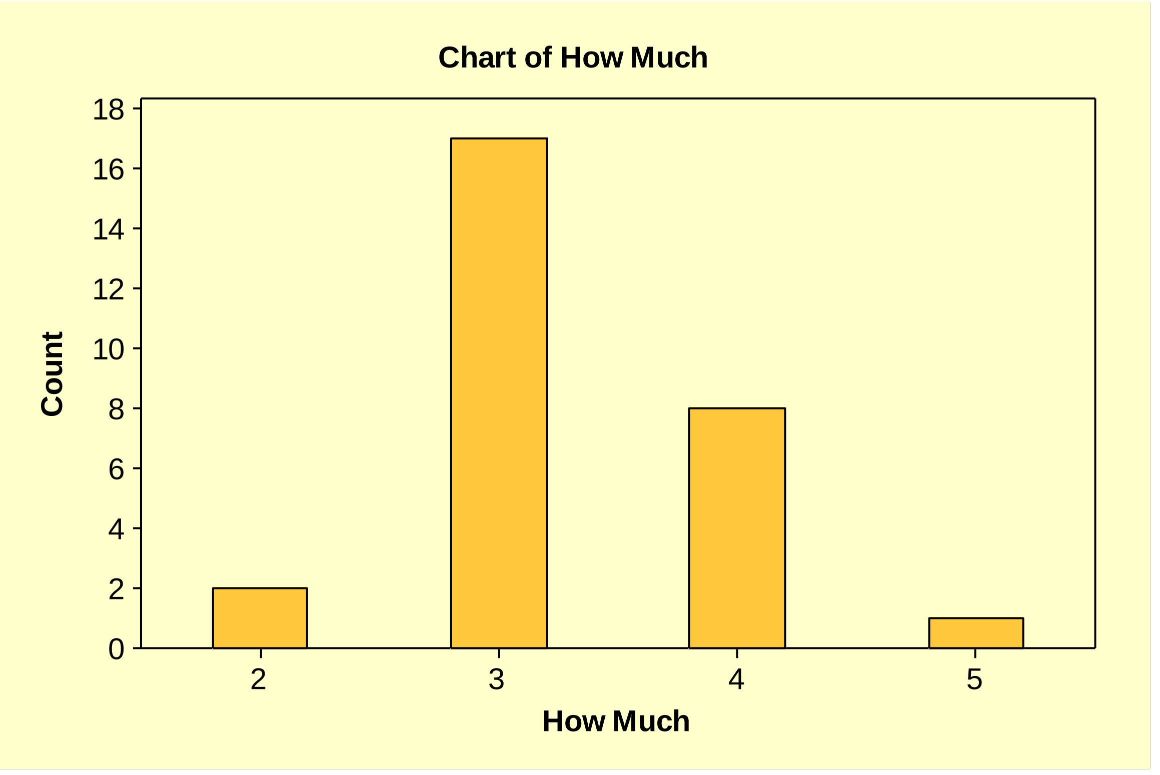
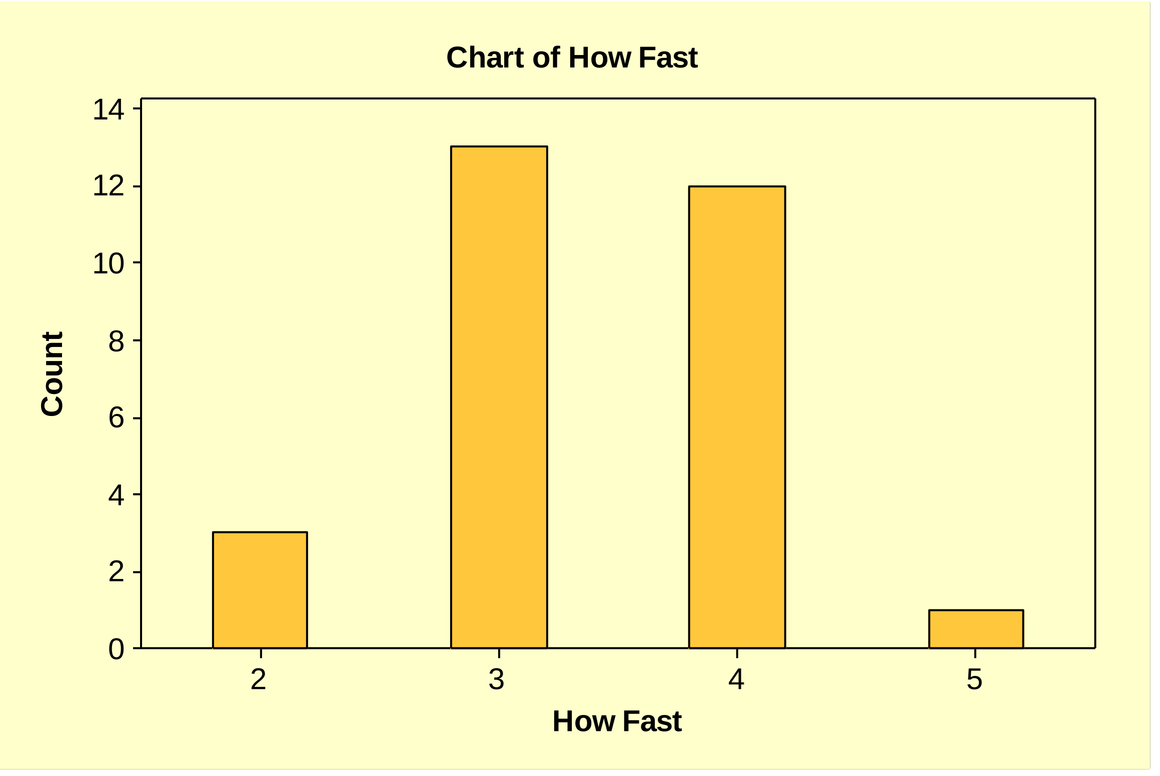
LNCT GROUP OF COLLEGES

