Probability Distribution

1. In scenario 2, the particle size distribution from the mill is:

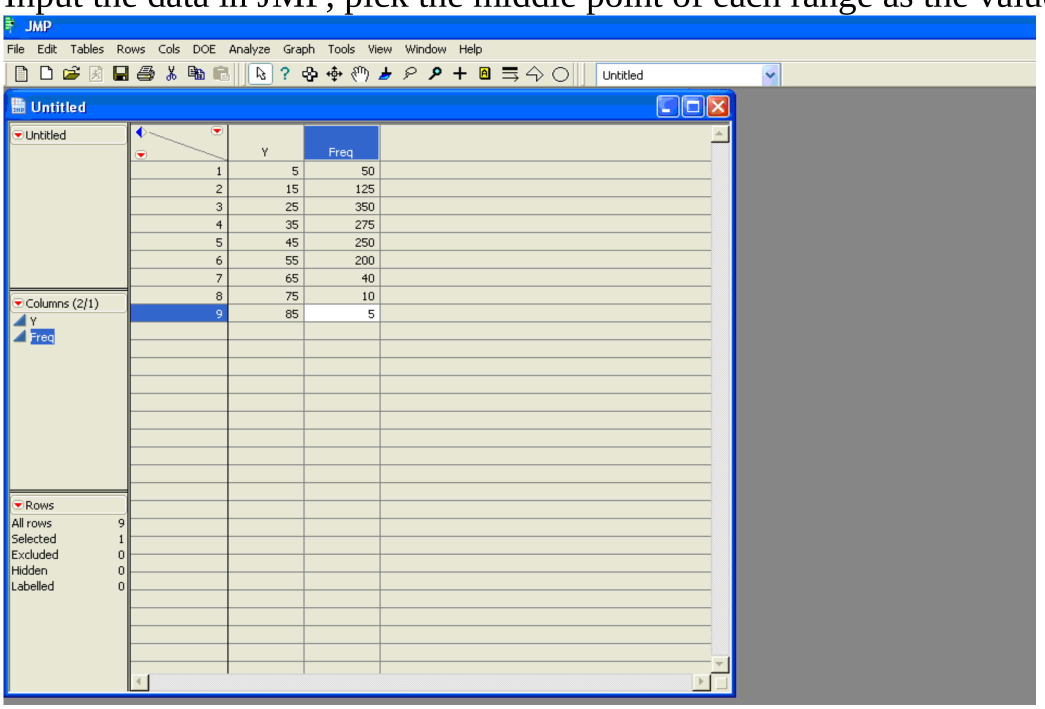
	Counts
<10 mm	50
11-20 mm	125
21-30 mm	350
31-40 mm	275
41-50 mm	250
51-60 mm	200
61-70 mm	40
71-80 mm	10
>81 mm	5

Use JMP to perform the following:

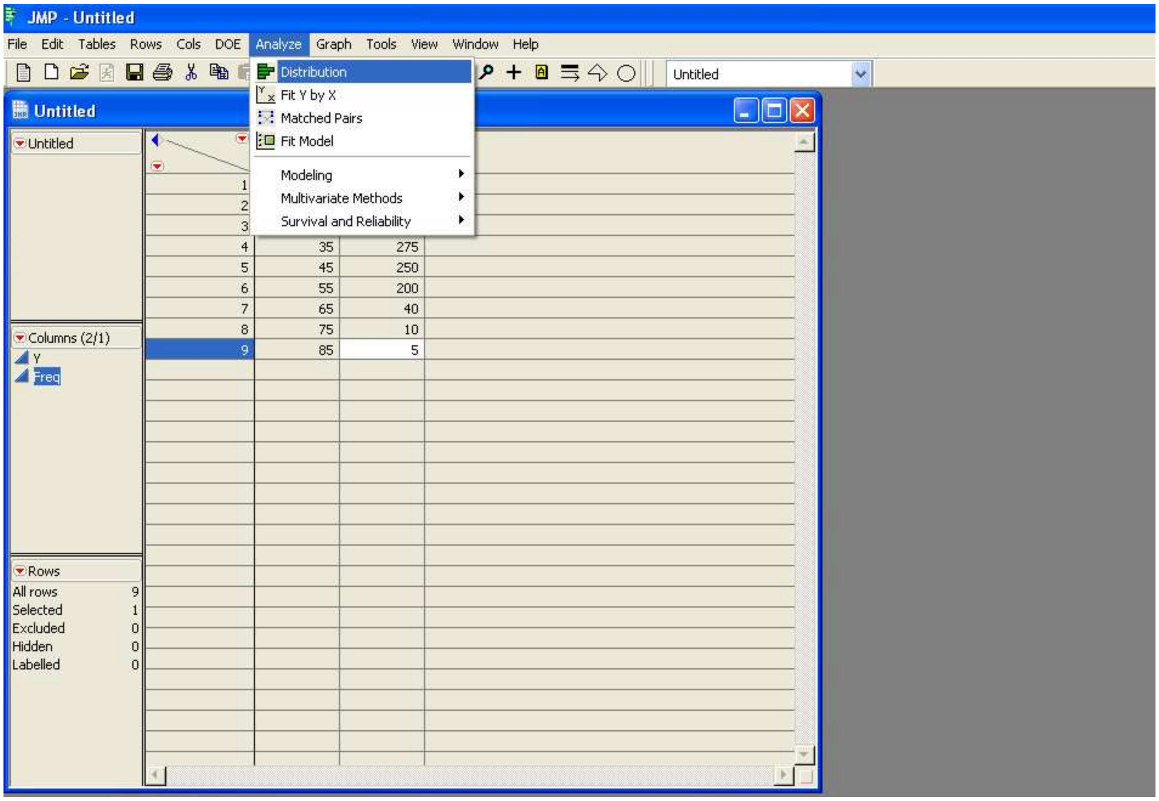
(1) Distribution of Counts Vs Size
(2) % Distribution Vs Size
(3) Mean
(4) Variance

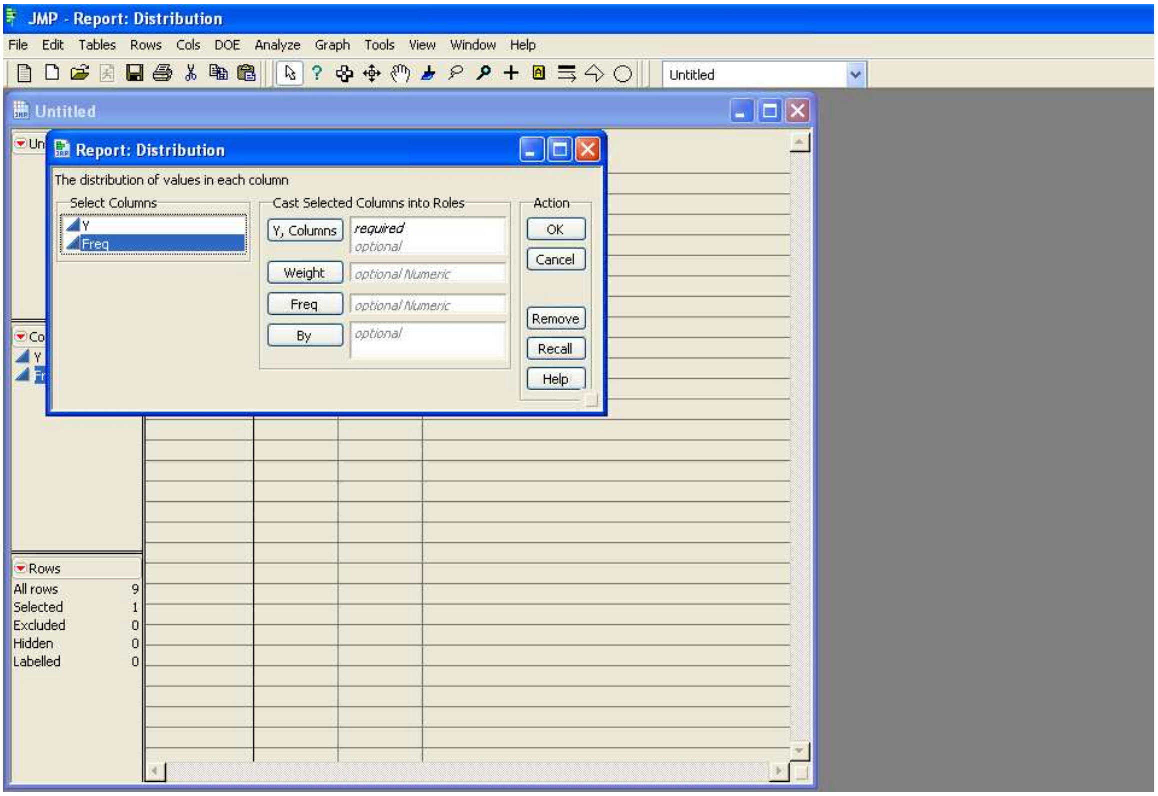
Solution:

Input the data in JMP, pick the middle point of each range as the value of Y:

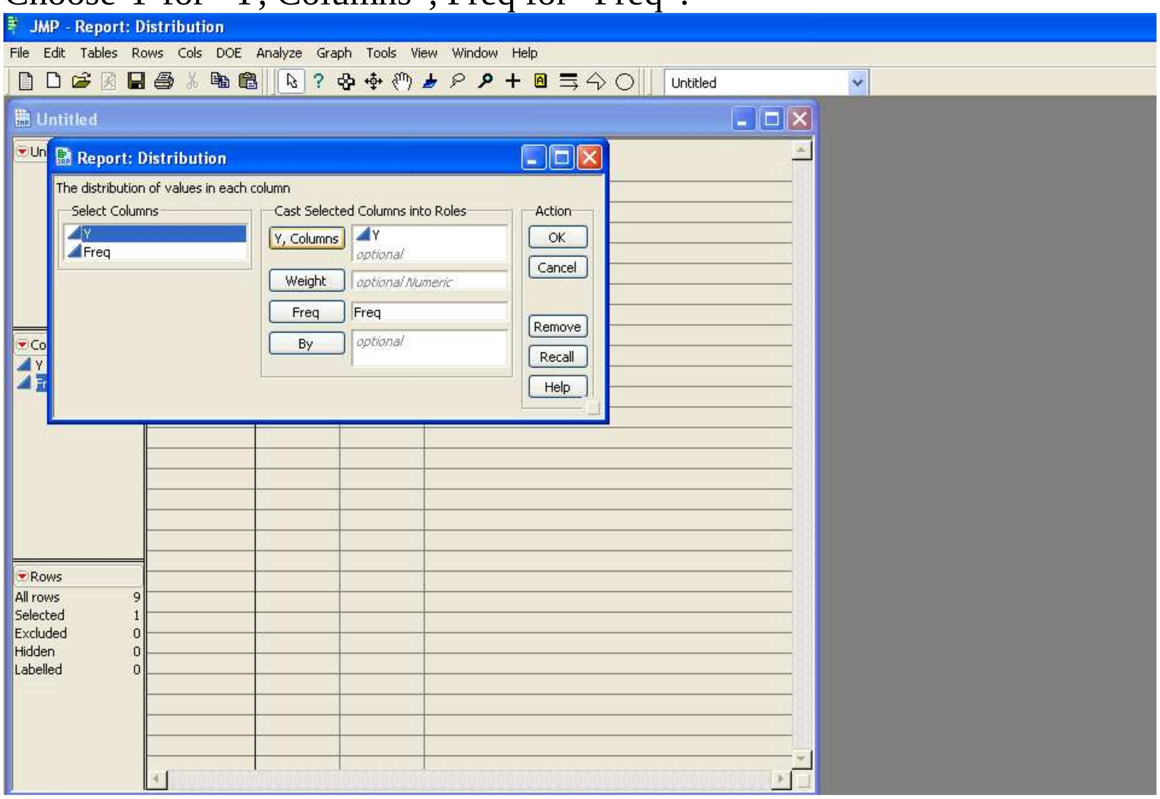


Choose “Distribution” in “Analyze”:

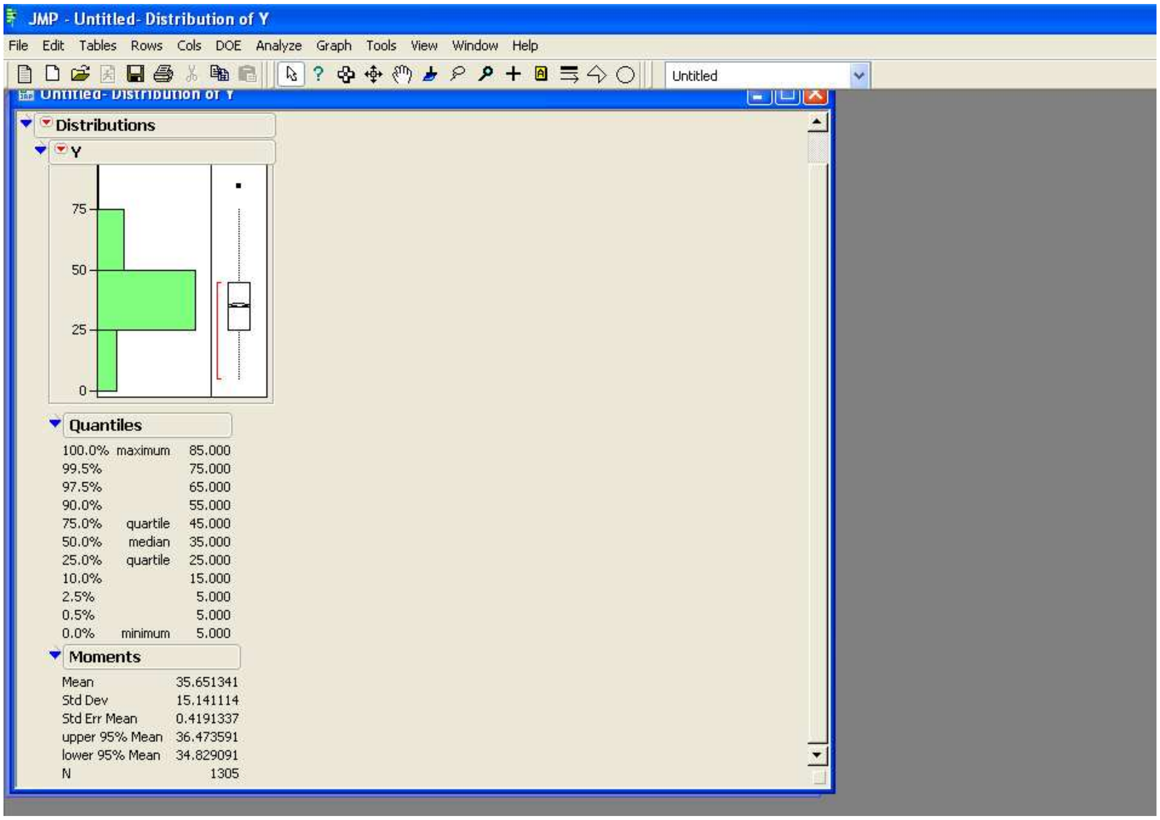




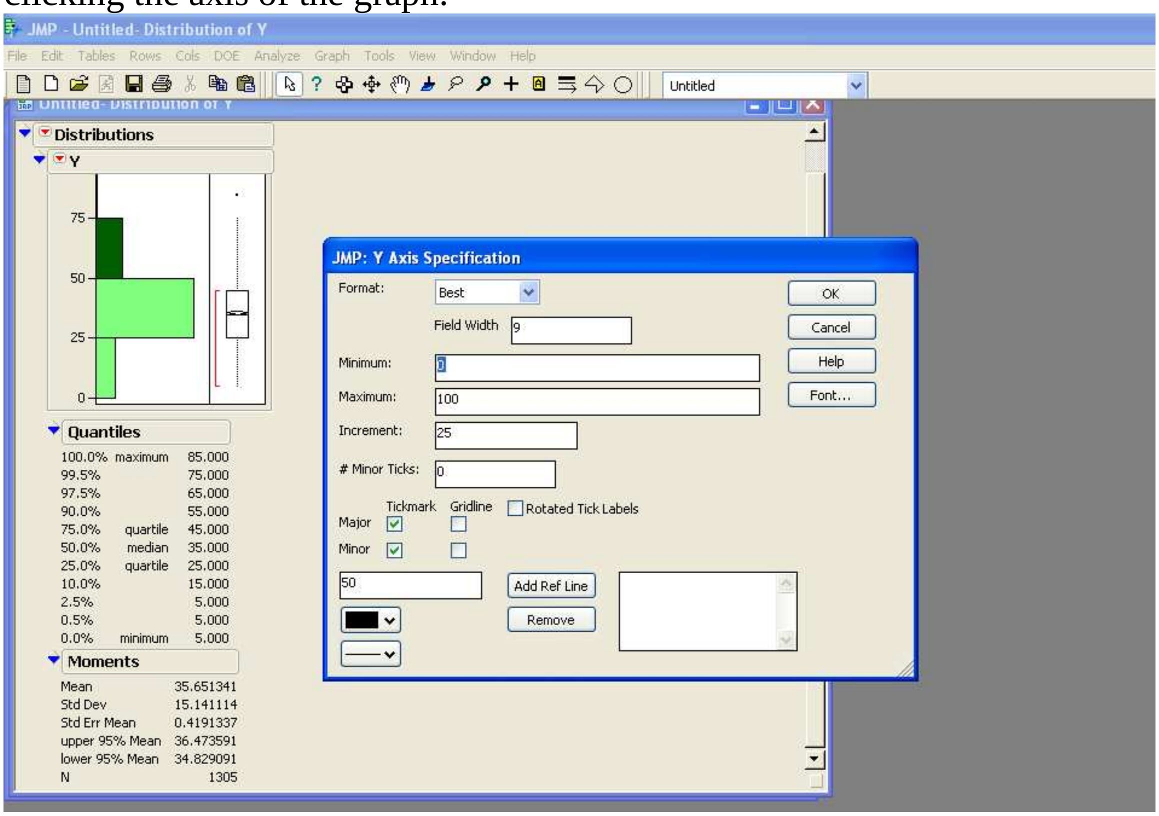
Choose Y for “Y, Columns”, Freq for “Freq”.



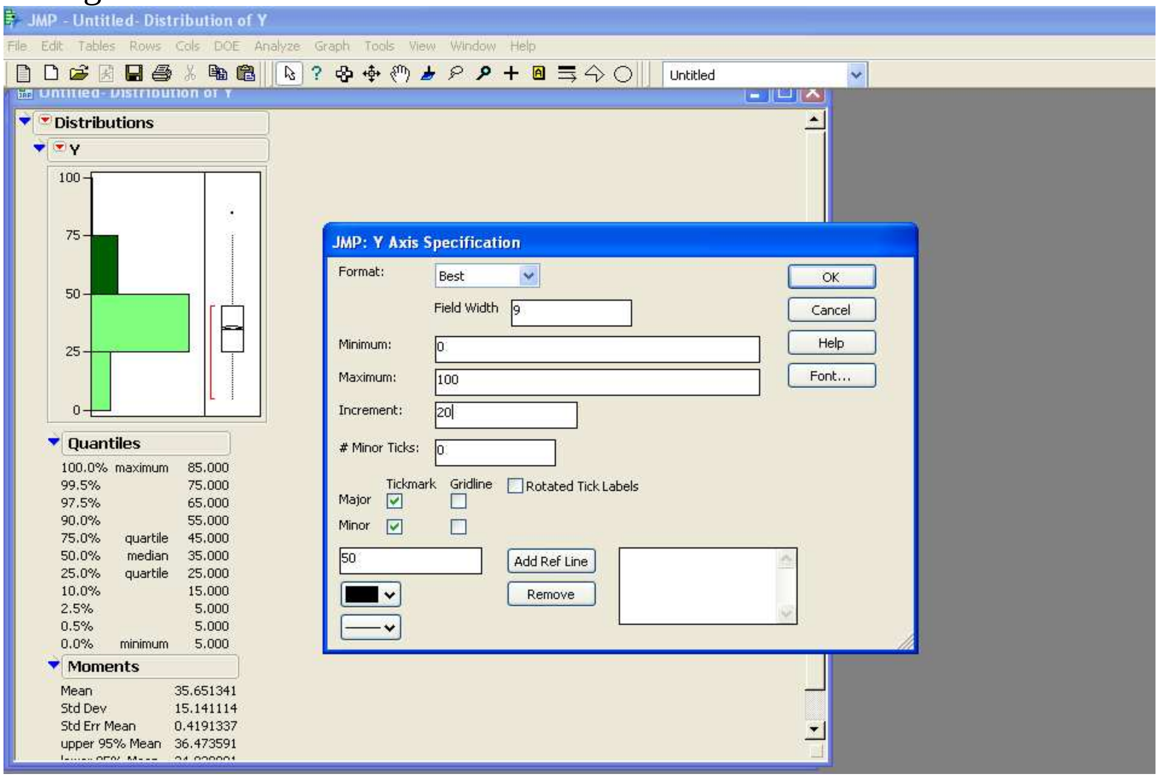
Click “Ok”:



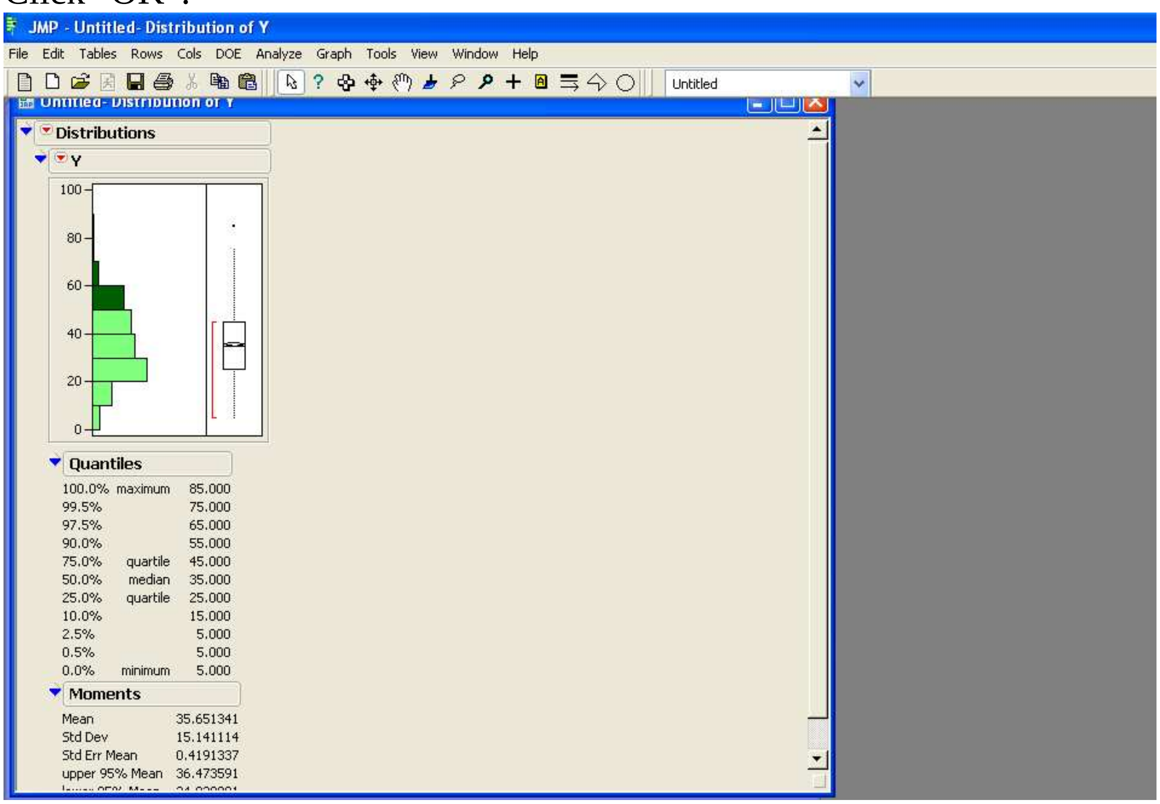
Here we could change the width of the columns in the graph by double clicking the axis of the graph:



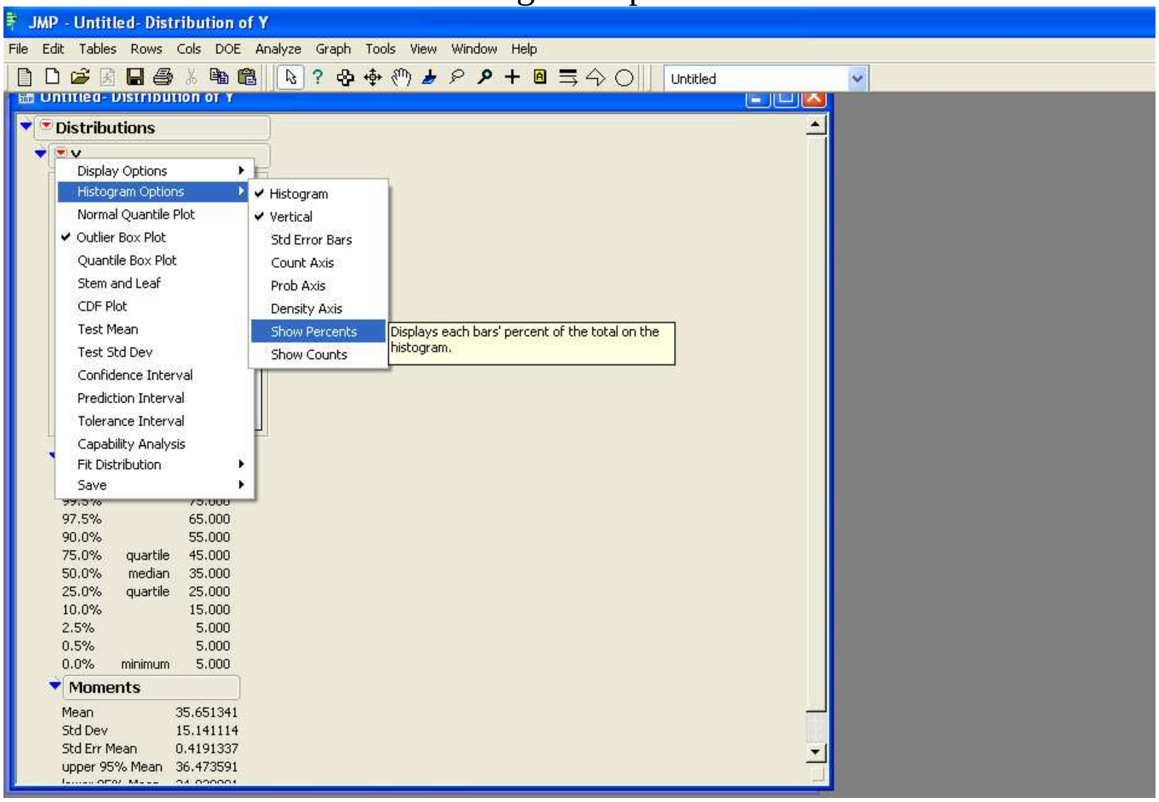
Change the Increment to 20:

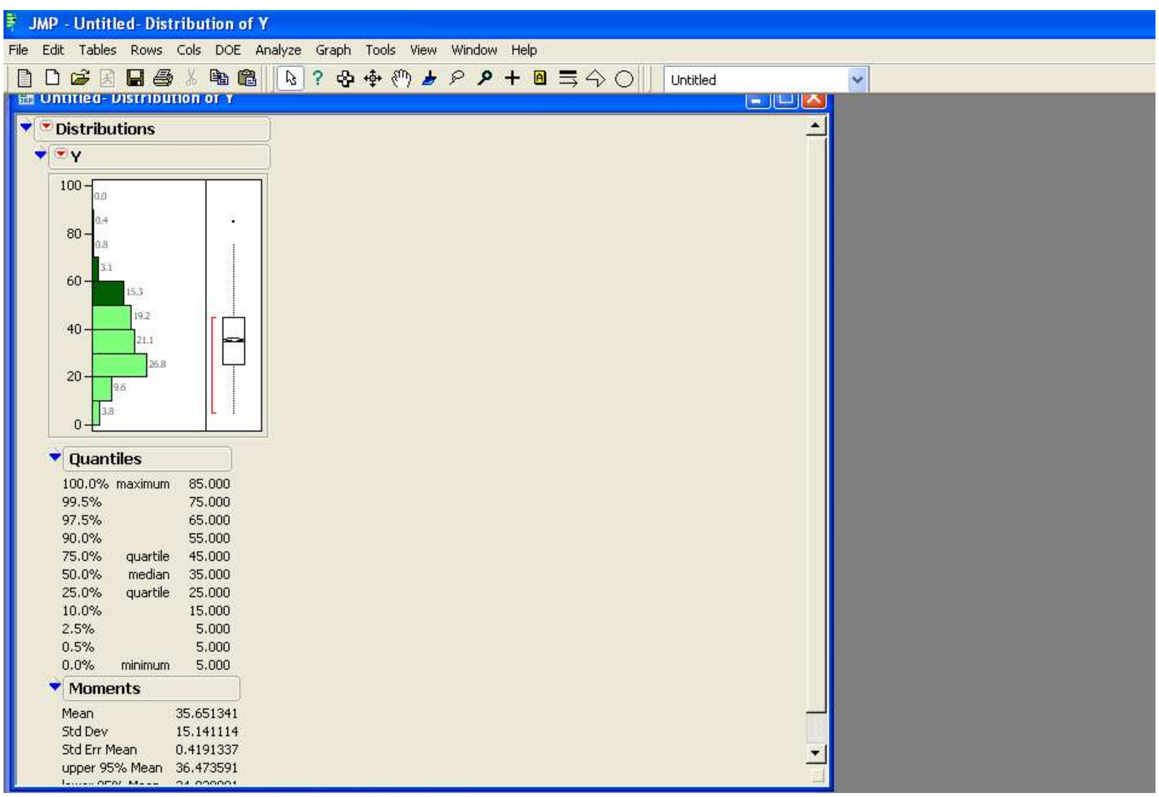


Click “OK”.



