500	9.2880	23.0790	115.6000	449.0700
2	1620	14.3560	35.7390	178.6900	691.9000
3	4887	4.7830	11.8490	58.5980	232.2700
4	3497	6.4890	16.4930	81.5360	264.6800
5	3472	6.6840	16.2780	82.1160	319.2200
6	3420	6.8550	16.8500	84.3430	327.9800
7	4880	4.7730	11.6380	58.5940	232.1900
8	4469	5.2200	12.9390	64.6080	256.0600
9	1624	14.3720	35.7140	178.7700	692.3900
10	4471	5.2550	13.0080	64.5610	255.8400
11	2611	9.0290	22.2280	111.1400	431.9700
12	1625	14.3020	35.2110	178.0600	688.0700
13	1640	13.9830	34.8060	174.3900	675.6900
14	2613	8.8890	22.0900	110.5400	430.0300
15	4472	5.2440	12.7420	64.4420	253.8700
16	1636	14.1800	35.1840	176.0000	681.8100
17	3486	6.6570	16.3820	82.6570	321.5200
18	2470	9.4120	23.3500	116.8800	454.5500
19	2610	8.7850	22.0730	110.4300	429.3700
20	4878	4.7660	11.8420	58.6120	232.3700
21	2473	9.4520	23.2860	116.6300	452.8400
22	1641	14.1270	35.1070	175.6700	680.8900
23	3424	6.8440	16.8820	83.9120	323.3900
24	3432	6.2570	16.6640	83.5590	324.8200

OXIDE CAPACITANCE EXAMPLE



C1 oxide thick
C2 Cap-100
C3 cap-250
C4 cap-1250
C5 cap-5000

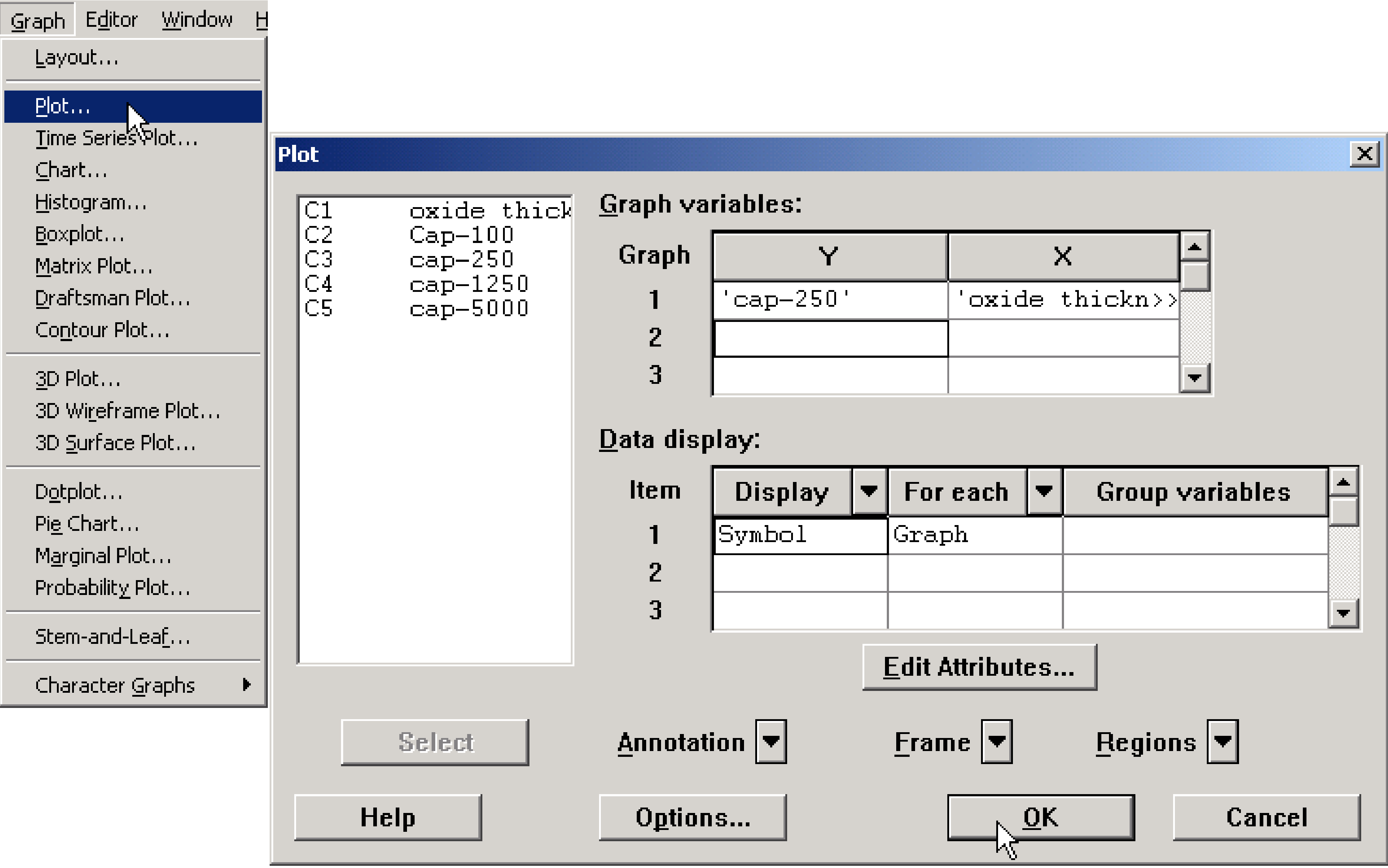
Variables: 'oxide thickness'-'cap-5000'

☑ Display p-values

☐ Store matrix (display nothing)

Correlations: oxide thickness, Cap-100, cap-250, cap-1250, cap-5000

```
       oxide th  Cap-100  cap-250 cap-1250
Cap-100  -0.948
          0.000

cap-250  -0.950    0.999
          0.000    0.000

cap-1250 -0.950    1.000    1.000
          0.000    0.000    0.000

cap-5000 -0.943    0.998    0.998    0.998
          0.000    0.000    0.000    0.000
```



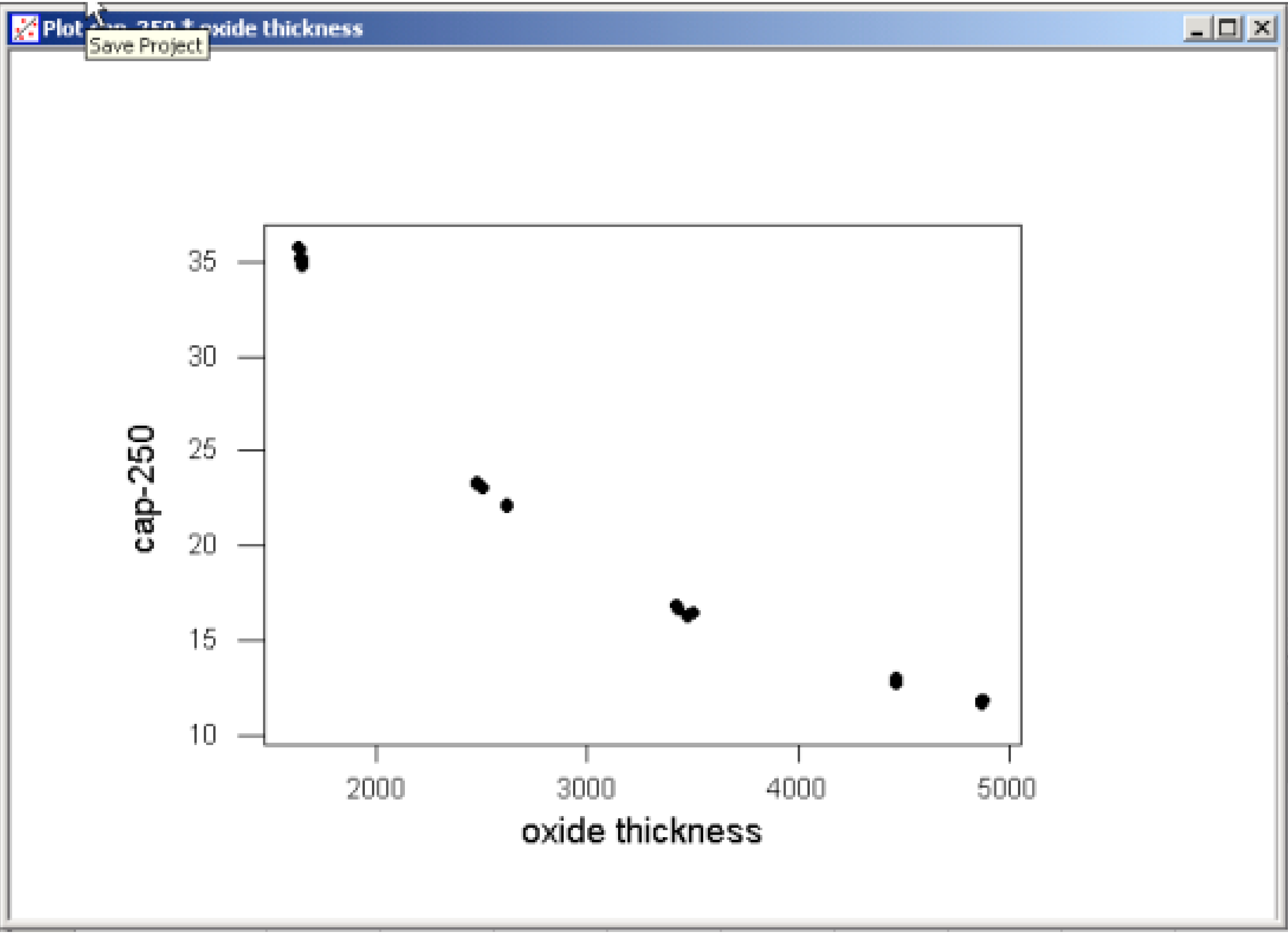
C1 oxide thick
C2 Cap-100
C3 cap-250
C4 cap-1250
C5 cap-5000

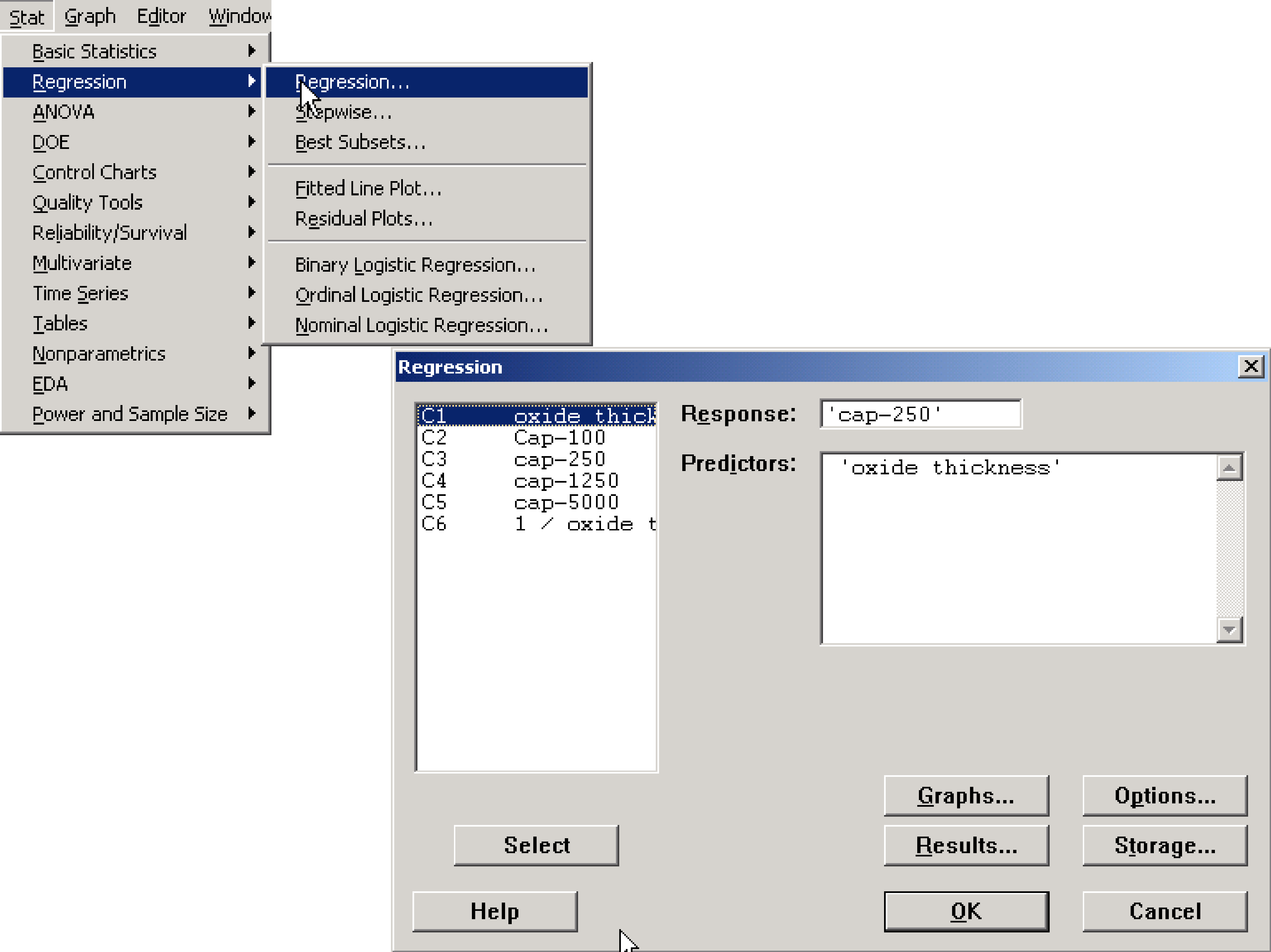
Graph variables:

Graph	Y	X
1	'cap-250'	'oxide thickn>>
2		
3		

Data display:

Item	Display	For each	Group variables
1	Symbol	Graph	
2			
3			


Save Project
oxide thickness
cap-250
35
30
25
20
15
10
2000
3000
4000
5000
oxide thickness



C1 oxide thick
C2 Cap-100
C3 cap-250
C4 cap-1250
C5 cap-5000
C6 1 / oxide t

Response: 'cap-250'

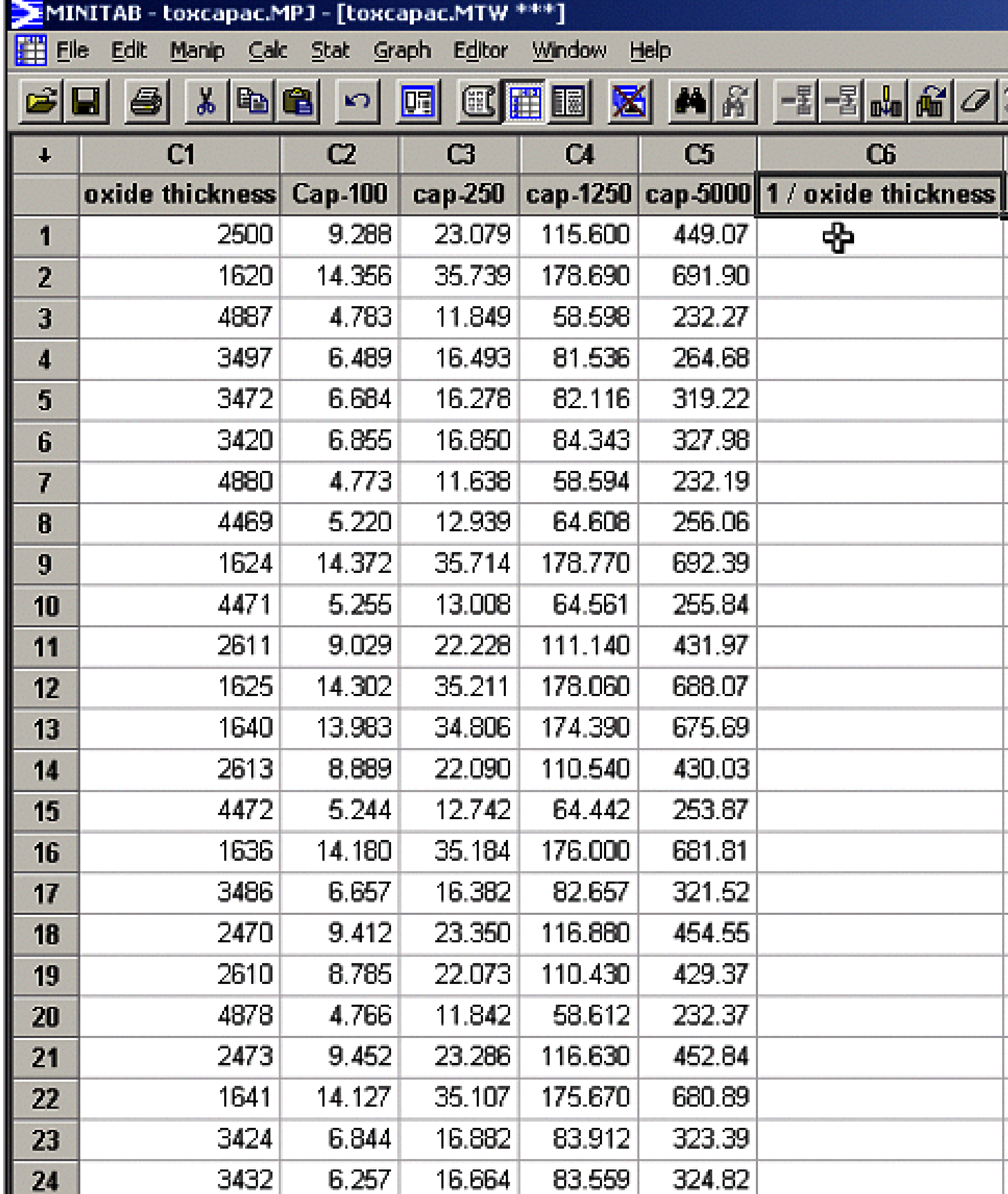
Predictors: 'oxide thickness'

Regression Analysis: cap-250 versus oxide thickness

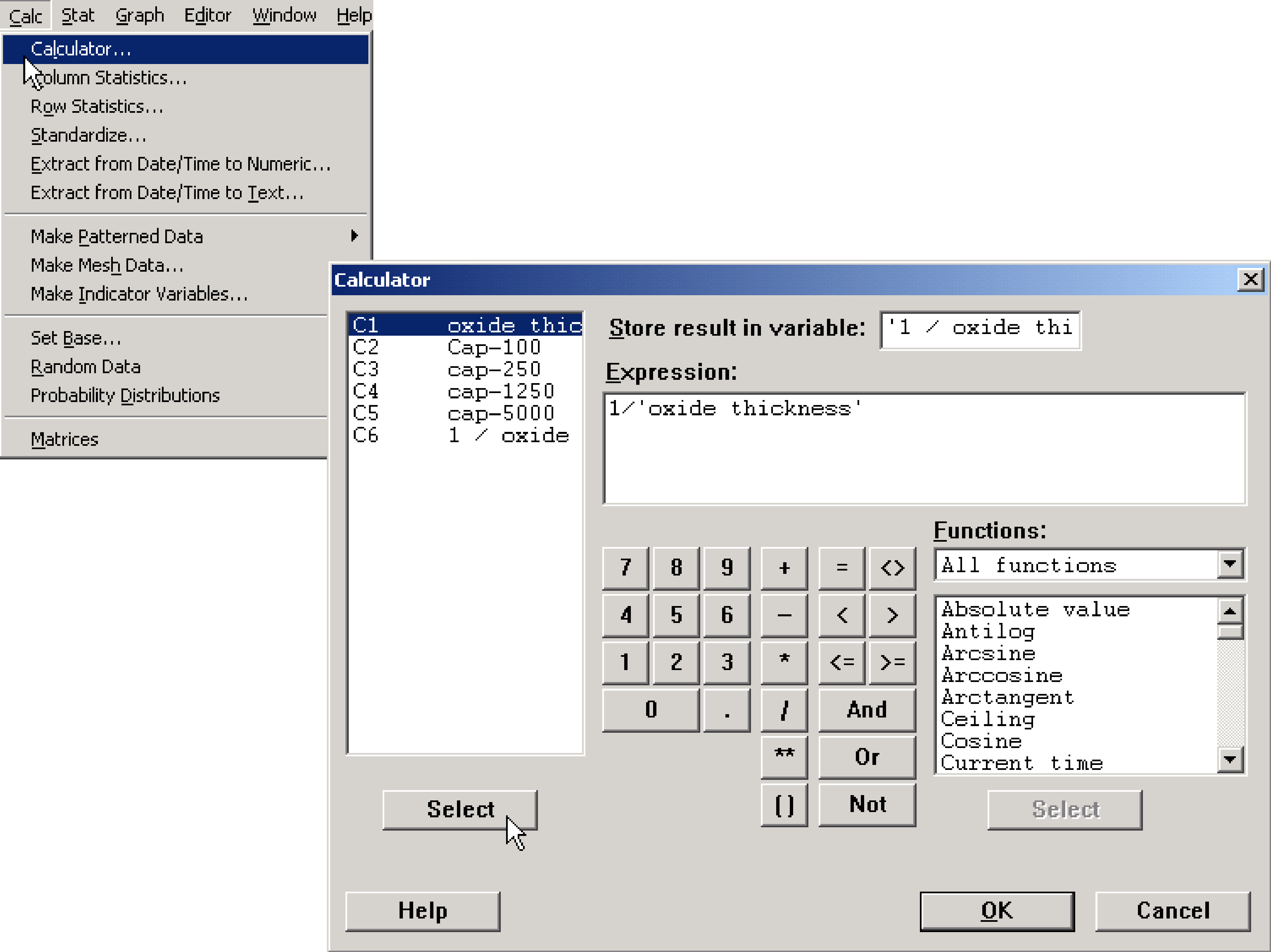
```
The regression equation is
cap-250 = 44.1 - 0.00727 oxide thickness

Predictor          Coef      SE Coef          T        P
Constant         44.108        1.676      26.32    0.000
oxide th     -0.0072735    0.0005111     -14.23    0.000

S = 2.833       R-Sq = 90.2%     R-Sq(adj) = 89.8%

Analysis of Variance

Source            DF           SS          MS         F         P
Regression         1       1625.4      1625.4    202.52     0.000
Residual Error    22        176.6         8.0
Total             23       1802.0
```