Name of Faculty: Dr. Amit Kumar Nayak
Designation: Professor
Department: Pharmacy
Subject: Biostatistics and Research Methodology (BP 801T)
Unit: V
Topic: Design and Analysis of Experiments

Minute Test: How Much



Minute Test: How Fast



Homework 2.2.1

A 2^2 experiment

Project:

optimisation of a chemical process yield

Factors (with levels):

operating temperature (Low, High)

catalyst (C1, C2)

Design:

Process run at all four possible combinations of factor levels, in duplicate, in random order.

Results

Standard Order	Run Order	Temperature	Catalyst	Yield
1	6	Low	1	60
2	8	High	1	72
3	1	Low	2	52
4	4	High	2	83
5	3	Low	1	54
6	7	High	1	68
7	2	Low	2	45
8	5	High	2	80

Calculating s and df

Temperature	Catalyst	Rep1	Rep2	s	df
Low	1	60	54	4.2	1
High	1	72	68	2.8	1
Low	2	52	45	4.9	1
High	2	83	80	2.1	1
			RMS	3.7	4

Calculation of t-statistic

Results (Temperature order)

Standard Order	Run Order	Temperature	Catalyst	Yield
3	1	Low	2	52
7	2	Low	2	45
5	3	Low	1	54
1	6	Low	1	60
4	4	High	2	83
8	5	High	2	80
6	7	High	1	68
2	8	High	1	72