To Show the percentage of each bar, click the hot spot left to “Y” and choose “Show Percents” in “Histogram Options”:





The mean and variance could be easily found in the output “Moments” below the graph. Here the Mean is 35.651341. The variance 15.141114.

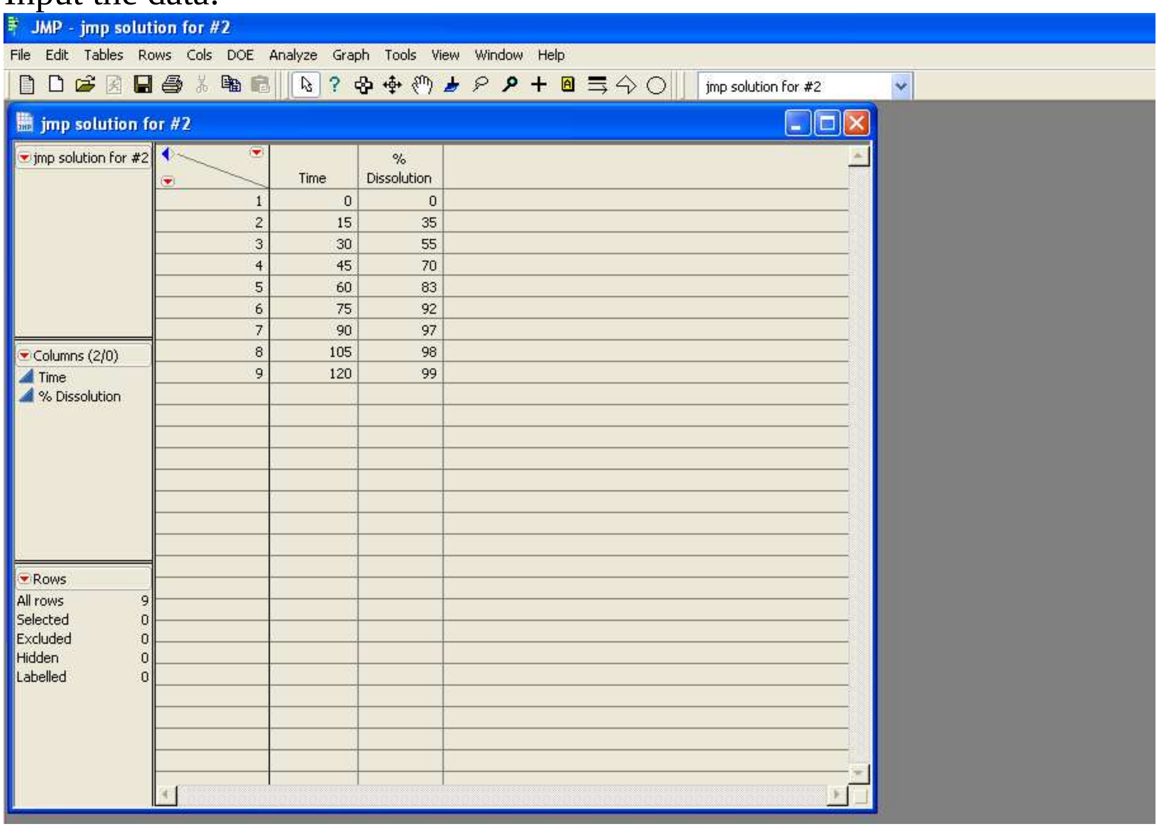
2. In scenario 2, the Percent Dissolution of tablets as a function of time is as the following:

Time	% Dissolution
0	0
15	35
30	55
45	70
60	83
75	92
90	97
105	98
120	99

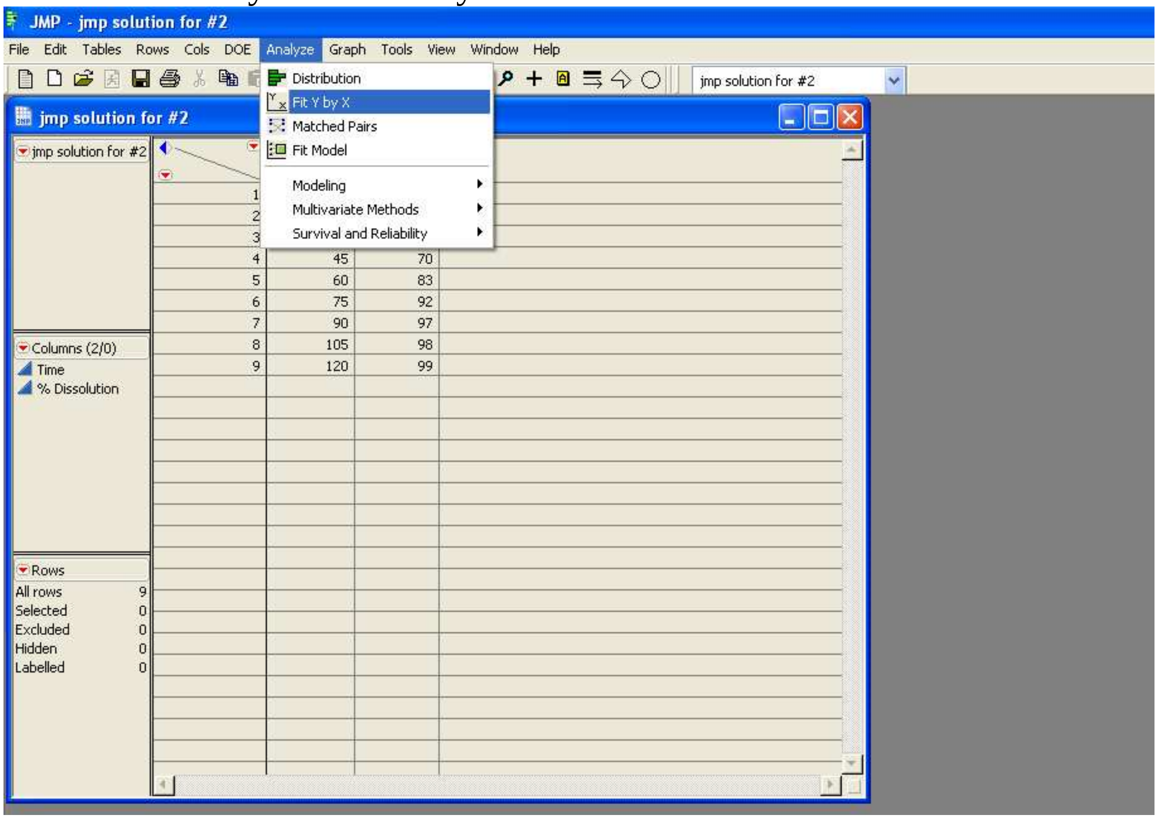
Use JMP to plot the Distribution and calculate the time at which 85% of the tablet has been dissolved.

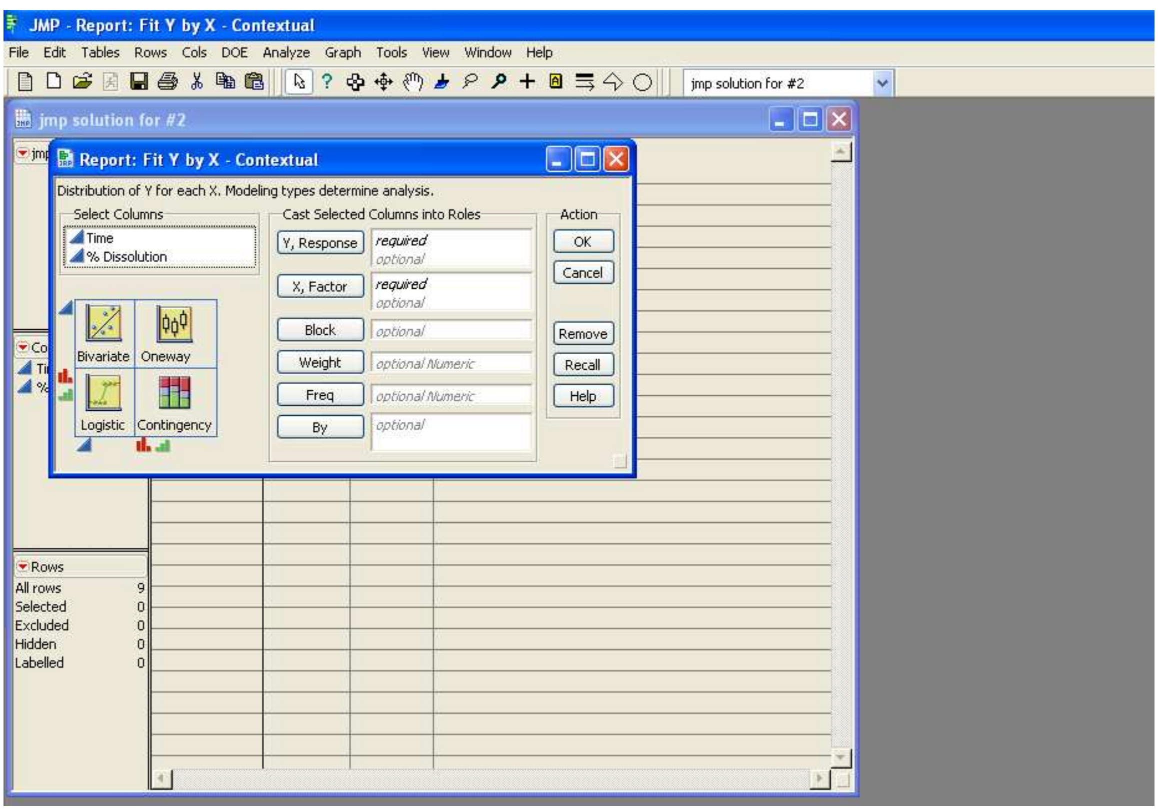
Solution:

Input the data:

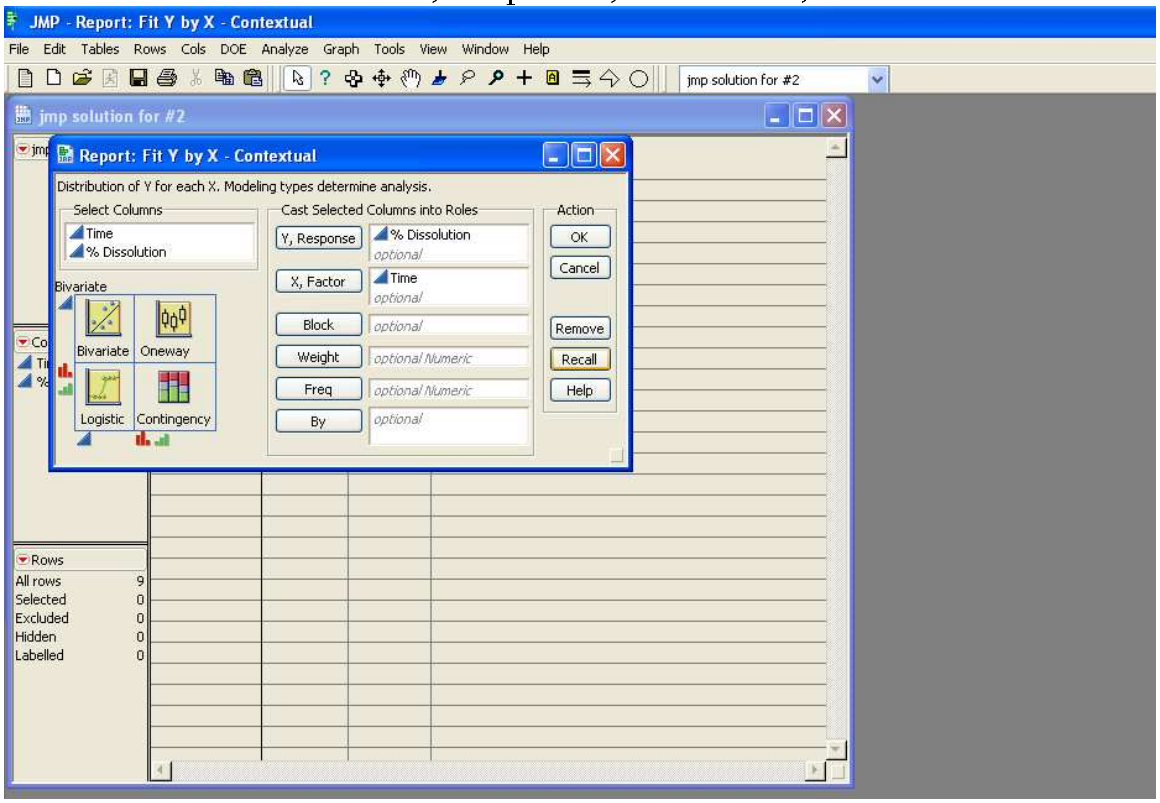


Choose “Fit Y by X” in “Analyze”:

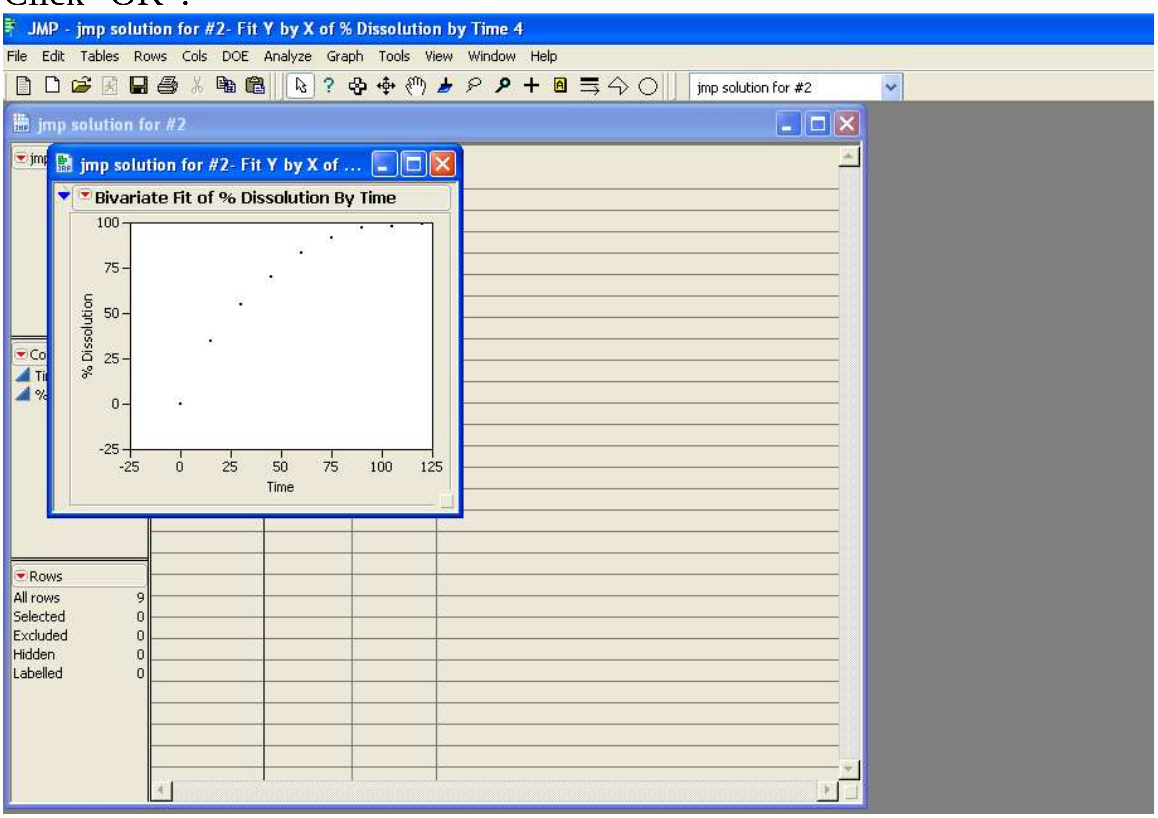




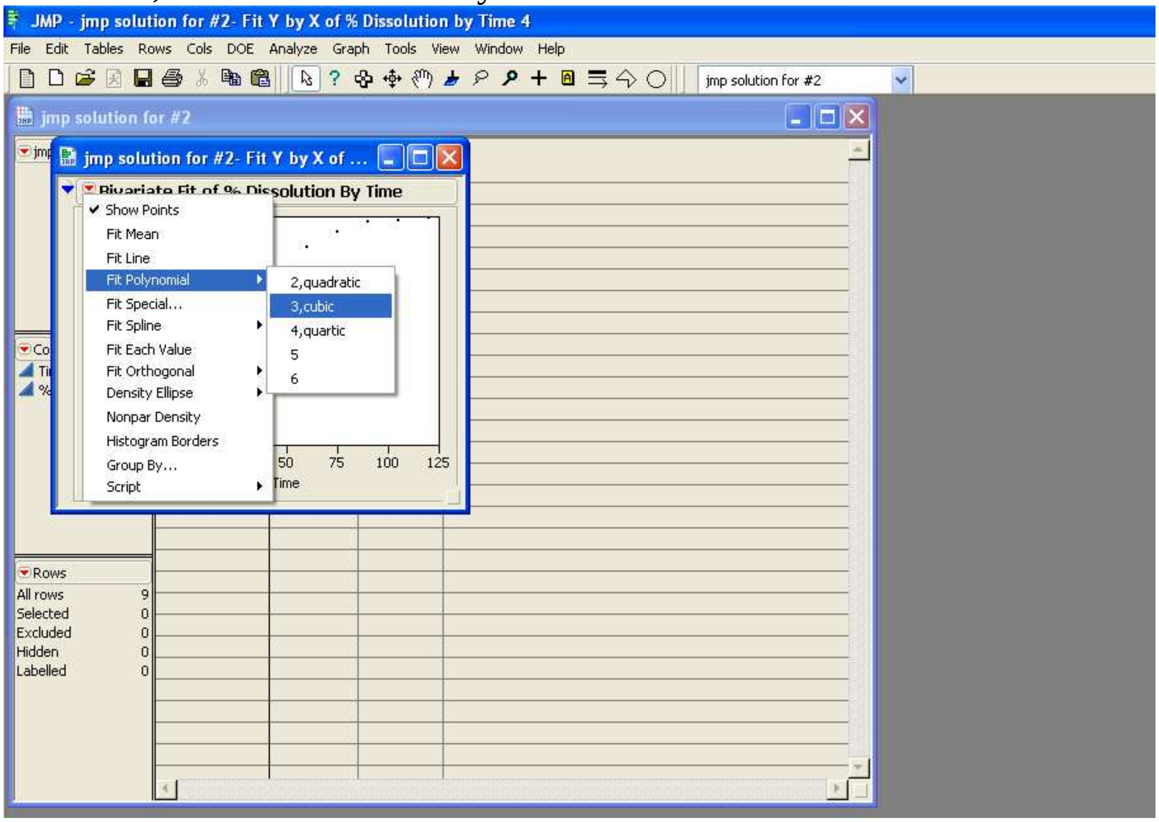
Choose % Dissolution as “Y, Response”, Time as “X, Factor”:

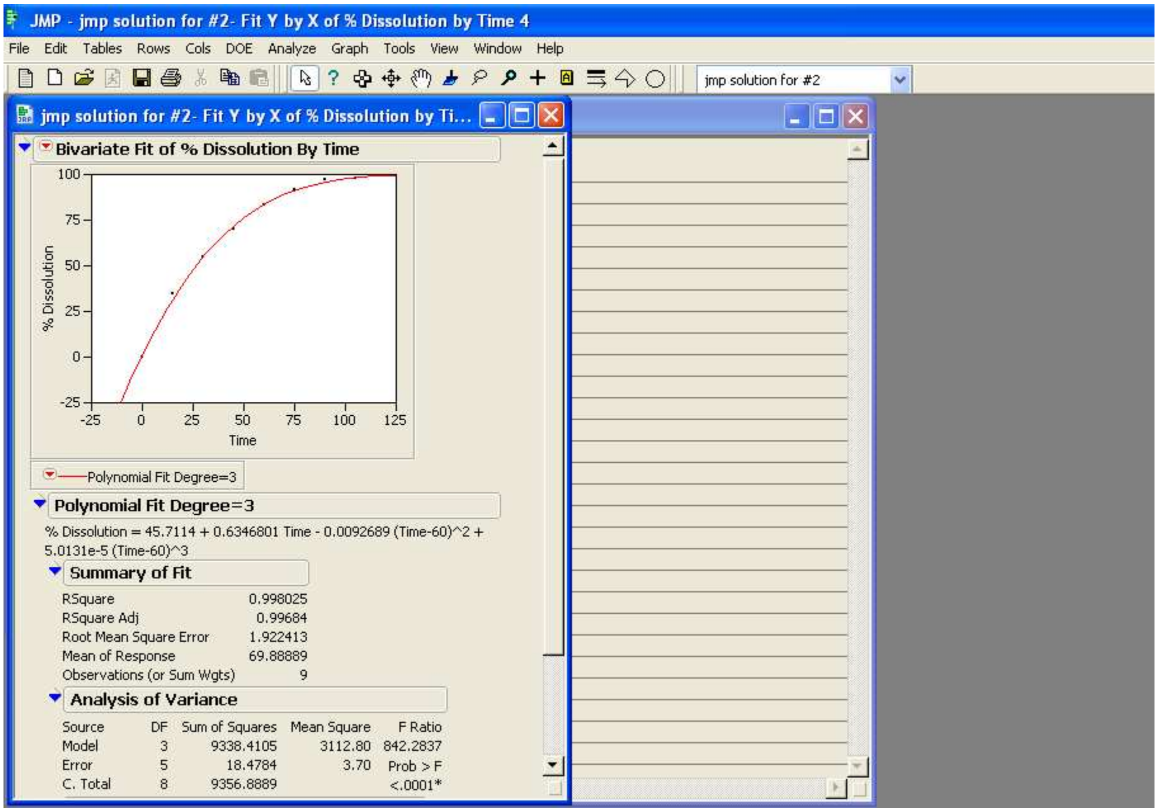


Click “OK”:



Click the hot spot left to “Bivariate Fit of % Dissolution By Time” and choose “3, cubic” from “Fit Polynomial”:





From the polynomial function JMP offered, we could calculate the time when Dissolution is 85%:

$x = 61.9584$

Comparison Tests

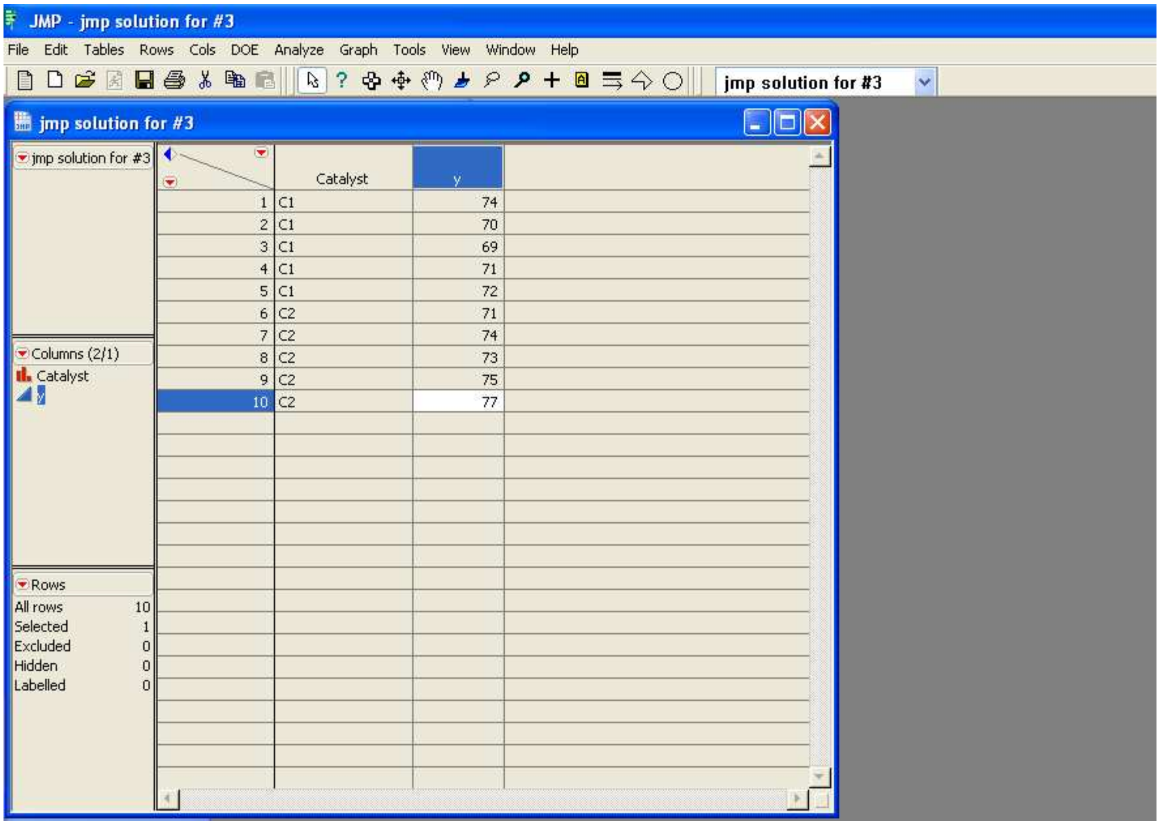
3. Two different catalysts are studied in the batch reactor. (Scenario 1) Differece runs are made with each catalyst and the yield of A measured after 1 hour. (all other factors held constant)

Catalyst C1	Catalyst C2
74	71
70	74
69	73
71	75
72	77

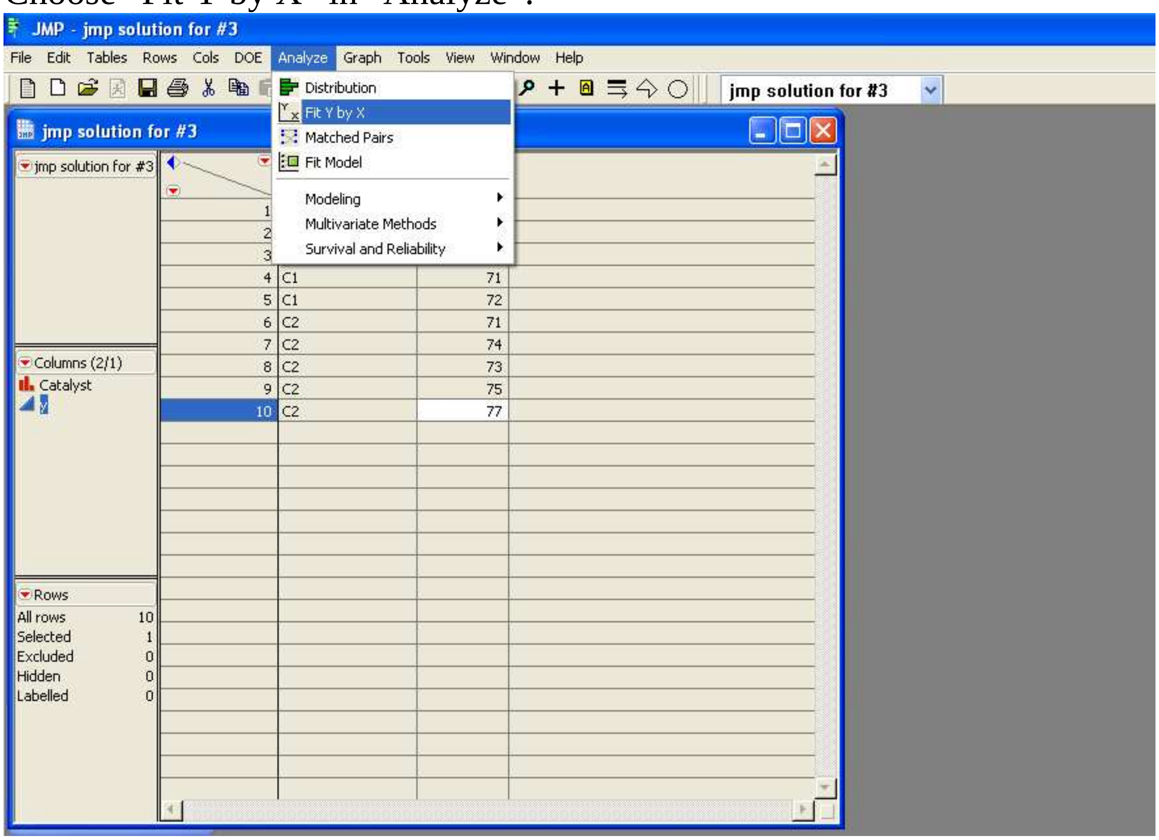
(1) Determine the mean and variance of each catalyst.
(2) Use the appropriate distribution to decide whether there is a difference at the 95% confidence level.
(3) At what level is there a difference between the two catalyst (p value).
(4) Use an F test to determine the level at which there is a difference between the variance of the yield between the catalysts.