Transform:

1 / oxide thickness



↓	C1	C2	C3	C4	C5	C6
	oxide thickness	Cap-100	cap-250	cap-1250	cap-5000	1 / oxide thickness
1	2500	9.288	23.079	115.600	449.07	
2	1620	14.356	35.739	178.690	691.90	
3	4887	4.783	11.849	58.598	232.27	
4	3497	6.489	16.493	81.536	264.68	
5	3472	6.684	16.278	82.116	319.22	
6	3420	6.855	16.850	84.343	327.98	
7	4880	4.773	11.638	58.594	232.19	
8	4469	5.220	12.939	64.608	256.06	
9	1624	14.372	35.714	178.770	692.39	
10	4471	5.255	13.008	64.561	255.84	
11	2611	9.029	22.228	111.140	431.97	
12	1625	14.302	35.211	178.060	688.07	
13	1640	13.983	34.806	174.390	675.69	
14	2613	8.889	22.090	110.540	430.03	
15	4472	5.244	12.742	64.442	253.87	
16	1636	14.180	35.184	176.000	681.81	
17	3486	6.657	16.382	82.657	321.52	
18	2470	9.412	23.350	116.880	454.55	
19	2610	8.785	22.073	110.430	429.37	
20	4878	4.766	11.842	58.612	232.37	
21	2473	9.452	23.286	116.630	452.84	
22	1641	14.127	35.107	175.670	680.89	
23	3424	6.844	16.882	83.912	323.39	
24	3432	6.257	16.664	83.559	324.82	



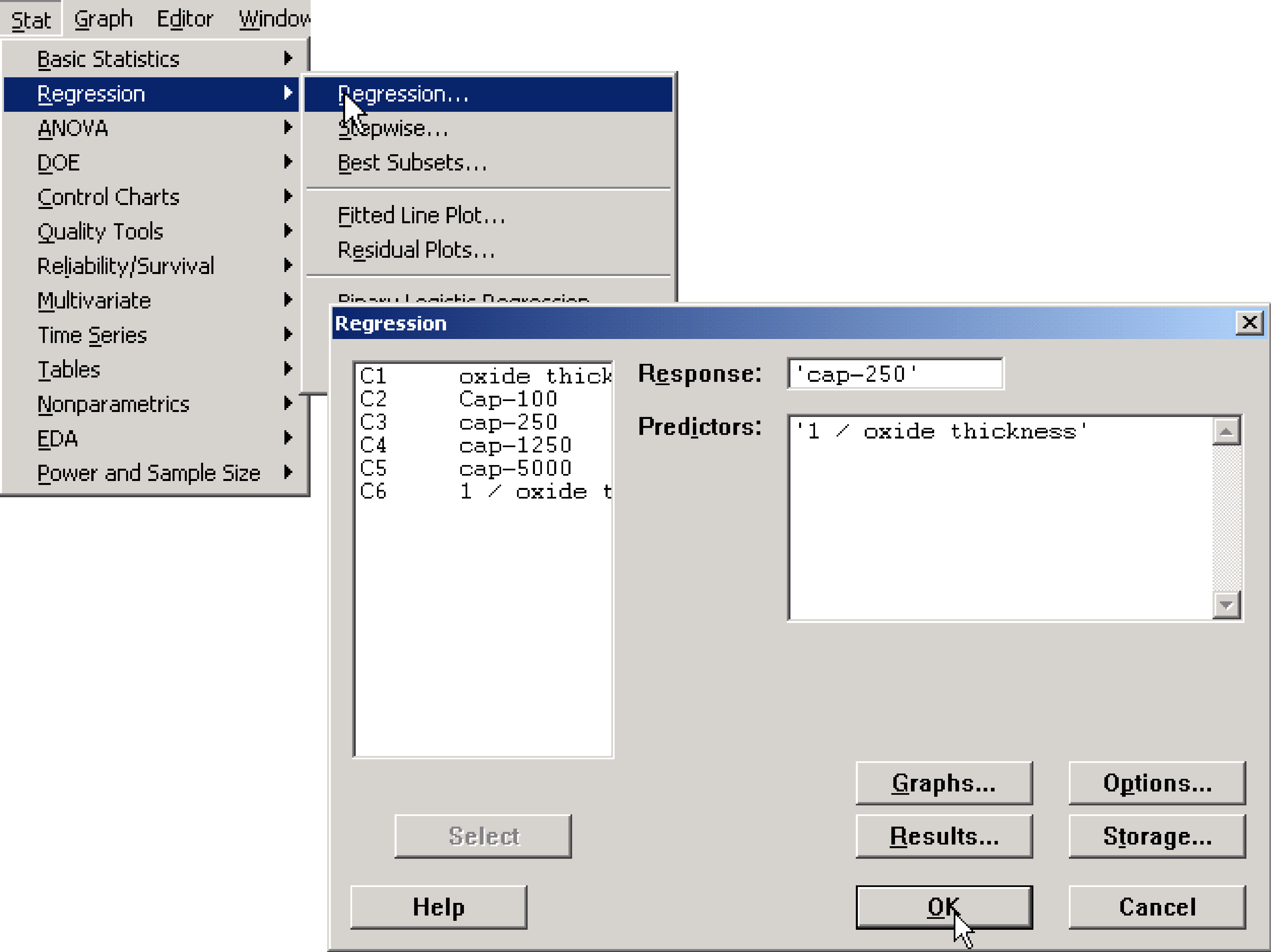
Calculator

C1	oxide thic
C2	Cap-100
C3	cap-250
C4	cap-1250
C5	cap-5000
C6	1 / oxide

Store result in variable: '1 / oxide thi

Expression: 1/'oxide thickness'

Functions: All functions



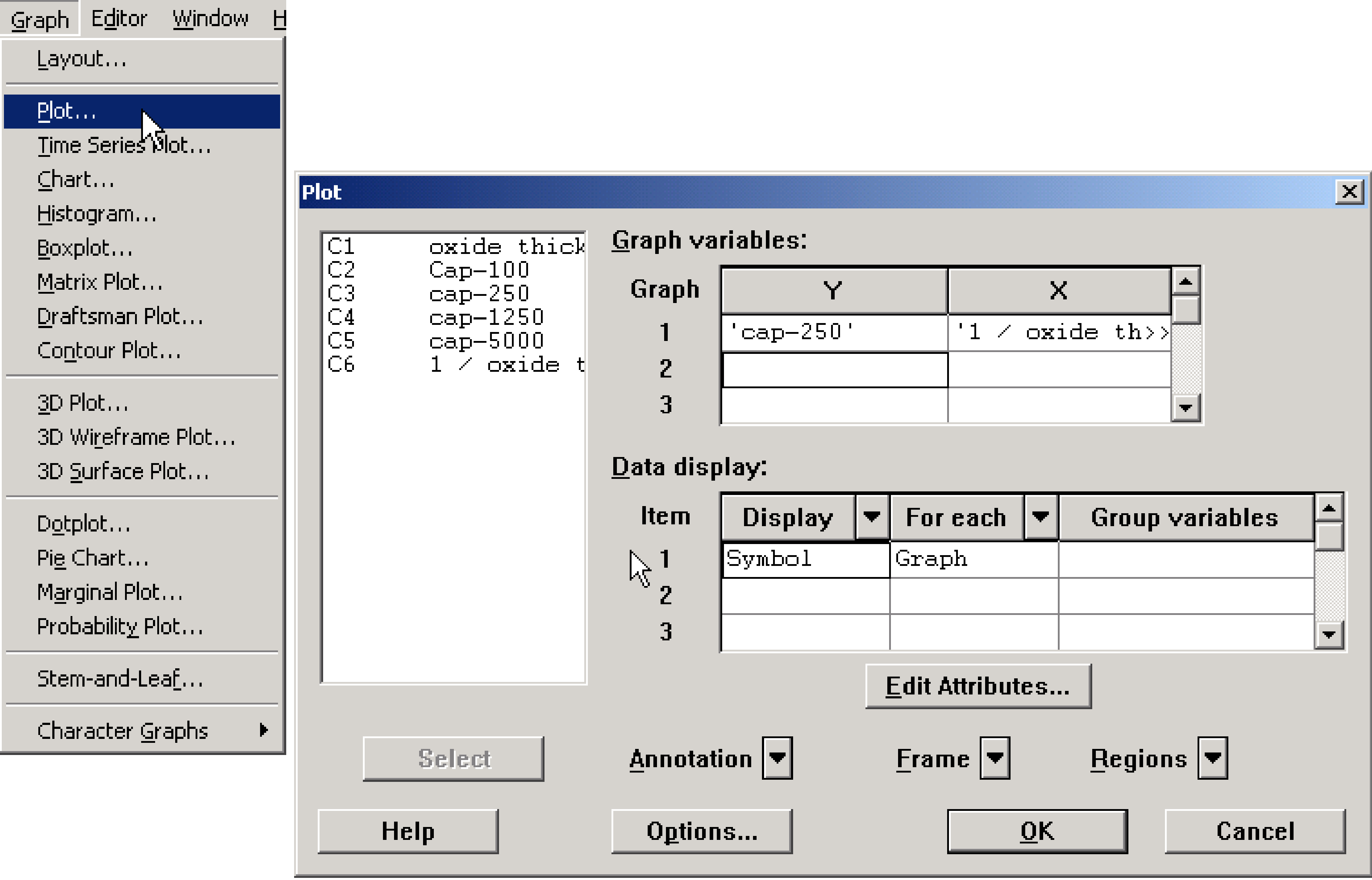
C1 oxide thick
C2 Cap-100
C3 cap-250
C4 cap-1250
C5 cap-5000
C6 1 / oxide t

Response: 'cap-250'

Predictors: '1 / oxide thickness'

Regression Analysis: cap-250 versus 1 / oxide thickness

```
The regression equation is

cap-250 = - 0.0304 + 57639 1 / oxide thickness


Predictor        Coef     SE Coef          T        P

Constant     -0.03036     0.08159      -0.37    0.713

1 / oxid      57639.3       200.8     287.02    0.000


S = 0.1479      R-Sq = 100.0%    R-Sq(adj) = 100.0%


Analysis of Variance


Source            DF          SS          MS
```



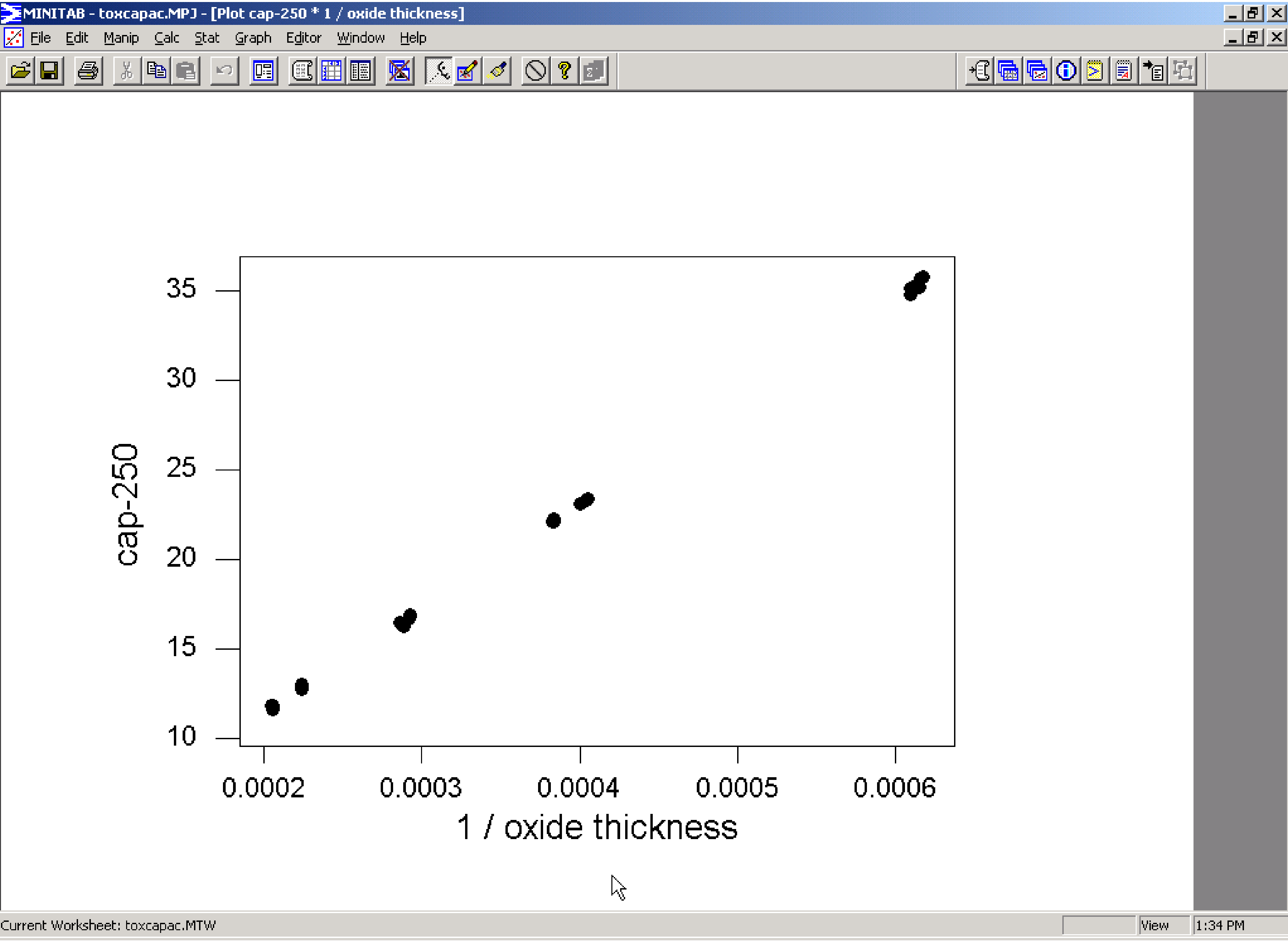
C1 oxide thick
C2 Cap-100
C3 cap-250
C4 cap-1250
C5 cap-5000
C6 1 / oxide t

Graph variables:

Graph	Y	X
1	'cap-250'	'1 / oxide th>>
2		
3		

Data display:

Item	Display	For each	Group variables
1	Symbol	Graph	
2			
3			


MINITAB - toxcapac.MPJ - [Plot cap-250 * 1 / oxide thickness]
File
Edit
Manip
Calc
Stat
Graph
Editor
Window
Help
35
30
25
20
15
10
cap-250
0.0002
0.0003
0.0004
0.0005
0.0006
1 / oxide thickness
Current Worksheet: toxcapac.MTW
View
1:34 PM

STATISTICS DECISION TREE

Multiple Input Variables

Compare Proportions

Chi-Square Test

Screening Experiments

Full Factorial
Fractional Factorial

Analysis of Experiments

ANOVA
Multiple Linear Regression

Response Surface Modeling

Box-Behnken Designs
Central Composite Designs
Multiple Linear Regression
Stepwise Regression
Contour Plots
3 D Mesh Plots

Model Response Distribution

Monte Carlo Simulation
Generation of System Moments

Optimization

Optimization of Expected Value:
Linear Programming
Non Linear Programming

EXPERIMENTAL DESIGN: COUNTER–EXAMPLE

Open the file “Simullab.xls”

New process involving

3 input variables

All 3 input variables (x1, x2, and x3) can vary between 0 and 100

RESPONSE (OUTPUT VARIABLE): YIELD

We do not know how any of these variables affect the yield, nor whether all three of the variables affect the yield.

Your job is: **to optimize the response, YIELD**

The goal is to approach 100% yield by optimizing the values of the three input variables.

SIMULLAB RESULTS

Run	Input X1		Input X2		Input X3		Yield
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							

SIMULLAB RESULTS

Run	Input X1	Input X2	Input X3	Yield
1	25	25	25	
2	75	25	25	
3	25	75	25	
4	75	75	25	
5	25	25	75	
6	75	25	75	
7	25	75	75	
8	75	75	75	
9	50	50	50	
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

Open Minitab
From the data screen, name column 4 (C4) "Yield".
Enter the yields for the 9 runs in column 4, in the order obtained

Pull down menu: Statistics, DOE (Design of Experiments), Fractional Factorial.
For "number of factors", enter 3.
For the "number of runs",enter 8. (This must be a power of 2]
Click into the box under "Store data matrix (blocks and factors) in:",
and type: c1-c3.

Click on the "Options" button; enter 1 in the box ,"Number of Center Points"
Click on the "OK" box. Click on the "OK" box again on the other screen.

Pull down the "Window" menu, "Data".
Low levels (25) are now represented with a (-1),
the high levels (75) with a (1), and the middle level (50) with a (0).
(This is the conventional way to represent the levels; the fractional factorial screen
allows the actual levels (25, 75, and 0) to be used instead)

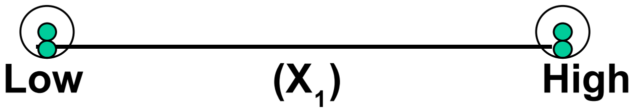
Pull down the menu "Stats", "Regression", "Regression".
Click on c4 "Yield", and click on the "Select" button.

Click on c1, hold down the button and move the mouse to highlight c1, c2, and c3
Let go of the button, and click on the "Select" button.

Click on the "OK" button.

FULL FACTORIAL EXPERIMENTAL DESIGN

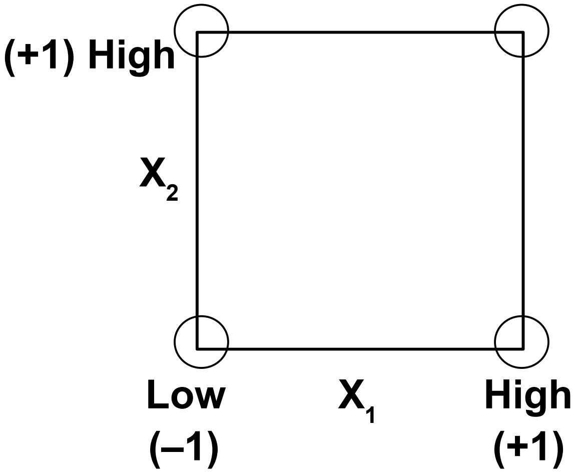
1. 1 independent variable (X_1), two levels:


Low (X_1) High

"Geometric representation"

- compare Y for $X_1\downarrow_{Low}$, $X_1\uparrow^{High}$

2. 2 independent variables (X_1, X_2), each having two levels:


(+1) High

X_2

Low (–1) X_1 High (+1)

	X_1	X_2
1	-1	-1
2	-1	+1
3	+1	-1
4	+1	+1

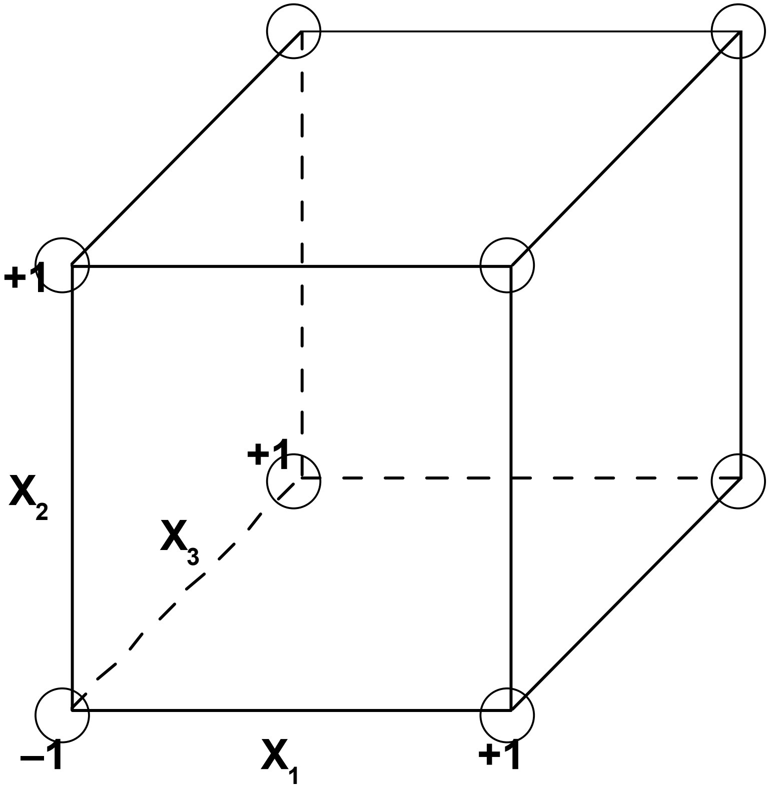
2^2

\# of factors or variables

\# of levels (High, Low)

- compare the mean $\overline{Y}$ for $X_1\downarrow_{Low}$ vs $X_1\uparrow^{High}$
- compare the mean $\overline{Y}$ for $X_2\downarrow_{Low}$ vs $X_2\uparrow^{High}$