52.75 75.75 23 2.6 8.8

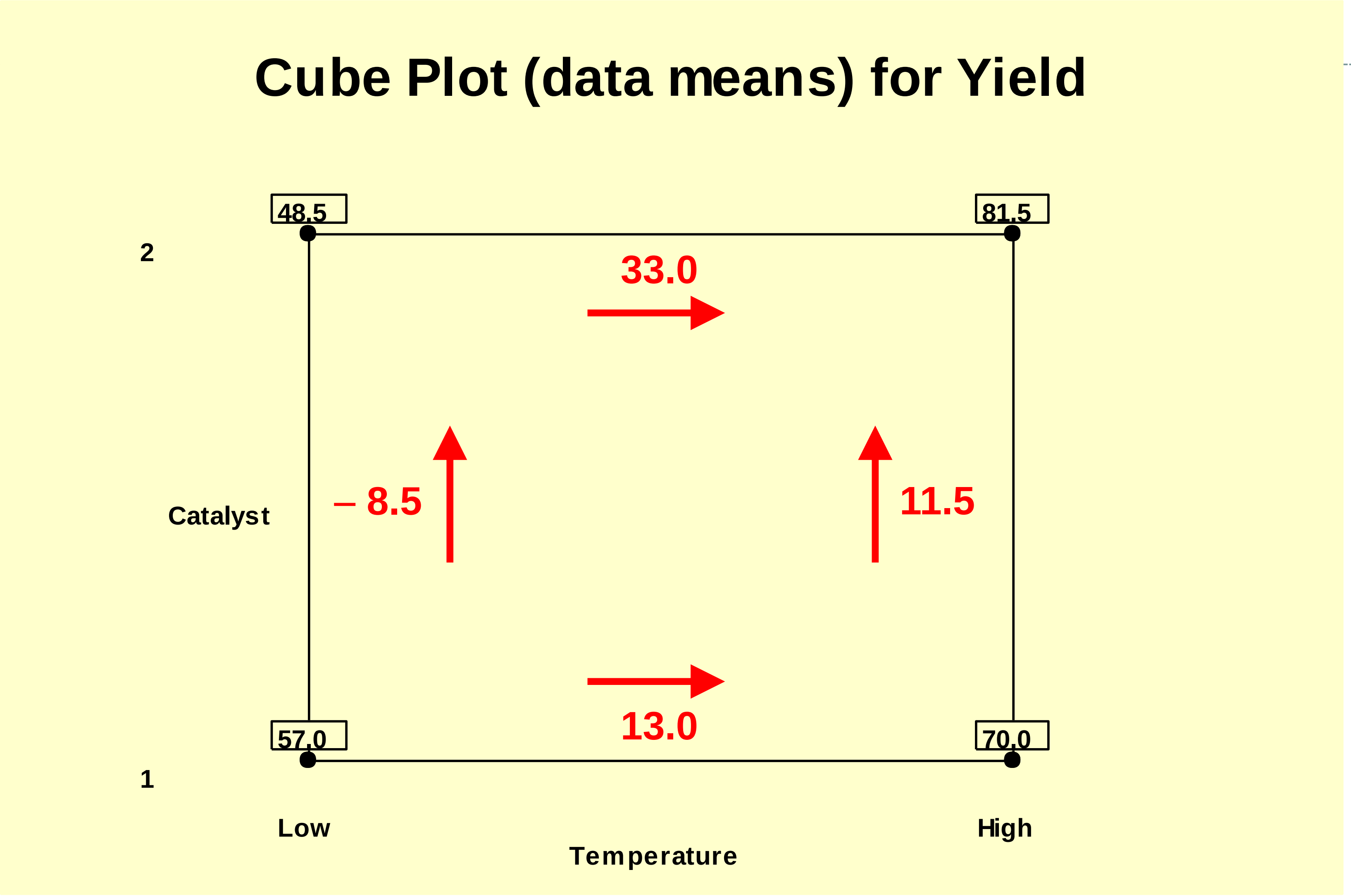
$\bar{Y}_{Low} =$ $\bar{Y}_{High} =$ $\bar{Y}_{High} - \bar{Y}_{Low} =$

$$SE(\bar{Y}_{High} - \bar{Y}_{Low}) = \sqrt{\frac{s^2}{4} + \frac{s^2}{4}} = \sqrt{2\frac{s^2}{4}} = \qquad t =$$

Homework 2.2.1

Test the statistical significance of and calculate confidence intervals for the Catalyst effect and the Temperature × Catalyst interaction.