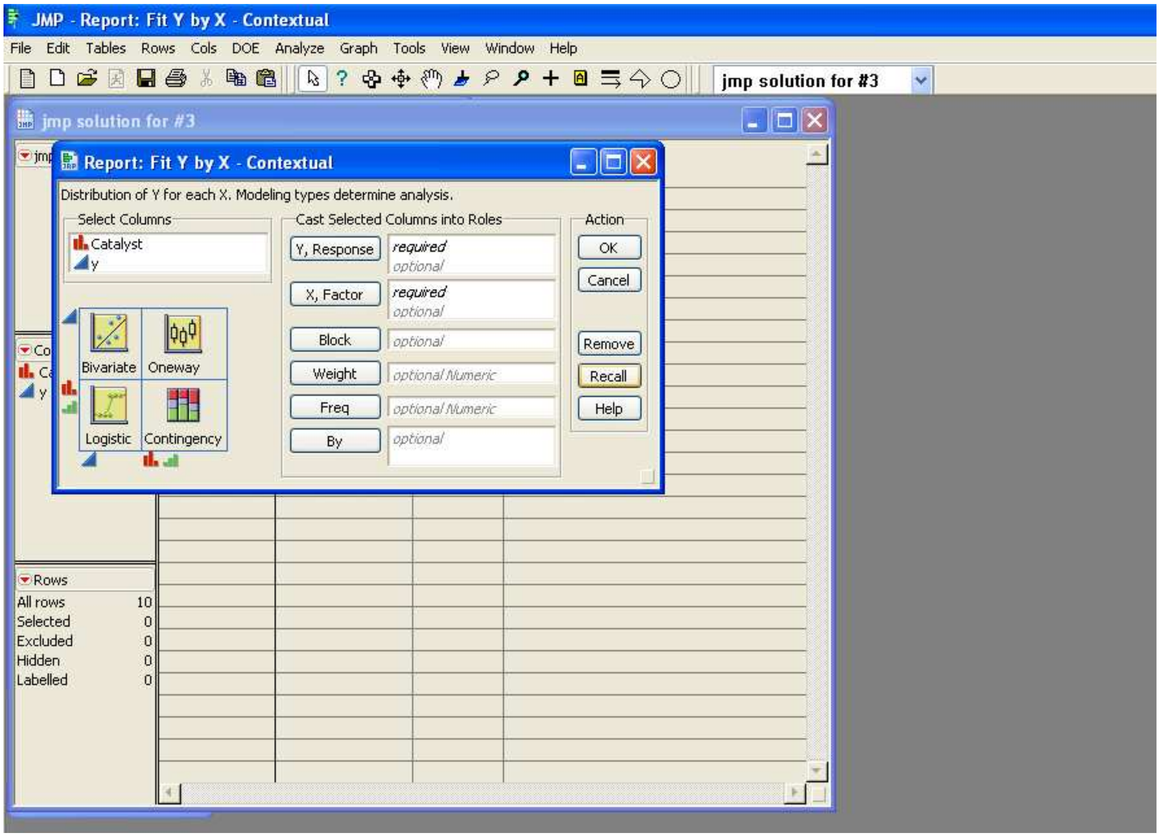
Solution:

Input the data. Here Catalyst is the type of Catalyst and its data type is "Character":

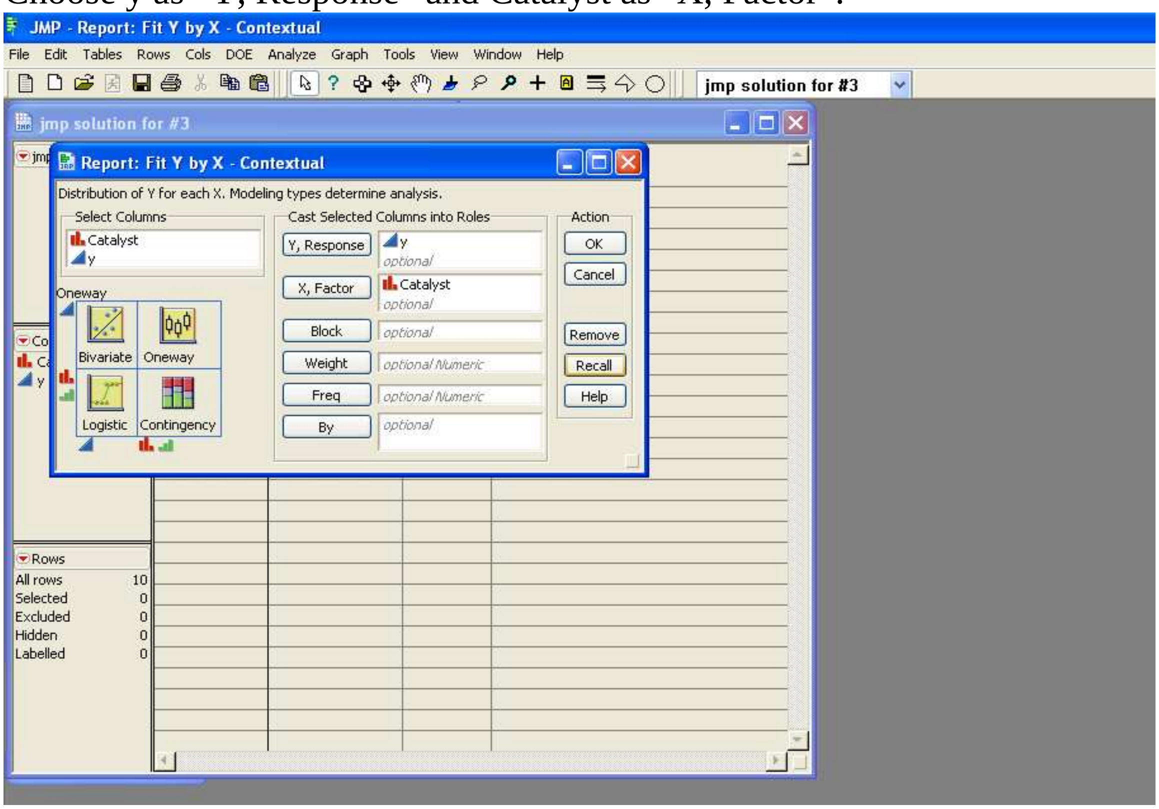


Choose “Fit Y by X” in “Analyze”:

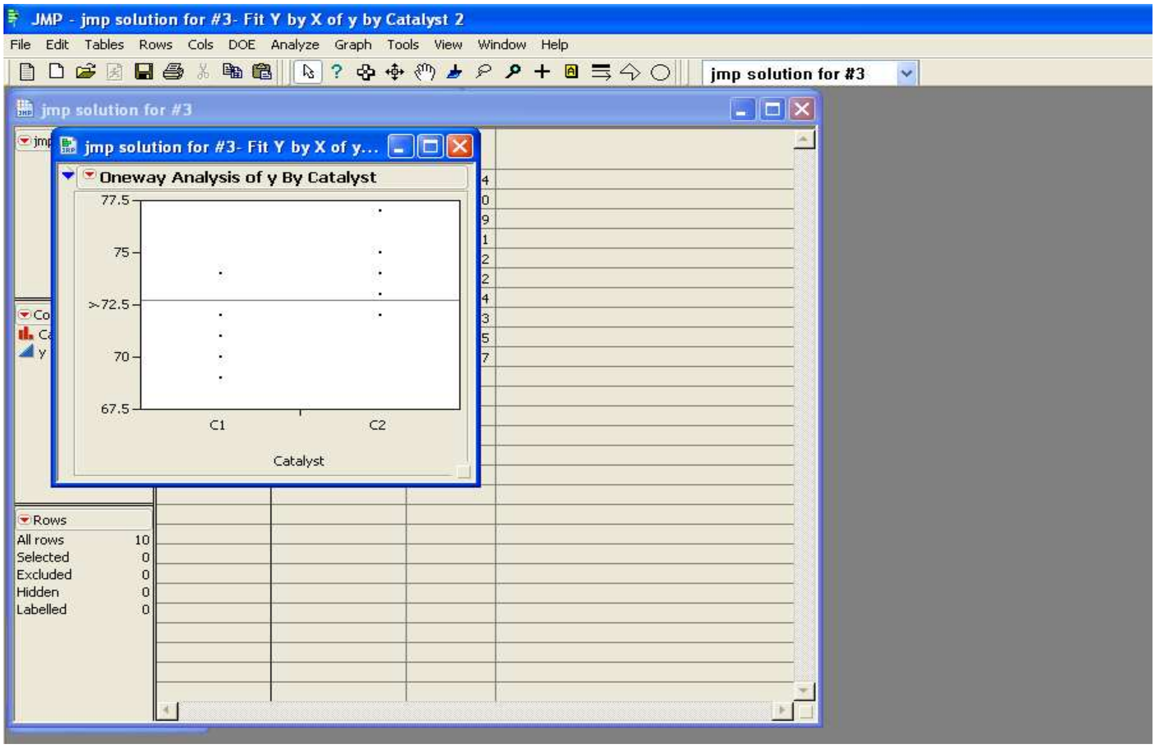




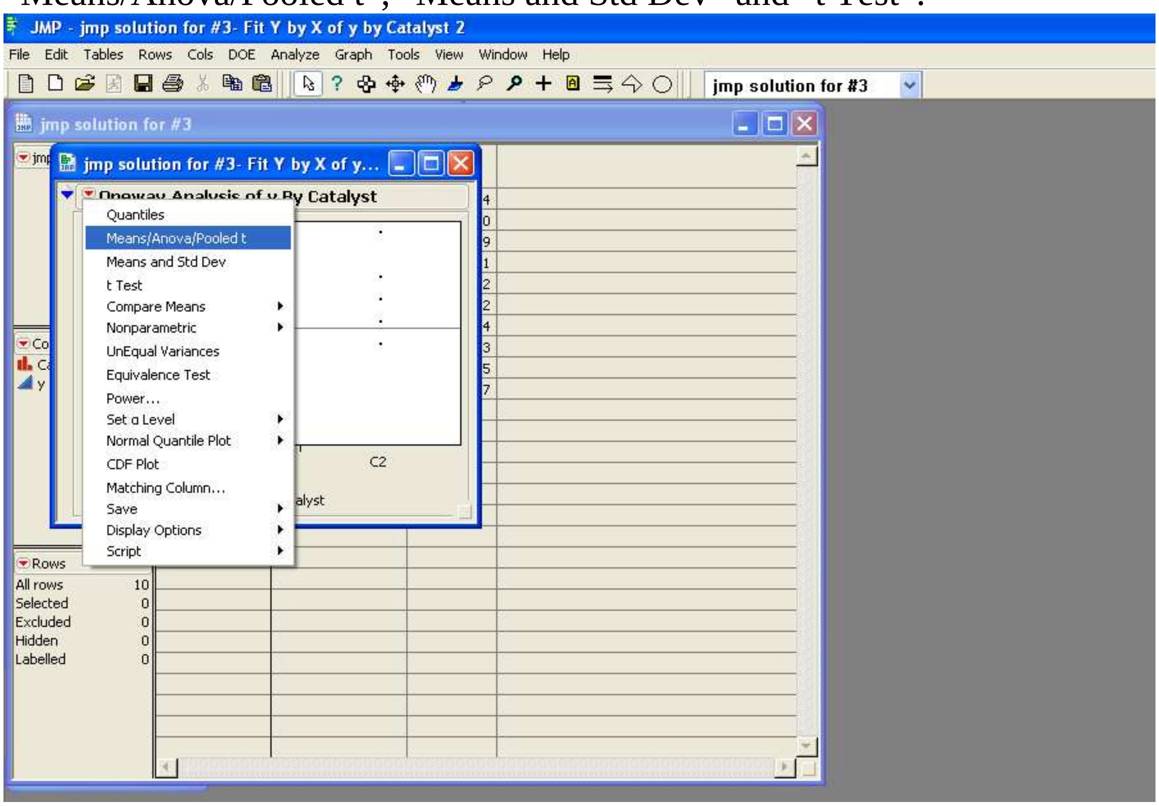
Choose y as “Y, Response” and Catalyst as “X, Factor”:

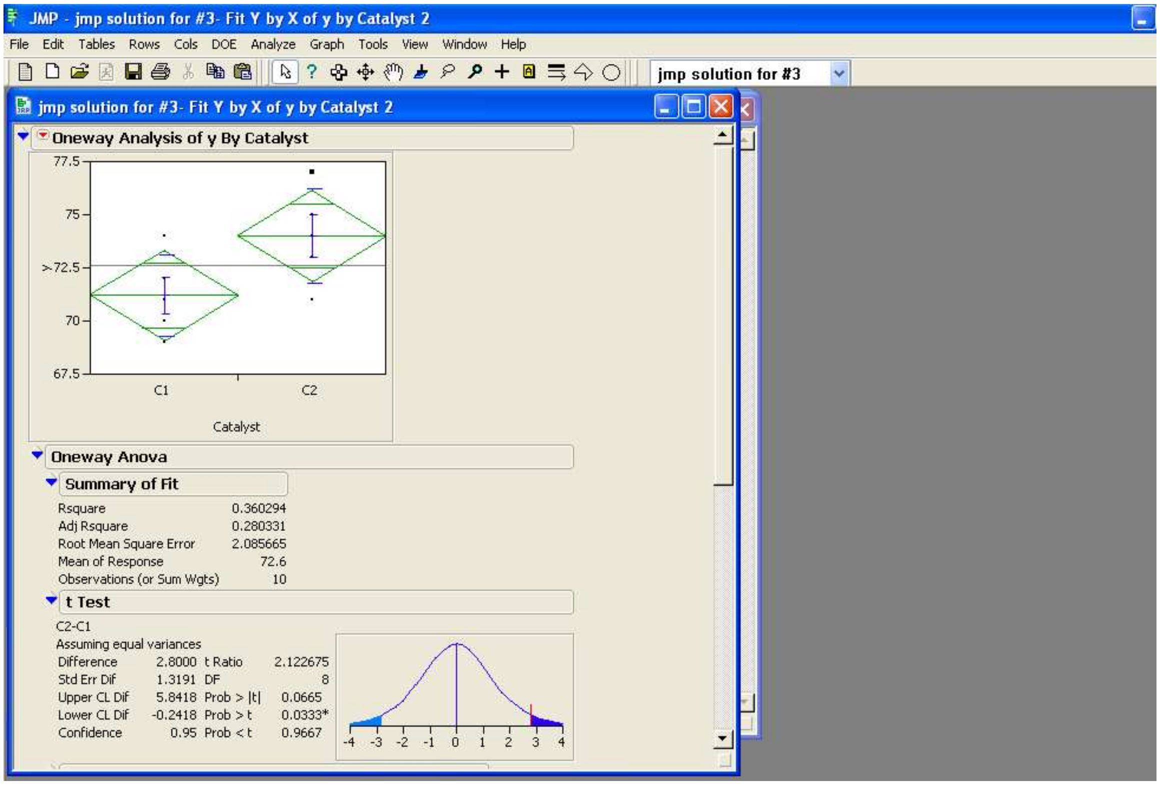


Click “OK”:



Click the hot spot left to “Oneway Analysis of y By Catalyst” and choose “Means/Anova/Pooled t”, “Means and Std Dev” and “t Test”:



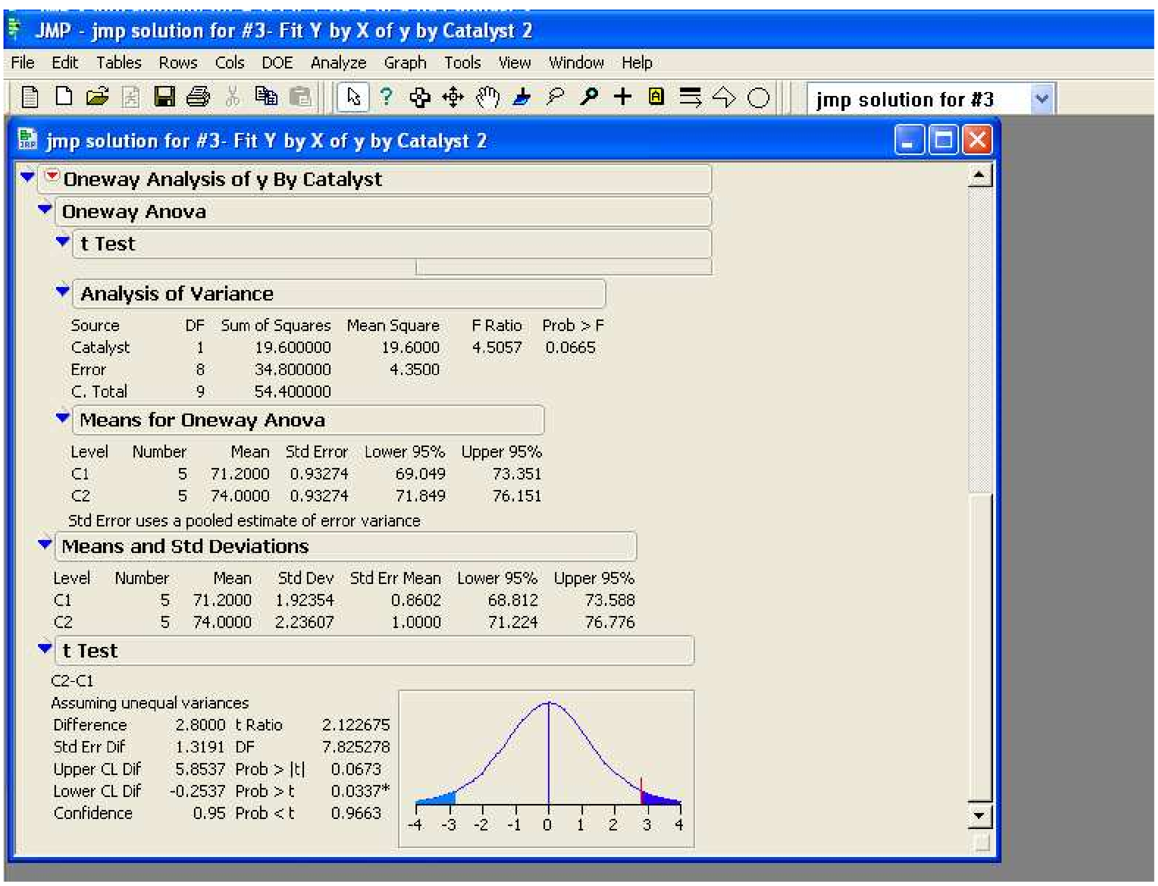


Rsquare	0.360294
Adj Rsquare	0.280331
Root Mean Square Error	2.085665
Mean of Response	72.6
Observations (or Sum Wgts)	10

C2-C1

Assuming equal variances

Difference	2.8000	t Ratio	2.122675
Std Err Dif	1.3191	DF	8
Upper CL Dif	5.8418	Prob > |t|	0.0665
Lower CL Dif	-0.2418	Prob > t	0.0333*
Confidence	0.95	Prob < t	0.9667



Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Catalyst	1	19.600000	19.6000	4.5057	0.0665
Error	8	34.800000	4.3500		
C. Total	9	54.400000			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
C1	5	71.2000	0.93274	69.049	73.351
C2	5	74.0000	0.93274	71.849	76.151

Std Error uses a pooled estimate of error variance

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
C1	5	71.2000	1.92354	0.8602	68.812	73.588
C2	5	74.0000	2.23607	1.0000	71.224	76.776

t Test

C2-C1

Assuming unequal variances

Difference	2.8000	t Ratio	2.122675
Std Err Dif	1.3191	DF	7.825278
Upper CL Dif	5.8537	Prob > |t|	0.0673
Lower CL Dif	-0.2537	Prob > t	0.0337*
Confidence	0.95	Prob < t	0.9663

(1) From the output, the mean and variance for C1 and C2 are 71.2, 1.92354 and 74, 2.23607.

(2)and (3). From the t test, there is a significant difference between the means and the p-value .0337.

(4) From “Analysis of Variance”, the p-value for F test is .0665, which is not significant.

Regression Analysis

4. Once the API is produced in a reactor described in Scenario 1, crystallization from solution is to separate the desired product $C(t_f)$ from $A(t_f)$ and $B(t_f)$ once the impurity $D(t_f)$ has been removed. In general for a pharmaceutical process crystallization may be used to achieve sufficient product purity, to minimize the filtration time or to achieve tablet stability when mixed with other crystals of other chemical species before forming a tablet. In this example we will dwell only on a single criterion filtration time In this example, based on the work of Togkalidou et al (2001), “Experimental Design and Inferential Modeling in Pharmaceutical Crystallization (AIChe Journal, Vo 27, No1), a pharmaceutical salt was crystallized in a baffled reactor, where the supersaturation was created by adding a less efficient solvent that was miscible in the original solvent. The details are not relevant for the example but the student is referred to the paper if more information about the crystallization process is required.

The following data were collected:

Experiment Number	Agitation(rpm)	Seed Amount (% of Batch)	Temperature (deg C)	Charge Time h	Filtration Time Min
1	2200	4	20	6	150
2	400	5	15	3	105
3	1300	3.5	15	9	165
4	2200	4	17.5	7.5	170
5	3100	3.5	17.5	7.5	90
6	2200	4	20	6	155
7	4000	5	20	6	50
8	400	3	20	6	280
9	1300	3.5	22.5	4.5	122
10	2200	4	22.5	4.5	100
11	3100	4.5	25	9	82
12	2200	4	20	6	145

Use Regression Analysis from JMP to determine a regression model and the conditions under which the filtration time is minimized.