FULL FACTORIAL
3 INDEPENDENT VARIABLES

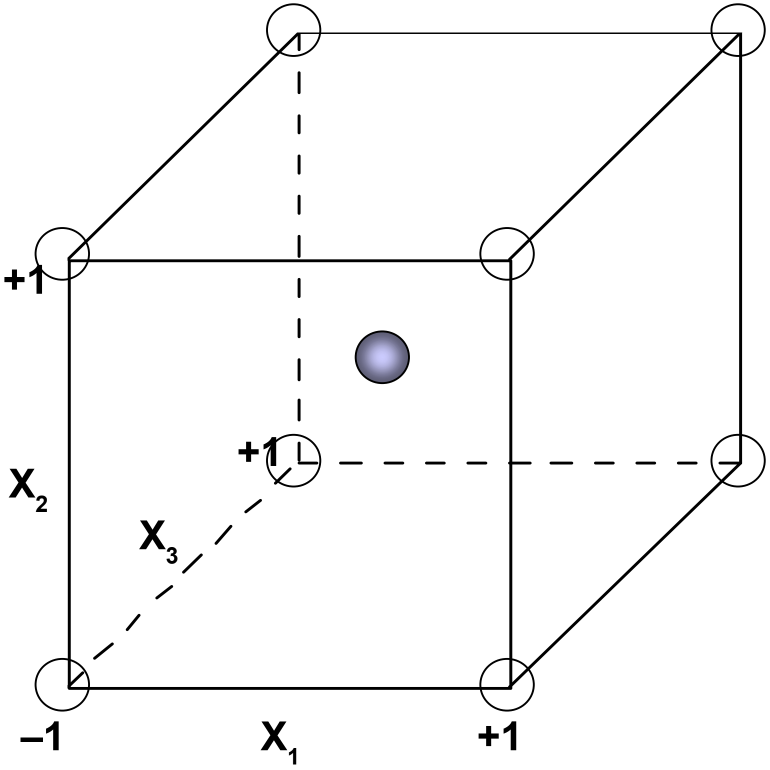


	X_1	X_2	X_3
1	-1	-1	-1
2	-1	-1	+1
3	-1	+1	-1
4	-1	+1	+1
5	+1	-1	-1
6	+1	-1	+1
7	+1	+1	-1
8	+1	+1	+1

2^3

FULL FACTORIAL CENTERPOINT

3 INDEPENDENT VARIABLES

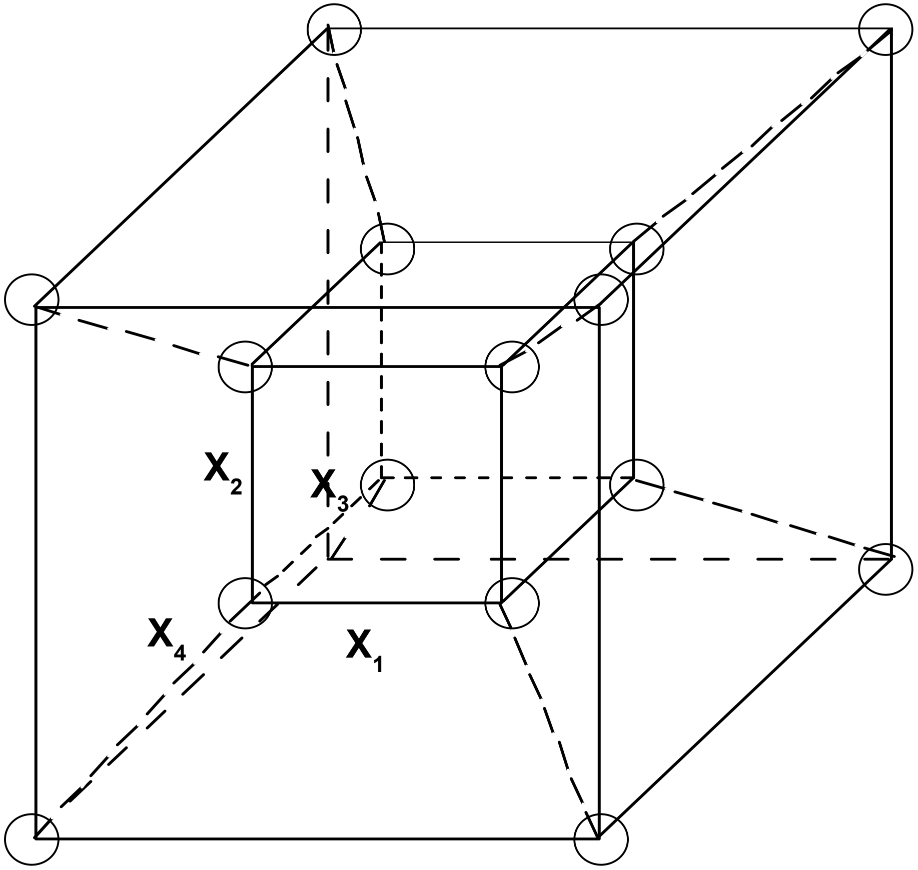


	X_1	X_2	X_3
1	-1	-1	-1
2	-1	-1	+1
3	-1	+1	-1
4	-1	+1	+1
5	+1	-1	-1
6	+1	-1	+1
7	+1	+1	-1
8	+1	+1	+1
9	0	0	0

2^3 + CP

FULL FACTORIAL

4 INDEPENDENT VARIABLES

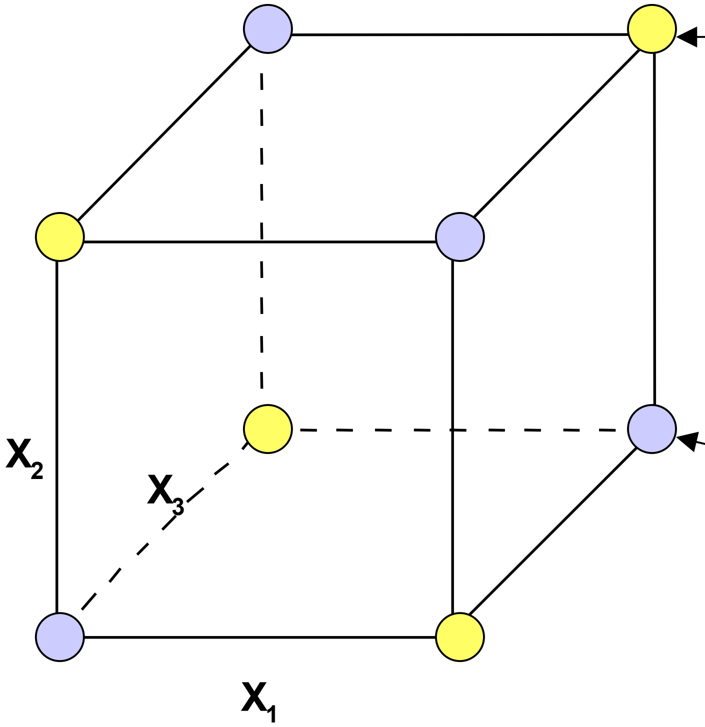


“Tesseract”

2^4

x1	x2	x3	x4
+1	+1	+1	+1
+1	+1	+1	-1
+1	+1	-1	+1
+1	+1	-1	-1
+1	-1	+1	+1
+1	-1	+1	-1
+1	-1	-1	+1
+1	-1	-1	-1
-1	+1	+1	+1
-1	+1	+1	-1
-1	+1	-1	+1
-1	+1	-1	-1
-1	-1	+1	+1
-1	-1	+1	-1
-1	-1	-1	+1
-1	-1	-1	-1

FRACTIONAL FACTORIAL

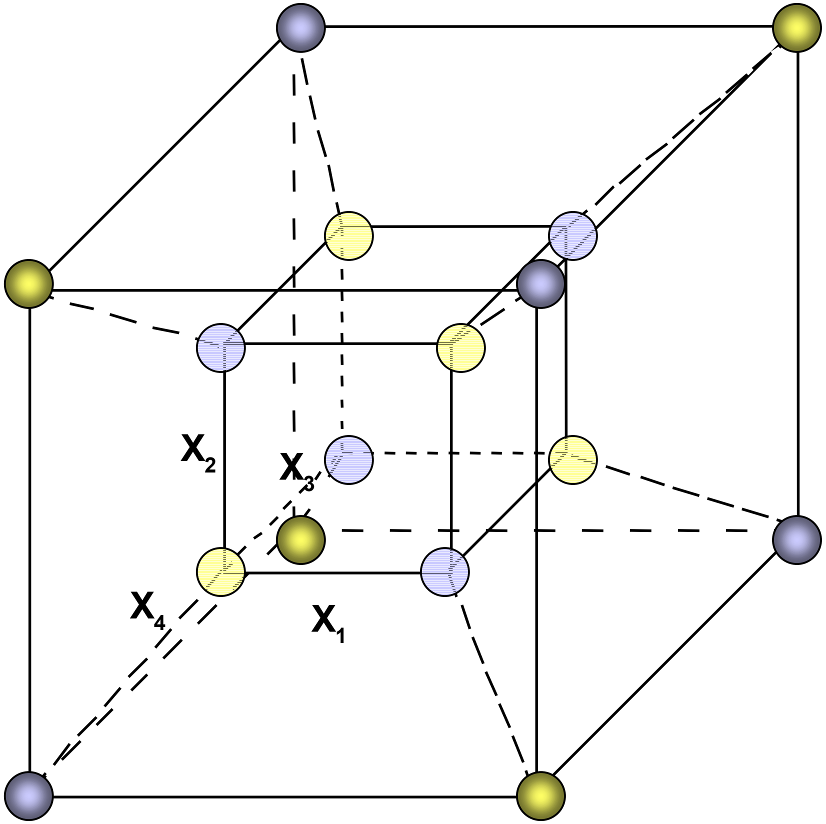


	X_1	X_2	$X_3 = X_1 \cdot X_2$
1	-1	-1	
2	-1	+1	
3	+1	-1	
4	+1	+1	

	X_1	X_2	$X_3 = - X_1 \cdot X_2$
1	-1	-1	
2	-1	+1	
3	+1	-1	
4	+1	+1	

2^{3-1}

FRACTIONAL FACTORIAL



	X_1	X_2	X_3	$X_4 = \pm X_1 \bullet X_2 \bullet X_3$
1				
2				
3				
4				
5				
6				
7				
8				

2^{4-1}

Design Resolution

A design of resolution R is one in which no p-factor effect is confounded with any other effect with less than R-p factors.

The resolution of a design is denoted by a subscript of R as a Roman numeral

A design of resolution R = III does not confound main effects with one another, but confounds main effects with two-factor interactions

A design of resolution R = IV does not confound main effects and two-factor interactions, but confounds two-factor interactions with other two-factor interactions.

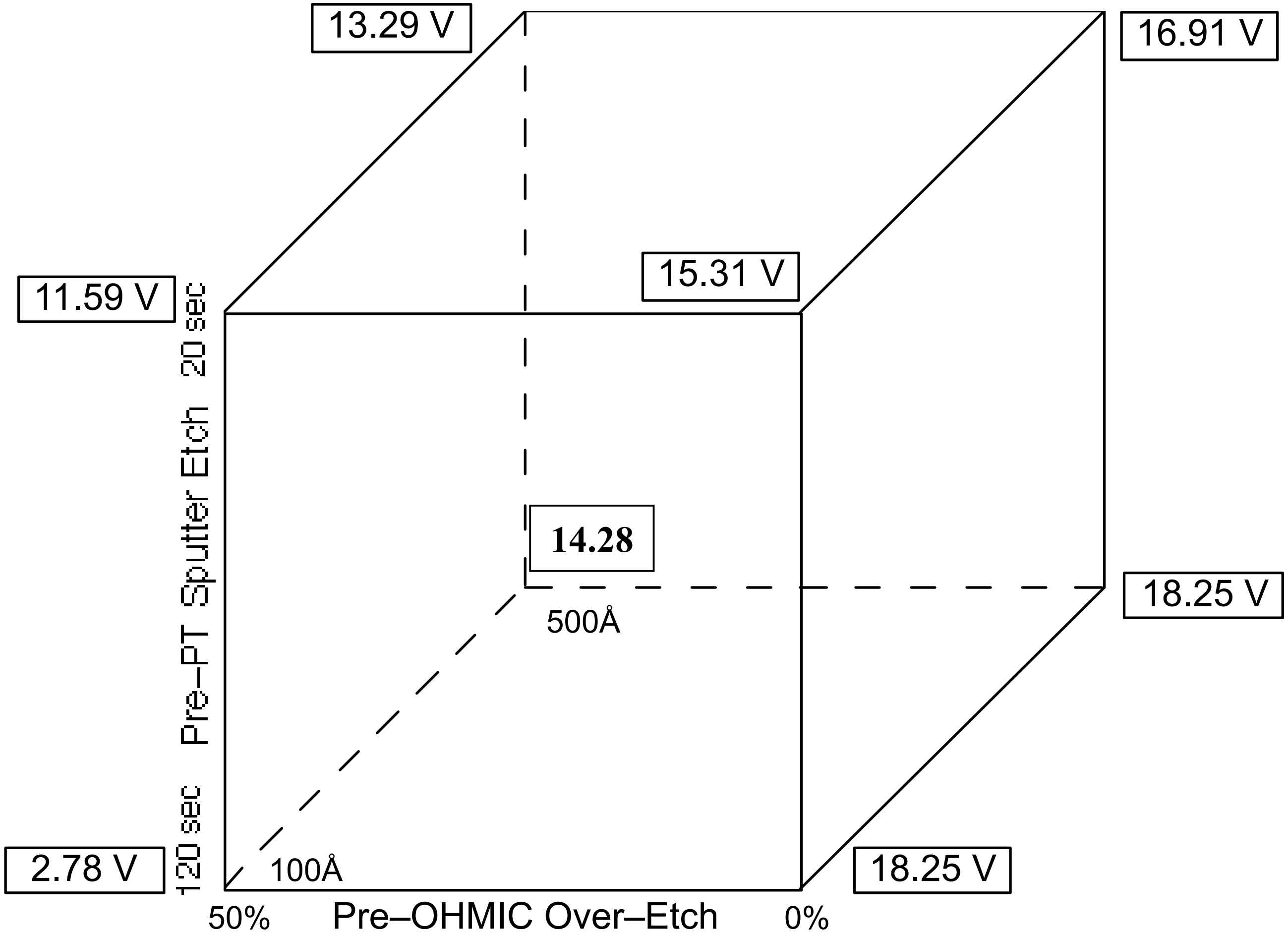
A design of resolution R = V does not confound main effects and two factor interactions with each other, but confounds two-factor interactions with three-factor interactions, and so on.

Factorial designs

Number of runs	Number of factors (variables)								
	2	3	4	5	6	7	8	9	10
4	2^2	2^{3-1}_{III} $x_3=x_1x_2$							
8		2^3	2^{4-1}_{IV} $x_4=x_1x_2x_3$	2^{5-2}_{III} $x_4=x_1x_2$ $x_5=x_1x_3$	2^{6-3}_{III} $x_4=x_1x_2$ $x_5=x_1x_3$ $x_6=x_2x_3$	2^{7-4}_{III} $x_4=x_1x_2$ $x_5=x_1x_3$ $x_6=x_2x_3$ $x_7=x_1x_2x_3$			
16			2^4	2^{5-1}_{V} $x_5=x_1x_2x_3x_4$	2^{6-2}_{IV} $x_5=x_1x_2x_3$ $x_6=x_2x_3x_4$	2^{7-3}_{IV} $x_5=x_1x_2x_3$ $x_6=x_2x_3x_4$ $x_7=x_1x_3x_4$	2^{8-4}_{IV} $x_5=x_2x_3x_4$ $x_6=x_1x_3x_4$ $x_7=x_1x_2x_3$ $x_8=x_1x_2x_4$	2^{9-5}_{III} $x_5=x_1x_2x_3$ $x_6=x_2x_3x_4$ $x_7=x_1x_3x_4$ $x_8=x_1x_2x_4$ $x_9=x_1x_2x_3x_4$	2^{10-6}_{III} $x_5=x_1x_2x_3$ $x_6=x_2x_3x_4$ $x_7=x_1x_3x_4$ $x_8=x_1x_2x_4$ $x_9=x_1x_2x_3x_4$ $x_{10}=x_1x_2$
32				2^5	2^{6-1}_{IV} $x_5=x_1x_2x_3x_4x_5$	2^{7-2}_{IV} $x_6=x_1x_2x_3x_4$ $x_7=x_1x_2x_4x_5$	2^{8-3}_{IV} $x_6=x_1x_2x_3$ $x_7=x_1x_2x_4$ $x_8=x_2x_3x_4x_5$	2^{9-4}_{IV} $x_6=x_2x_3x_4x_5$ $x_7=x_1x_3x_4x_5$ $x_8=x_1x_2x_4x_5$ $x_9=x_1x_2x_3x_5$	2^{10-5}_{IV} $x_6=x_1x_2x_3x_4$ $x_7=x_1x_2x_3x_5$ $x_8=x_1x_2x_4x_5$ $x_9=x_1x_3x_4x_5$ $x_{10}= x_2x_3x_4x_5$

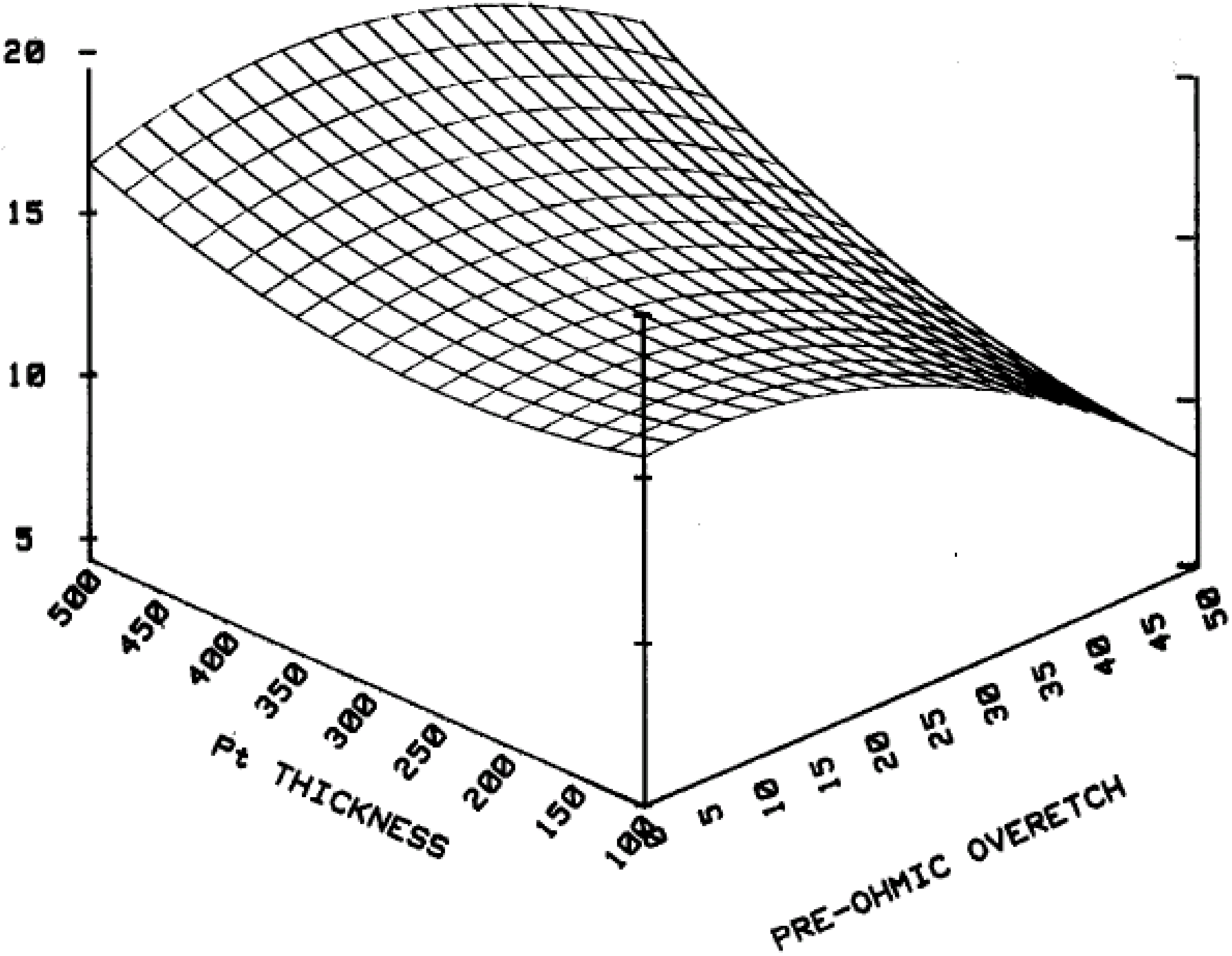
*adapted from Box, Hunter and Hunter, Statistics for Experimenters: An Introduction to Design, Data analysis, and Model Building, John Wiley and sons, 1978, p 410.