Interaction illustrated



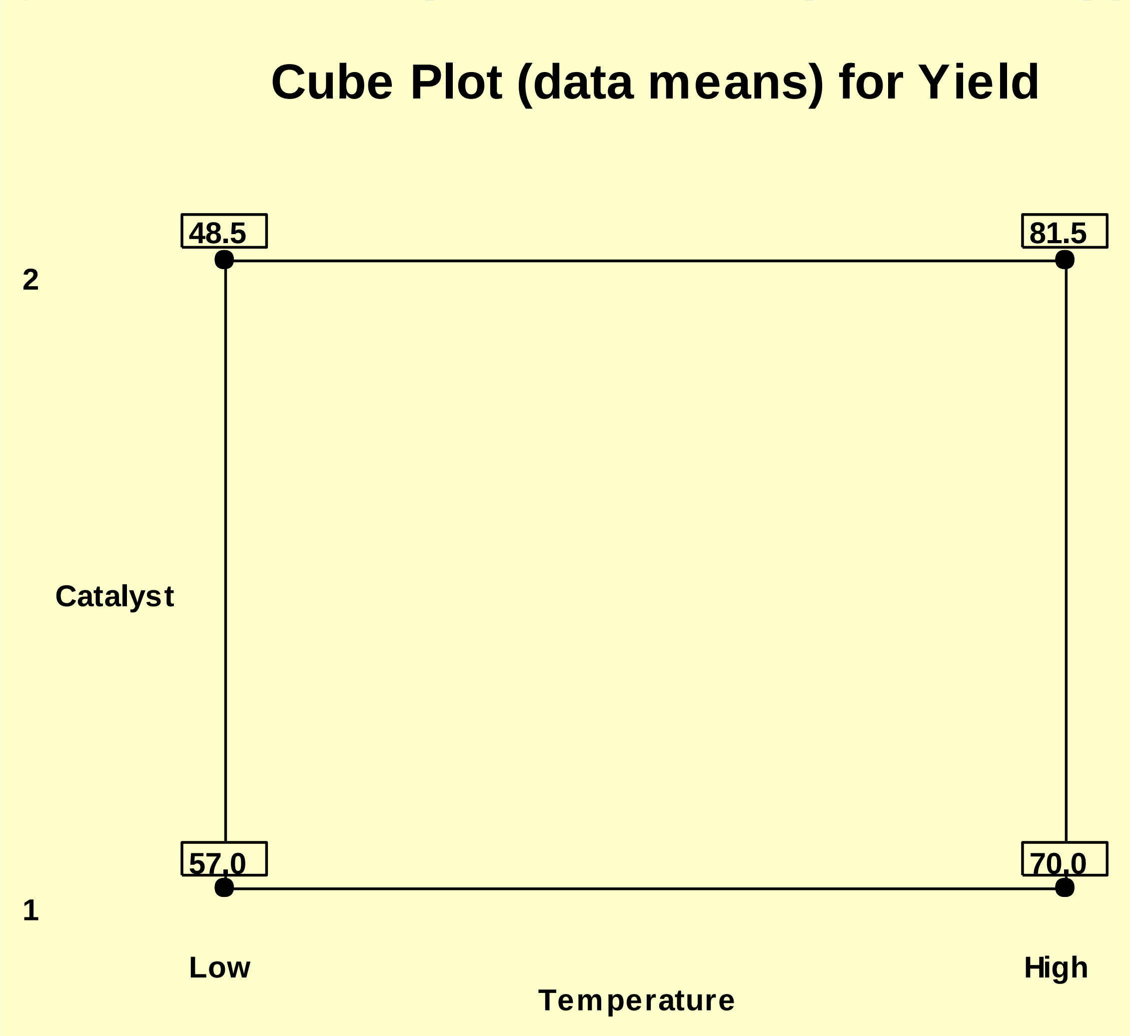
Application
Finding the optimum

More Minitab results

Least Squares Means for Yield

	Mean	SE Mean
Temperature		
Low	52.75	1.854
High	75.75	1.854
Catalyst		
1	63.50	1.854
2	65.00	1.854
Temperature*Catalyst		
Low 1	57.00	2.622
High 1	70.00	2.622
Low 2	48.50	2.622
High 2	81.50	2.622

Optimum operating conditions



Highest yield achieved

with Catalyst 2

at High temperature.

Estimated yield: 81.5%

95% confidence interval:

$81.5 \pm 2.78 \times 2.622$,

i.e., 81.5 ± 7.3,

i.e., (74.2 , 88.8)

Part 2 Wine tasting measurement

As part of a project to develop a GC method for analysing trace compounds in wine without the need for prior extraction of the compounds, a synthetic mixture of aroma compounds in ethanol-water was prepared.

The effects of two factors, Injection volume and Solvent flow rate, on GC measured peak areas given by the mixture were assessed using a 2^2 factorial design with 3 replicate measurements at each design point. The results are shown in the table that follows.

What conclusions can be drawn from these data? Display results numerically and graphically. Check model assumptions by using appropriate residual plots.