Solution:

(1) Run a regression model with all four factors in the model using the steps as showed in the JMP tutorial S2E4 and S2E5:

Summary of Fit

RSquare	0.700772
RSquare Adj	0.529785
Root Mean Square Error	40.42272
Mean of Response	134.5
Observations (or Sum Wgts)	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	26787.025	6696.76	4.0984
Error	7	11437.975	1634.00	Prob > F
C. Total	11	38225.000		0.0506

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	5	11387.975	2277.60	91.1038
Pure Error	2	50.000	25.00	Prob > F
Total Error	7	11437.975		0.0109*
				Max RSq
				0.9987

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	313.02873	164.7935	1.90	0.0993
Agitation	-0.032988	0.016613	-1.99	0.0874
Seed Amount	-40.78237	25.80307	-1.58	0.1580
Temperature	0.5069102	4.592396	0.11	0.9152
Charge Time	6.7678051	8.777356	0.77	0.4659

(2) Remove the most insignificant term by comparing the p-values. Temperature is eliminated and the model is run again:

Summary of Fit

RSquare	0.700252
RSquare Adj	0.587846
Root Mean Square Error	37.84489
Mean of Response	134.5
Observations (or Sum Wgts)	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	26767.116	8922.37	6.2297
Error	8	11457.884	1432.24	Prob > F
C. Total	11	38225.000		0.0173*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	6	11407.884	1901.31	76.0526
Pure Error	2	50.000	25.00	Prob > F
Total Error	8	11457.884		0.0130*
				Max RSq
				0.9987

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	325.80338	109.8334	2.97	0.0180*
Agitation	-0.032151	0.013839	-2.32	0.0487*
Seed Amount	-41.53415	23.3008	-1.78	0.1125
Charge Time	6.5187692	7.941494	0.82	0.4355

(3) Once again, remove the most insignificant term, Change Time. Run the model again:

Summary of Fit

RSquare	0.675006
RSquare Adj	0.602785
Root Mean Square Error	37.15271
Mean of Response	134.5
Observations (or Sum Wgts)	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	25802.085	12901.0	9.3464
Error	9	12422.915	1380.3	Prob > F
C. Total	11	38225.000		0.0064*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	4	8728.415	2182.10	2.9532
Pure Error	5	3694.500	738.90	Prob > F
Total Error	9	12422.915		0.1330
				Max RSq
				0.9033

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	390.22516	75.43199	5.17	0.0006*
Agitation	-0.025807	0.01127	-2.29	0.0478*
Seed Amount	-50.70497	20.07366	-2.53	0.0325*

Both the Agitation and Seed Amount are significant at .05 level. The result regression equation is:

Filtration Time= 390.22516 -.025807 Agitation – 50.70497 Seed Amount

By comparing the sign of the coefficient, the filtration time would be minimized when Agitation is set at its maximum value of 4000 and Seed Amount at 5. At these values the filtration time is 33.47231

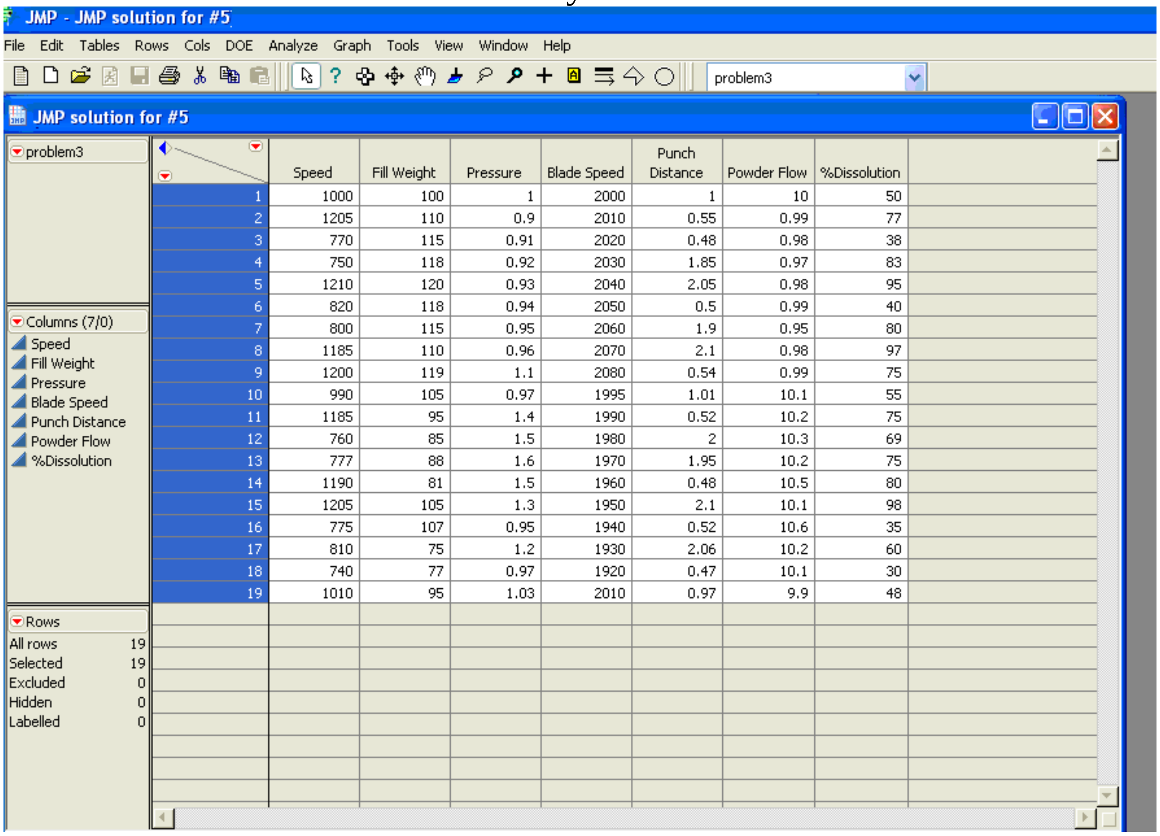
5. A study was launched to determine the effect of several factors on the %Dissolution after 60 minutes of a new product from the Tabletting machine in Scenario 2. The following data were obtained:

Expt Number	Speed (Rpm)	Fill Weight (kg)	Pressure (Ton)	Blade Speed (rpm)	Punch Distance (mm)	Powder Flow (kg/hr)	% Diss
1	1000	100	1	2000	1	10	50
2	1205	110	.90	2010	.55	.99	77
3	770	115	.91	2020	.48	.98	38
4	750	118	.92	2030	1.85	.97	83
5	1210	120	.93	2040	2.05	.98	95
6	820	118	.94	2050	.5	.99	40
7	800	115	.95	2060	1.9	.95	80
8	1185	110	.96	2070	2.1	.98	97
9	1200	119	1.1	2080	.54	.99	75
10	990	105	.97	1995	1.01	10.1	55
11	1185	95	1.4	1990	.52	10.2	75
12	760	85	1.5	1980	2.0	10.3	69
13	777	88	1.6	1970	1.95	10.2	75
14	1190	81	1.5	1960	.48	10.5	80
15	1205	105	1.3	1950	2.1	10.1	98
16	775	107	.95	1940	.52	10.6	35
17	810	75	1.2	1930	2.06	10.2	60
18	740	77	.97	1920	.47	10.1	30
19	1010	95	1.03	2010	.97	9.9	48

(1) Determine the extent of correlation between the various factors.
(2) Build a regression model relating the %Dissolution to the factors.
 i)Use Standard Regression
 ii)Use Stepwise Regression
 iii) Why are results in ii) different than in i)

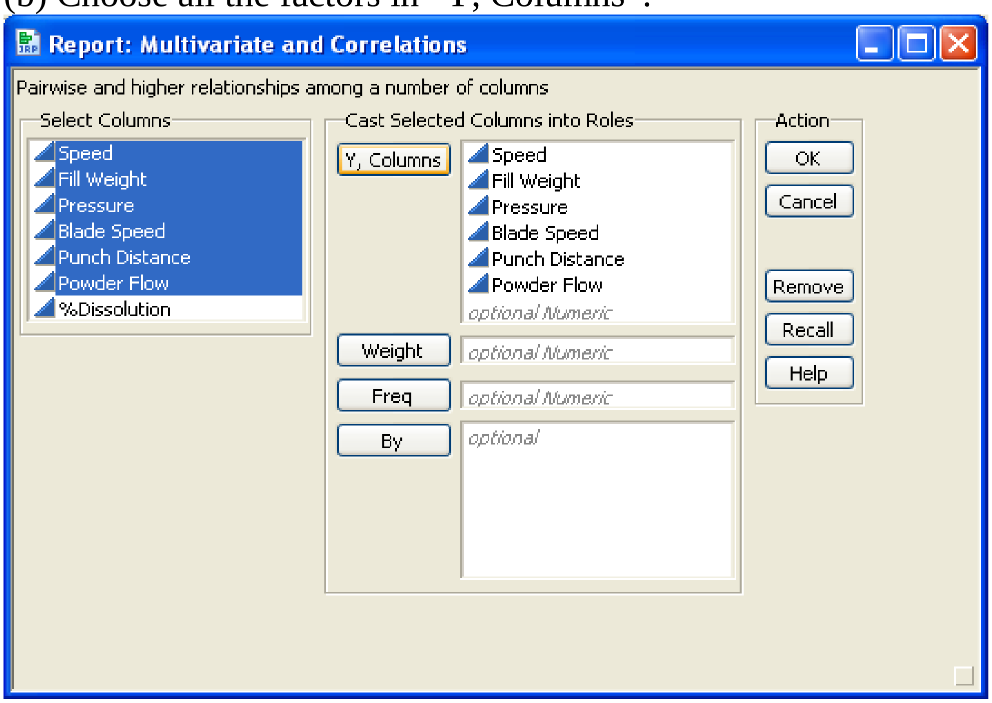
Solution:
(1)

(a) To acquire the correlation between the factors, choose “Multivariate” from “Multivariate Method” in “Analyze”:

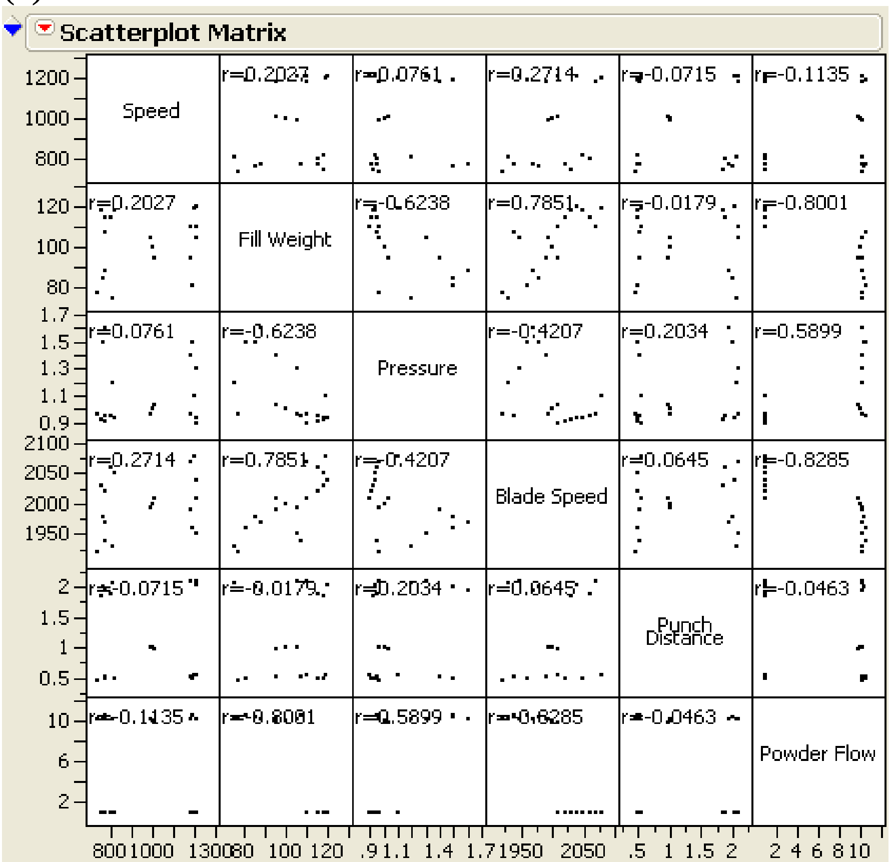


	Speed	Fill Weight	Pressure	Blade Speed	Punch Distance	Powder Flow	%Dissolution
1	1000	100	1	2000	1	10	50
2	1205	110	0.9	2010	0.55	0.99	77
3	770	115	0.91	2020	0.48	0.98	38
4	750	118	0.92	2030	1.85	0.97	83
5	1210	120	0.93	2040	2.05	0.98	95
6	820	118	0.94	2050	0.5	0.99	40
7	800	115	0.95	2060	1.9	0.95	80
8	1185	110	0.96	2070	2.1	0.98	97
9	1200	119	1.1	2080	0.54	0.99	75
10	990	105	0.97	1995	1.01	10.1	55
11	1185	95	1.4	1990	0.52	10.2	75
12	760	85	1.5	1980	2	10.3	69
13	777	88	1.6	1970	1.95	10.2	75
14	1190	81	1.5	1960	0.48	10.5	80
15	1205	105	1.3	1950	2.1	10.1	98
16	775	107	0.95	1940	0.52	10.6	35
17	810	75	1.2	1930	2.06	10.2	60
18	740	77	0.97	1920	0.47	10.1	30
19	1010	95	1.03	2010	0.97	9.9	48

(b) Choose all the factors in “Y, Columns”:



(c) Click “OK”:



Note the following pair of factors are highly correlated:
Fill Weight and Blade Speed.
Fill Weight and Powder Flow.
Blade Speed and Powder Flow
Fill Weight and Pressure

(2)

i. Standard Regression:

Summary of Fit

RSquare	0.920888
RSquare Adj	0.881331
Root Mean Square Error	7.411549
Mean of Response	66.31579
Observations (or Sum Wgts)	19

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	7672.9325	1278.82	23.2805
Error	12	659.1727	54.93	Prob > F
C. Total	18	8332.1053		<.0001*

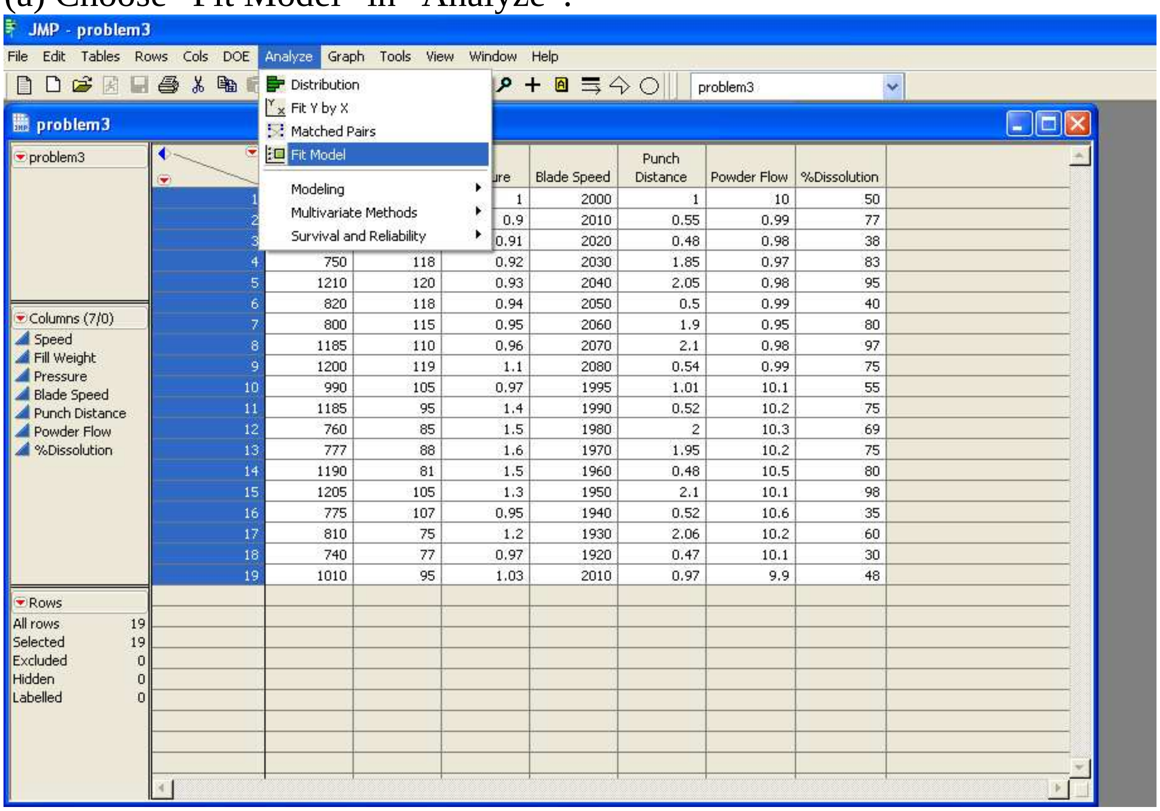
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	113.97311	142.6448	0.80	0.4398
Speed	0.0626565	0.009696	6.46	<.0001*
Fill Weight	0.2232583	0.231023	0.97	0.3529
Pressure	38.856114	10.82738	3.59	0.0037*
Blade Speed	-0.090549	0.075411	-1.20	0.2530
Punch Distance	17.220037	2.584938	6.66	<.0001*
Powder Flow	-2.188226	0.794755	-2.75	0.0175*

Based on the analysis, Fill Weight and Blade Speed are unimportant. This is not surprising since they are correlated with Powder Flow in Part (1).

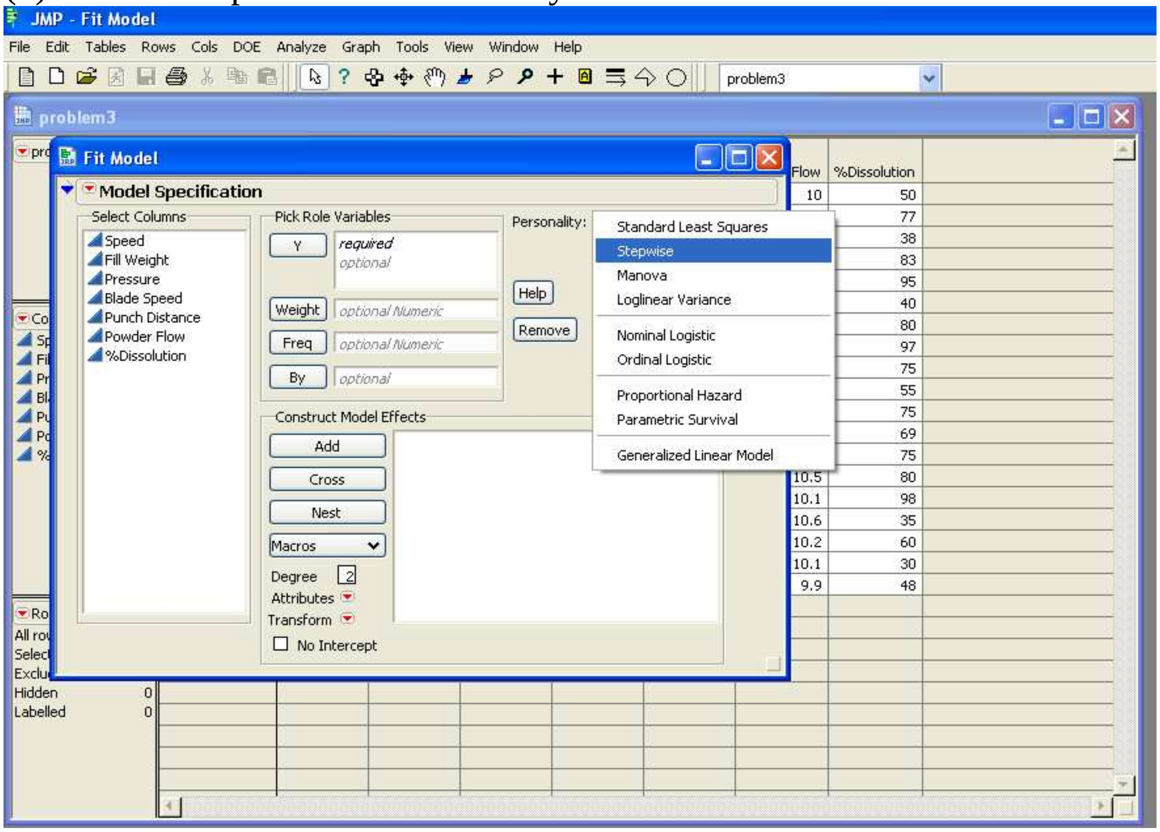
ii) Stepwise Regression

(a) Choose “Fit Model” in “Analyze”:

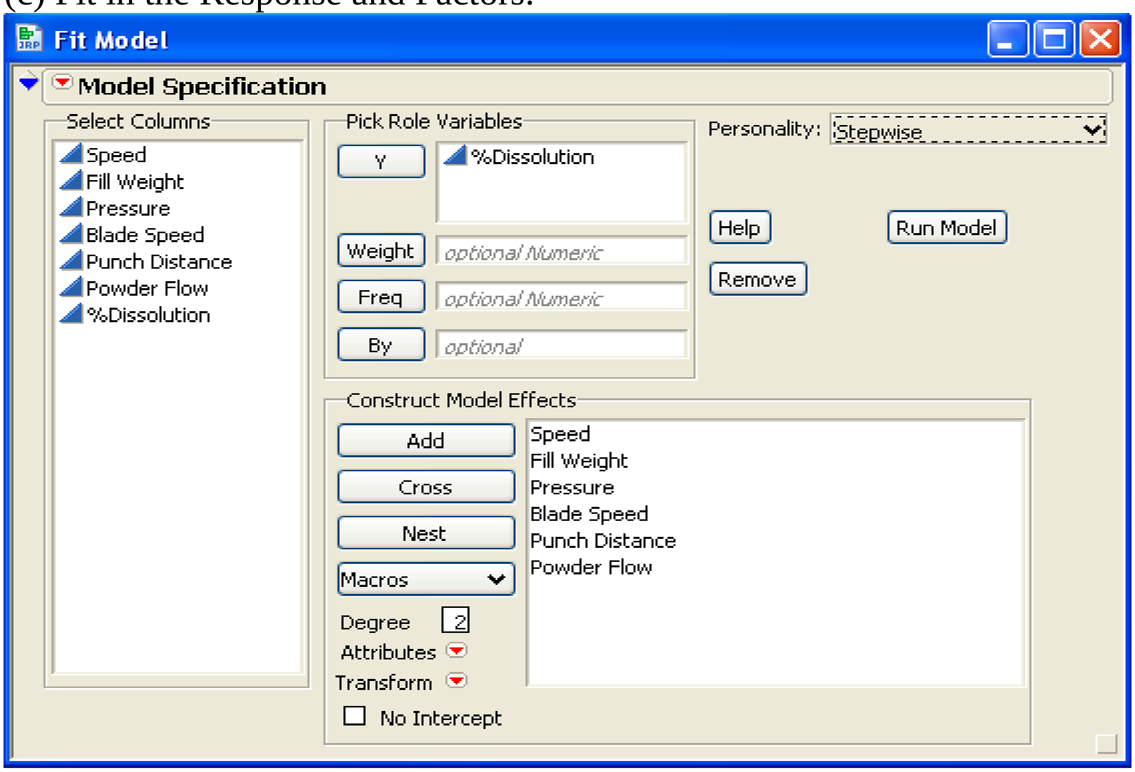


			ure	Blade Speed	Punch Distance	Powder Flow	%Dissolution
1			1	2000	1	10	50
2			0.9	2010	0.55	0.99	77
3			0.91	2020	0.48	0.98	38
4	750	118	0.92	2030	1.85	0.97	83
5	1210	120	0.93	2040	2.05	0.98	95
6	820	118	0.94	2050	0.5	0.99	40
7	800	115	0.95	2060	1.9	0.95	80
8	1185	110	0.96	2070	2.1	0.98	97
9	1200	119	1.1	2080	0.54	0.99	75
10	990	105	0.97	1995	1.01	10.1	55
11	1185	95	1.4	1990	0.52	10.2	75
12	760	85	1.5	1980	2	10.3	69
13	777	88	1.6	1970	1.95	10.2	75
14	1190	81	1.5	1960	0.48	10.5	80
15	1205	105	1.3	1950	2.1	10.1	98
16	775	107	0.95	1940	0.52	10.6	35
17	810	75	1.2	1930	2.06	10.2	60
18	740	77	0.97	1920	0.47	10.1	30
19	1010	95	1.03	2010	0.97	9.9	48

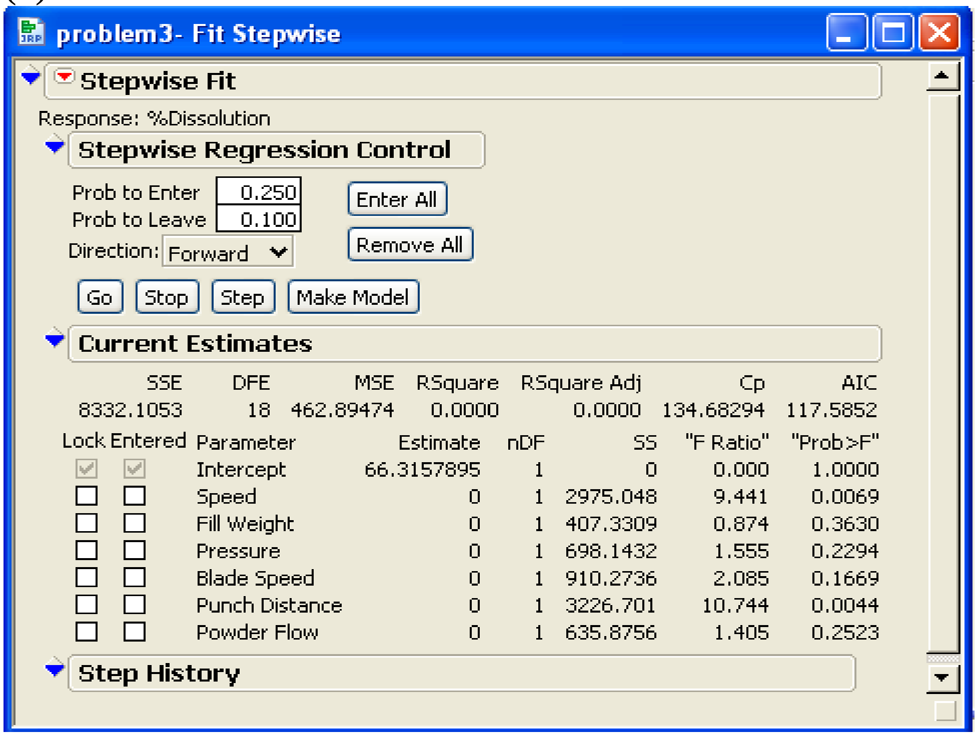
(b) Select "Stepwise" in "Personality":



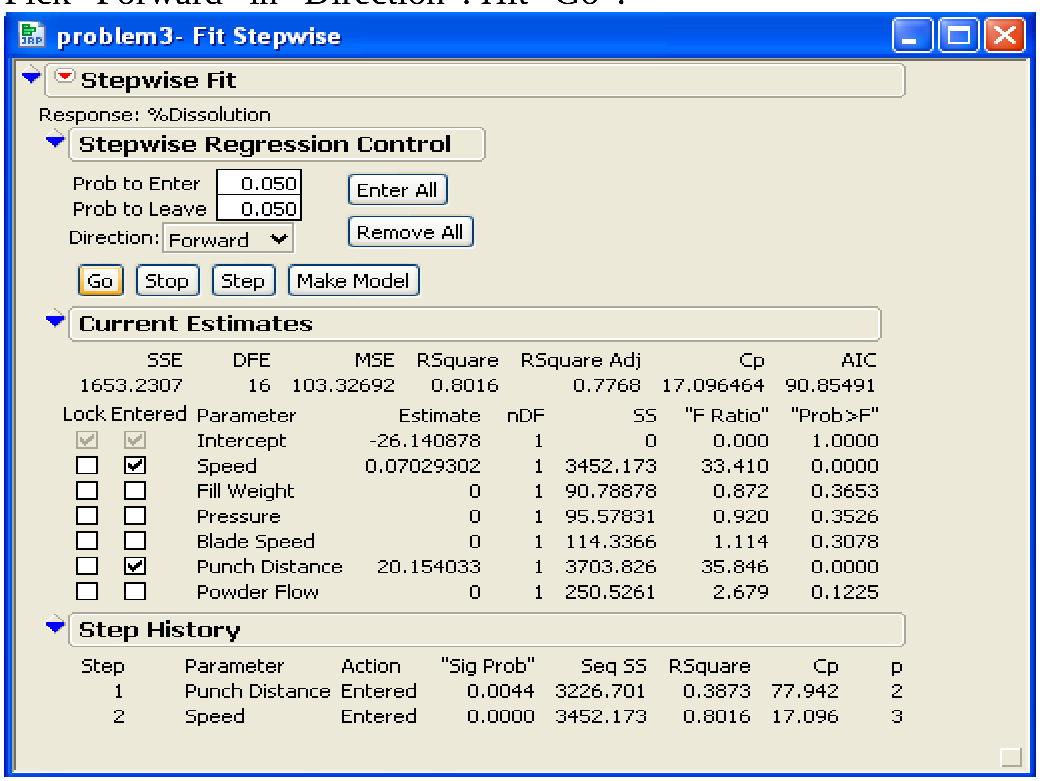
(c) Fit in the Response and Factors:



(d) Hit “Run Model”:

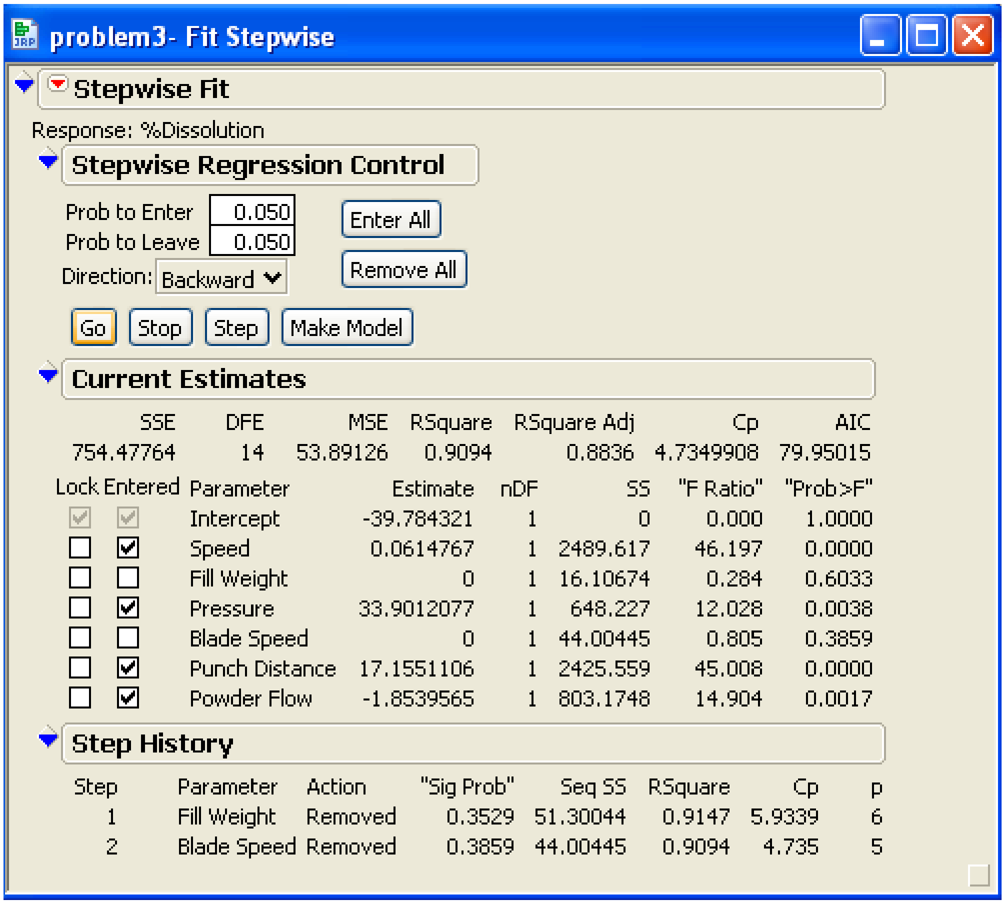


(e) Now we may choose either forward selection or backward selection.
To do forward selection, input .05 as the α Entry level and Exit level.
Pick “Forward” in “Direction”. Hit “Go”:



Punch Distance and Speed are kept in the final model.