SCHOTTKY REVERSE VOLTAGE AT 100nA
(COMBINED RESULTS FROM 3 FRACTIONAL FACTORIAL (2^{5-1}) EXPERIMENTS)



CORPORATE TECHNOLOGY REVIEW

SCHOTTKY DIODE RESPONSE SURFACE



STATISTICS DECISION TREE

Multiple Input Variables

Compare Proportions
Chi-Square Test

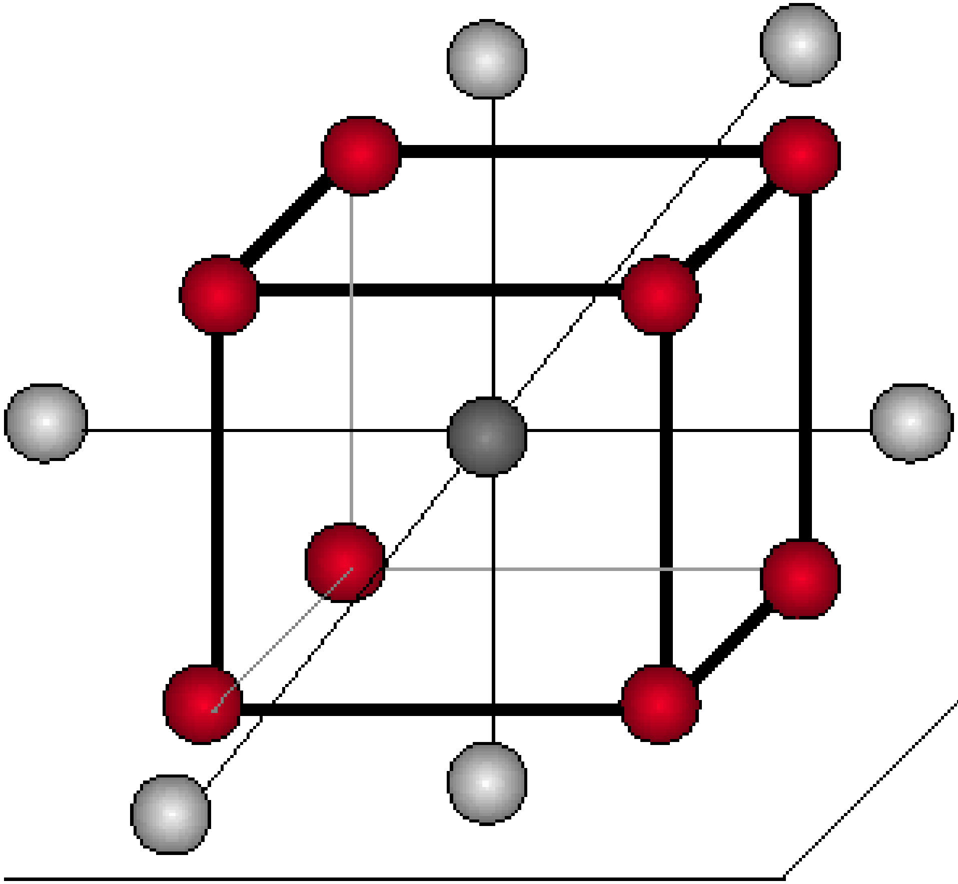
Screening Experiments
Full Factorial
Fractional Factorial

Analysis of Experiments
ANOVA
Multiple Linear Regression

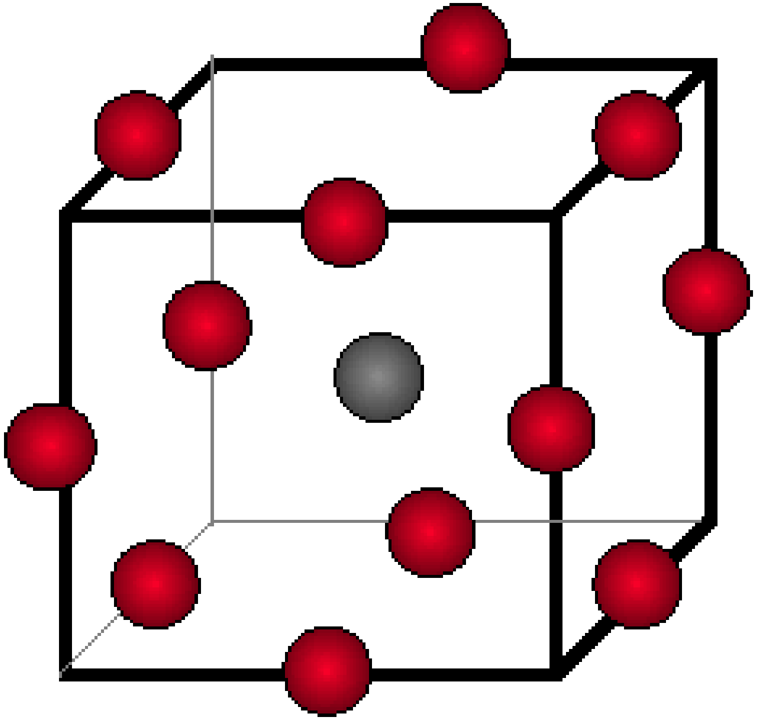
Response Surface Modeling
Box-Behnken Designs
Central Composite Designs
Multiple Linear Regression
Stepwise Regression
Contour Plots
3 D Mesh Plots

Model Response Distribution
Monte Carlo Simulation
Generation of System Moments

Optimization
Optimization of Expected Value:
Linear Programming
Non Linear Programming



CENTRAL
COMPOSITE
DESIGN



BOX–
BEHNKEN
DESIGN

CENTRAL COMPOSITE DESIGNS

- Each factor varies over five levels
- Used for fitting 2nd order response surface models
- Typically smaller than Box-Behnken designs
- Built upon two-level fractional factorials
- Rotatable

CENTRAL COMPOSITE DESIGNS

GENERAL STRUCTURE:

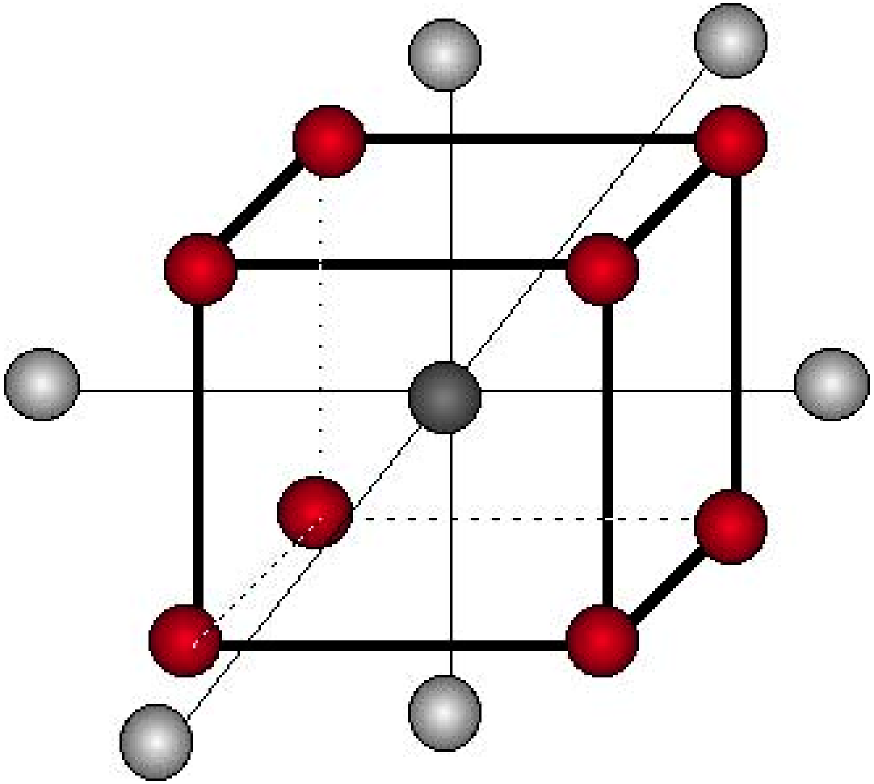
2^{n-k} Fractional Factorial

+

Star points

+

Centerpoints



CENTRAL COMPOSITE DESIGNS

Construction for n factors

- Select a resolution V two-level fractional factorial for n factors
- Generate 2 x n star points

$$\begin{array}{ccccc}
\alpha & o & o & \ldots\ldots\ldots\ldots & o \\
-\alpha & o & o & \ldots\ldots\ldots\ldots & o \\
o & \alpha & o & \ldots\ldots\ldots\ldots & o \\
o & -\alpha & o & \ldots\ldots\ldots\ldots & o \\
 & & . & & \\
 & & . & & \\
 & & . & & \\
o & o & o & \ldots\ldots\ldots\ldots & \alpha \\
o & o & o & \ldots\ldots\ldots & -\alpha
\end{array}$$

where α^4 = m and m = number of runs , 2^{n-k} in the fractional factorial design.

$$\alpha = \sqrt[4]{(2^{n-k})}$$

for 4 runs in 2^{n-k},	$\alpha = 1.414$
for 8 runs,	$\alpha = 1.68$;
for 16 runs,	$\alpha = 2$

- Add one or more centerpoints; for example: o, o, o

CENTRAL COMPOSITE DESIGNS FOR TWO FACTORS

N = 9

RUN	X_1	X_2
1	–	–
2	+	–
3	–	+
4	+	+
5	1.414	0
6	-1.414	0
7	0	1.414
8	0	-1.414
9	0	0

CENTRAL COMPOSITE DESIGNS FOR THREE FACTORS

N = 15

RUN	X_1	X_2	X_3
1	–	–	–
2	+	–	–
3	–	+	–
4	+	+	–
5	–	–	+
6	+	–	+
7	–	+	+
8	+	+	+
9	1.682	0	0
10	-1.682	0	0
11	0	1.682	0
12	0	-1.682	0
13	0	0	1.682
14	0	0	-1.682
15	0	0	0

CENTRAL COMPOSITE DESIGNS FOR FOUR FACTORS

N = 25

RUN	X_1	X_2	X_3	X_4
1	–	–	–	–
2	+	–	–	–
3	–	+	–	–
4	+	+	–	–
5	–	–	+	–
6	+	–	+	–
7	–	+	+	–
8	+	+	+	–
9	–	–	–	+
10	+	–	–	+
11	–	+	–	+
12	+	+	–	+

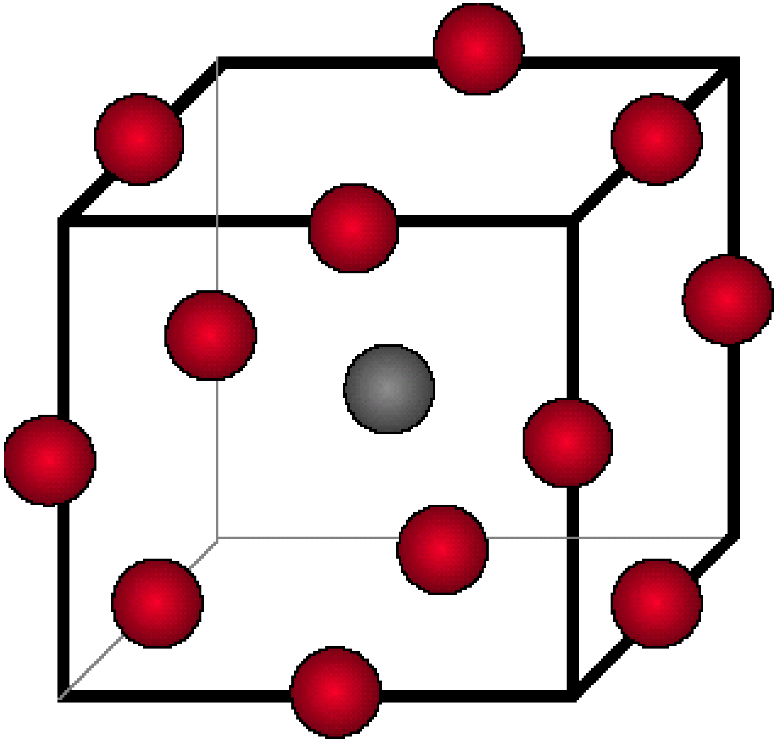
RUN	X_1	X_2	X_3	X_4
13	–	–	+	+
14	+	–	+	+
15	–	+	+	+
16	+	+	+	+
17	2	0	0	0
18	-2	0	0	0
19	0	2	0	0
20	0	-2	0	0
21	0	0	2	0
22	0	0	-2	0
23	0	0	0	2
24	0	0	0	-2
25	0	0	0	0

CENTRAL COMPOSITE DESIGNS
FOR FIVE FACTORS
N = 27

RUN	X_1	X_2	X_3	X_4	X_5
1	–	–	–	–	+
2	+	–	–	–	–
3	–	+	–	–	–
4	+	+	–	–	+
5	–	–	+	–	–
6	+	–	+	–	+
7	–	+	+	–	+
8	+	+	+	–	–
9	–	–	–	+	–
10	+	–	–	+	+
11	–	+	–	+	+
12	+	+	–	+	–
13	–	–	+	+	+
14	+	–	+	+	–
15	–	+	+	+	–
16	+	+	+	+	+
17	2	0	0	0	0
18	-2	0	0	0	0
19	0	2	0	0	0
20	0	-2	0	0	0
21	0	0	2	0	0
22	0	0	-2	0	0
23	0	0	0	2	0
24	0	0	0	-2	0
25	0	0	0	0	2
26	0	0	0	0	-2
27	0	0	0	0	0

BOX–BEHNKEN DESIGNS

- Each factor is varied over three levels
- Used for fitting 2nd order response surface models
- Alternative to central composite designs



BOX–BEHNKEN
3 LEVEL, 2 FACTOR
N = 9

CELL	X_1	X_2
1	–	–
2	0	–
3	+	–
4	–	0
5	0	0
6	+	0
7	–	+
8	0	+
9	+	+

BOX-BEHNKEN DESIGN FOR 3 FACTORS
N = 15

RUN	X_1	X_2	X_3
1	–	–	0
2	+	–	0
3	–	+	0
4	+	+	0
5	–	0	–
6	+	0	–
7	–	0	+
8	+	0	+
9	0	–	–
10	0	+	–
11	0	–	+
12	0	+	+
13	0	0	0
14	0	0	0
15	0	0	0

BOX-BEHNKEN DESIGN FOR 4 FACTORS

N = 27

RUN	X_1	X_2	X_3	X_4
1	–	–	0	0
2	+	–	0	0
3	–	+	0	0
4	+	+	0	0
5	0	0	–	–
6	0	0	+	–
7	0	0	–	+
8	0	0	+	+
9	0	0	0	0
10	–	0	0	–
11	+	0	0	–
12	–	0	0	+
13	+	0	0	+

RUN	X_1	X_2	X_3	X_4
14	0	–	–	0
15	0	+	–	0
16	0	–	+	0
17	0	+	+	0
18	0	0	0	0
19	–	–	–	0
20	+	–	–	0
21	–	+	+	0
22	+	+	+	0
23	0	0	0	–
24	0	0	0	–
25	0	0	0	+
26	0	0	0	+
27	0	0	0	0

BOX–BEHNKEN DESIGN FOR 5 FACTORS

N = 46

RUN	X_1	X_2	X_3	X_4	X_5
1	-	-	0	0	0
2	+	-	0	0	0
3	-	+	0	0	0
4	+	+	0	0	0
5	0	0	-	-	0
6	0	0	+	-	0
7	0	0	-	+	0
8	0	0	+	+	0
9	0	-	0	0	-
10	0	+	0	0	-
11	0	-	0	0	+
12	0	+	0	0	+
13	-	0	-	0	0
14	+	0	-	0	0
15	-	0	+	0	0
16	+	0	+	0	0
17	0	0	0	-	-
18	0	0	0	+	-
19	0	0	0	-	+
20	0	0	0	+	+
21	0	0	0	0	0
22	0	0	0	0	0
23	0	0	0	0	0
24	0	-	-	0	0
25	0	+	-	0	0
26	0	-	+	0	0
27	0	+	+	0	0
28	-	0	0	-	0
29	+	0	0	-	0
30	-	0	0	+	0
31	+	0	0	+	0
32	0	0	-	0	-
33	0	0	+	0	-
34	0	0	-	0	+
35	0	0	+	0	+
36	-	0	0	0	-
37	+	0	0	0	-
38	-	0	0	0	+
39	+	0	0	0	+
40	0	-	0	-	0
41	0	+	0	-	0
42	0	-	0	+	0
43	0	+	0	+	0
44	0	0	0	0	0
45	0	0	0	0	0
46	0	0	0	0	0

BOX–BEHNKEN DESIGN FOR 6 FACTORS

N = 54

RUN	X_1	X_2	X_3	X_4	X_5	X_6
1	-	-	0	-	0	0
2	+	-	0	-	0	0
3	-	+	0	-	0	0
4	+	+	0	-	0	0
5	-	-	0	+	0	0
6	+	-	0	+	0	0
7	-	+	0	+	0	0
8	+	+	0	+	0	0
9	0	-	-	0	-	0
10	0	+	-	0	-	0
11	0	-	+	0	-	0
12	0	+	+	0	-	0
13	0	-	-	0	+	0
14	0	+	-	0	+	0
15	0	-	+	0	+	0
16	0	+	+	0	+	0
17	0	0	-	-	0	-
18	0	0	+	-	0	-
19	0	0	-	+	0	-
20	0	0	+	+	0	-
21	0	0	-	-	0	+
22	0	0	+	-	0	+
23	0	0	-	+	0	+
24	0	0	+	+	0	+
25	0	0	0	0	0	0
26	0	0	0	0	0	0
27	0	0	0	0	0	0
28	-	0	0	-	-	0
29	+	0	0	-	-	0
30	-	0	0	+	-	0
31	+	0	0	+	-	0
32	-	0	0	-	+	0
33	+	0	0	-	+	0
34	-	0	0	+	+	0
35	+	0	0	+	+	0
36	-	0	-	0	0	-
37	+	0	-	0	0	-
38	-	0	+	0	0	-
39	+	0	+	0	0	-
40	-	0	-	0	0	+

BOX–BEHNKEN DESIGN FOR 6 FACTORS (Cont'd)

N = 54

RUN	X_1	X_2	X_3	X_4	X_5	X_6
41	+	0	-	0	0	+
42	-	0	+	0	0	+
43	+	0	+	0	0	+
44	0	-	0	0	-	-
45	0	+	0	0	-	-
46	0	-	0	0	+	-
47	0	+	0	0	+	-
48	0	-	0	0	-	+
49	0	+	0	0	-	+
50	0	-	0	0	+	+
51	0	+	0	0	+	+
52	0	0	0	0	0	0
53	0	0	0	0	0	0
54	0	0	0	0	0	0

BOX–BEHNKEN DESIGN FOR 7 FACTORS

N = 62

RUN	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	0	0	0	-	-	-	0
2	0	0	0	+	-	-	0
3	0	0	0	-	+	-	0
4	0	0	0	+	+	-	0
5	0	0	0	-	-	+	0
6	0	0	0	+	-	+	0
7	0	0	0	-	+	+	0
8	0	0	0	+	+	+	0
9	-	0	0	0	0	-	-
10	+	0	0	0	0	-	-
11	-	0	0	0	0	+	-
12	+	0	0	0	0	+	-
13	-	0	0	0	0	-	+
14	+	0	0	0	0	-	+
15	-	0	0	0	0	+	+
16	+	0	0	0	0	+	+
17	0	-	0	0	-	0	-
18	0	+	0	0	-	0	-
19	0	-	0	0	+	0	-
20	0	+	0	0	+	0	-
21	0	-	0	0	-	0	+
22	0	+	0	0	-	0	+
23	0	-	0	0	+	0	+
24	0	+	0	0	+	0	+
25	-	-	0	-	0	0	0
26	+	-	0	-	0	0	0
27	-	+	0	-	0	0	0
28	+	+	0	-	0	0	0
29	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0
32	-	-	0	+	0	0	0
33	+	-	0	+	0	0	0
34	-	+	0	+	0	0	0
35	+	+	0	+	0	0	0
36	-	0	-	0	-	0	0
37	+	0	-	0	-	0	0
38	-	0	+	0	-	0	0
39	+	0	+	0	-	0	0
40	-	0	-	0	+	0	0

BOX–BEHNKEN DESIGN FOR 7 FACTORS (Cont’d)

N = 62

RUN	X_1	X_2	X_3	X_4	X_5	X_6	X_7
41	+	0	-	0	+	0	0
42	-	0	+	0	+	0	0
43	+	0	+	0	+	0	0
44	0	0	-	-	0	0	-
45	0	0	+	-	0	0	-
46	0	0	-	+	0	0	-
47	0	0	+	+	0	0	-
48	0	0	-	-	0	0	+
49	0	0	+	-	0	0	+
50	0	0	-	+	0	0	+
51	0	0	+	+	0	0	+
52	0	-	-	0	0	-	0
53	0	+	-	0	0	-	0
54	0	-	+	0	0	-	0
55	0	+	+	0	0	-	0
56	0	-	-	0	0	+	0
57	0	+	-	0	0	+	0
58	0	-	+	0	0	+	0
59	0	+	+	0	0	+	0
60	0	0	0	0	0	0	0
61	0	0	0	0	0	0	0
62	0	0	0	0	0	0	0

BOX–BEHNKEN DESIGN FOR 9 FACTORS

N = 130

RUN	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1–8	±1	0	0	±1	0	0	±1	0	0
9–16	0	±1	0	0	±1	0	0	±1	0
17–24	0	0	±1	0	0	±1	0	0	±1
25	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0
27–34	±1	±1	±1	0	0	0	0	0	0
35–42	0	0	0	±1	±1	±1	0	0	0
43–50	0	0	0	0	0	0	±1	±1	±1
51	0	0	0	0	0	0	0	0	0
52	0	0	0	0	0	0	0	0	0
53–60	±1	0	0	0	±1	0	0	0	±1
61–68	0	0	±1	±1	0	0	0	±1	0
69–76	0	±1	0	0	0	±1	±1	0	0
77	0	0	0	0	0	0	0	0	0
78	0	0	0	0	0	0	0	0	0

BOX–BEHNKEN DESIGN FOR 9 FACTORS (Cont'd)

N = 130

RUN	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
79–86	±1	0	0	0	0	±1	0	±1	0
87–94	0	±1	0	±1	0	0	0	0	±1
95–102	0	0	±1	0	±1	0	±1	0	0
103	0	0	0	0	0	0	0	0	0
104	0	0	0	0	0	0	0	0	0
105–112	±1	0	0	±1	0	0	±1	0	0
113–120	0	±1	0	0	±1	0	0	±1	0
121–128	0	0	±1	0	0	±1	0	0	±1
129	0	0	0	0	0	0	0	0	0
130	0	0	0	0	0	0	0	0	0

NUMBER OF RUNS REQUIRED TO FIT A FULL QUADRATIC MODEL

NO. OF INDEPENDENT VARIABLES (FACTORS)	NO. OF COEFFICIENTS IN FULL QUADRATIC	NO. OF TRIALS IN FULL THREE-LEVEL FACTORIAL	NO. RUNS IN BOX-BEHNKEN DESIGN	CENTRAL COMPOSITE
2	6	9	9	9
3	10	27	15	15
4	15	81	27	25
5	21	243	46	27
6	28	729	54	46
7	36	2187	62	80

SMALL COMPOSITE DESIGNS*

Composite designs for fitting second-order models in k factors all contain cube portions of resolution at least V, plus axial points, plus center points.