Peak areas for GC study

Solvent flow rate, mL/min	Injection volume, μL	
	100	200
400	13.1	126.5
	15.3	118.5
	17.7	122.1
200	48.8	134.5
	42.1	135.4
	39.2	128.6

Organising the data for analysis

Design Point	Volume	Flow Rate	Peak Area		
1	100	400	13.1	15.3	17.7
2	200	400	126.5	118.5	122.1
3	100	200	48.8	42.1	39.2
4	200	200	134.5	135.4	128.6

Organising the data for analysis

Design Point	Volume	Flow Rate	Peak Area			Mean	SD
1	100	400	13.1	15.3	17.7	15.37	2.30
2	200	400	126.5	118.5	122.1	122.37	4.01
3	100	200	48.8	42.1	39.2	43.37	4.92
4	200	200	134.5	135.4	128.6	132.83	3.69

$s^2 = \text{average}(SD^2)$

$= (2.30^2 + 4.01^2 + 4.92^2 + 3.69^2) / 4$

$= 14.798$

$s = 3.85$

$df(s) = \text{sum}[df(SD)]$

$= 2 + 2 + 2 + 2$

$= 8$

Organising the data for calculation

Design Point	Volume	Flow Rate	Peak Area			Mean	SD
1	100	400	13.1	15.3	17.7	15.37	2.30
2	200	400	126.5	118.5	122.1	122.37	4.01
3	100	200	48.8	42.1	39.2	43.37	4.92
4	200	200	134.5	135.4.	128.6	132.83	3.69

Generic notation

Design Point	A	B	Peak Area			Mean	SD
1	–	–	13.1	15.3	17.7	15.37	2.30
2	+	–	126.5	118.5	122.1	122.37	4.01
3	–	+	48.8	42.1	39.2	43.37	4.92
4	+	+	134.5	135.4	128.6	132.83	3.69

The design matrix

Design Point	Factor A	Factor B
1	–	–
2	+	–
3	–	+
4	+	+

- Prior to the experiment, the *rows* designate the *design points*, the sets of conditions under which the process is to be run.
- After the experiment, the *columns* designate the *contrasts*, the combinations of design point means which measure the main effects of the factors.

Calculating interaction effects; the *extended* design matrix

The extended design matrix

Design Point	A	B	AB	Mean
1	–	–	+	$\bar{Y}_1$ = 15.37
2	+	–	–	$\bar{Y}_2$ = 122.37
3	–	+	–	$\bar{Y}_3$ = 43.37
4	+	+	+	$\bar{Y}_4$ = 132.83

AB Interaction = ½(A effect at high B − A effect at low B)

$$= \tfrac{1}{2}[(\bar{Y}_4 - \bar{Y}_3) - (\bar{Y}_2 - \bar{Y}_1)]$$

$$= \tfrac{1}{2}(\bar{Y}_1 - \bar{Y}_2 - \bar{Y}_3 + \bar{Y}_4)$$

Check: AB = A × B

Part 3 3 factors each at 2 levels, a 2^3 experiment

An experiment to investigate the effects on yield of a chemical process of changes to operating Temperature, raw material Concentration and type of Catalyst was conducted in a pilot plant set up for experimentation. Details were as follows.

Factor settings and codes

Temperature, T (ºC)		Concentration, C (%)		Catalyst, K	
160	180	20	40	A	B
–	+	–	+	–	+

Design

Standard Order	Temperature T(°C)	Concentration C(%)	Catalyst K
1	160	20	A
2	160	20	A
3	180	20	A
4	180	20	A
5	160	40	A
6	160	40	A
7	180	40	A
8	180	40	A
9	160	20	B
10	160	20	B
11	180	20	B
12	180	20	B
13	160	40	B
14	160	40	B
15	180	40	B
16	180	40	B

Design

Standard Order	Temperature T(°C)	Concentration C(%)	Catalyst K	Run Order
1	160	20	A	6
2	160	20	A	13
3	180	20	A	2
4	180	20	A	4
5	160	40	A	1
6	160	40	A	16
7	180	40	A	5
8	180	40	A	10
9	160	20	B	8
10	160	20	B	12
11	180	20	B	9
12	180	20	B	14
13	160	40	B	3
14	160	40	B	11
15	180	40	B	7
16	180	40	B	15