(f) To do backward selection, input .05 as the α Entry level and Exit level. Pick “Backward” in “Direction”. Hit “Enter All” and “Go”:



Speed, Pressure, Punch Distance and Powder Flow are in the final model.

iii) However, the results are different because of the correlations among the factors.

Single Factor Experiments

6. Completely Randomized Design

In a study to determine the effect of roller speed on roller gap in a roller compactor (Scenario 2), five replicates of the Roller Gap in mm were measured at five different values of roll speed (rpm) where the experiments were run in random order. The following data were obtained:

Roll Speed (rpm)	Roller gap (mm)				
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

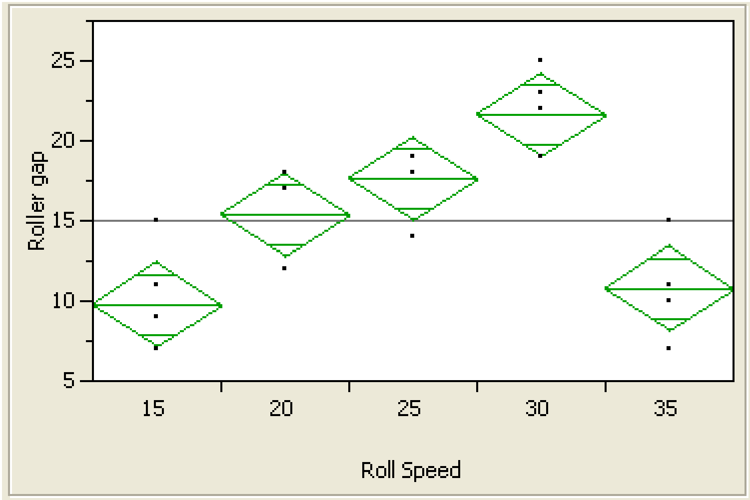
(1) Does roller speed affect roller gap at the 95% confidence level? Perform an ANOVA.

(2) Using a multiple range test at 95% confidence which levels are different from one another?

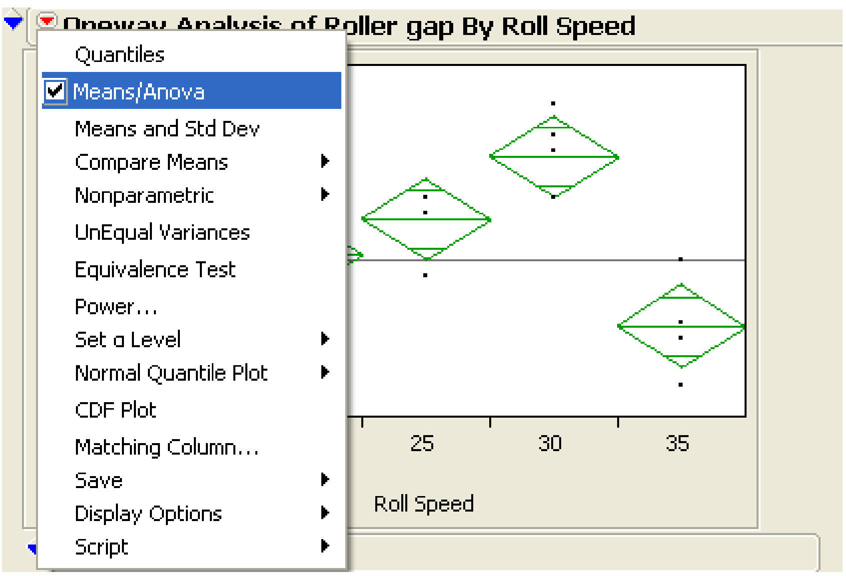
(3) Find a suitable regression model between roller gap and roll speed if one exists.

(4) Compare the results of (2) and (3).

Solution:
(1) Choose "Fit Y by X" in "analyze" with Roller gap as Y and Roll speed as X.



Choose "means/Anova" in hot spot aside "Oneway analysis of Roller gap by Roller speed":



Analysis of Variance

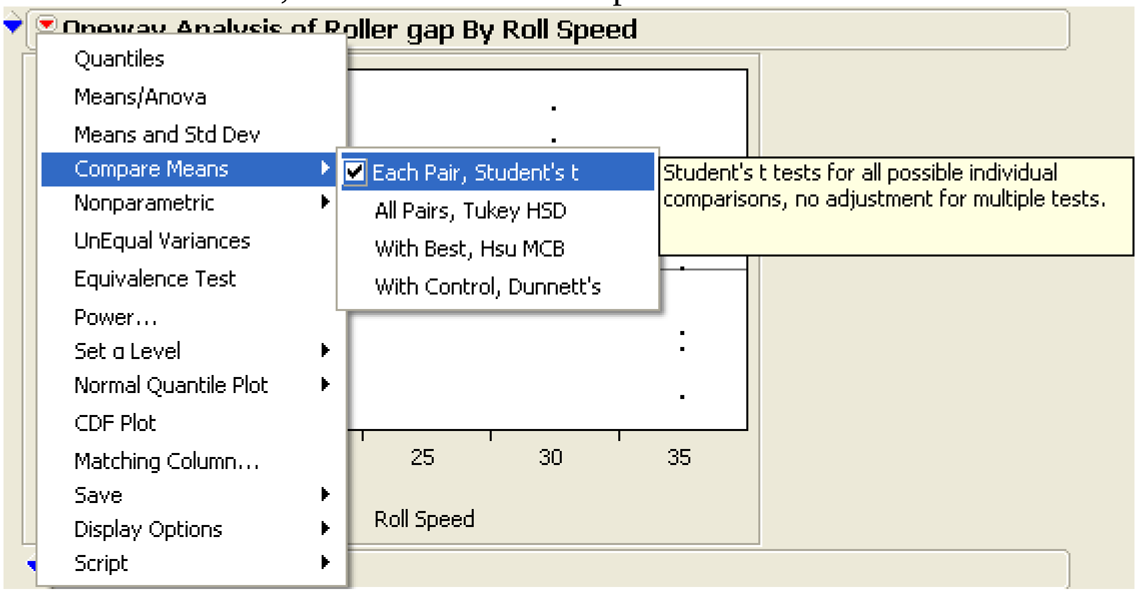
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Roll Speed	4	475.76000	118.940	14.7568	<.0001*
Error	20	161.20000	8.060		
C. Total	24	636.96000			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
15	5	9.8000	1.2696	7.152	12.448
20	5	15.4000	1.2696	12.752	18.048
25	5	17.6000	1.2696	14.952	20.248
30	5	21.6000	1.2696	18.952	24.248
35	5	10.8000	1.2696	8.152	13.448

Yes, roller speed affects roller gap at the 95% confidence level since the p value is <.0001.

(2)
Choose “each Pair, Student’s t” in “Compare Means”:



Comparisons for each pair using Student's t

t	Alpha
2.08596	0.05

Abs(Dif)-LSD

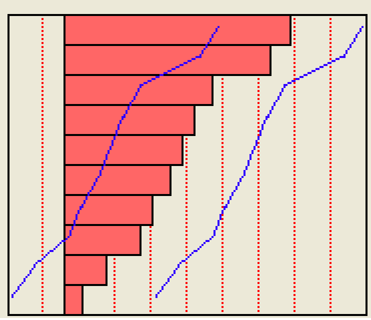
	30	25	20	35	15
30	-3.7455	0.2545	2.4545	7.0545	8.0545
25	0.2545	-3.7455	-1.5455	3.0545	4.0545
20	2.4545	-1.5455	-3.7455	0.8545	1.8545
35	7.0545	3.0545	0.8545	-3.7455	-2.7455
15	8.0545	4.0545	1.8545	-2.7455	-3.7455

Positive values show pairs of means that are significantly different.

Level				Mean
30	A			21.600000
25		B		17.600000
20		B		15.400000
35			C	10.800000
15			C	9.800000

Levels not connected by same letter are significantly different.

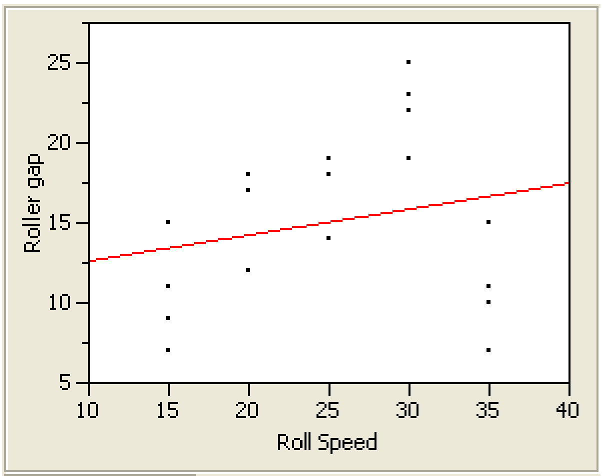
Level	- Level	Difference	Lower CL	Upper CL	p-Value
30	15	11.80000	8.05455	15.54545	<.0001*
30	35	10.80000	7.05455	14.54545	<.0001*
25	15	7.80000	4.05455	11.54545	0.0003*
25	35	6.80000	3.05455	10.54545	0.0012*
30	20	6.20000	2.45455	9.94545	0.0025*
20	15	5.60000	1.85455	9.34545	0.0054*
20	35	4.60000	0.85455	8.34545	0.0186*
30	25	4.00000	0.25455	7.74545	0.0375*
25	20	2.20000	-1.54545	5.94545	0.2347
35	15	1.00000	-2.74545	4.74545	0.5838



By the analysis, Level 30 in Group A is different from level 25 and 20 in group B. Level 25 and 20 in group B are different from 35 and 15 in group C.

(3)
Firstly, fit a first order linear model:
Let roll speed be X, roller gap be Y

$Y = \beta_0 + \beta_1 X + \varepsilon$



Summary of Fit

RSquare	0.052782
RSquare Adj	0.011599
Root Mean Square Error	5.121735
Mean of Response	15.04
Observations (or Sum Wgts)	25

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	33.62000	33.6200	1.2816
Error	23	603.34000	26.2322	Prob > F
C. Total	24	636.96000		0.2693

Lack Of Fit

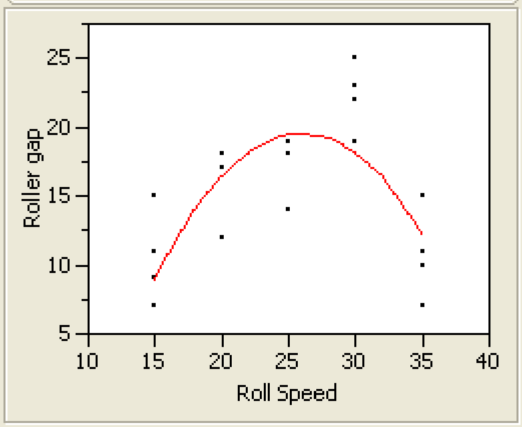
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	3	442.14000	147.380	18.2854
Pure Error	20	161.20000	8.060	Prob > F
Total Error	23	603.34000		<.0001*
				Max RSq
				0.7469

There is a significant lack of fit at the .05 level. Then try a second order model:
Let roll speed be X, roller gap be Y

$Y = \beta_0 + \beta_1 X + + \beta_2 X^2 + \varepsilon$

Choose “fit model” in “analyze”. Then add Roll speed and Roll speed*Roll speed as factors. (To add Roll speed*Roll speed, click Roll speed in the added factor area, then click cross, then click Roll speed in the Select Columns.)

Regression Plot



Actual by Predicted Plot

Summary of Fit

RSquare	0.591614
RSquare Adj	0.554488
Root Mean Square Error	3.438589
Mean of Response	15.04
Observations (or Sum Wgts)	25

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	376.83429	188.417	15.9353
Error	22	260.12571	11.824	Prob > F
C. Total	24	636.96000		<.0001*

Lack Of Fit

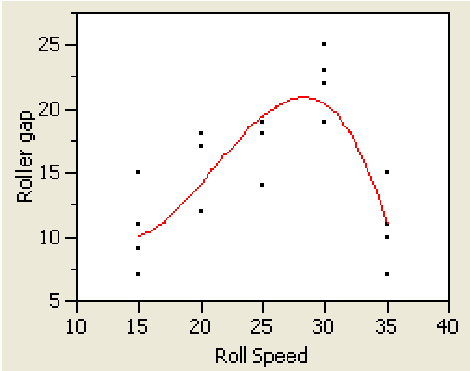
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	2	98.92571	49.4629	6.1368
Pure Error	20	161.20000	8.0600	Prob > F
Total Error	22	260.12571		0.0084*
				Max RSq
				0.7469

There is still a significant lack of fit. Then try a third order model.
Let roll speed be X, roller gap be Y

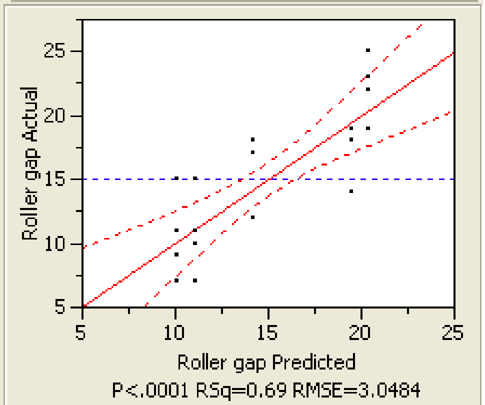
$Y = \beta_0 + \beta_1 X + + \beta_2 X^2 + \beta_3 X^3 + \varepsilon$

The third order term, Roll speed* Roll speed* Roll speed is added the similar way as the second order term Roll speed* Roll speed.

Regression Plot



Actual by Predicted Plot



Summary of Fit

RSquare	0.69363
RSquare Adj	0.649862
Root Mean Square Error	3.048385
Mean of Response	15.04
Observations (or Sum Wgts)	25

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	441.81429	147.271	15.8482
Error	21	195.14571	9.293	Prob > F
C. Total	24	636.96000		<.0001*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	1	33.94571	33.9457	4.2116
Pure Error	20	161.20000	8.0600	Prob > F
Total Error	21	195.14571		0.0535
				Max RSq
				0.7469

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.781429	6.545893	-0.12	0.9061
Roll Speed	0.81	0.259063	3.13	0.0051*
(Roll Speed-25)*(Roll Speed-25)	-0.088571	0.014574	-6.08	<.0001*
(Roll Speed-25)*(Roll Speed-25)*(Roll Speed-25)	-0.0076	0.002874	-2.64	0.0152*

There is no significant lack of fit. We can conclude a cubic model is adequate to describe the data.

Randomized Block Design

7. A study was conducted to determine effect of Roll Speed (rpm) on ribbon uniformity (dimensionless) in a roller compactor (Scenario 2).. Six different replicates were conducted on six batches of material from a blending operation. The order of selecting the samples was from the blenders were randomized as was the order of running the experiments. The data from this completely randomized block design is shown below:

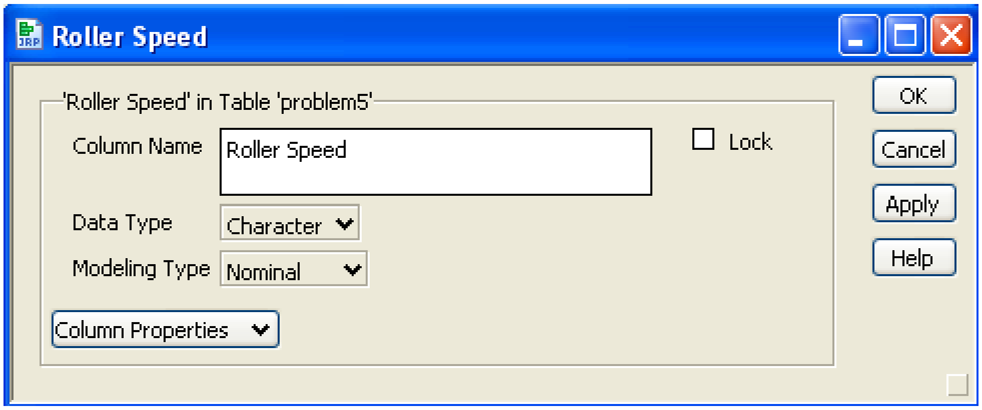
Batch Number

Roll Speed (rpm)	1	2	3	4	5	6
10	.78	.80	.81	.75	.77	.78
16	.85	.85	.92	.86	.81	.83
23	.93	.92	.95	.89	.89	.83
31	1.14	.97	.98	.88	.86	.83
40	.97	.86	.78	.76	.76	.75

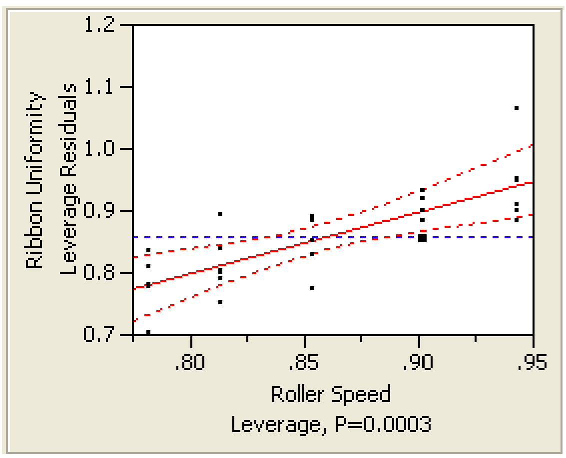
(1) Does Roll Speed affect the ribbon uniformity? Is the between batch variation significant?
(2) Determine the regression equation between roller uniformity and roll speed. Compare the results with a)
(3) Are the residuals from this experiment normally distributed?

Solution:

(1) In JMP, double click the tab of “Roll speed” and choose the data type as “Character”:



Use “Fit model” in “Analyze” as in the previous problems:



Summary of Fit

RSquare	0.742294
RSquare Adj	0.626327
Root Mean Square Error	0.053526
Mean of Response	0.858667
Observations (or Sum Wgts)	30

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	0.16504667	0.018339	6.4009
Error	20	0.05730000	0.002865	Prob > F
C. Total	29	0.22234667		0.0003*

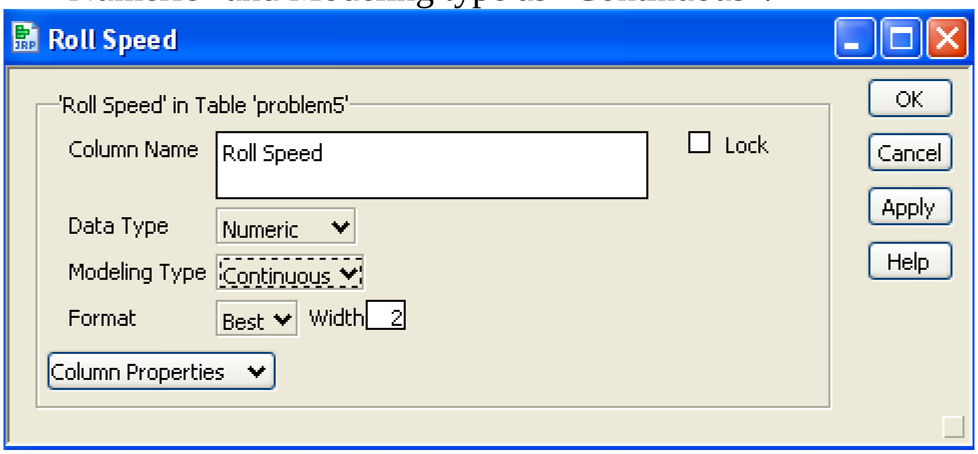
Parameter Estimates

Effect Tests

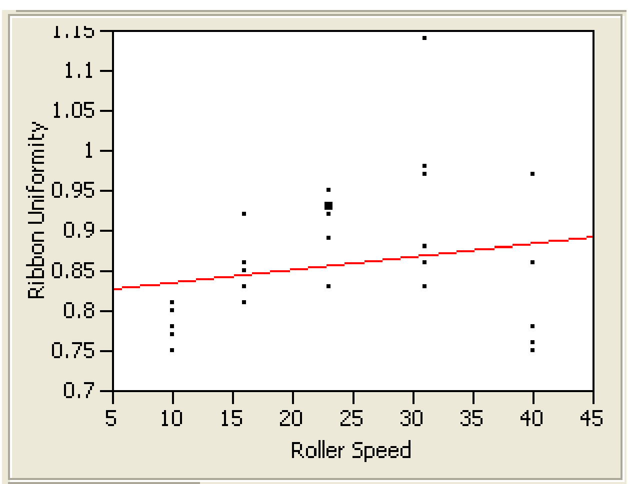
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Roll Speed	4	4	0.10218000	8.9162	0.0003*
Batch	5	5	0.06286667	4.3886	0.0074*

Roll Speed affects the ribbon uniformity at the .05 level since the p value is .0003. There is significant variation between the Batches at the .05 level since the p value is .0074.

(2) Double click the tab of "Roll speed" and choose the data type as "Numeric" and Modeling type as "Continuous":



Use "Fit Y by X" in "Analyze" as in the previous problems:



Summary of Fit

RSquare	0.042576
RSquare Adj	0.008383
Root Mean Square Error	0.087194
Mean of Response	0.858667
Observations (or Sum Wgts)	30

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	3	0.09271330	0.030904	6.4295
Pure Error	25	0.12016667	0.004807	Prob > F
Total Error	28	0.21287997		0.0022*
				Max RSq
				0.4596

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.00946670	0.009467	1.2451
Error	28	0.21287997	0.007603	Prob > F
C. Total	29	0.22234667		0.2740

Parameter Estimates

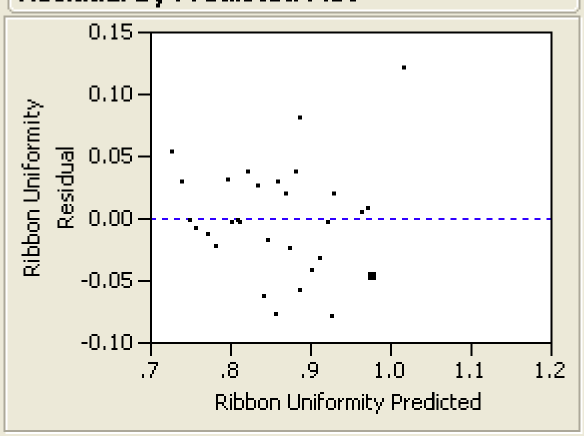
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.818596	0.039281	20.84	<.0001*
Roller Speed	0.0016696	0.001496	1.12	0.2740

The Roll speed is not significant in this model which has a significant lack of fit in this linear regression model. Comparing the results with (a), the Batch effect has been lumped in with experimental error dramatically increasing its size and limiting the suitability of the regression analysis. It is necessary to remove the batch effect to get an effective model.

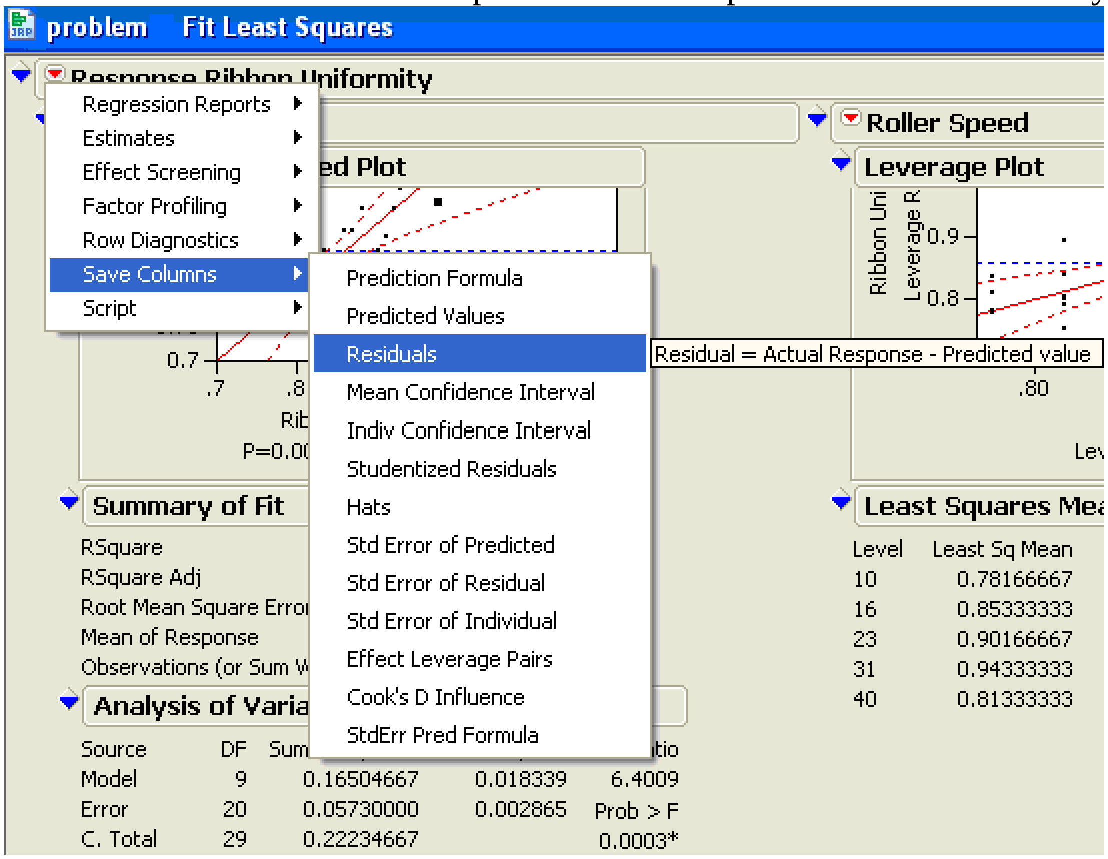
(3)

In (1), get the residual plot from the results:

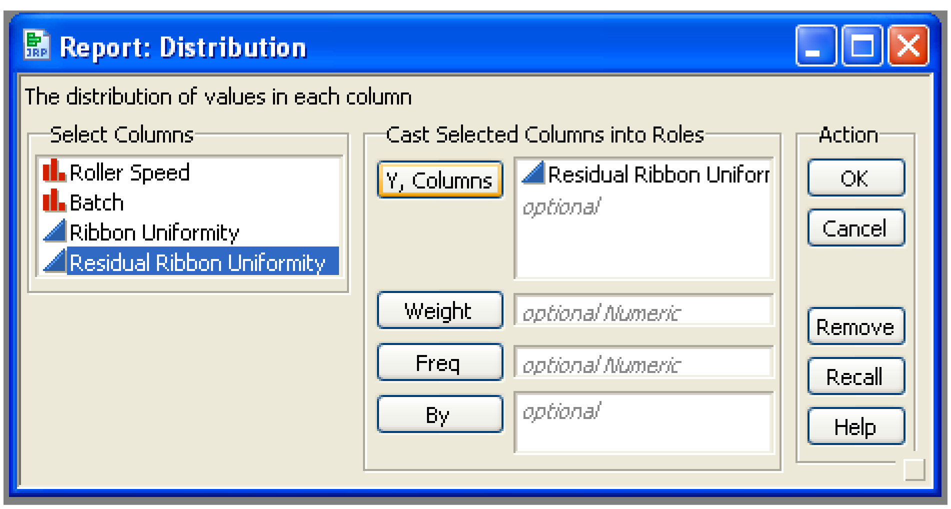
Residual by Predicted Plot

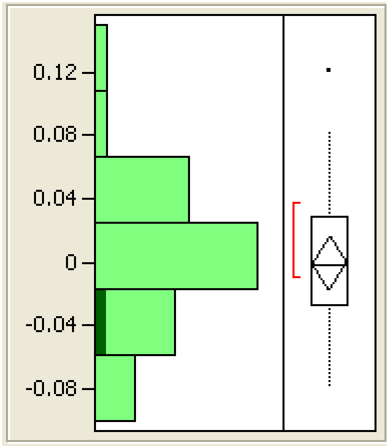


To further check its normality, save the residual by choosing “Residuals” in “Save Columns” from the hot spot aside the Response Ribbon Uniformity:



Then we analyze it in “Distribution”:





The residuals are normally distributed.