There must be at least one point for each coefficient --> 1/2(k + 1)(k + 2) points.

Hartley (1959) showed that the cube portion of the composite design doesn't need to be resolution V - it can be as low as resolution III if two-factor interactions aren't aliased with two-factor interactions.
Two-factor interactions can be aliased with main effects, because the star portion provides additional information on the main effects.

This allows much Composite Designs. Westlake (1965) took this idea further by finding even smaller cubes for the k = 5, 7, and 9 cases.

The following table shows the numbers of points in various suggested designs.

***Box and Draper, Empirical Model-Building and Response Surfaces, John Wiley and sons, 1987, page 521.**

Designs Requiring Only A Small Number Of Runs

Points Needed By Some Small Composite Designs

Factors, k	2	3	4	5	6	7	8	9
Coefficients $\frac{1}{2}(k + 1)(k + 2)$	6	10	15	21	28	36	45	55
Points in Box-Hunter (1957) designs	8	14	24	26	44	78	80	146
Hartley's number of points	6	10	16	26	28	46	48	82
Westlake's number of points	—	—	--	22	--	40	--	62

RSM PROCEDURE – SHORT EXPLANATION

1. Design **Experimental Matrix**

   Possibilities: Factorial with centerpoint
   Box-Behnken
   Central composite design

2. Run experimental matrix; collect data

3. Analyze data using **Multiple Linear Regression**, with second order equation:

Main Effects	**Second Order Effects**	**Interactions**
X1	X1*X1	X1*X2
X2	X2*X2	X2*X3
X3	X3*X3	X1*X3

Response = A + B*X1 + C*X2 + D*X3
+ E*X1^2 + F*X2^2 + G*X3^2
+ I*X1*X2 + J*X2*X3 + K*X1*K3

4. Generate Response Surfaces, using model from multiple linear regression

   - Contour plots
   - Mesh plots

TWO–LEVEL CENTRAL COMPOSITE DESIGN

EXAMPLE: P CHANNEL VT VS I² DOSES

INPUT VARIABLES		RESPONSE
Subs Implant Dose (E11)	Blanket Implant Dose (E11)	Vt–p (mV)
2.71	1.84	– 1088
16.2	1.84	– 1282
2.71	7.44	– 577
16.2	7.44	– 881
23.6	3.70	– 1257
1.87	3.70	– 913
6.64	9.94	– 402
6.64	1.38	– 1187
6.64	3.70	– 1012

```
MTB > regress c10 5 c1 c2 c11 c22 c12

The regression equation is
Vtp = 1175 + 16.7 Sub l2 - 70.7 Blnkt l2 - 0.253 C11 - 2.68 C22 + 1.46 C12

Predictor        Coef      Stdev     t-ratio      p
Constant      1175.15       8.00      146.95  0.000
Sub l2        16.6748     0.9889       16.86  0.000
Blnkt l2      -70.729      2.686      -26.33  0.000
C11          -0.25334    0.03424       -7.40  0.005
C22           -2.6755     0.2211      -12.10  0.001
C12           1.46004    0.09203       15.87  0.001
```

Student's t-test to check if each slope (coefficient) is zero

Alpha risk that each slope is actually zero, & the non-zero value is due to chance alone

```
s = 3.588     R-sq = 100.%    R-sq(adj) = 100.%
```

% of Y variance attributed to variance of the input variables:

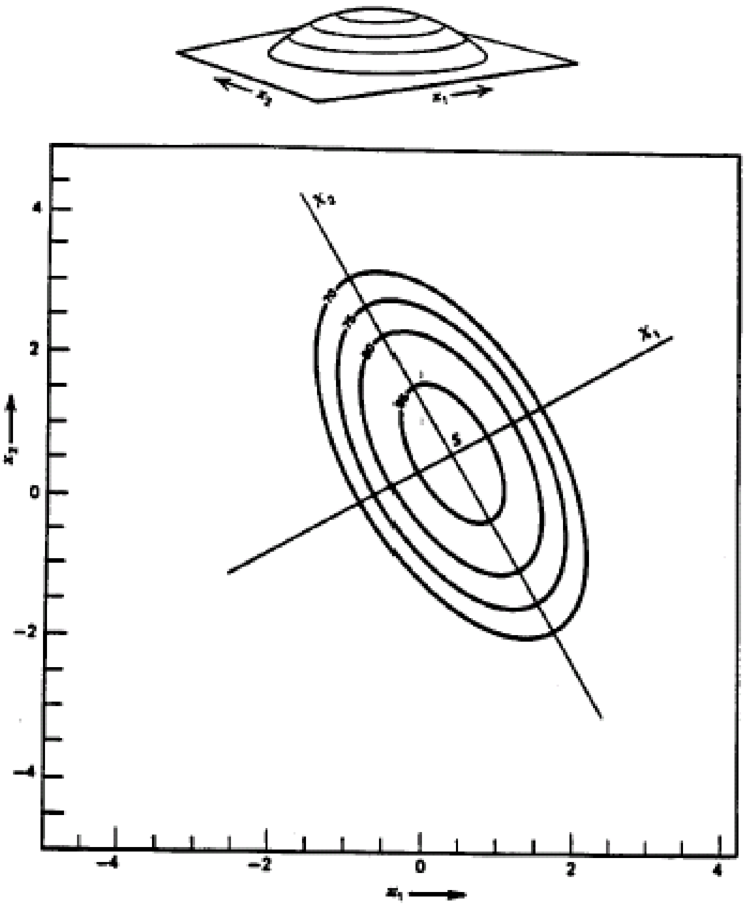
$$\frac{(\partial Y/\partial x_1 * S_{x1})^2 + \ldots + (\partial Y/\partial x_n * S_{xn})^2}{\text{Variance of Y}}$$

```
Analysis of Variance

SOURCE       DF       SS        MS         F        p
Regression    5   728788    145758  11324.46    0.000
Error         3       39        13
Total         8   728826
```

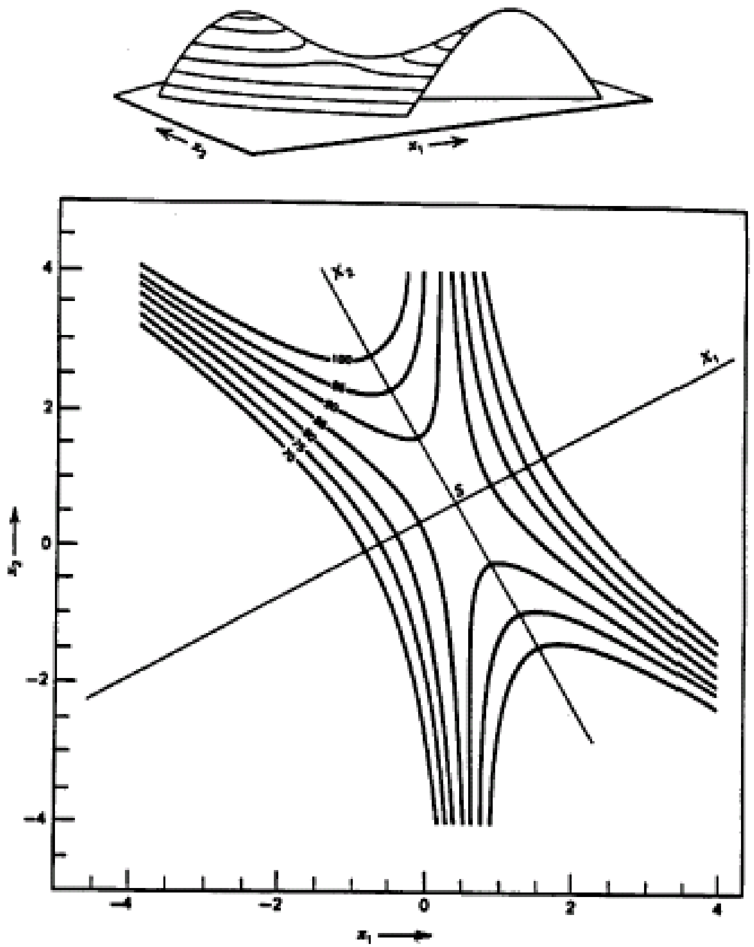
Test hypothesis that at least one slope is not zero.

```
SOURCE       DF   SEQ SS
Sub           1   163238
Blnkt l2      1   558420
C11           1      287
C22           1     3602
C12           1     3240
```

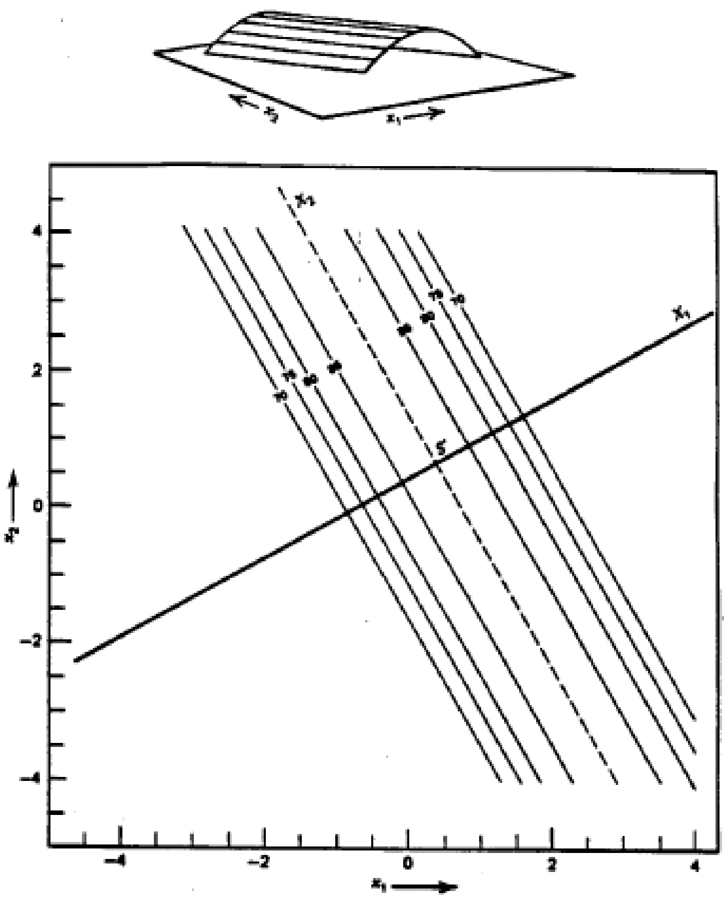
Attempt to attribute sum of squares (like variance) to each input variable.
May be misleading.



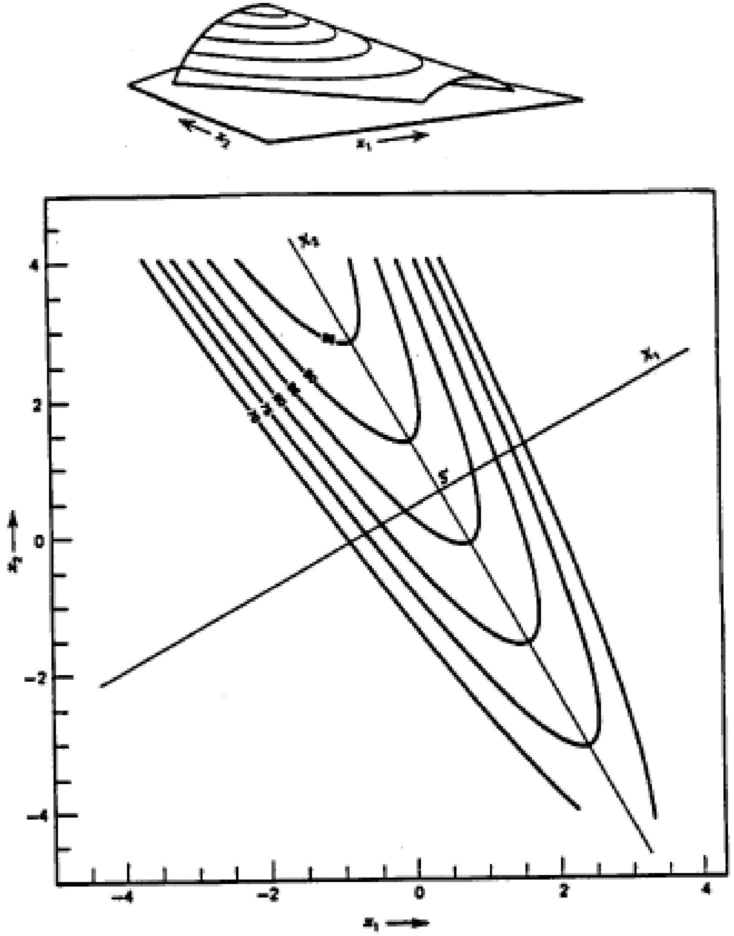
Dome / Simple maximum



Saddle / Minimax



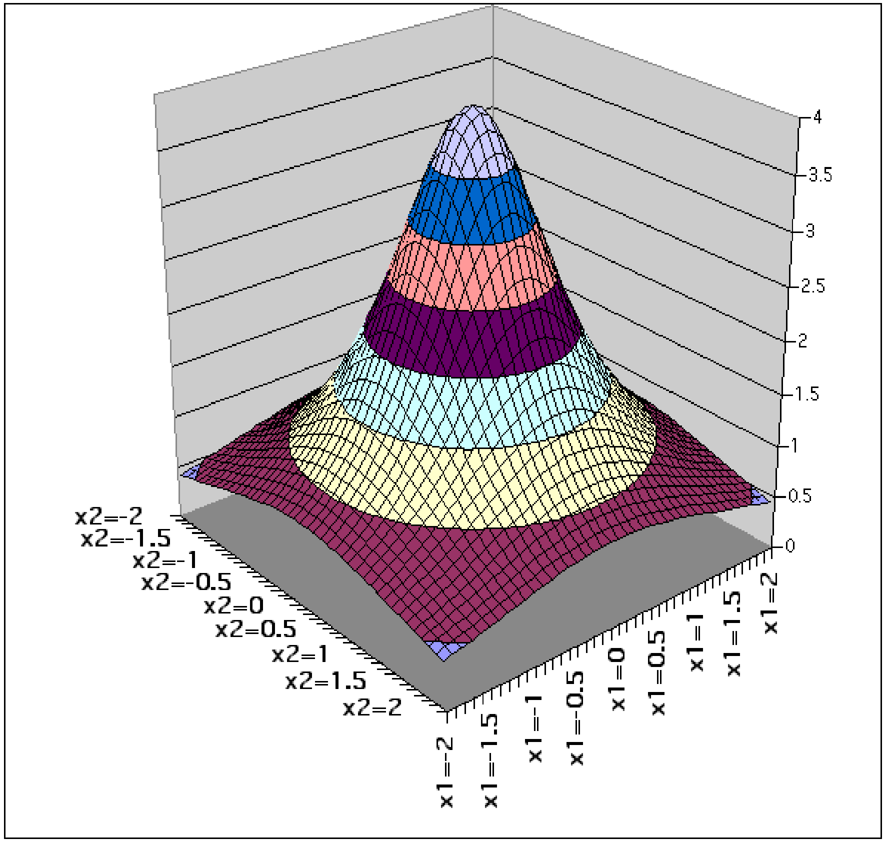
Stationary Ridge



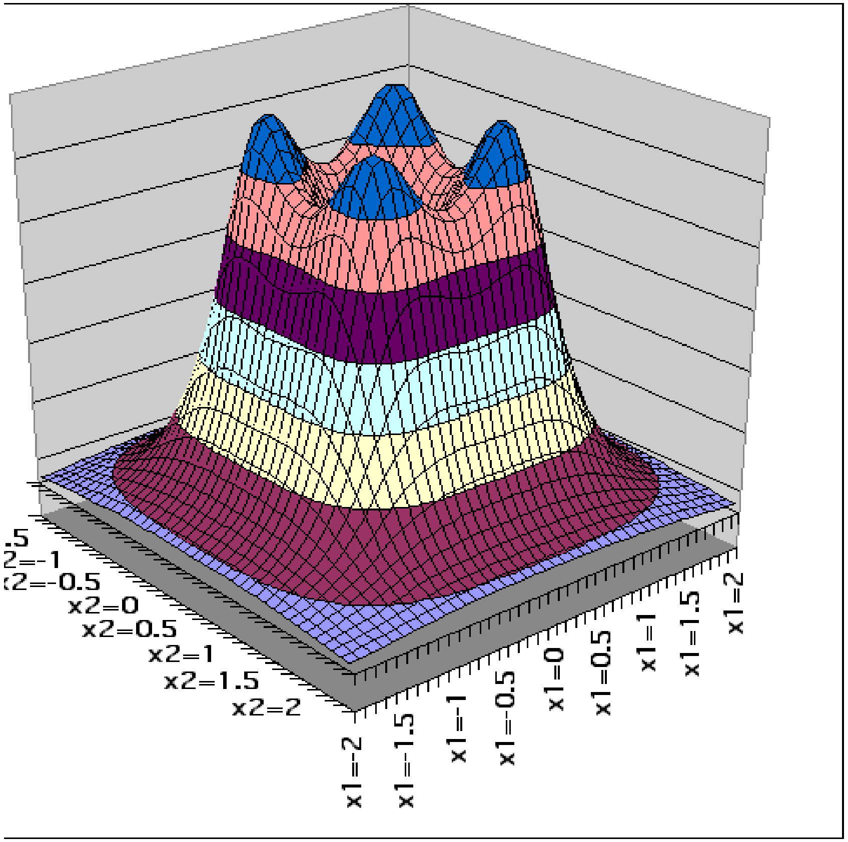
Rising Ridge

“...The [3^2 factorial] design generates four pockets of high information which seemingly have little to do with the needs of an experimenter.

... it is possible to choose designs of second and higher orders for which the information contours are spherical. Equivalently, these rotatable designs have the property that the variances and covariances of the effects remain unaffected by rotation.”

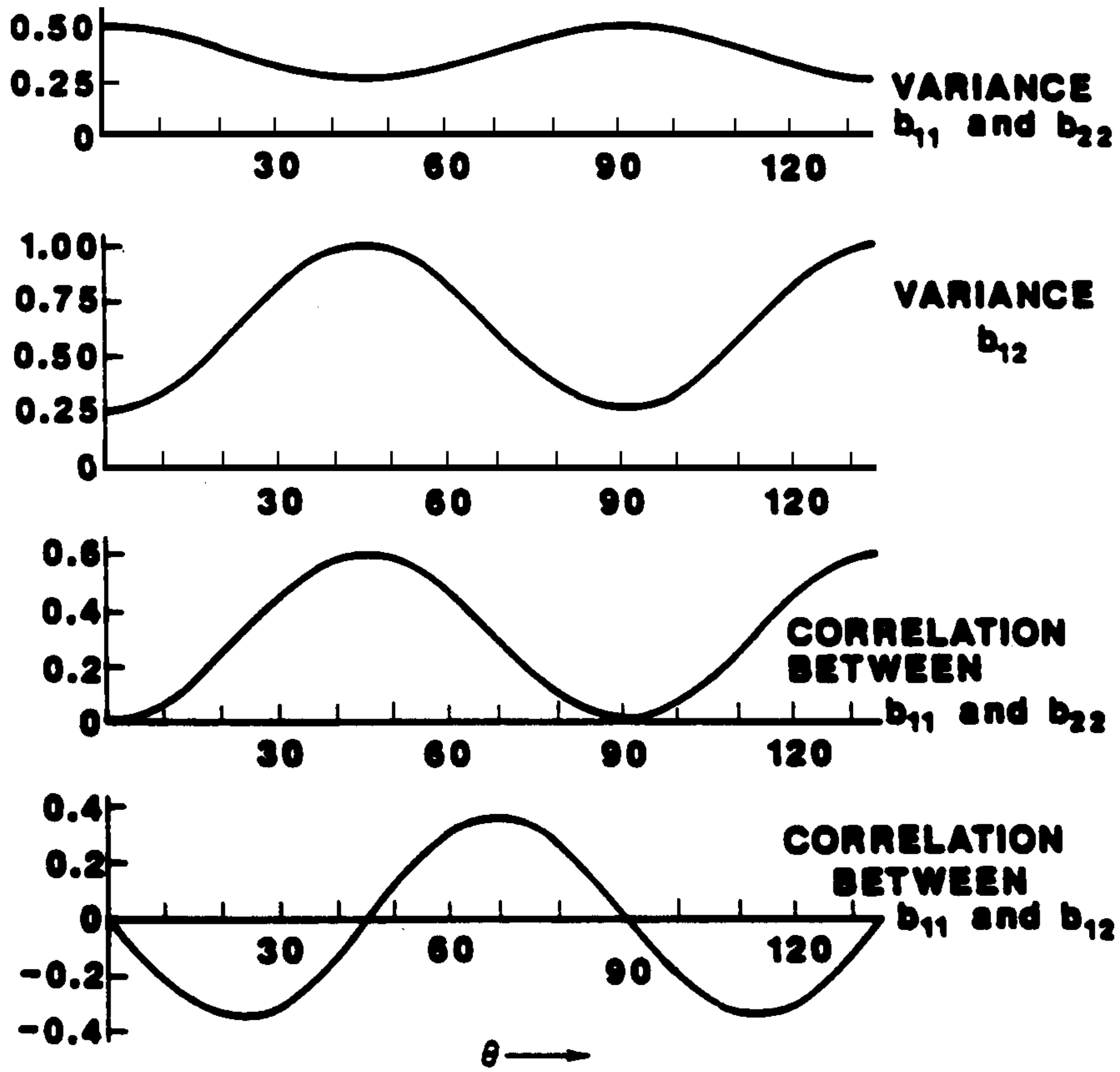


2^2 Design - Rotatable



3^2 Design - Not Rotatable

***Box and Draper, Empirical Model-Building and Response Surfaces, John Wiley and sons, 1987, page 484.**



Variances and correlations between, second-order coefficients estimated from a 3^2 factorial design rotated through various angles

CENTRAL COMPOSITE DESIGNS FOR TWO FACTORS

N = 9

RUN	X_1 EMITTER DOSE	X_2 EMITTER ANNEAL TIME	EMITTER RS
1	6.3E15	21	31.45
2	1E16	21	28.47
3	6.3E15	34	26.96
4	1E16	34	23.62
5	1.2E16	27	25.07
6	5E15	27	31.79
7	7.9E15	42	23.2
8	7.9E15	15	33.52
9	7.9E15	27	27.0

THREE FACTOR CENTRAL COMPOSITE DESIGN EXPERIMENT FOR Nch THRESHOLD.