Design

Standard Order	Temperature T(°C)	Concentration C(%)	Catalyst K	Run Order
5	160	40	A	1
3	180	20	A	2
13	160	40	B	3
4	180	20	A	4
7	180	40	A	5
1	160	20	A	6
15	180	40	B	7
9	160	20	B	8
11	180	20	B	9
8	180	40	A	10
14	160	40	B	11
10	160	20	B	12
2	160	20	A	13
12	180	20	B	14
16	180	40	B	15
6	160	40	A	16

Results

Standard Order	Temperature T(°C)	Concentration C(%)	Catalyst K	Run Order	Yield
5	160	40	A	1	50
3	180	20	A	2	74
13	160	40	B	3	46
4	180	20	A	4	70
7	180	40	A	5	69
1	160	20	A	6	59
15	180	40	B	7	79
9	160	20	B	8	50
11	180	20	B	9	81
8	180	40	A	10	67
14	160	40	B	11	44
10	160	20	B	12	54
2	160	20	A	13	61
12	180	20	B	14	85
16	180	40	B	15	81
6	160	40	A	16	58

Results, in standard order

T	C	K	Yield		Mean	SD
–	–	–	59	61	60	1.41
+	–	–	74	70	72	2.83
–	+	–	50	58	54	5.66
+	+	–	69	67	68	1.41
–	–	+	50	54	52	2.83
+	–	+	81	85	83	2.83
–	+	+	46	44	45	1.41
+	+	+	79	81	80	1.41

Calculating effects, the extended design matrix

Design Point	T	C	K	TC	TK	CK	TCK	Mean
1	–	–	–	+	+	+	–	60
2	+	–	–	–	–	+	+	72
3	–	+	–	–	+	–	+	54
4	+	+	–	+	–	–	–	68
5	–	–	+	+	–	–	+	52
6	+	–	+	–	+	–	–	83
7	–	+	+	–	–	+	–	45
8	+	+	+	+	+	+	+	80

3-factor interaction measures

the change in any 2fi when the third factor changes.

e.g., the change in T×C between low and high K.

Report results separately at all 2^3 combinations

Calculating s

T	C	K	Yield		Mean	SD	Variance =½(diff)^2
–	–	–	59	61	60	1.41	2
+	–	–	74	70	72	2.83	8
–	+	–	50	58	54	5.66	32
+	+	–	69	67	68	1.41	2
–	–	+	50	54	52	2.83	8
+	–	+	81	85	83	2.83	8
–	+	+	46	44	45	1.41	2
+	+	+	79	81	80	1.41	2

Total	64
s^2	8
s	2.83

Exercise 3.1.1

Calculate the t-ratio for the T effect, the TC 2-factor interaction and the TCK 3-factor interaction.

What conclusions do you draw?