Optimization Problem.

8. The product uniformity y from a continuous blender in scenario 2 is related to the tilt(deg) T by the relationship:

$$Y = \begin{cases} 100-(20.5-T)^2+\varepsilon, & \text{if } Y>0 \\ 0, & \text{if } Y\leq 0 \end{cases}$$

It is clear from the above relationship that the maximum uniformity is obtained at T=20.5

Show how (1) dichotomous search and (2) golden section search can be used to search out this optimum over the region $0 \le T \le 50$ where the measurement error at any point is

$\varepsilon \sim N(0,.25)$

The smallest difference in T which can be detected is 2 degree.

(Hint: Program the relationship in Excel using the available random number generator)

Solution:

In excel, input Y = 100-(20.5-D2)^2+0.5*RAND() as the uniformity generator.

(1) Dichotomous search:

Step	Working interval		middle point	T	Y
1	0	50	25	24	88.13416
	0	50	25	26	69.75396
2	0	26	13	12	27.75556
	0	26	13	14	57.95248
3	12	26	19	18	93.89717
	12	26	19	20	100.2358
4	18	26	22	21	100.0315
	18	26	22	23	94.03289

Note in step 1 since Y(26) < Y(24), the optimum cannot lie in the interval (26,50) which is dropped. The rest steps are similar.

Since the smallest detectable difference is 2, we find the maximum is close to (20, 21) as expected.

(2) Golden section method:

Step	Working interval		T	Y
1	0	50	19.1	98.42316
	0	50	30.9	0

2	0	30.9	11.8	24.34522
3	11.8	30.9	23.6	90.42856
4	11.8	23.6	16.3	82.70831
5	16.3	23.6	20.8	100.2592

In step 1, by gold section ratio, 50*.618 =30.9, 50*.382=19.1. Since the uniformity is greater at 19.1 than at 30.9, the interval (30.9, 50) cannot contain the optimum. The next experiment is located at 11.8 symmetrically with the (0, 30.9) interval. (30.9*.382=11.8)

Since only smallest detectable difference is 2, we find the maximum is close to 20.8 as expected.

Comparing these two methods, Dichotomous search requires 8 runs while Golden section only 6.

Factorial Experimentation

9. A study is conducted to assess the effect of Pressure (Ton) and Punch Distance (mm) on percent dissolution of a new API after 80 minutes in a Tablet Press in Scenario 2. Three different replicates were taken at random at three pressures and two Punch Distances The data are as follows:

	Pressure (Ton)		
Punch Distance (mm)	.75	1	1.5
1	74,64,50	73,61,44	78,85,92
2	92,86,68	98,73,88	66,45,85

(1) Build a mathematical model to describe the mathematical relationship between %Dissolution and (Pressure, Punch Distance).

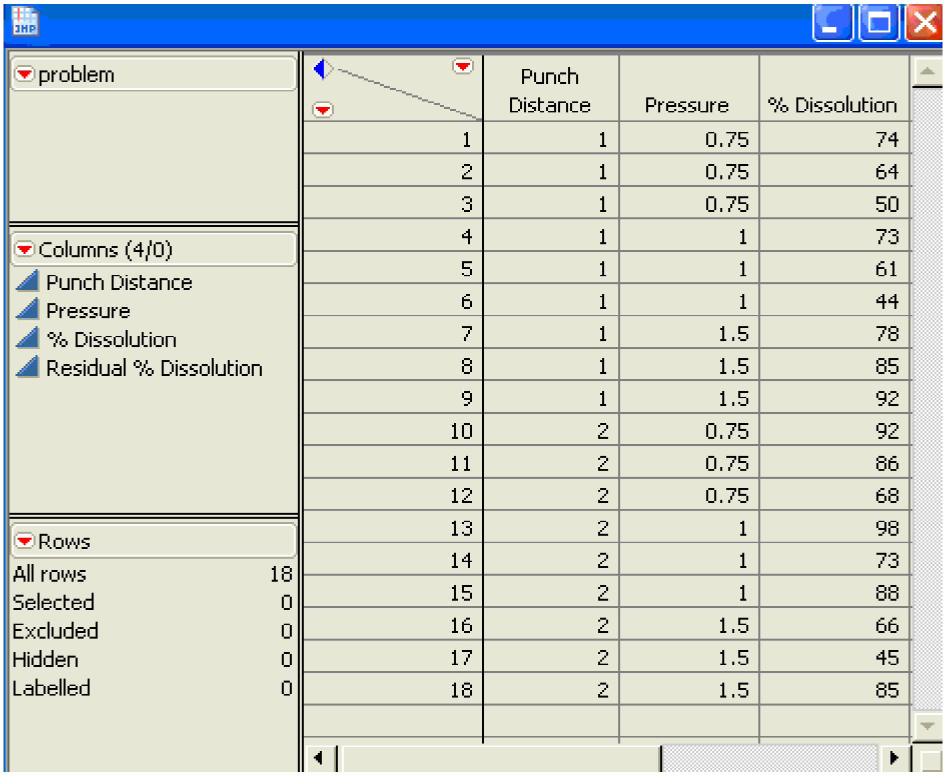
(2) Analyze the residuals from this experiment.

Solution:

(1) (a) The mathematical model for a 2*3 full factorial experiment is:

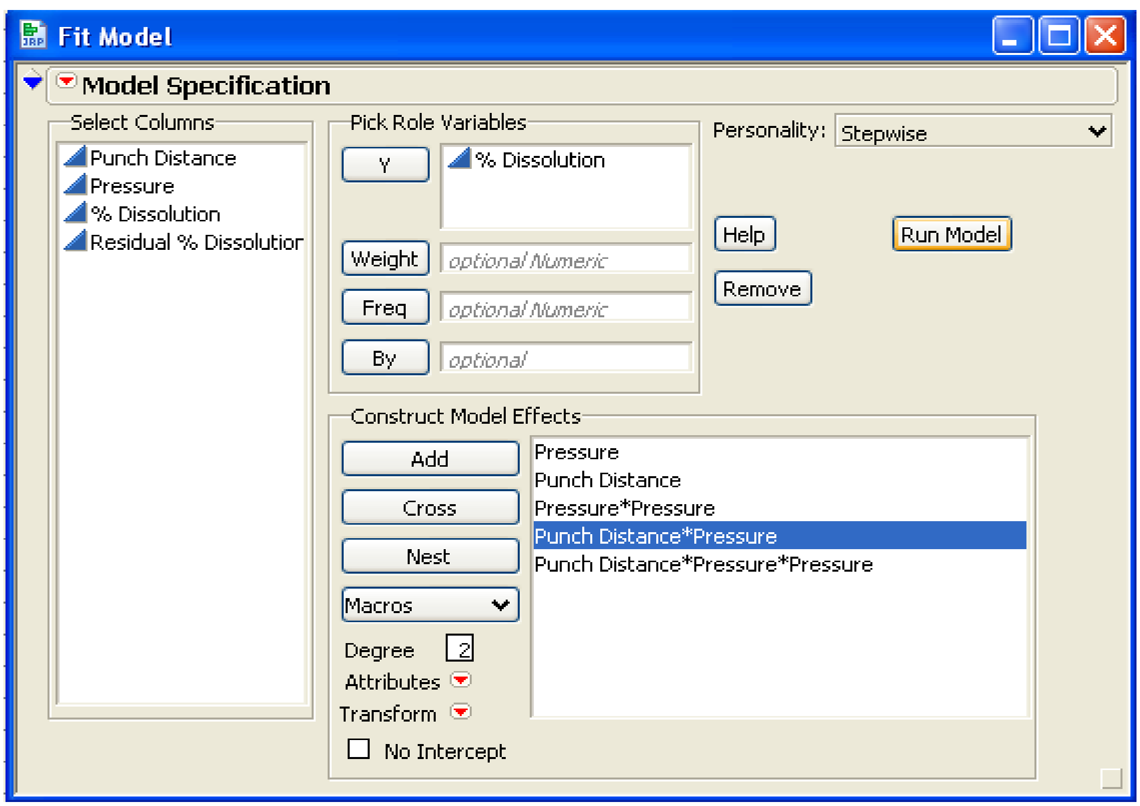
$$Y = \beta_0 + \beta_1 P + \beta_2 D + \beta_3 PD + \beta_4 P^2 + \beta_5 P^2 D$$

(b) Input the data in JMP:

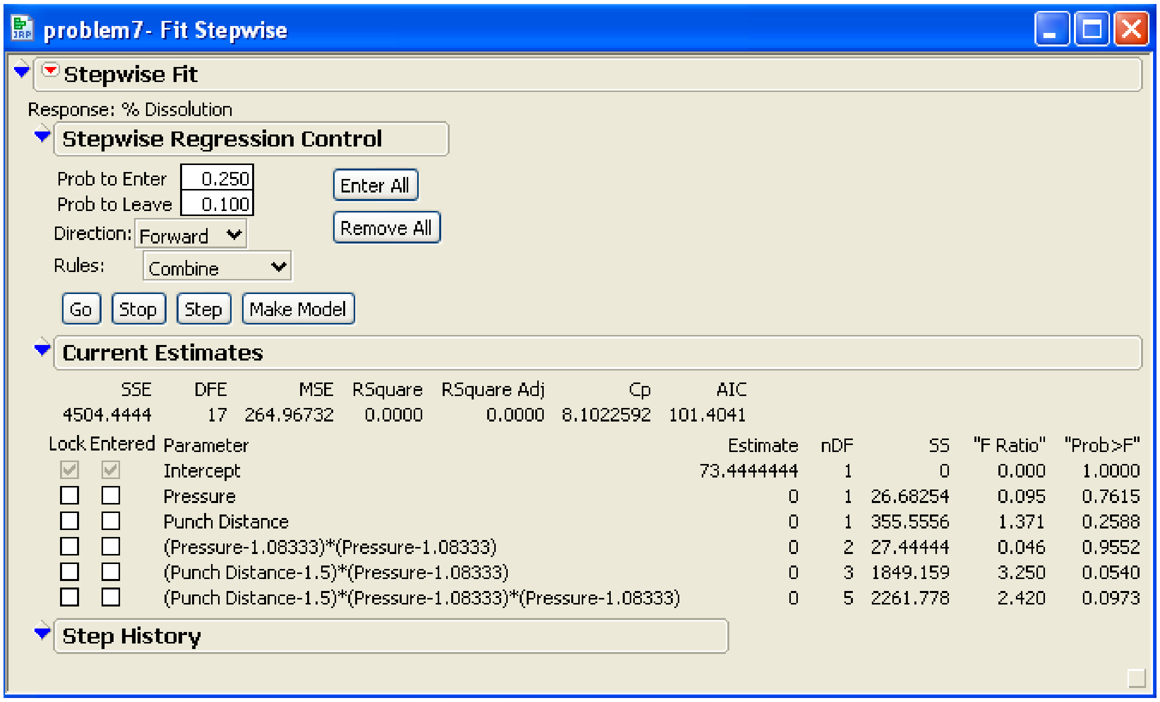


	Punch Distance	Pressure	% Dissolution
1	1	0.75	74
2	1	0.75	64
3	1	0.75	50
4	1	1	73
5	1	1	61
6	1	1	44
7	1	1.5	78
8	1	1.5	85
9	1	1.5	92
10	2	0.75	92
11	2	0.75	86
12	2	0.75	68
13	2	1	98
14	2	1	73
15	2	1	88
16	2	1.5	66
17	2	1.5	45
18	2	1.5	85

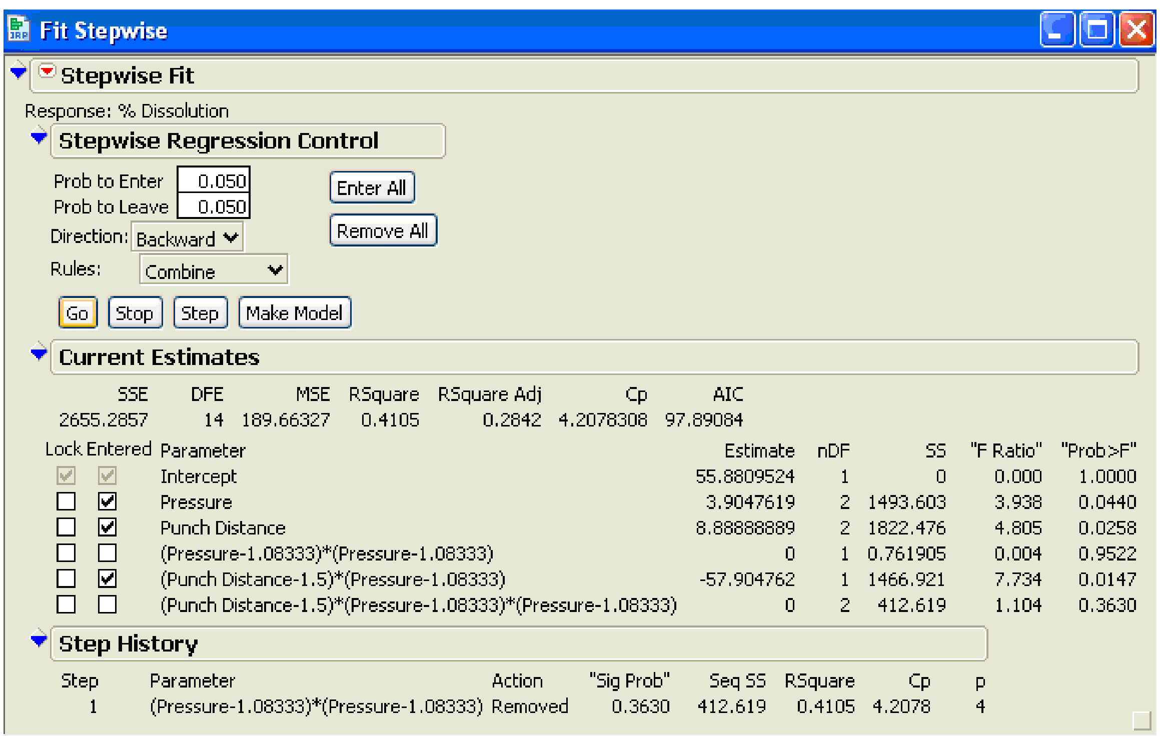
(c) Use stepwise regression. Input the response and all the factors as in the mathematical model in (a).



(c) Hit “Run Model”:



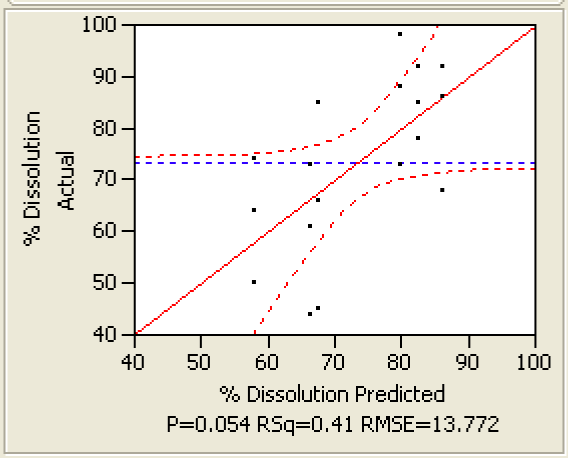
(d) Input .05 as Entry and Exit α level. Choose “Backward” in “Direction”. Hit “Enter All” and “Go”:



Now since the interaction term is significant, for the sake of easy explanation, we keep both main effects from the interaction in the model.

(e) Hit “Make Model” and run the model:

Actual by Predicted Plot



Summary of Fit

Analysis of Variance

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	2	412.6190	206.310	1.1039
Pure Error	12	2242.6667	186.889	Prob > F
Total Error	14	2655.2857		0.3630
				Max RSq
				0.5021

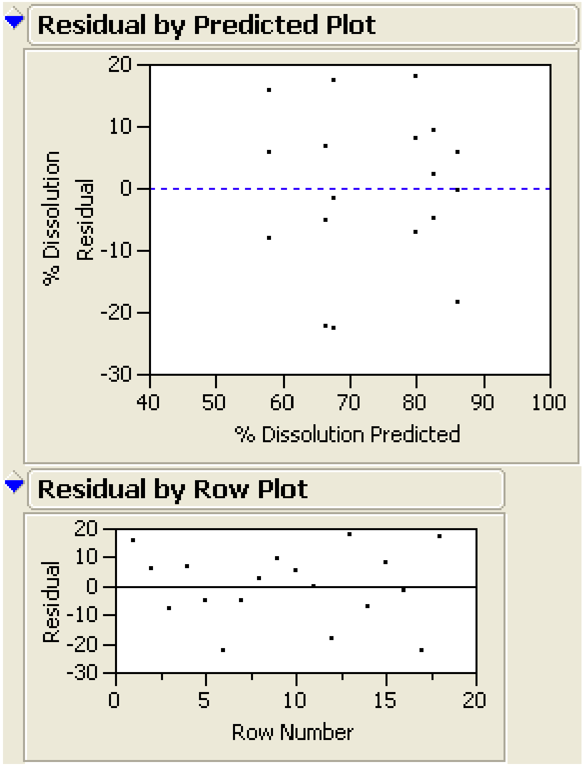
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	55.880952	15.25003	3.66	0.0026*
Pressure	3.9047619	10.41052	0.38	0.7132
Punch Distance	8.8888889	6.492102	1.37	0.1925
(Punch Distance-1.5)*(Pressure-1.08333)	-57.90476	20.82105	-2.78	0.0147*

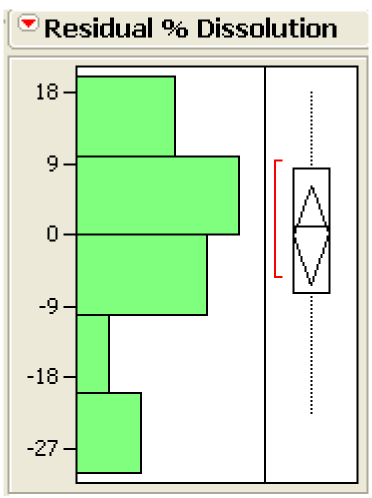
The final model is:

Y = 55.88 + 3.90P + 8.89D -57.90PD

(2) Get the residual plots in the analysis in (1):



The residuals seem randomly scattered.



But its normality needs further test.

10. Design a full factorial experiment to determine the effect of Tilt, Speed, Load and Inlet powder flow on the uniformity and density in a series of batch runs in a continuous blender in scenario 2. Consider the following cases:

(a) All factors at two levels.

(b) All factors at three levels.

(c) Tilt at 2 levels, Speed at three levels, load at four levels and inlet powder flow at 2 levels.

(1) For each of these cases give the following:

i) the actual experiments that must be run.

ii) the mathematical model

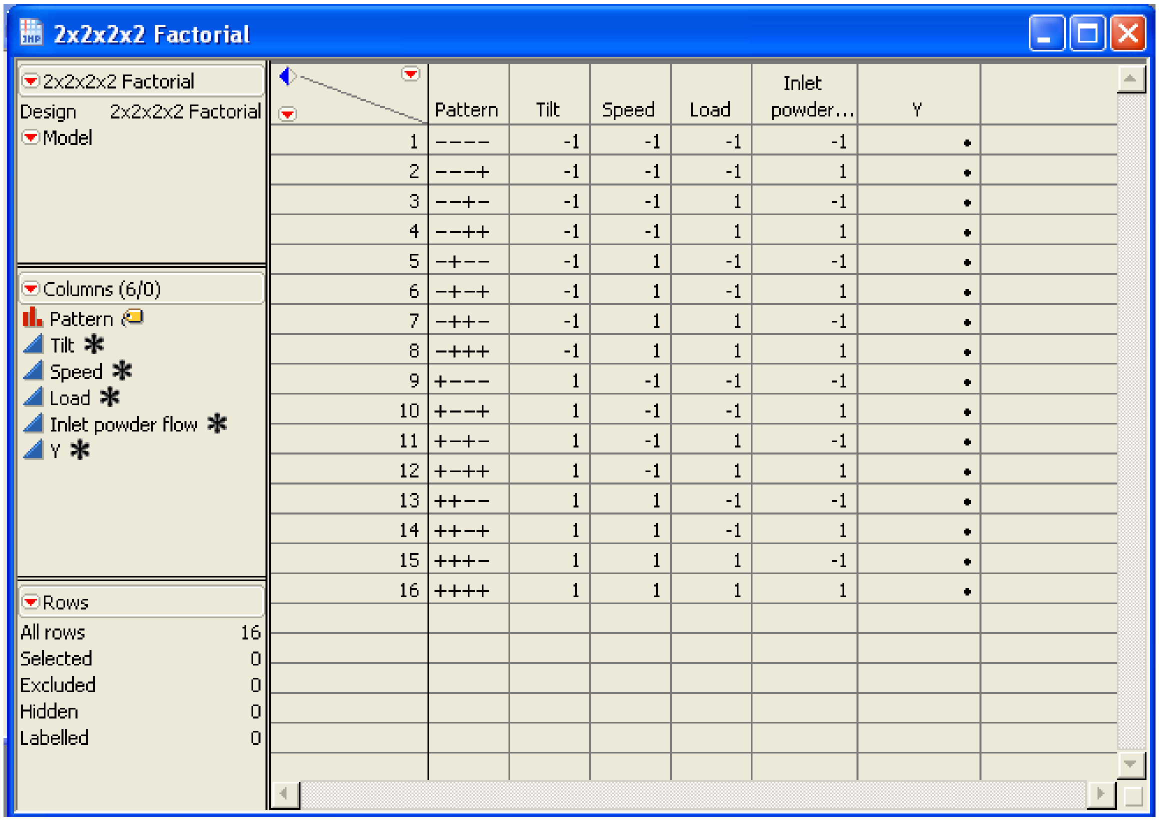
(2) Describe the role of replication, randomization and blocking

Solution:

(1)

(a)

i) Use “Full Factorial Design” in “DOE”, input the factors and levels. Hit “Make Table”:



	Pattern	Tilt	Speed	Load	Inlet powder…	Y
1	−−−−	-1	-1	-1	-1	•
2	−−−+	-1	-1	-1	1	•
3	−−+−	-1	-1	1	-1	•
4	−−++	-1	-1	1	1	•
5	−+−−	-1	1	-1	-1	•
6	−+−+	-1	1	-1	1	•
7	−++−	-1	1	1	-1	•
8	−+++	-1	1	1	1	•
9	+−−−	1	-1	-1	-1	•
10	+−−+	1	-1	-1	1	•
11	+−+−	1	-1	1	-1	•
12	+−++	1	-1	1	1	•
13	++−−	1	1	-1	-1	•
14	++−+	1	1	-1	1	•
15	+++−	1	1	1	-1	•
16	++++	1	1	1	1	•

The mathematical model is:

(Where T is for Tilt, S is for Speed, L is for Load, I is for Inlet powder flow)

$$Y = \mu + T + S + L + I + TS + TL + TI + SL + SI + LI + TSL + TSI + TLI + SLI + TSLI + \varepsilon$$

(b) Use the same method as in (a)(The table is copied from JMP):

Pattern	Tilt	Speed	Load	Inlet Powder Flow
1111	1	1	1	1
1112	1	1	1	2
1113	1	1	1	3
1121	1	1	2	1
1122	1	1	2	2

1123	1	1	2	3
1131	1	1	3	1
1132	1	1	3	2
1133	1	1	3	3
1211	1	2	1	1
1212	1	2	1	2
1213	1	2	1	3
1221	1	2	2	1
1222	1	2	2	2
1223	1	2	2	3
1231	1	2	3	1
1232	1	2	3	2
1233	1	2	3	3
1311	1	3	1	1
1312	1	3	1	2
1313	1	3	1	3
1321	1	3	2	1
1322	1	3	2	2
1323	1	3	2	3
1331	1	3	3	1
1332	1	3	3	2
1333	1	3	3	3
2111	2	1	1	1
2112	2	1	1	2
2113	2	1	1	3
2121	2	1	2	1
2122	2	1	2	2
2123	2	1	2	3
2131	2	1	3	1
`2132	2	1	3	2
2133	2	1	3	3
2211	2	2	1	1
2212	2	2	1	2
2213	2	2	1	3
2221	2	2	2	1
2222	2	2	2	2
2223	2	2	2	3
2231	2	2	3	1
2232	2	2	3	2
2233	2	2	3	3

2311	2	3	1	1
2312	2	3	1	2
2313	2	3	1	3
2321	2	3	2	1
2322	2	3	2	2
2323	2	3	2	3
2331	2	3	3	1
2332	2	3	3	2
2333	2	3	3	3
3111	3	1	1	1
3112	3	1	1	2
3113	3	1	1	3
3121	3	1	2	1
3122	3	1	2	2
3123	3	1	2	3
3131	3	1	3	1
3132	3	1	3	2
3133	3	1	3	3
3211	3	2	1	1
3212	3	2	1	2
3213	3	2	1	3
3221	3	2	2	1
3222	3	2	2	2
3223	3	2	2	3
3231	3	2	3	1
3232	3	2	3	2
3233	3	2	3	3
3311	3	3	1	1
3312	3	3	1	2
3313	3	3	1	3
3321	3	3	2	1
3322	3	3	2	2
3323	3	3	2	3
3331	3	3	3	1
3332	3	3	3	2
3333	3	3	3	3

.