RUN	Subs Dose	P–well Dose	Blkt Dose	Vtn (mV)
1	(–)	(–)	(–)	429
2	(+)	(–)	(–)	342
3	(–)	(+)	(–)	833
4	(+)	(+)	(–)	776
5	(–)	(–)	(+)	609
6	(+)	(–)	(+)	537
7	(–)	(+)	(+)	962
8	(+)	(+)	(+)	910
9	(+1.682)	(0)	(0)	523
10	(–1.682)	(0)	(0)	669
11	(0)	(+1.682)	(0)	1037
12	(0)	(–1.682)	(0)	369
13	(0)	(0)	(+1.682)	860
14	(0)	(0)	(–1.682)	569
15	(0)	(0)	(0)	644

CENTRAL COMPOSITE DESIGNS FOR FOUR FACTORS

N = 17

Level	-2	-1	0	+1	+2
Base Dose	9.10	9.9	11	12.1	12.8
Base Energy	123	130	140	150	167
Base anneal	13	20	30	40	47
SI Etched	50	100	150	200	250

RUN	**X_1 BASE DOSE**	**X_2 BASE ENERGY**	**X_3 BASE FOR ANNEAL TIME**	**$X_4 = X_1 \bullet X_2 \bullet X_3$ SI ETCHED**	**Hfe**
1	-	-	-	-	114
2	+	-	-	+	106
3	-	+	-	+	71
4	+	+	-	-	52
5	-	-	+	+	129
6	+	-	+	-	86
7	-	+	+	-	63
8	+	+	+	+	56
9	2	0	0	0	64
10	-2	0	0	0	96
11	0	2	0	0	42
12	0	-2	0	0	138
13	0	0	2	0	76
14	0	0	-2	0	78
15	0	0	0	2	86
16	0	0	0	-2	69
17	0	0	0	0	77

STATISTICS DECISION TREE

Multiple Input Variables

Compare Proportions

Chi-Square Test

Screening Experiments

Full Factorial
Fractional Factorial

Analysis of Experiments

ANOVA
Multiple Linear Regression

Response Surface Modeling

Box-Behnken Designs
Central Composite Designs
Multiple Linear Regression
Stepwise Regression
Contour Plots
3 D Mesh Plots

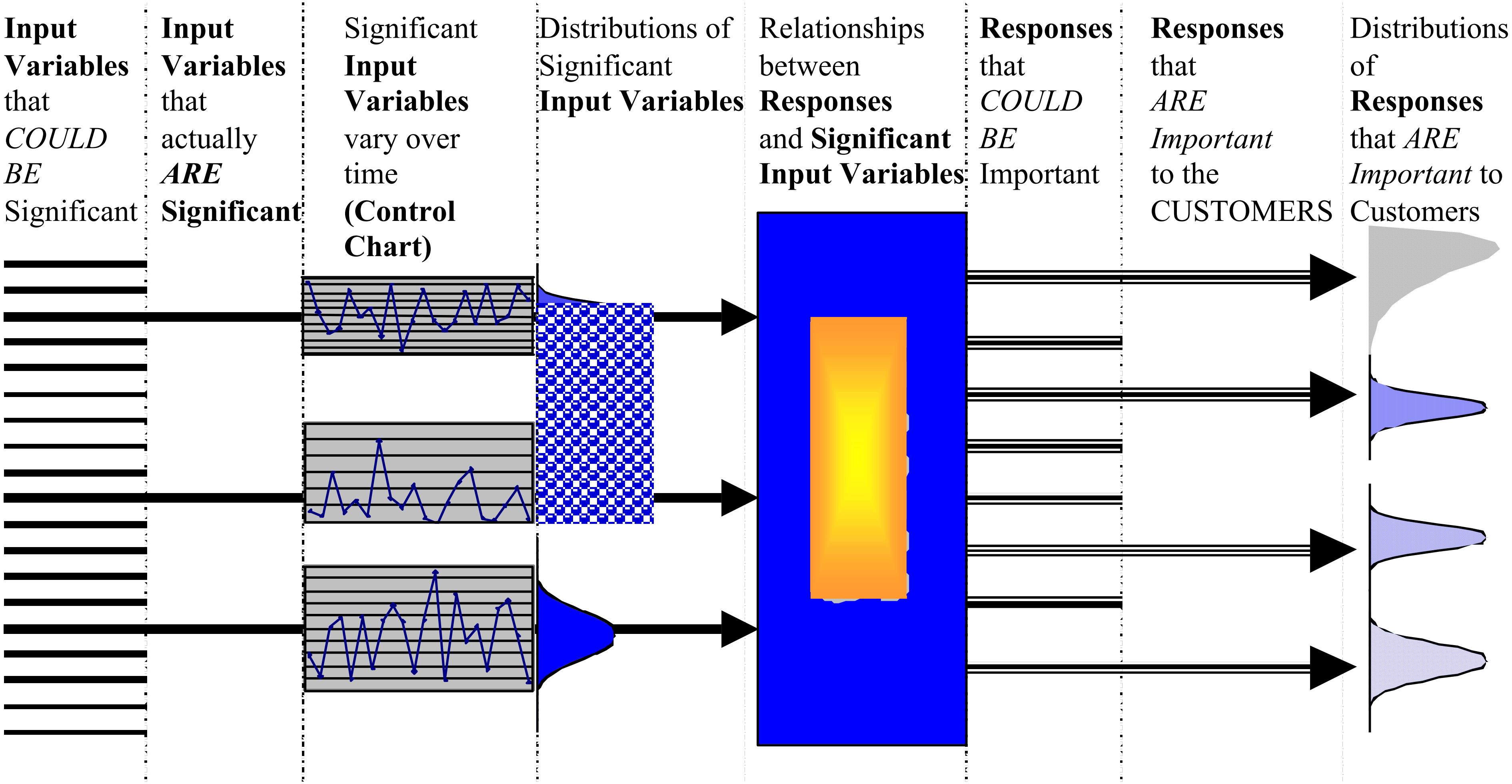
Model Response Distribution

Monte Carlo Simulation
Generation of System Moments

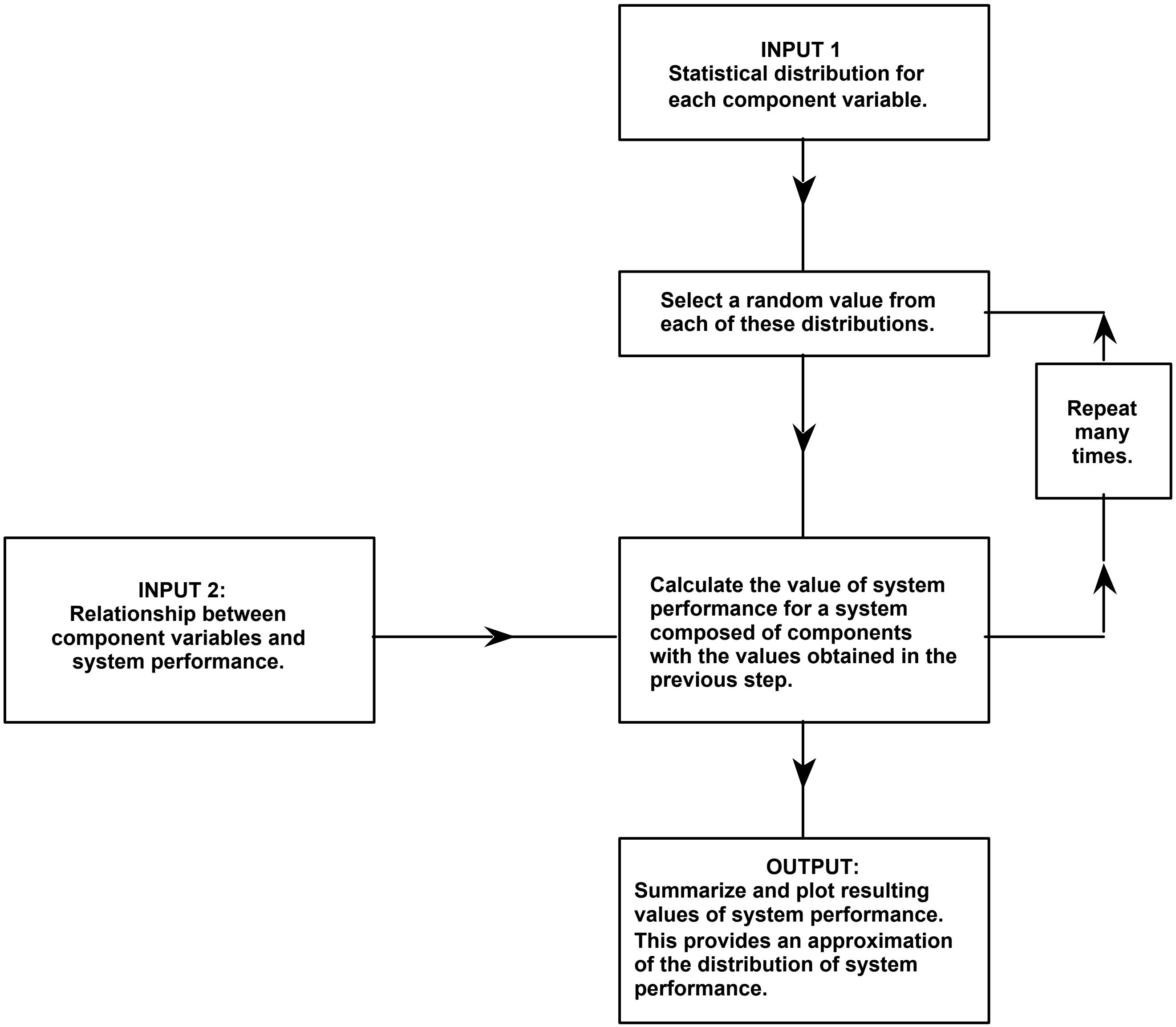
Optimization

Optimization of Expected Value:
Linear Programming
Non Linear Programming
Yield Surface Modeling™

KNOWLEDGE OF A SYSTEM



FLOW CHART OF
MONTE CARLO SIMULATION METHOD.



INPUT 1
Statistical distribution for each component variable.

Select a random value from each of these distributions.

Repeat many times.

INPUT 2:
Relationship between component variables and system performance.

Calculate the value of system performance for a system composed of components with the values obtained in the previous step.

OUTPUT:
Summarize and plot resulting values of system performance. This provides an approximation of the distribution of system performance.

```
MTB >  Name C1  ‘RS’
MTB >  Random 1000 C1;
SUBC >  Normal 196  3.55.
MTB >  Name C2  ‘CD’
MTB >  Random 1000 C2;
SUBC >  Normal 7.492  .371.
MTB >  Name C3  ‘R’
MTB >  Let C3 = C1*50/C2
MTB >  DESC  C3
```

```
MTB >  Histo C3

Histogram of R   N  =  1000
Each * represents 10 obs

Midpoint   Count

  1100          1   *
  1150          8   *
  1200         82   ********
  1250        200   ********************
  1300        299   ******************************
  1350        234   ************************
  1400        124   *************
  1450         39   ****
  1500         10   *
  1550          3   *
```

```
MTB >
MTB >  Desc  C1 - C3


          N       Mean     Median    TRMean     StDev    SeMean
       1000     195.87     195.67    195.87      3.55      0.11
       1000     7.4738     7.4643    7.4721    0.3628    0.0115
       1000     1313.4     1311.2    1312.4      67.3       2.1

        Min        Max         Q1        Q3
     184.43     206.66     193.35    198.42
     6.4271     8.5193     7.2322    7.7163
     1117.8     1546.2     1268.1    1355.7
```

Generation of system moments method:

Mean of R:

R-bar = RS-bar * L / CD-bar = 196*50 / 7.492 = 1308.1

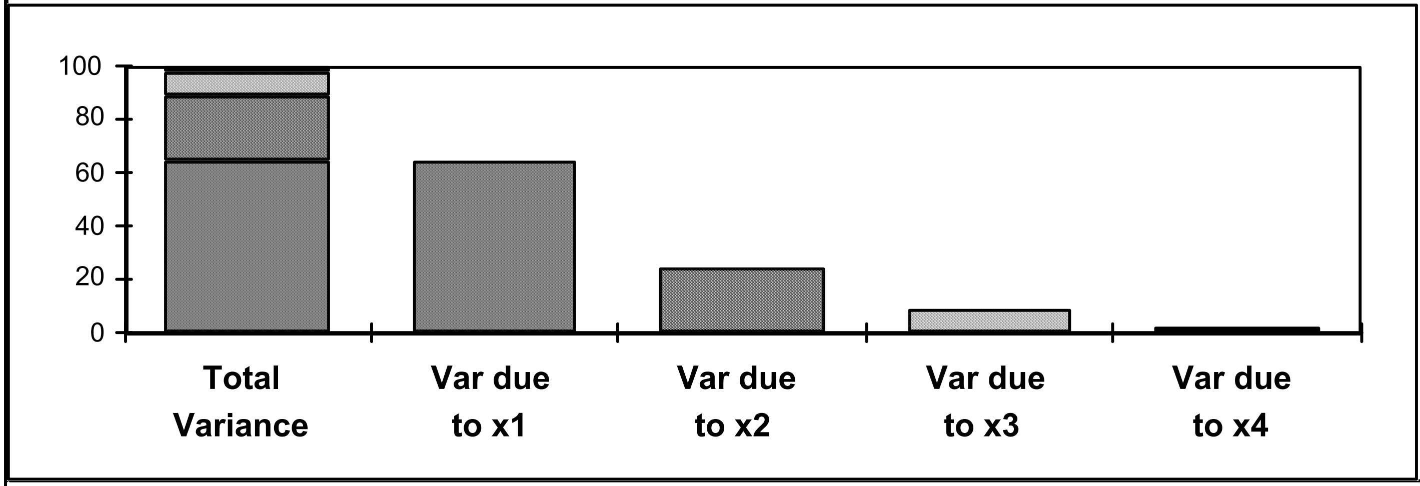
COMPARISON OF METHODS

Monte Carlo simulation has more intuitive appeal than does the generation of system moments and consequently is easier to understand. The desired precision can be obtained by conducting sufficient trials. Also, the **Monte Carlo method is very flexible** and can be applied to many highly complex situations for which the method of generation of system moments becomes too difficult. This is especially true when there are interrelationships between the component variables.

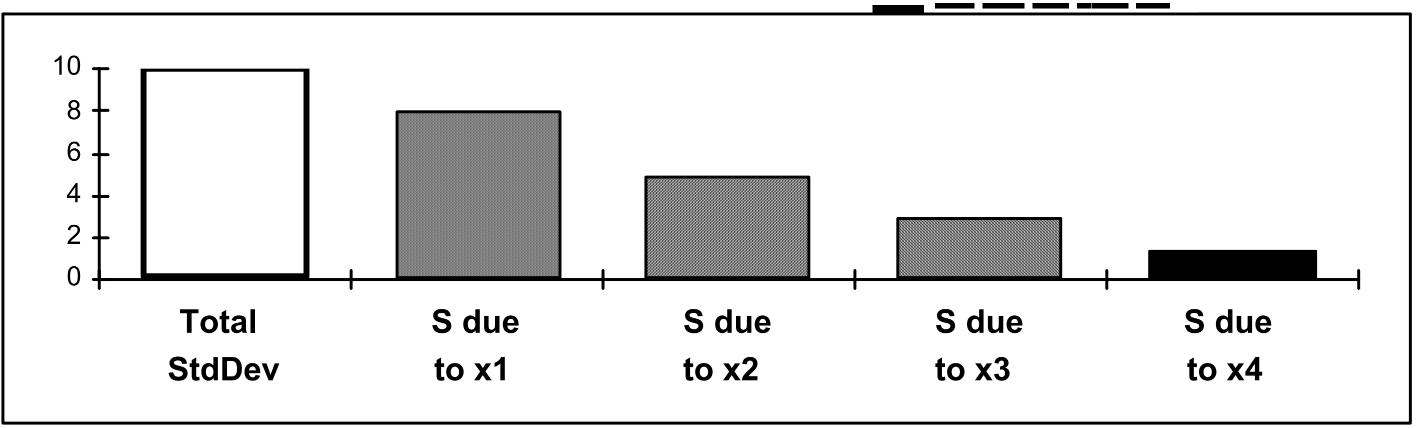
A major drawback of the Monte Carlo method is that there is frequently no way of determining whether any of the variables are dominant or more important than others. Furthermore, if a change is made in one variable, the entire simulation must be redone. Also, **the method generally requires developing a complex computer program**; and if a large number of trials are required, a **great deal of computer time may be needed** to obtain the necessary answers.

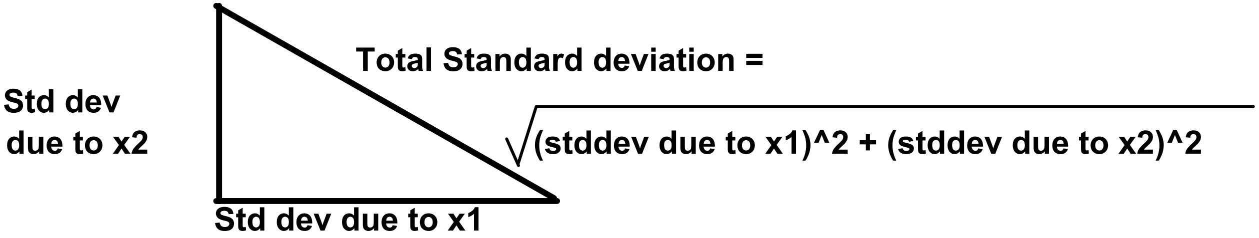
Consequently, the generation of system moments, in conjunction with a Pearson or Johnson distribution approximation, is **sometimes the most economical approach**. Although the precision of the answers usually cannot be easily assessed for this method, the results of the study suggest that this approach often does provide an adequate approximation. In addition, the **generation of system moments allows us to analyze the relative importance of each component variable by examining the magnitude of its partial derivative.**

ADDITIVITY OF VARIANCES



STANDARD DEVIATIONS COMBINE





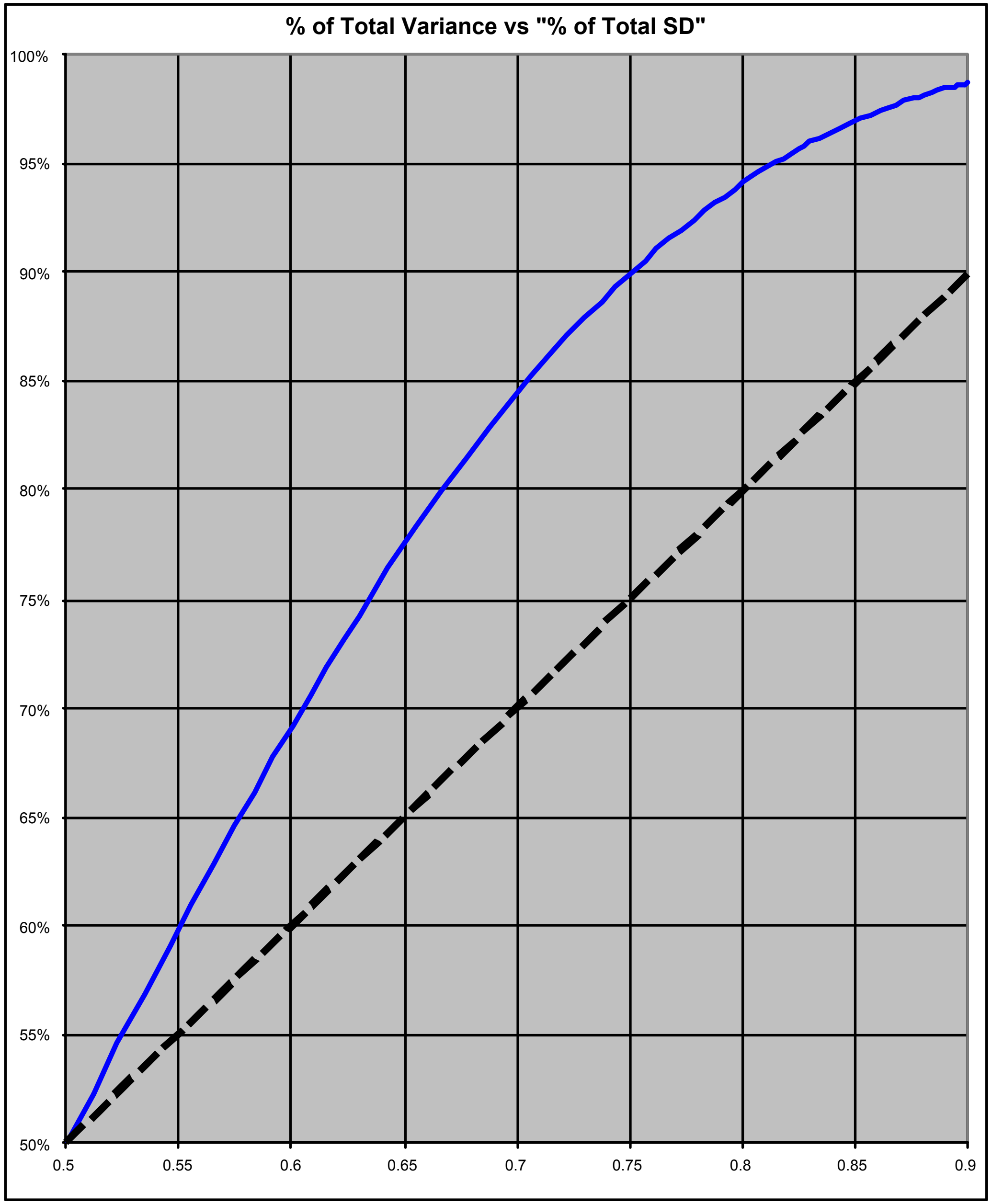
Total Standard deviation = $\sqrt{(\text{stddev due to x1})^2 + (\text{stddev due to x2})^2}$

Since Variances add,

the larger standard deviation's

impact is MAGNIFIED.