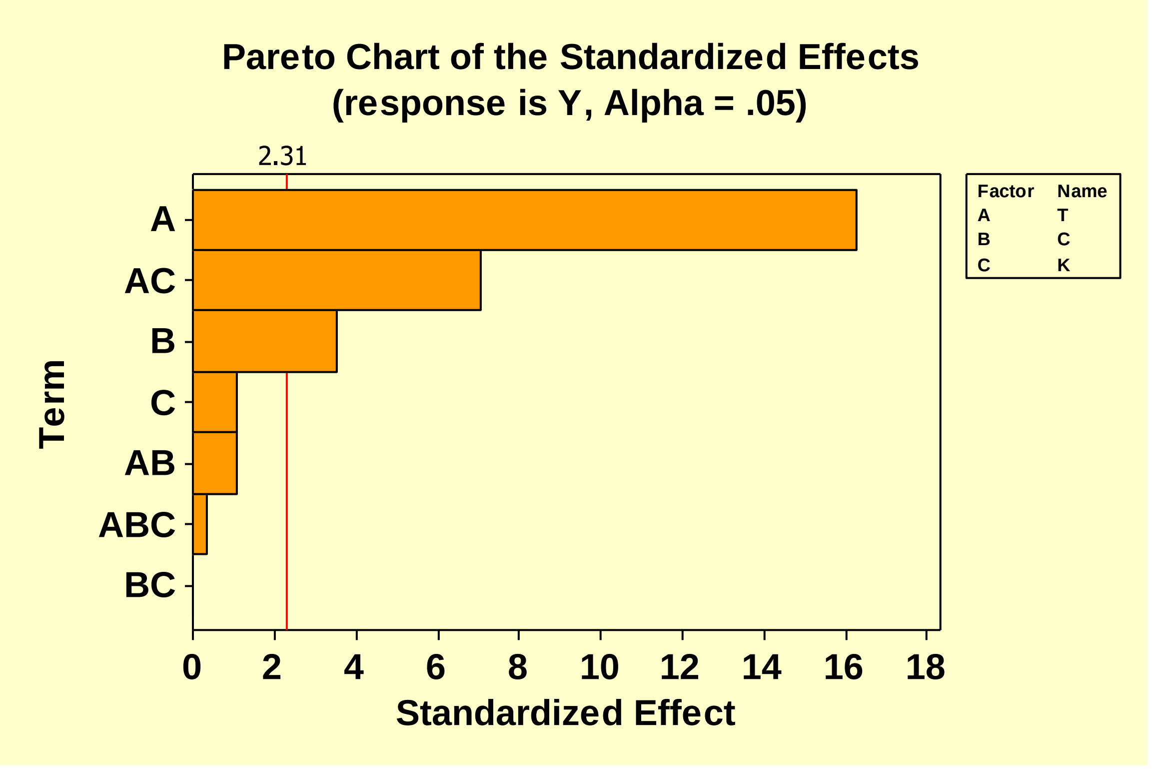
Minitab analysis

Estimated Effects for Yield

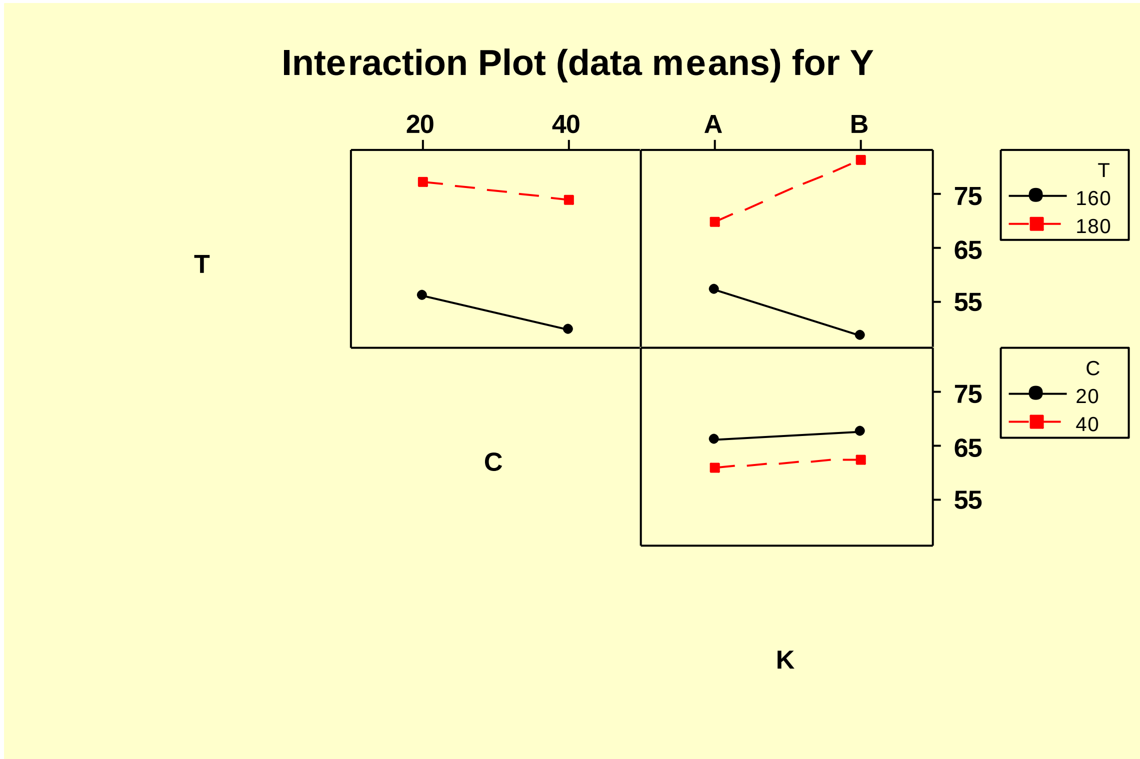
Term	Effect	SE	T	P
T	23.0	1.414	16.26	0.000
C	-5.0	1.414	-3.54	0.008
K	1.5	1.414	1.06	0.320
T*C	1.5	1.414	1.06	0.320
T*K	10.0	1.414	7.07	