The mathematical model is:

(Where T is for Tilt, S is for Speed, L is for Load, I is for Inlet powder flow)

$Y = \mu + T + S + L + I + TS + TL + TI + SL + SI + LI + TSL + TSI + TLI + SLI + TSLI + T^2 + S^2 + L^2 + I^2 + T^2S + TS^2 + T^2S^2 + T^2L + TL^2 + T^2L^2 + T^2I + TI^2 + T^2I^2 + S^2L + SL + SL^2 + S^2I + SI^2 + S^2I^2 + L^2I + LI^2 + L^2I^2 + T^2SL + TS^2L + TSL^2 + T^2S^2L + T^2SL^2 + TS^2L^2 + T^2S^2L^2 + T^2SI + TS^2I + TSI^2 + T^2S^2I + TS^2I^2 + T^2SI^2 + T^2S^2I^2 + T^2LI + TL^2I + TLI^2 + T^2L^2I + T^2LI^2 + TL^2I^2 + T^2L^2I^2 + S^2LI + SL^2I + SLI^2 + S^2L^2I + S^2LI^2 + SL^2I^2 + S^2L^2I^2 + T^2SLI + TS^2LI + TSL^2I + TSLI^2 + T^2S^2LI + T^2SL^2I + T^2SLI^2 + TS^2L^2I + TS^2LI^2 + TSL^2I^2 + T^2S^2L^2I + T^2S^2LI^2 + T^2SL^2I^2 + TS^2L^2I^2 + T^2S^2L^2I^2 + \varepsilon$

(c) Use the same method as in (a) (The table is copied from JMP):

Pattern	Tilt	Speed	Load	Inlet Powder Flow
−11−	-1	1	1	-1
−11+	-1	1	1	1
−12−	-1	1	2	-1
−12+	-1	1	2	1
−13−	-1	1	3	-1
−13+	-1	1	3	1
−14−	-1	1	4	-1
−14+	-1	1	4	1
−21−	-1	2	1	-1
−21+	-1	2	1	1
−22−	-1	2	2	-1
−22+	-1	2	2	1
−23−	-1	2	3	-1
−23+	-1	2	3	1
−24−	-1	2	4	-1
−24+	-1	2	4	1
−31−	-1	3	1	-1
−31+	-1	3	1	1
−32−	-1	3	2	-1
−32+	-1	3	2	1

−33−	-1	3	3	-1
−33+	-1	3	3	1
−34−	-1	3	4	-1
−34+	-1	3	4	1
+11−	1	1	1	-1
+11+	1	1	1	1
+12−	1	1	2	-1
+12+	1	1	2	1
+13−	1	1	3	-1
+13+	1	1	3	1
+14−	1	1	4	-1
+14+	1	1	4	1
+21−	1	2	1	-1
+21+	1	2	1	1
+22−	1	2	2	-1
+22+	1	2	2	1
+23−	1	2	3	-1
+23+	1	2	3	1
+24−	1	2	4	-1
+24+	1	2	4	1
+31−	1	3	1	-1
+31+	1	3	1	1
+32−	1	3	2	-1
+32+	1	3	2	1
+33−	1	3	3	-1
+33+	1	3	3	1
+34−	1	3	4	-1
+34+	1	3	4	1

The mathematical model is:

(Where T is for Tilt, S is for Speed, L is for Load, I is for Inlet powder flow)

$$Y = \mu + T + S + L + I + TS + TL + TI + SL + SI + LI + TSL + TSI + TLI + SLI + TSLI + S^2 + L^2 + L^3 + TS^2 + TL^2 + TL^3 + S^2L^2 + S^2L^3 + S^2I + L^2I + L^3I + TS^2L + TSL^2 + TSL^3 + TS^2L^2 + TS^2L^3 + TS^2I + TL^2I + TL^3I + S^2LI + SL^2I + SL^3I + S^2L^2I + S^2L^3I + TS^2LI + TSL^2I + TSL^3I + TS^2L^2I + TSL^3I + \varepsilon$$

(2) Replication provides the estimate of pure error. Randomization is necessary for conclusions drawn from the experiment to be correct, unambiguous and defensible. Randomization eliminates the batch effects. Blocking may show the batch effects.

Fractional Factorial Experiments with two levels

11. In the investigation of the conditions of filtration during the preparation of an API, the objective was to improve the quality of the product. Four factors were examined:

A. Concentration of liquor when filtered (concentrated v. dilute)

B. Effect of Liquor Storage (fresh vs old). The liquor was either filtered immediately or kept a week before filtration.

C. Presence or absence of an anti-frothing agent.

D. Temperature of Filtration (high vs low)

It was considered unlikely that large interactions would exist between these factors so that a ½ replicate of a 2^4 factorial was selected with defining contrast D=ABC. The purity of the product was recorded in the table below:

Run No..	A	B	C	D	Purity
1	-1	-1	-1	-1	107
2	1	-1	-1	+1	114
3	-1	1	-1	1	122
4	+1	+1	-1	-1	130
5	-1	-1	1	1	106
6	1	-1	+1	-1	121
7	-1	+1	+1	-1	120
8	1	1	1	1	132

Determine:

(1) The pattern of aliases for the experiment.

(2) The main effects and interactions

(3) If the error in the measurements is 2 units, which factors are significant?

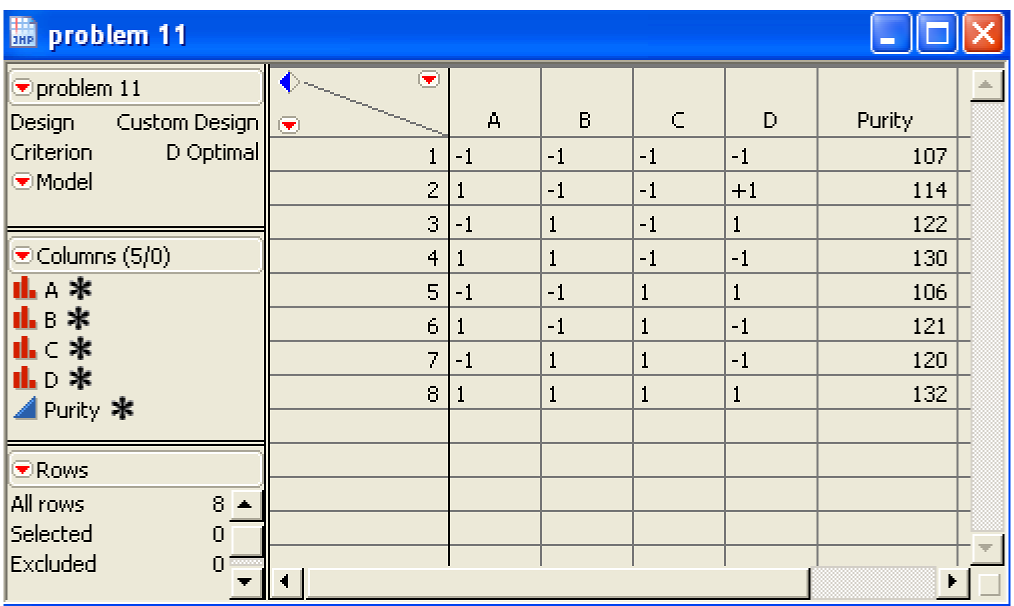
Solution”

(1) A= BCD, B=ACD, C=ABD, D=ABC.

and

AB=CD, AC=BD, AD=BC

(2) Input the data in the JMP:



	A	B	C	D	Purity
1	-1	-1	-1	-1	107
2	1	-1	-1	+1	114
3	-1	1	-1	1	122
4	1	1	-1	-1	130
5	-1	-1	1	1	106
6	1	-1	1	-1	121
7	-1	1	1	-1	120
8	1	1	1	1	132

Run "Fit Model" in "Analyze" with main effects A, B, C and interactions AB, AC and ABC as factors:

Summary of Fit

RSquare	0.992991
RSquare Adj	0.950935
Root Mean Square Error	2.12132
Mean of Response	119
Observations (or Sum Wgts)	8

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	637.50000	106.250	23.6111
Error	1	4.50000	4.500	Prob > F
C. Total	7	642.00000		0.1562

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	119	0.75	158.67	0.0040*
A[-1]	-5.25	0.75	-7.00	0.0903
B[-1]	-7	0.75	-9.33	0.0680
C[-1]	-0.75	0.75	-1.00	0.5000
A[-1]*B[-1]	-0.25	0.75	-0.33	0.7952
A[-1]*C[-1]	1.5	0.75	2.00	0.2952
A[-1]*B[-1]*C[-1]	0.5	0.75	0.67	0.6257

(3) Calculate the Z statistic and check the Z value as:

Term	Estimate	error	Z statistic	Prob>|Z|
A	-5.25	2	-2.625	0.0087
B	-7	2	-3.5	0.0005
C	-0.75	2	-0.375	0.7077
AB	-0.25	2	-0.125	0.9005
AC	1.5	2	0.75	0.4533
ABC	0.5	2	0.25	0.8026

Main effects A and B are significant at .05 level.

12. O.L. Davies. The following experiment was conducted in a batch reactor (Scenario 1) to investigate conditions affecting the yield of an API. Five factors were investigated with the following levels:

Factors	Level	
A A/B Feed ratio	Low	High
B Amount of Liquid Catalyst	Concentrated	Dilute
C Amount of Anti-foaming agent	None	Some
D Time of Reaction	Short	Fast
E Agitation	Slow	Fast

Setting the signs of D=-AE and C=+AB, the following Percent Yield data were obtained (the analysis for each run was repeated)

Design of Experiment and Product Yield

Run No	A	B	C	D	E	Yield
1	-1	-1	+1	-1	-1	53.1,54.6
2	+1	-1	-1	+1	-1	49.3,48.4
3	-1	+1	-1	-1	-1	50.1,51.4
4	+1	+1	+1	+1	-1	68.3,67.4
5	-1	-1	+1	+1	+1	73.4,75.3
6	+1	-1	-1	-1	+1	79.7,78.0
7	-1	+1	-1	+1	+1	84.5,86.4
8	+1	+1	+1	-1	+1	81.3,80.4

(1) What are the defining contrasts?

(2) Determine the pattern of aliases.

(3) What are the significant main effects and interactions?

(4) Is there a significant lack of fit?

(5) Based on this data what is the optimal way to run the reaction?

Solution:

(1) D=-AE and C=+AB

The defining contrasts are:

I = -ADE = ABC = -BCDE

(2) A = -DE = BC = -ABCDE

B= -ABDE = AC = -CDE

C=-ACDE = AB = -BDE

D= -AE = ABCD = -BCE

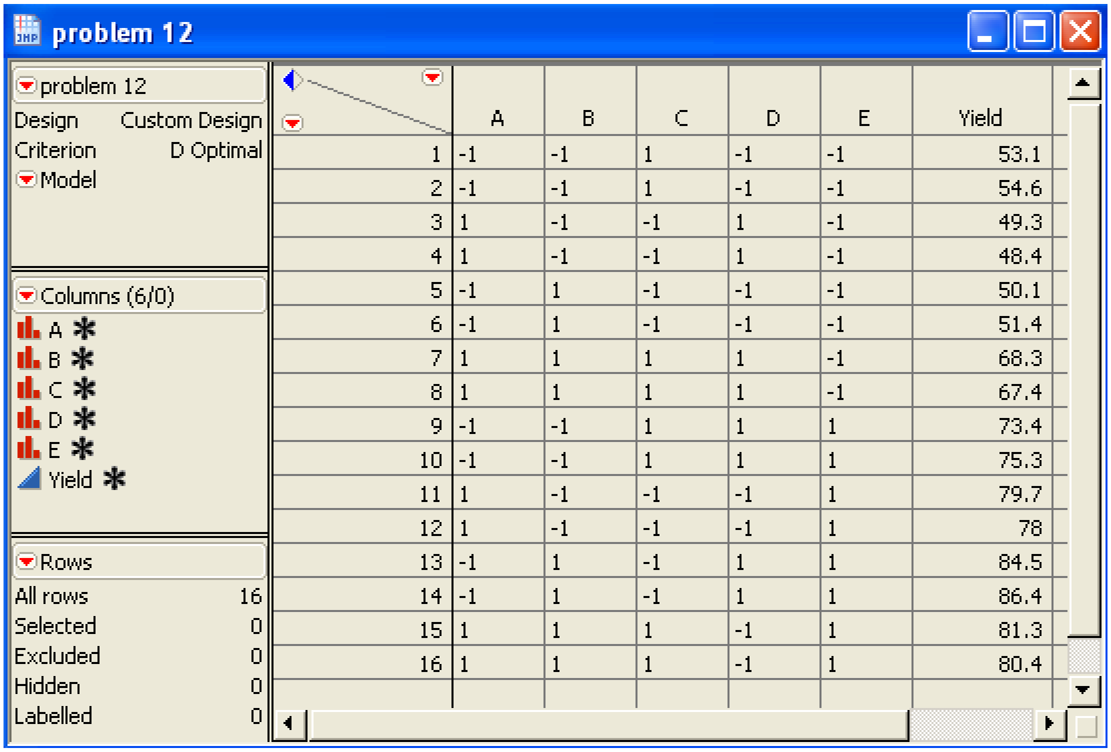
E= -AD= ABCE = -BCD

BD = -ABE = ACD = -CE

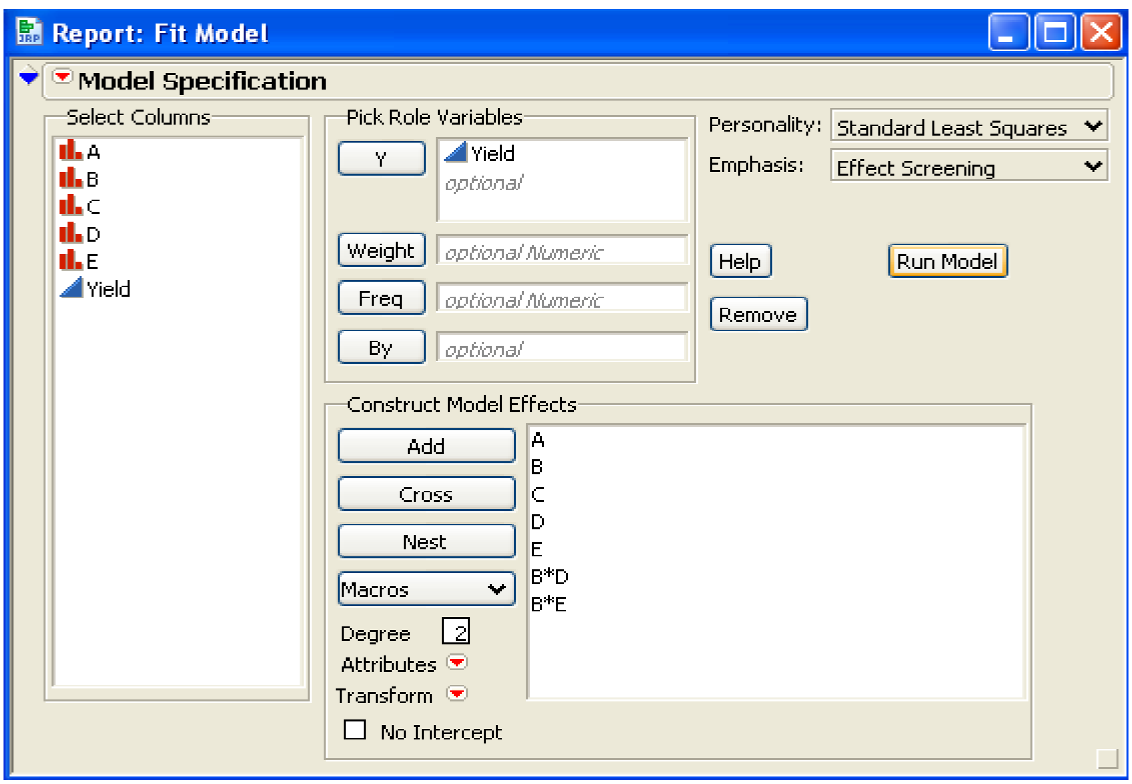
BE = - ABD = ACE = -CD

(3)

(a) Input the data in the JMP:



(b) Input the response and the factors:



(c) Run the model:

Summary of Fit

RSquare	0.997244
RSquare Adj	0.994833
Root Mean Square Error	1.014889
Mean of Response	67.6
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	2981.8400	425.977	413.5700
Error	8	8.2400	1.030	Prob > F
C. Total	15	2990.0800		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	67.6	0.253722	266.43	<.0001*
A[-1]	-1.5	0.253722	-5.91	0.0004*
B[-1]	-3.625	0.253722	-14.29	<.0001*
C[-1]	-1.625	0.253722	-6.40	0.0002*
D[-1]	-1.525	0.253722	-6.01	0.0003*
E[-1]	-12.275	0.253722	-48.38	<.0001*
B[-1]*D[-1]	3.9	0.253722	15.37	<.0001*
B[-1]*E[-1]	-0.35	0.253722	-1.38	0.2051

All the main effects are significant on the .05 level. BD interaction is also significant on the .05 level.

(4) Remove BE interaction, run the model again:

Summary of Fit

RSquare	0.996589
RSquare Adj	0.994315
Root Mean Square Error	1.064581
Mean of Response	67.6
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	2979.8800	496.647	438.2176
Error	9	10.2000	1.133	Prob > F
C. Total	15	2990.0800		<.0001*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	1	1.960000	1.96000	1.9029
Pure Error	8	8.240000	1.03000	Prob > F
Total Error	9	10.200000		0.2051
				Max RSq
				0.9972

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	67.6	0.266145	254.00	<.0001*
A[-1]	-1.5	0.266145	-5.64	0.0003*
B[-1]	-3.625	0.266145	-13.62	<.0001*
C[-1]	-1.625	0.266145	-6.11	0.0002*
D[-1]	-1.525	0.266145	-5.73	0.0003*
E[-1]	-12.275	0.266145	-46.12	<.0001*
B[-1]*D[-1]	3.9	0.266145	14.65	<.0001*

There is no significant lack of fit on the .05 level.

(5) To maximize the yield, all the main effects should be run on the low level.

13. In the batch reaction API yield study described in scenario 1, it was decided to make a series of runs including temperate as well as the other five factors. Based on their previous success they were allowed to conduct 16 runs.

(1) Design a fractional factorial experiment which is a ¼ fraction of a 2^6 full factorial experiment which maximizes the probability of testing for the significant of main effect and two factor interactions.

(2) What are the defining contrasts and pattern of aliases for this design.

(3) List the considerations in deciding which fraction to run.

Solution:

(1) A Resolution IV design with generators E = ABC and F = BCD is:

Run	A	B	C	D	E=ABC	F=BCD
1	-	-	-	-	-	-
2	+	-	-	-	+	-
3	-	+	-	-	+	+
4	+	+	-	-	-	+
5	-	-	+	-	+	+
6	+	-	+	-	-	+
7	-	+	+	-	-	-
8	+	+	+	-	+	-
9	-	-	-	+	-	+
10	+	-	-	+	+	+
11	-	+	-	+	+	-
12	+	+	-	+	-	-
13	-	-	+	+	+	-
14	+	-	+	+	-	-
15	-	+	+	+	-	+
16	+	+	+	+	+	+

(2) Generators:

E = ABC and F = BCD

The defining contrasts are:

I = ABCE = BCDF = ADEF

The aliases pattern are:

A = BCE = DEF = ABCDF

B = ACE = CDF = ABDEF

C = ABE = BDF = ACDEF

D = BCF = AEF = ABCDE

E = ABC = ADF = BCDEF

F = BCD = ADE = ABCEF

AB = CE = ACDF = BDEF

AC = BE = ABDF = CDEF

AD = EF = BCDE = ABCF

AE = BC = DF = ABCDEF

AF = DE = BCEF = ABCD

BD = CF = ACDE = ABEF

BF = CD = ACEF = ABDE

ABD = CDE = ACF = BEF

ACD = BDE = ABF = CEF

(3) All fractions have the same extent of confounding between main effects and interactions. Frequently several experiments are already available and it is wise to select for the fraction in which the greatest number of existing experiments has been run. Another consideration is the actual level of the experiments. Run the easiest ones. For example, the run with all the factors at their highest level might be difficult. Carefully go over the potential difficulties before selecting the fraction.

Response Surface Modeling and Optimization

14. An experiment was run in a batch reactor to determine the effect of temperature and reaction time on the yield of the API. These factors are coded as x1= (temperature -300deg)/50deg and x2=(time-10hrs)/5 hours. The following coded data was obtained where the yield is in percent

Run No	X1	X2	Yield (%)
1	-1	0	78.03
2	1	0	80.4
3	0	0	80.1
4	0	0	80.95
5	0	-1	80.3
6	0	1	80.08
7	0	0	80.97
8	-1.4142	-1.4142	74.38
9	-1.4142	1.4142	74.87
10	1.4142	-1.4142	75.68
11	1.4142	1.4142	78.13
12	0	0	80.44

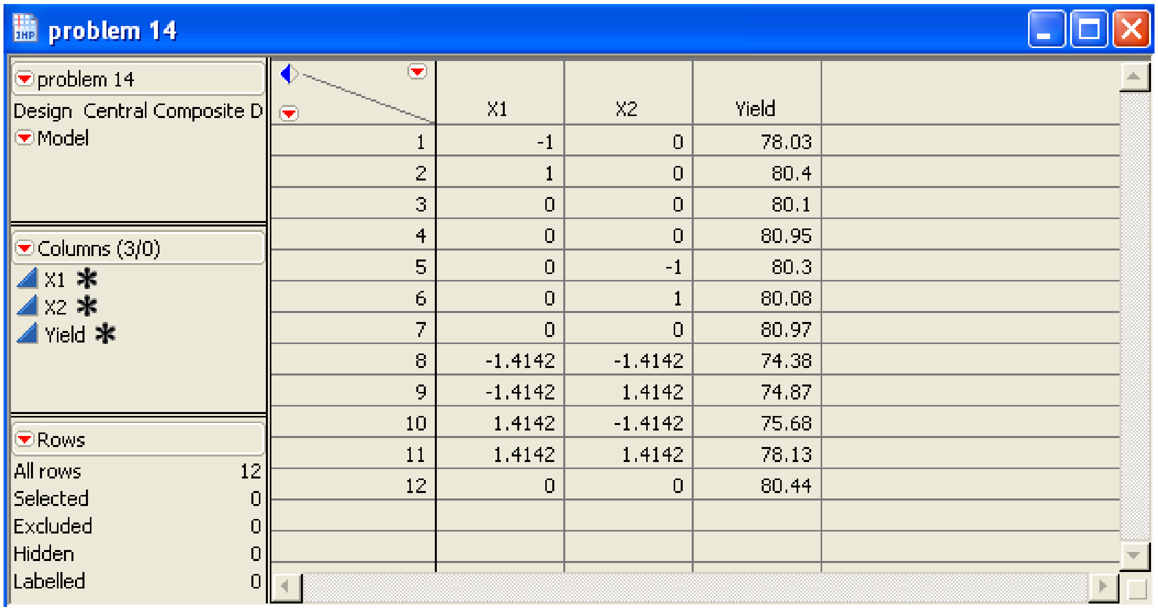
(1) Fit a response surface model to the data. Is the model adequate to describe the data?

(2) Plot the yield response curve. What recommendations would you make about the operating conditions for the reactor?

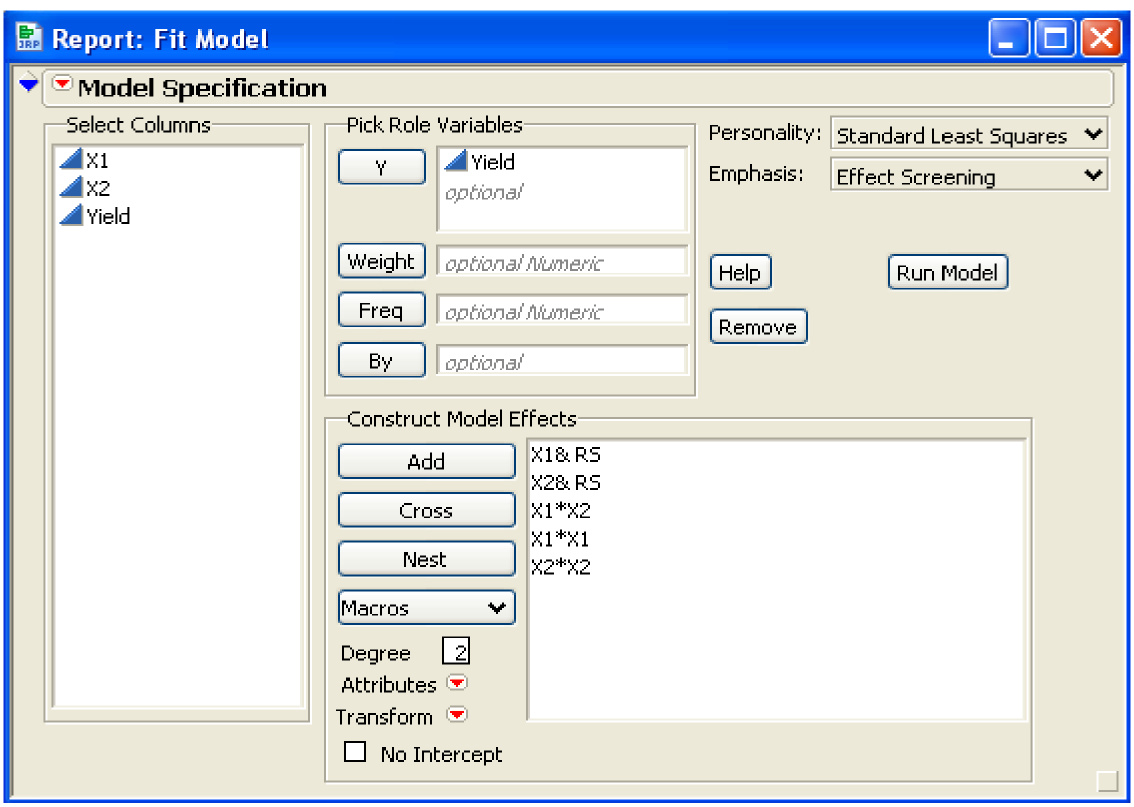
Solution

(1)

(a) Input the data:



(b) Run script in “Model”:



(3) Run model:

Summary of Fit

RSquare	0.975376
RSquare Adj	0.954855
Root Mean Square Error	0.519592
Mean of Response	78.69417
Observations (or Sum Wgts)	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	64.162636	12.8325	47.5321
Error	6	1.619856	0.2700	Prob > F
C. Total	11	65.782492		<.0001*

Lack Of Fit

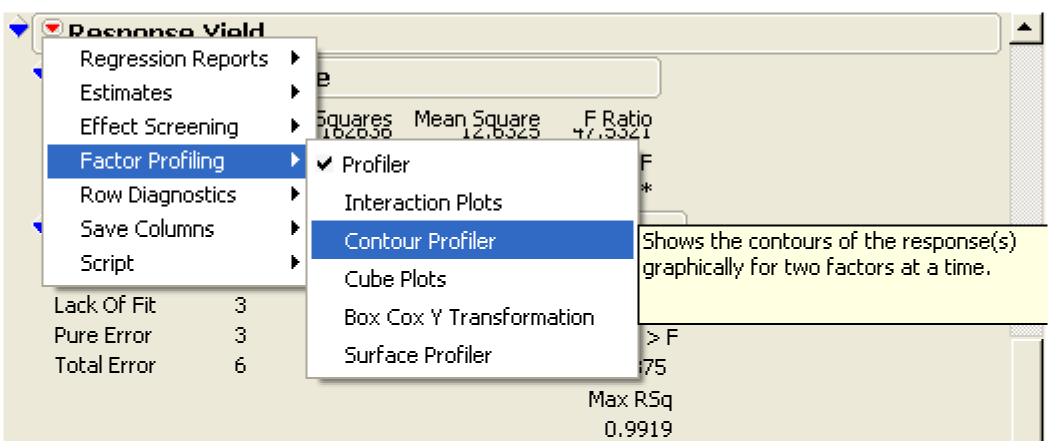
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	3	1.0857557	0.361919	2.0329
Pure Error	3	0.5341000	0.178033	Prob > F
Total Error	6	1.6198557		0.2875
				Max RSq
				0.9919