This is the theoretical basis for the concept of the **RED X**



GENERATION OF SYSTEM MOMENTS/

Propagation of Errors

- Derived from a multivariate Taylor series expansion of P = f(X1, X2, ... Xn)
- Retaining the terms up to third order, and assuming that the component variables (process factors) are **uncorrelated**:

$$\boxed{[S(P)]^2 = \sum_{i=1}^{n} \left[\frac{\partial P}{\partial X_i} \cdot S(X_i)\right]^2 + \sum_{i=1}^{n} \left(\frac{\partial P}{\partial X_i}\right)\left(\frac{\partial 2P}{\partial X_i^{\,2}}\right) \mu_3 (X_i)}$$

Where: S(P) = Standard deviation of device parameter P
S(Xi) = Standard deviation of process factor Xi
μ3(Xi) = Third central moment of process factor Xi

Neglecting the last term, the variance of device parameter P can be partitioned into the variance due to each process factor:

$$\boxed{[S(P_i)]^2 = \left[\frac{\partial P}{\partial X_i} \cdot S(X_i)\right]^2}$$

GENERATIONS OF SYSTEM MOMENTS METHOD: SD OF RESISTOR VALUE

dr/dRs = L / CD = 50 / 7.492 = 6.674

Variance of R due to Rs = [(dR/dRs)*Srs]^2 = (6.674*3.55)^2 = 561.3

dR/dCD = - Rs * L / (CD)^2 = - (196)*(50) / (7.492)^2 = -174.6

Variance of R due to CD = [(dR/dCD)*Scd]^2 = (-174.6*.371)^2 = 4195.7

Variance of R = 561.3 + 4195.7 = 4757;

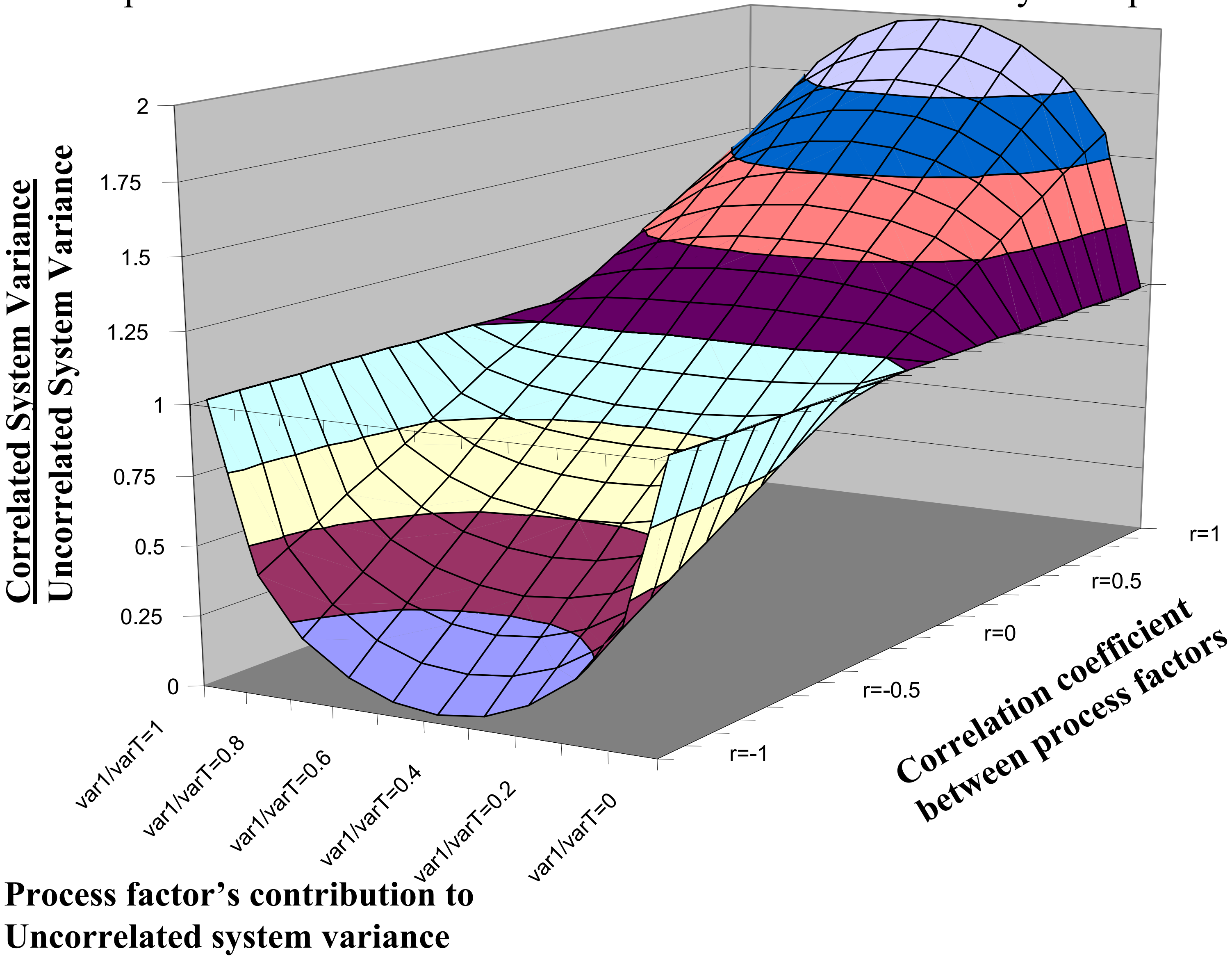
Stdev of R = Sqrt (4757) = 69.0

Relative importance of input parameters for resistor variability:

4195.7 / 4757 $\Rightarrow$ 88% of resistor variance is due to CD variability

Propagation of Errors assumes that the process factors are **uncorrelated**.

If the process factors are correlated, the impact on the system variance can vary from offsetting the uncorrelated variance to doubling it depending on the **positive** or **negative correlation** between the process factors and the amount of variance contributed by each process factor.



THRESHOLD VOLTAGE (V_t) VARIANCE EXAMPLE - A CASE STUDY

Model:

$$Vt = \phi ms\,(Nd) + 2\,\phi f\,(Nd) - \frac{Qb\,(Nd)}{Co\,(tox)} - \frac{Qi}{Co\,(tox)}$$

Where:

$\phi ms\,(Nd)$ = The metal-semiconductor work function difference

$\phi f\,(Nd)$ = The Fermi potential

$Qb\,(Nd)$ = The charge per unit area in the surface depletion region at inversion

$Co\,(tox)$ = The gate oxide capacitance per unit area

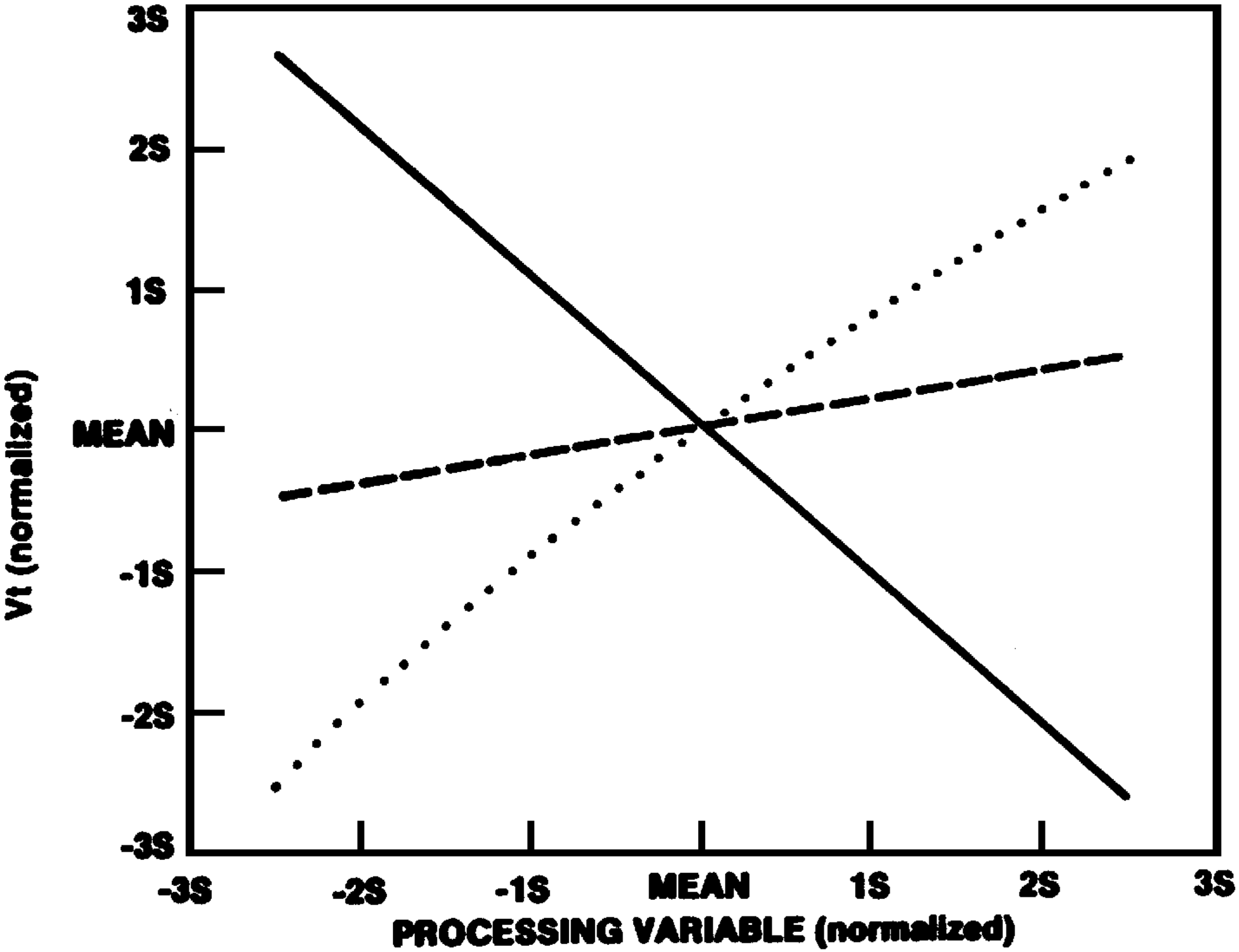
Process Factors:

Nd = The doping concentration in the channel,

tox = The gate oxide thickness, and

Qi = The oxide/interface charge per unit area

- Distributions obtained from CV plots of test pattern capacitors



Normalized threshold voltage as a function of the normalized processing variables:

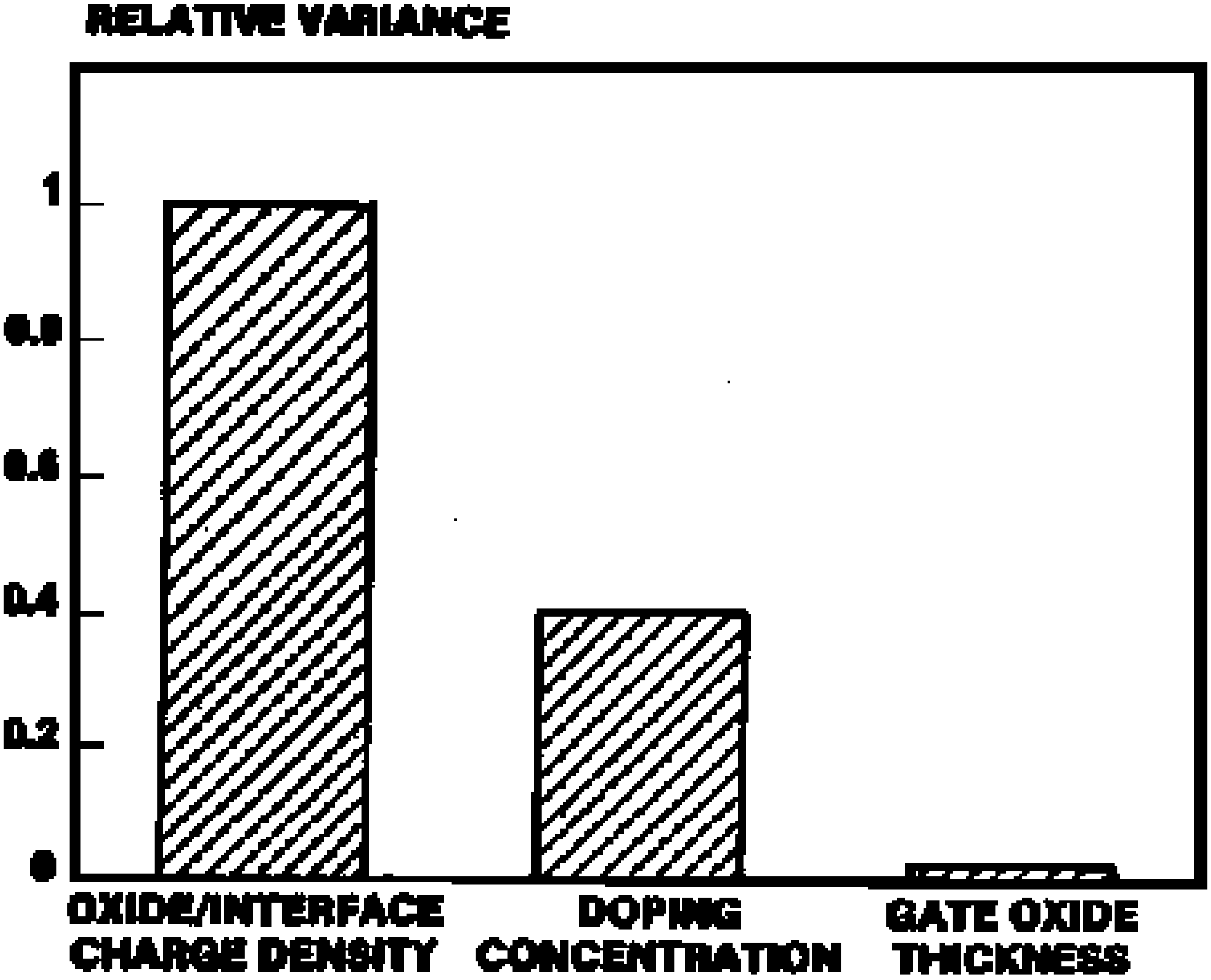
Gate oxide thickness - - - - - - - - - - -

Oxide/interface charge density ————

Concentration of doping in the channel region • • • • • • • • • • • •

VT VARIANCE EXAMPLE

RELATIVE VARIANCE ATTRIBUTED TO SOURCES



BVceo Variance Example -
Second Case Study

Model:

$$BV_{CEO} = \frac{V\infty C^2}{4\sqrt{\beta + 1}}\left[\frac{2W_E}{W_{CEO}C} - \left(\frac{W_E}{W_{CEO}C}\right)^2\right]$$

Where:

BV_{CEO} = Collector-emitter brackdown voltage

$W_{CEO} = W\infty / 8\sqrt{\beta + 1}$

$W\infty = 3.60 \times 10^3 \left(\frac{V\infty}{N_D}\right)^{1/2}$

$V\infty = 60 \left(\frac{N_D}{10^{16}}\right)^{-3/4}$

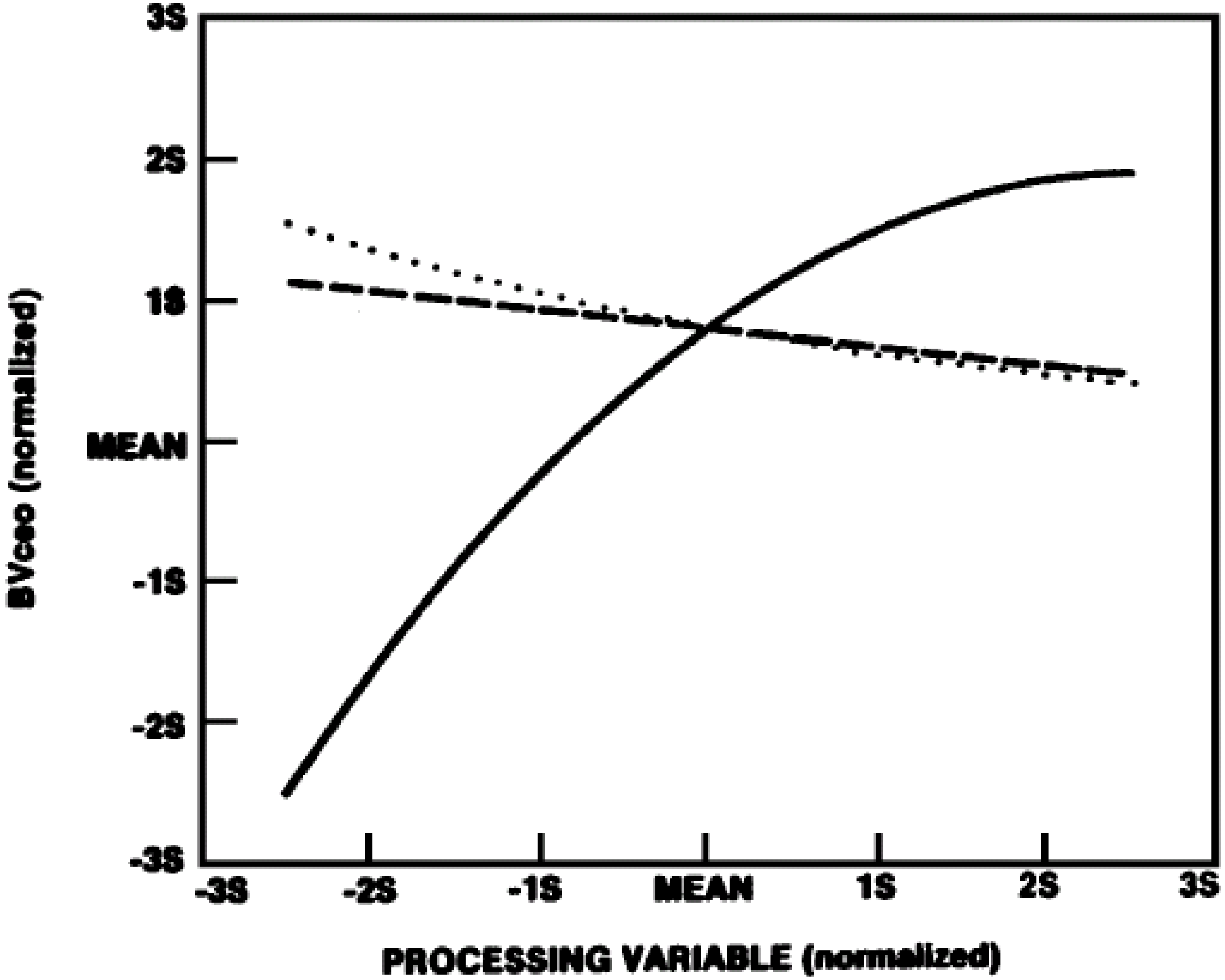
C = A semi-empirical two dimensional correction factor, between 0 and 1

Process Factors:

N_D = The doping concentration of the epitaxial layer

W_E = Intrinsic thickness of epitaxial layer (base to subcollector)