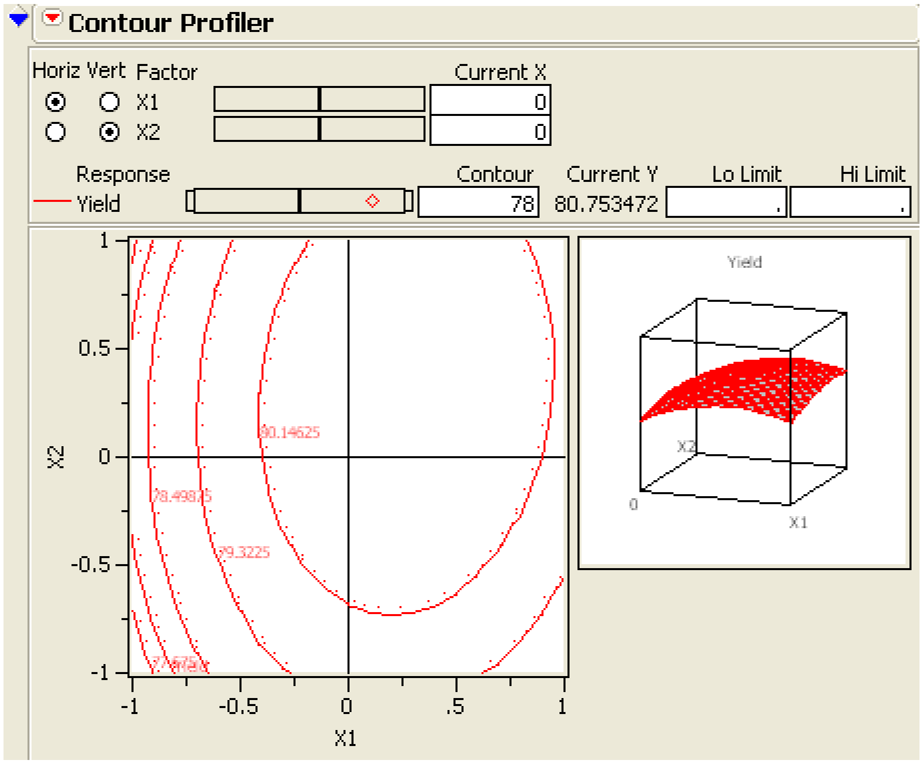
Parameter Estimates

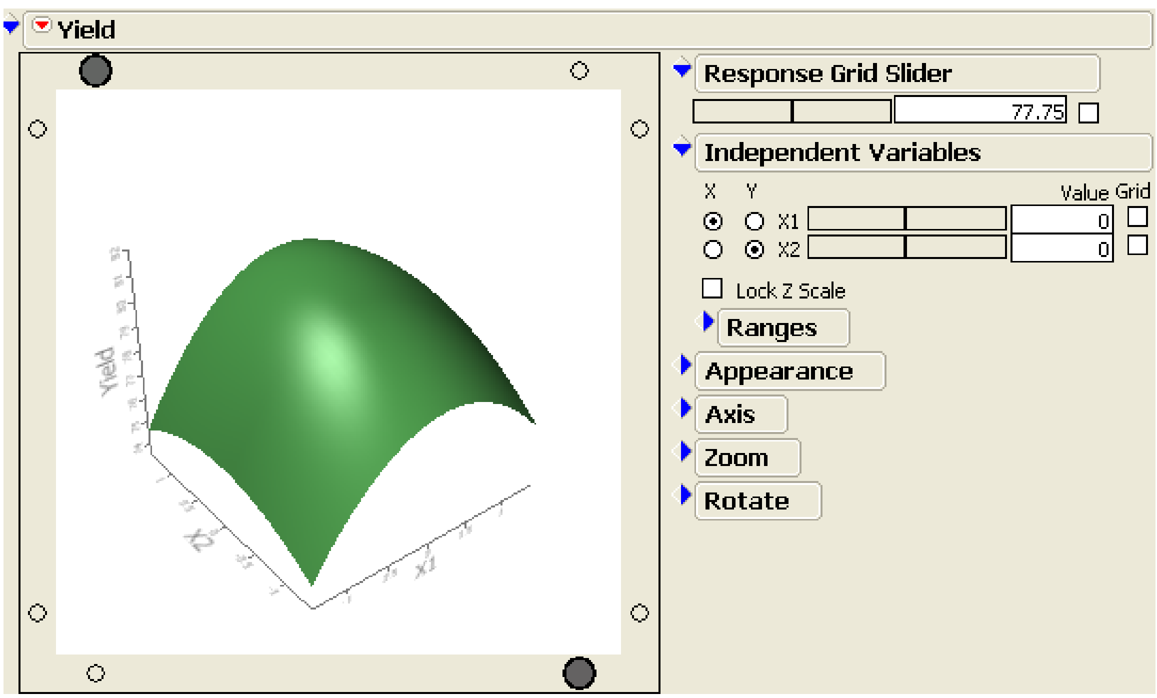
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	80.753472	0.210074	384.41	<.0001*
X1	0.8818887	0.164311	5.37	0.0017*
X2	0.3937808	0.164311	2.40	0.0535
X1*X2	0.2450047	0.129901	1.89	0.1082
X1*X1	-1.723102	0.274376	-6.28	0.0008*
X2*X2	-0.748102	0.274376	-2.73	0.0343*

Since the p-value of lack of fit test is large than .05, the model is adequate.

(2) Choose “Contour Profiler” and “Surface Profiler” in “Factor Profiling” by clicking the hot spot aside the “Response Yield”:




Contour Profiler
Horiz Vert Factor
Current X
X1
X2
0
0
Response
Contour
Current Y
Lo Limit
Hi Limit
Yield
78
80.753472
X2
X1
1
0.5
0
-0.5
-1
-1
-0.5
0
.5
1
80.14625
78.49875
79.3225
Yield
X2
X1
0


Yield
Response Grid Slider
77.75
Independent Variables
X
Y
Value
Grid
X1
X2
0
0
Lock Z Scale
Ranges
Appearance
Axis
Zoom
Rotate
Yield
X2
X1

Response Surface

Coef

	X1	X2	Yield
X1	-1.723102	0.2450047	0.8818887
X2	.	-0.748102	0.3937808

Solution

Variable	Critical Value
X1	0.2778471
X2	0.3086843

Solution is a Maximum

Predicted Value at Solution 80.936764

The solution is a maximum. The maximum will be reached at:

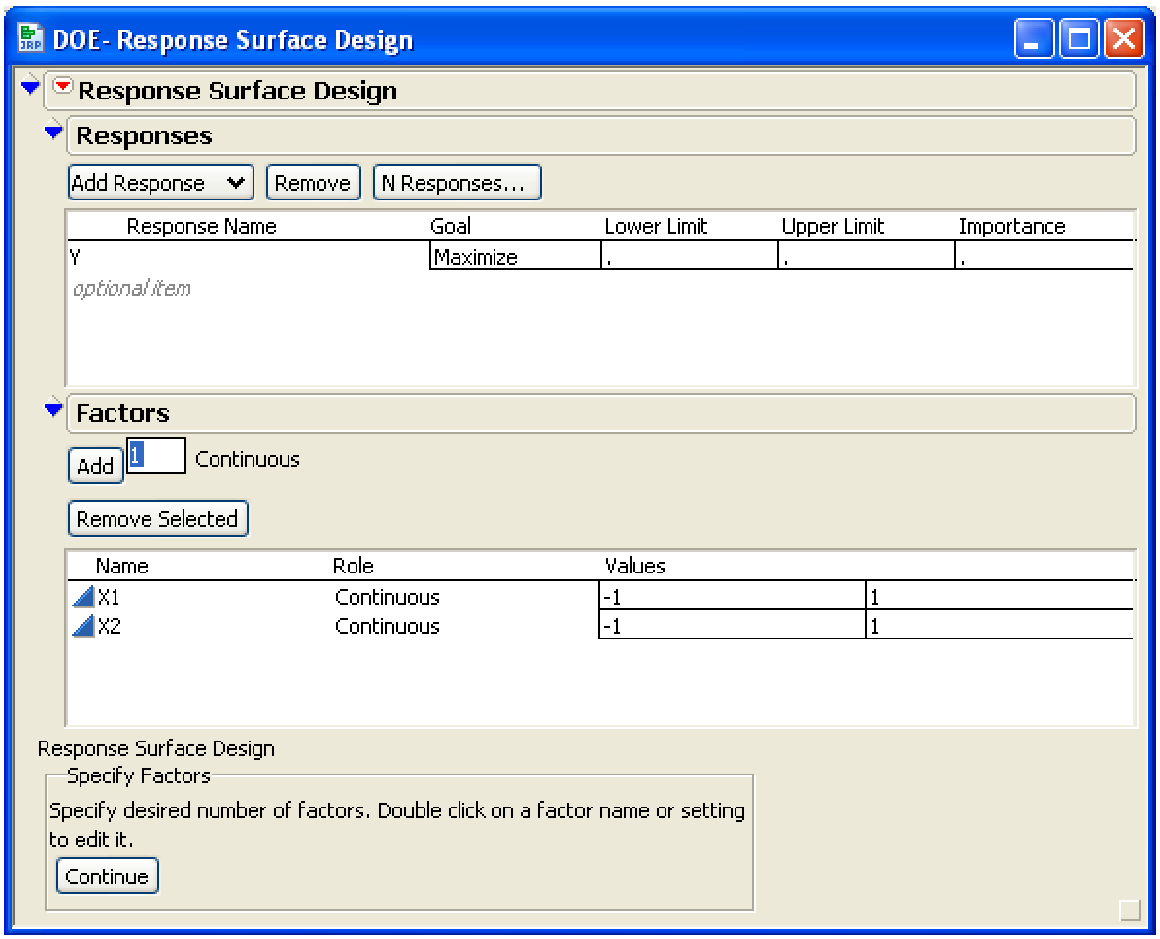
X1=.278, X2 =.309

15. Design a Central Composite Design, a Three Level Factorial Design and a Box Behnken design for generating a response surface for yield in a batch reactor system(Scenario 1) where the effect of temperature, termination time and agitation rate are to be investigated. Compare the features of the three designs in terms of the number of runs required.

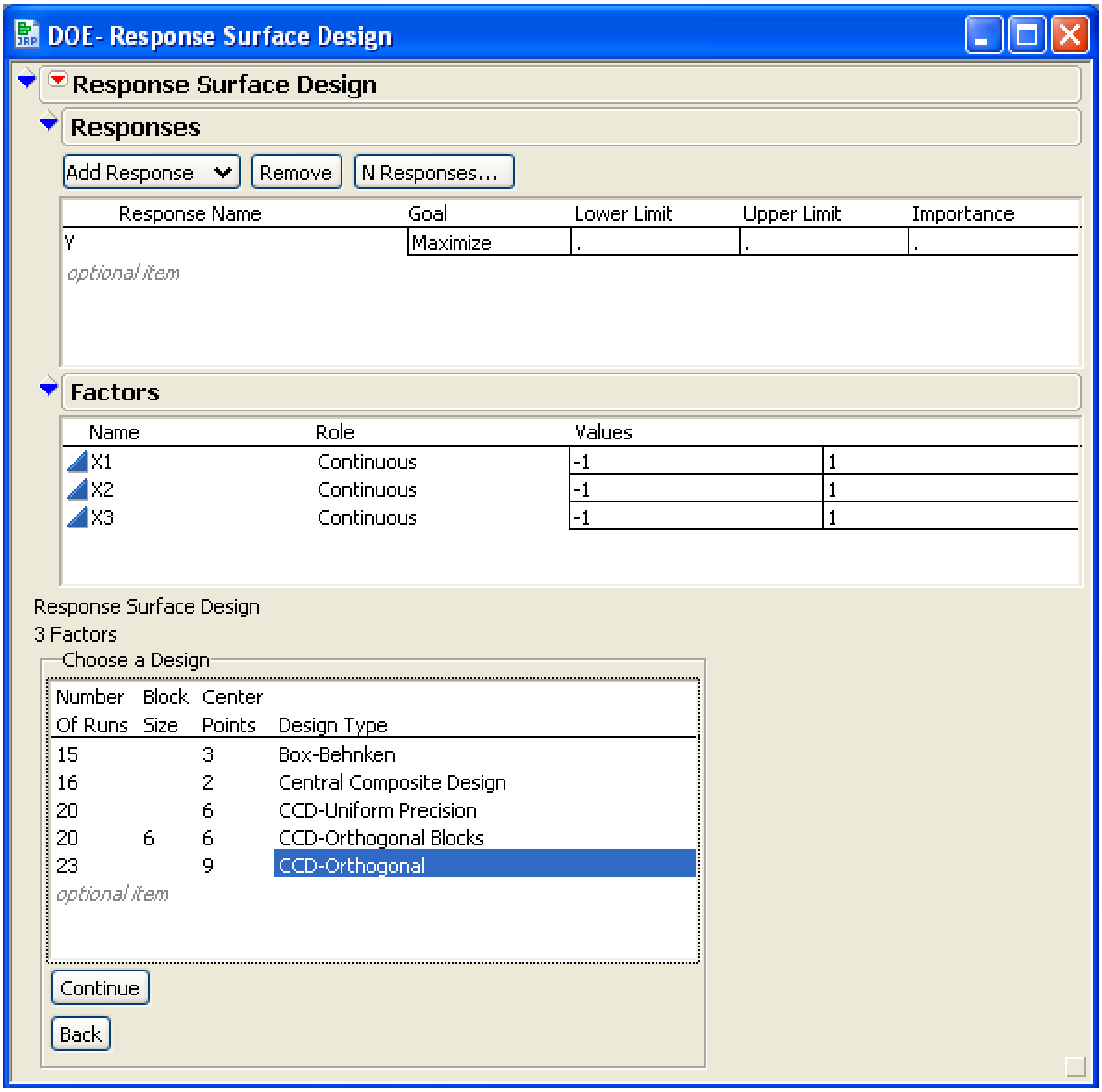
Solution

Let X1 = Temperature, X2 = Termination time, X3 = Agitation rate and Y = Yield:

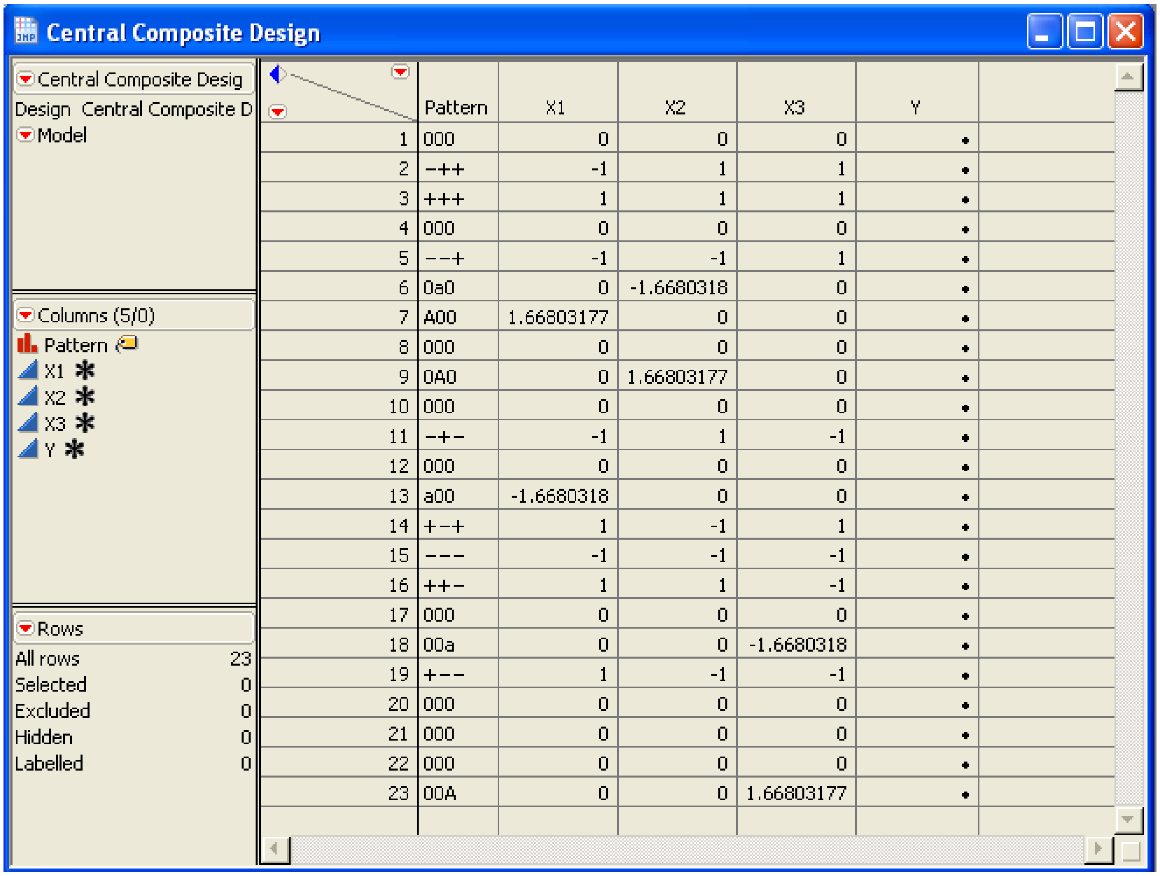
(1) CCD. Choose “Response Surface Design” in “DOE”.



Input factors and continue. Choose CCD-Orthogonal:



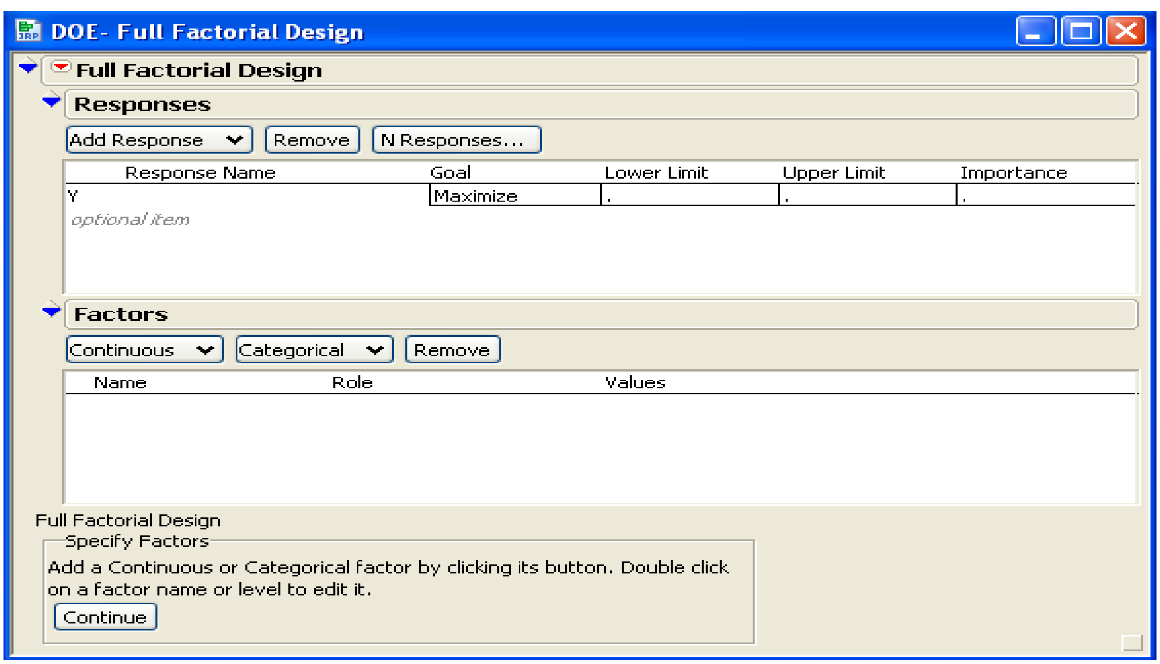
Continue. Make the table:

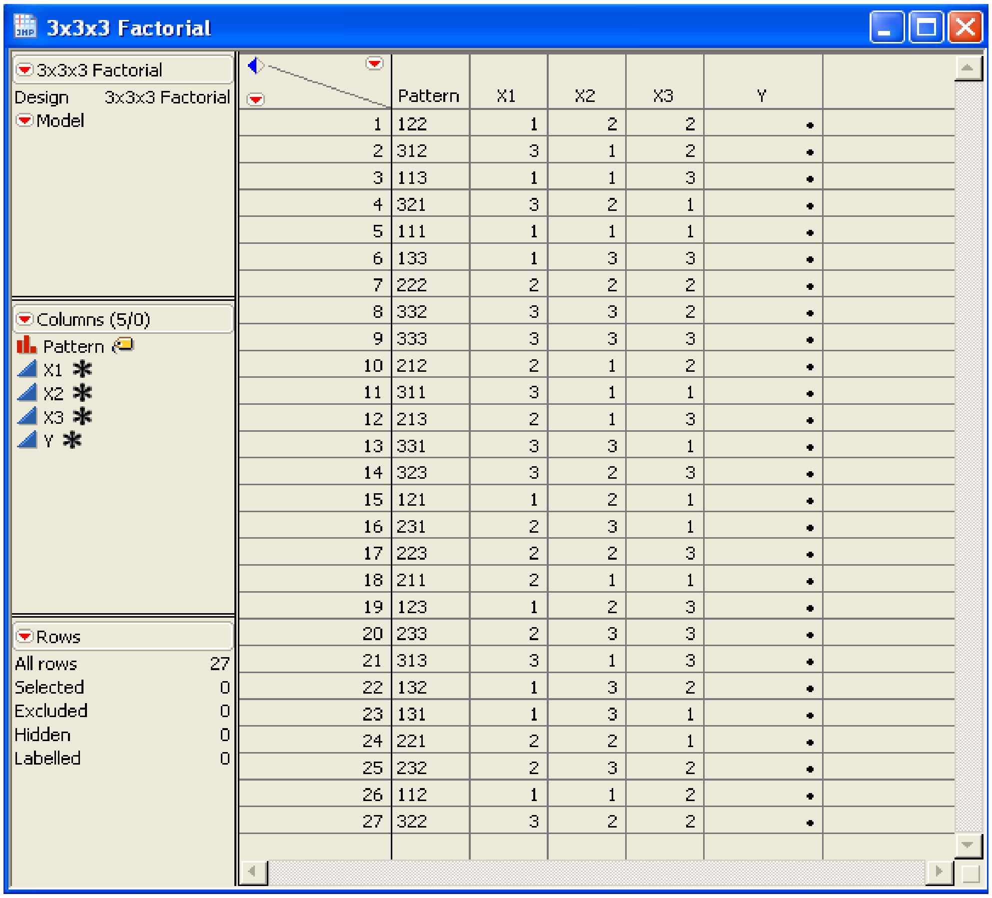


	Pattern	X1	X2	X3	Y
1	000	0	0	0	•
2	−++	-1	1	1	•
3	+++	1	1	1	•
4	000	0	0	0	•
5	−−+	-1	-1	1	•
6	0a0	0	-1.6680318	0	•
7	A00	1.66803177	0	0	•
8	000	0	0	0	•
9	0A0	0	1.66803177	0	•
10	000	0	0	0	•
11	−+−	-1	1	-1	•
12	000	0	0	0	•
13	a00	-1.6680318	0	0	•
14	+−+	1	-1	1	•
15	−−−	-1	-1	-1	•
16	++−	1	1	-1	•
17	000	0	0	0	•
18	00a	0	0	-1.6680318	•
19	+−−	1	-1	-1	•
20	000	0	0	0	•
21	000	0	0	0	•
22	000	0	0	0	•
23	00A	0	0	1.66803177	•

(2) 3 level factorial design

Choose “Full Factorial Design” in “DOE”:





3x3x3 Factorial

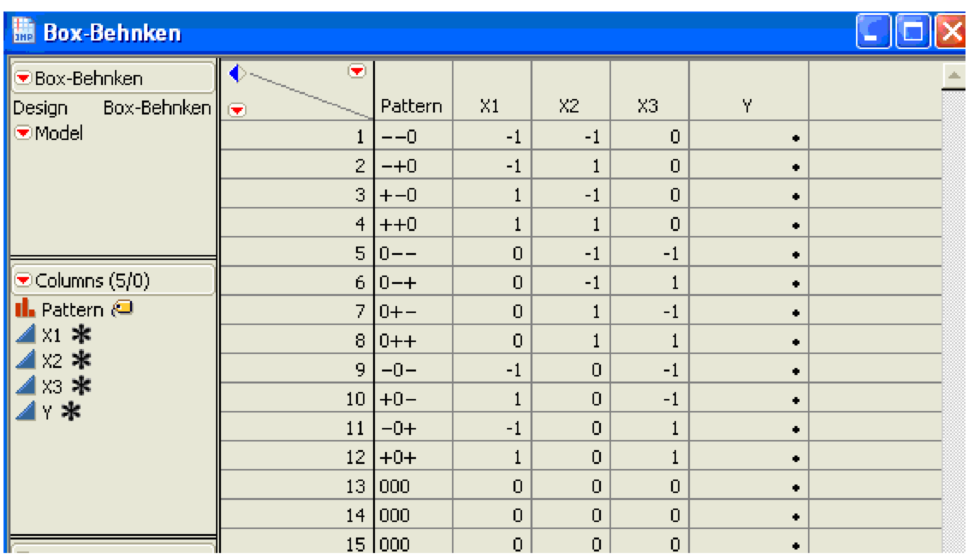
3x3x3 Factorial
Design 3x3x3 Factorial
Model

Columns (5/0)
Pattern
X1
X2
X3
Y

Rows
All rows 27
Selected 0
Excluded 0
Hidden 0
Labelled 0

	Pattern	X1	X2	X3	Y
1	122	1	2	2	•
2	312	3	1	2	•
3	113	1	1	3	•
4	321	3	2	1	•
5	111	1	1	1	•
6	133	1	3	3	•
7	222	2	2	2	•
8	332	3	3	2	•
9	333	3	3	3	•
10	212	2	1	2	•
11	311	3	1	1	•
12	213	2	1	3	•
13	331	3	3	1	•
14	323	3	2	3	•
15	121	1	2	1	•
16	231	2	3	1	•
17	223	2	2	3	•
18	211	2	1	1	•
19	123	1	2	3	•
20	233	2	3	3	•
21	313	3	1	3	•
22	132	1	3	2	•
23	131	1	3	1	•
24	221	2	2	1	•
25	232	2	3	2	•
26	112	1	1	2	•
27	322	3	2	2	•

(3) Box- Behnken Design:



Box-Behnken

Box-Behnken
Design Box-Behnken
Model

Columns (5/0)
Pattern
X1
X2
X3
Y

	Pattern	X1	X2	X3	Y
1	−−0	-1	-1	0	•
2	−+0	-1	1	0	•
3	+−0	1	-1	0	•
4	++0	1	1	0	•
5	0−−	0	-1	-1	•
6	0−+	0	-1	1	•
7	0+−	0	1	-1	•
8	0++	0	1	1	•
9	−0−	-1	0	-1	•
10	+0−	1	0	-1	•
11	−0+	-1	0	1	•
12	+0+	1	0	1	•
13	000	0	0	0	•
14	000	0	0	0	•
15	000	0	0	0	•

Compare these three designs, the Box-Behnken has the minimum runs.