β = NPN current gain



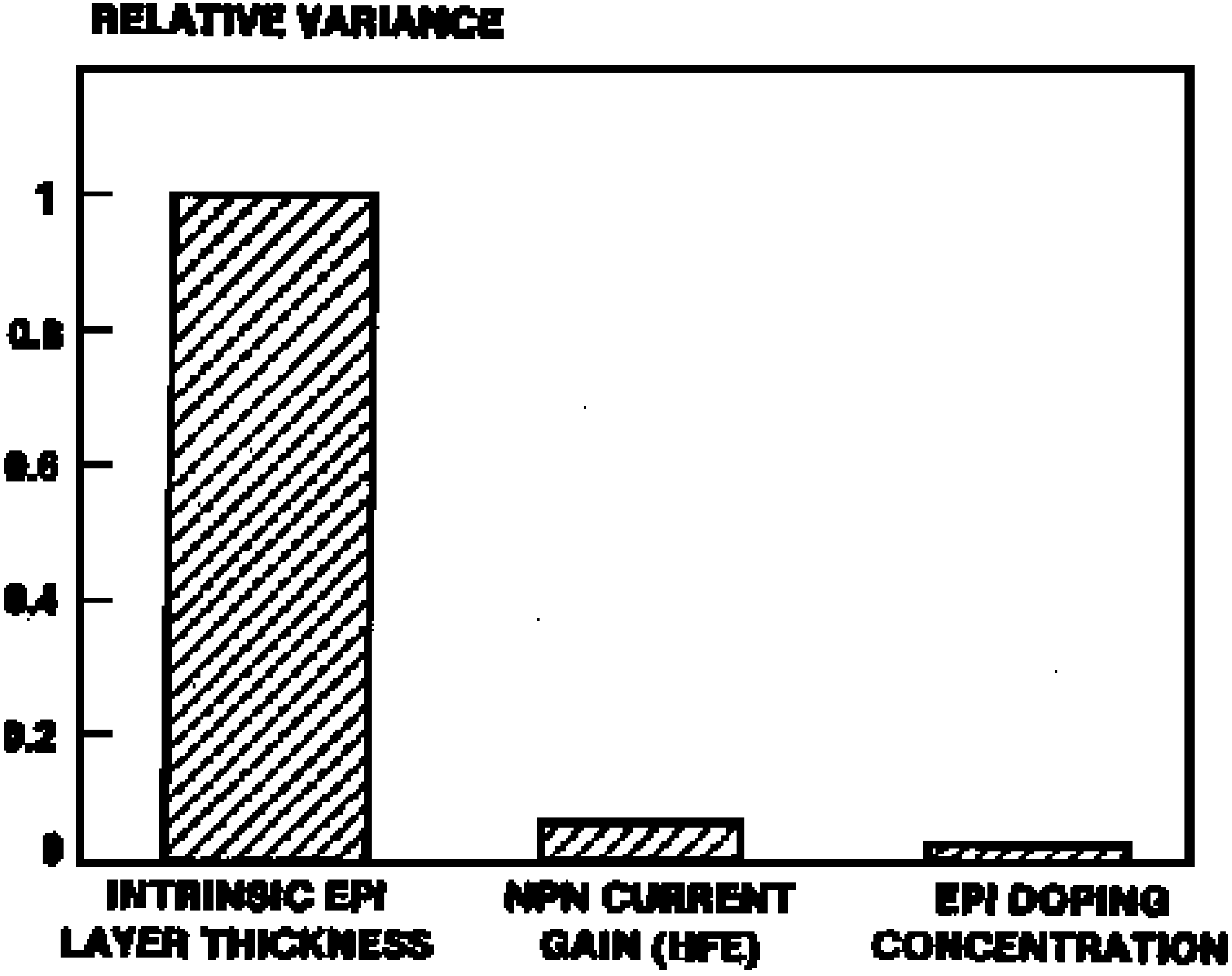
Normalized BVceo as a function of the normalize processing variables:

• • • • • • • • • • • • • NPN current gain (Hfe)

__________ Intrinsic epitaxy layer thickness

– – – – – – – – – – Concentration of doping in the epitaxial layer

BVCEO VARIANCE EXAMPLE
RELATIVE VARIANCE ATTRIBUTED TO SOURCES



DATA vs DATA

One Input Variable

Compare Variability
- F-ratio test (two levels)
- Bartlett's test (multiple levels)
- Cochran's test (multiple levels)

Compare Means
- Student's T Test (two levels)
- ANOVA (multiple levels)
- Nested ANOVA (multiple levels)

Compare Medians
- Mann-Whitney (two levels)
- Kruskal-Wallis (multiple levels)

Study Source of Variation
- Y vs X plot
- Correlation Coefficient
- Linear Regression

Compare Proportions
- Proportion Test
- Chi-Square Test

Multiple Input Variables

Compare Proportions
- Chi-Square Test

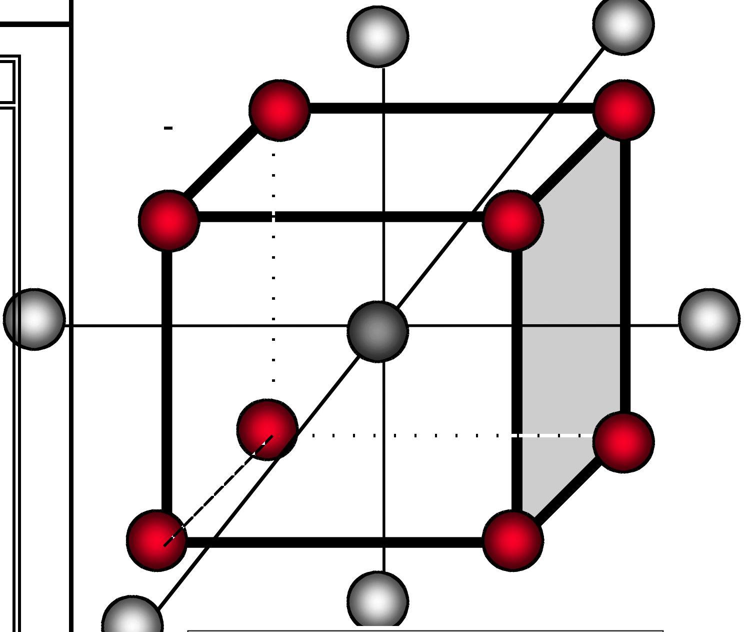
Screening Experiments
- Full Factorial
- Fractional Factorial

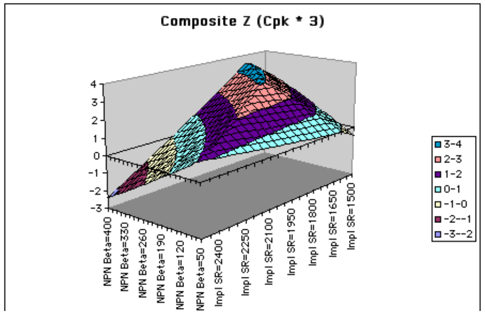
Analysis of Experiments
- ANOVA
- Multiple Linear Regression

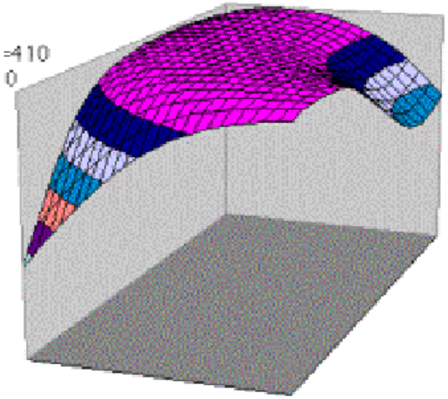
Response Surface Modeling
- Box-Behnken Designs
- Central Composite Designs
- Multiple Linear Regression
- Stepwise Regression
- Contour Plots
- 3 D Mesh Plots

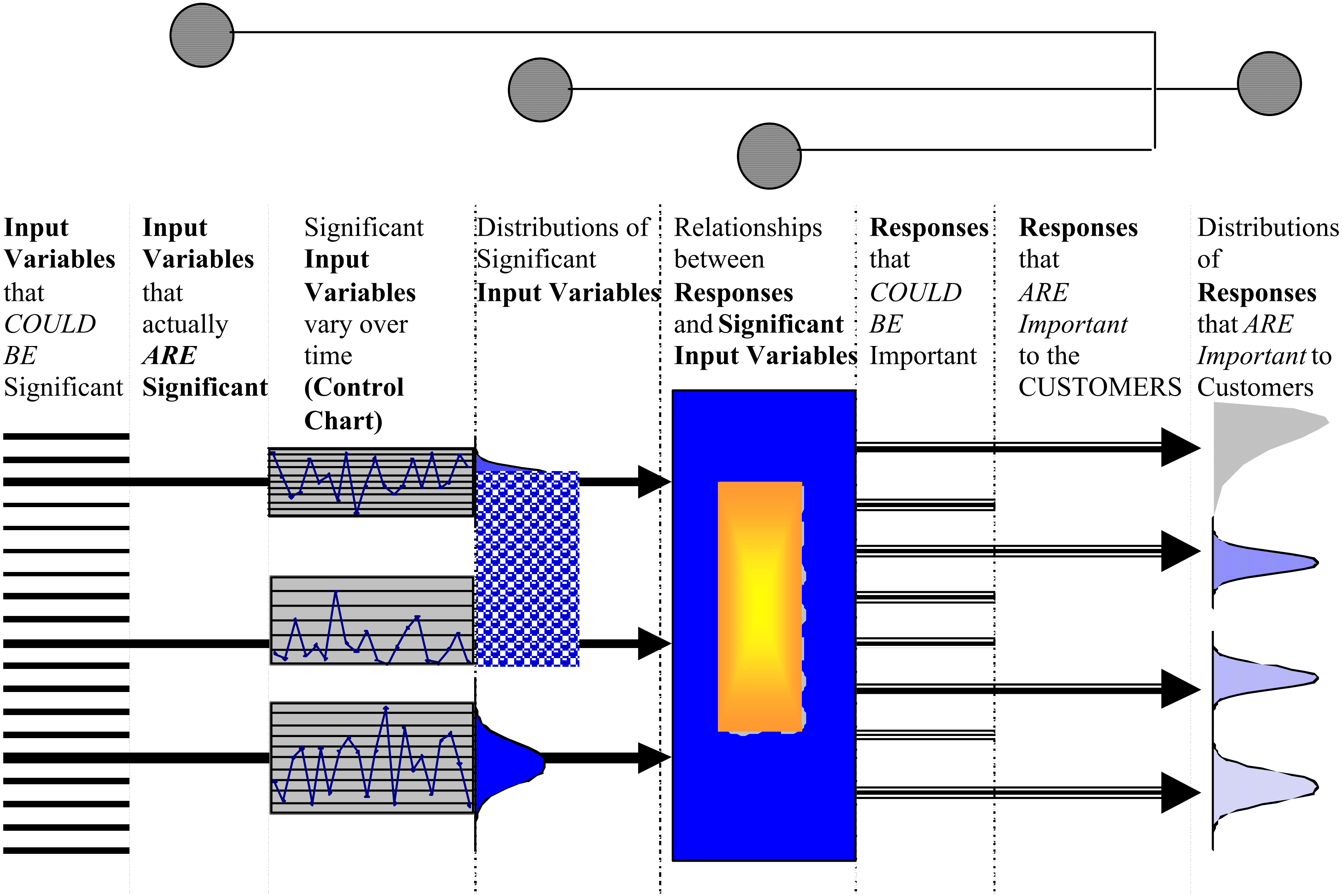
Model Response Distribution
- Monte Carlo Simulation
- Generation of System Moments

Optimization
- Optimization of Expected Value:
- Linear Programming
- Non Linear Programming
- Yield Surface Modeling™









Input Variables that *COULD BE* Significant

Input Variables that actually ***ARE*** **Significant**

Significant **Input Variables** vary over time **(Control Chart)**

Distributions of Significant **Input Variables**

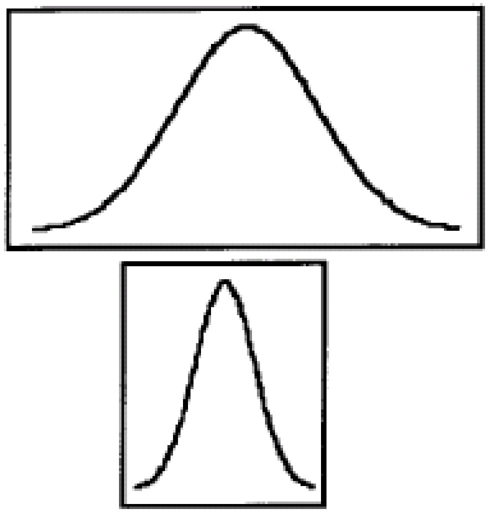
Relationships between **Responses** and **Significant Input Variables**

Responses that *COULD BE* Important

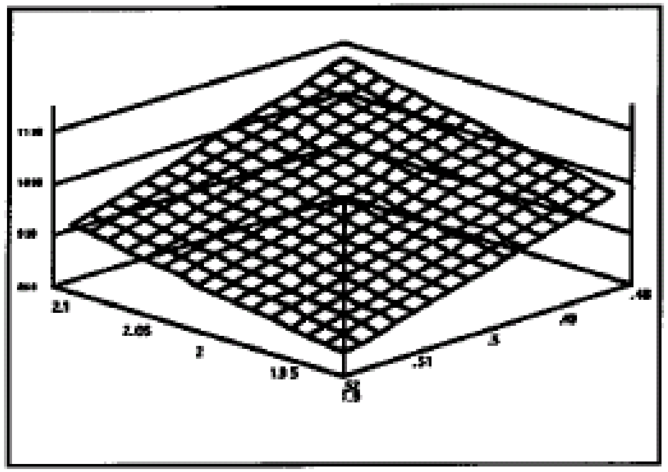
Responses that *ARE Important* to the CUSTOMERS

Distributions of **Responses** that *ARE Important* to Customers

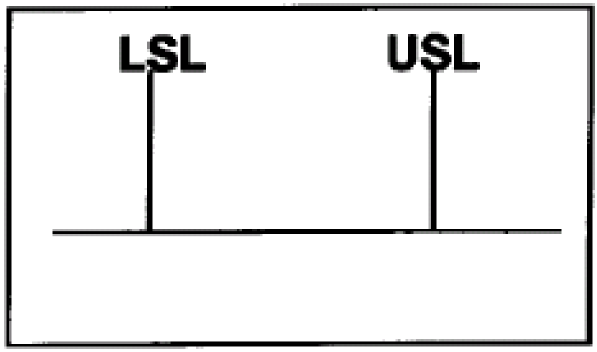
DISTRIBUTIONS



RESPONSE SURFACE MODELING

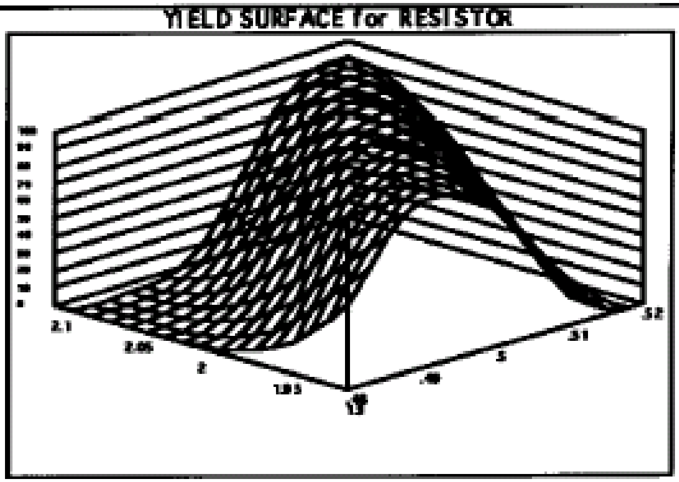


SPEC LIMITS

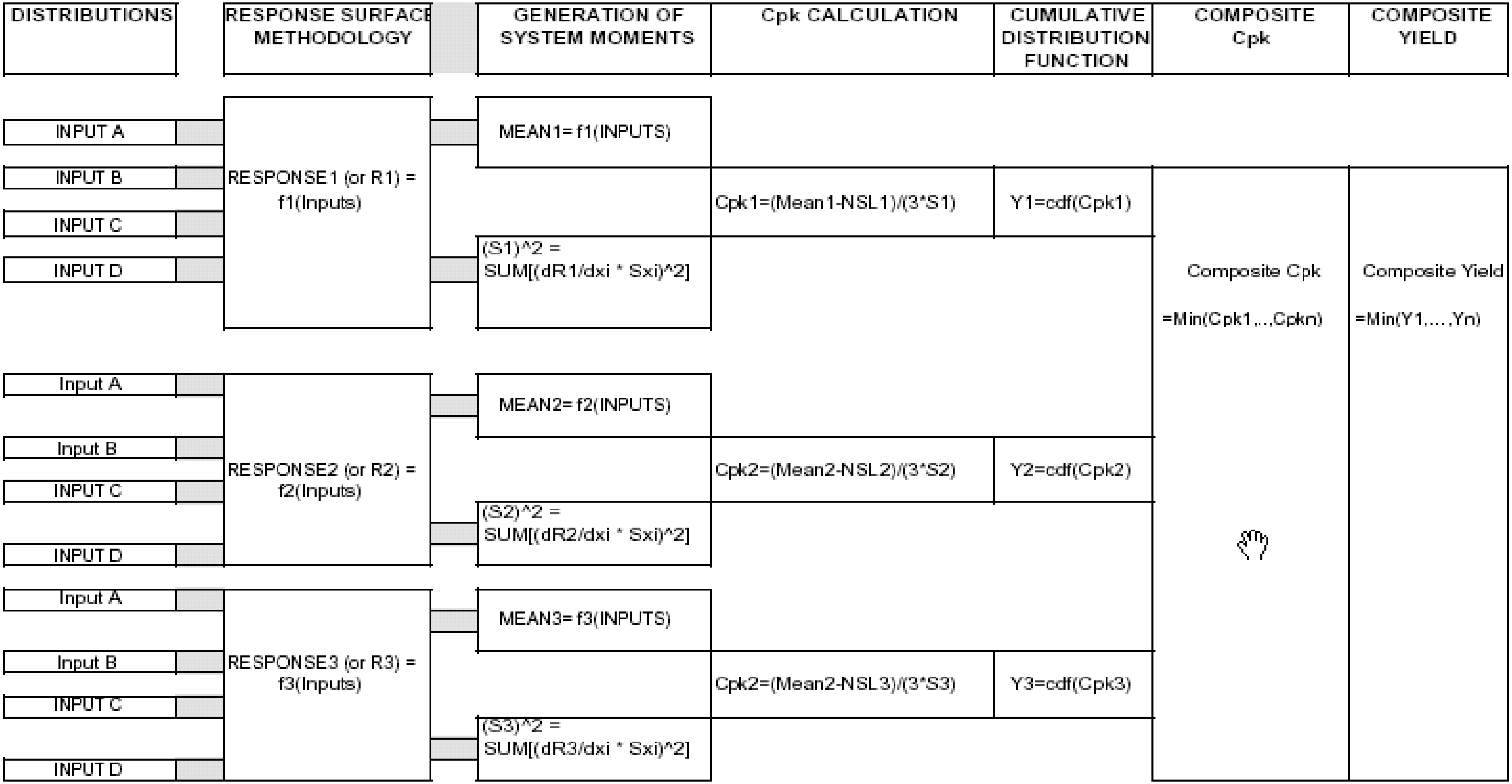


INPUT VARIABLES + **RELATIONSHIPS: INPUTS -> OUTPUTS** + **OUTPUT VARIABLES**

Yield = Surface Modeling


YIELD SURFACE for RESISTOR

Yield Surface Modeling™ - Overview of Method



DISTRIBUTIONS

RESPONSE SURFACE METHODOLOGY

GENERATION OF SYSTEM MOMENTS

Cpk CALCULATION

CUMULATIVE DISTRIBUTION FUNCTION

COMPOSITE Cpk

COMPOSITE YIELD

INPUT A

INPUT B

INPUT C

INPUT D

RESPONSE1 (or R1) = f1(Inputs)

MEAN1= f1(INPUTS)

(S1)^2 = SUM[(dR1/dxi * Sxi)^2]

Cpk1=(Mean1-NSL1)/(3*S1)

Y1=cdf(Cpk1)

Input A

Input B

INPUT C

INPUT D

RESPONSE2 (or R2) = f2(Inputs)

MEAN2= f2(INPUTS)

(S2)^2 = SUM[(dR2/dxi * Sxi)^2]

Cpk2=(Mean2-NSL2)/(3*S2)

Y2=cdf(Cpk2)

Input A

Input B

INPUT C

INPUT D

RESPONSE3 (or R3) = f3(Inputs)

MEAN3= f3(INPUTS)

(S3)^2 = SUM[(dR3/dxi * Sxi)^2]

Cpk2=(Mean2-NSL3)/(3*S3)

Y3=cdf(Cpk3)

Composite Cpk

=Min(Cpk1,..,Cpkn)

Composite Yield

=Min(Y1,...,Yn)

Statistics Decision Tree

DATA

Look at Distribution

Histogram
Stem-and-Leaf

Describe Distribution-Moments

Mean
Standard Deviation/Variance
Skewness
Kurtosis

Determine Type of Distribution

Normal
Beta
Gamma
Exponential
Log Normal
General: Pearson Distributions

Test - Type of Distribution

Normal Probability Plot
Correlation Test for Normality
Chi Square Test for Distribution

Compare Distribution to Limits

Cp
Cpk
Variance from Target

DATA vs TIME

Look at Trend versus Time

Trend Chart

Model Distribution vs Time

Time Series Modeling
Autocorrelation
Partial Autocorrelation
Moving Average
EWMA
AR
MA
ARIMA

Study Sources -Time Variation

Gauge Capability
Variance Components Analysis

Compare Trend to Limits

Control Charts
X-Bar
R, S
Individuals
Moving R
EWMA

DATA vs DATA

One Input Variable

Compare Variability

F-ratio test (two levels)
Bartlett's test (multiple levels)
Cochran's test (multiple levels)

Compare Means

Student's T Test (two levels)
ANOVA (multiple levels)
Nested ANOVA (multiple levels)

Compare Medians

Mann-Whitney (two levels)
Kruskal-Wallis (multiple levels)

Study Source of Variation

Y vs X plot
Correlation Coefficient
Linear Regression

Compare Proportions

Proportion Test
Chi-Square Test

Multiple Input Variables

Compare Proportions

Chi-Square Test

Screening Experiments

Full Factorial
Fractional Factorial

Analysis of Experiments

ANOVA
Multiple Linear Regression

Response Surface Modeling

Box-Behnken Designs
Central Composite Designs
Multiple Linear Regression
Stepwise Regression
Contour Plots
3 D Mesh Plots

Model Response Distribution

Monte Carlo Simulation
Generation of System Moments

Optimization

Optimization of Expected Value:
Linear Programming
Non Linear Programming
Yield Surface Modeling™