Minitab analysis

Lecture 2.2

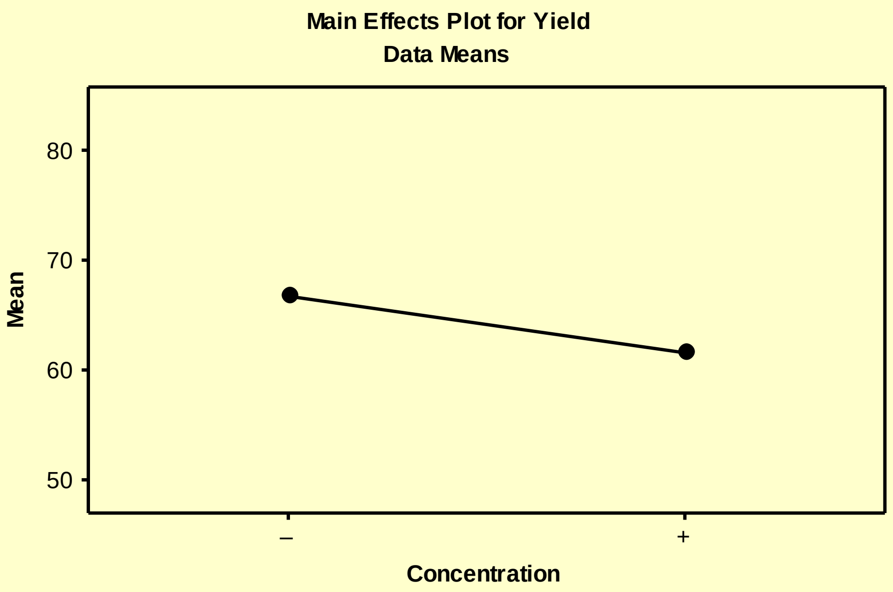


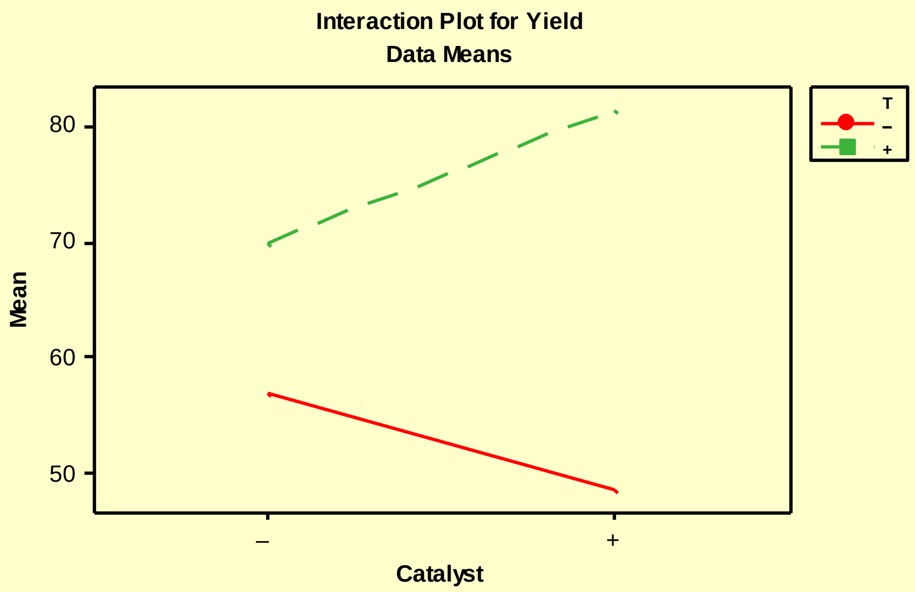
Minitab analysis



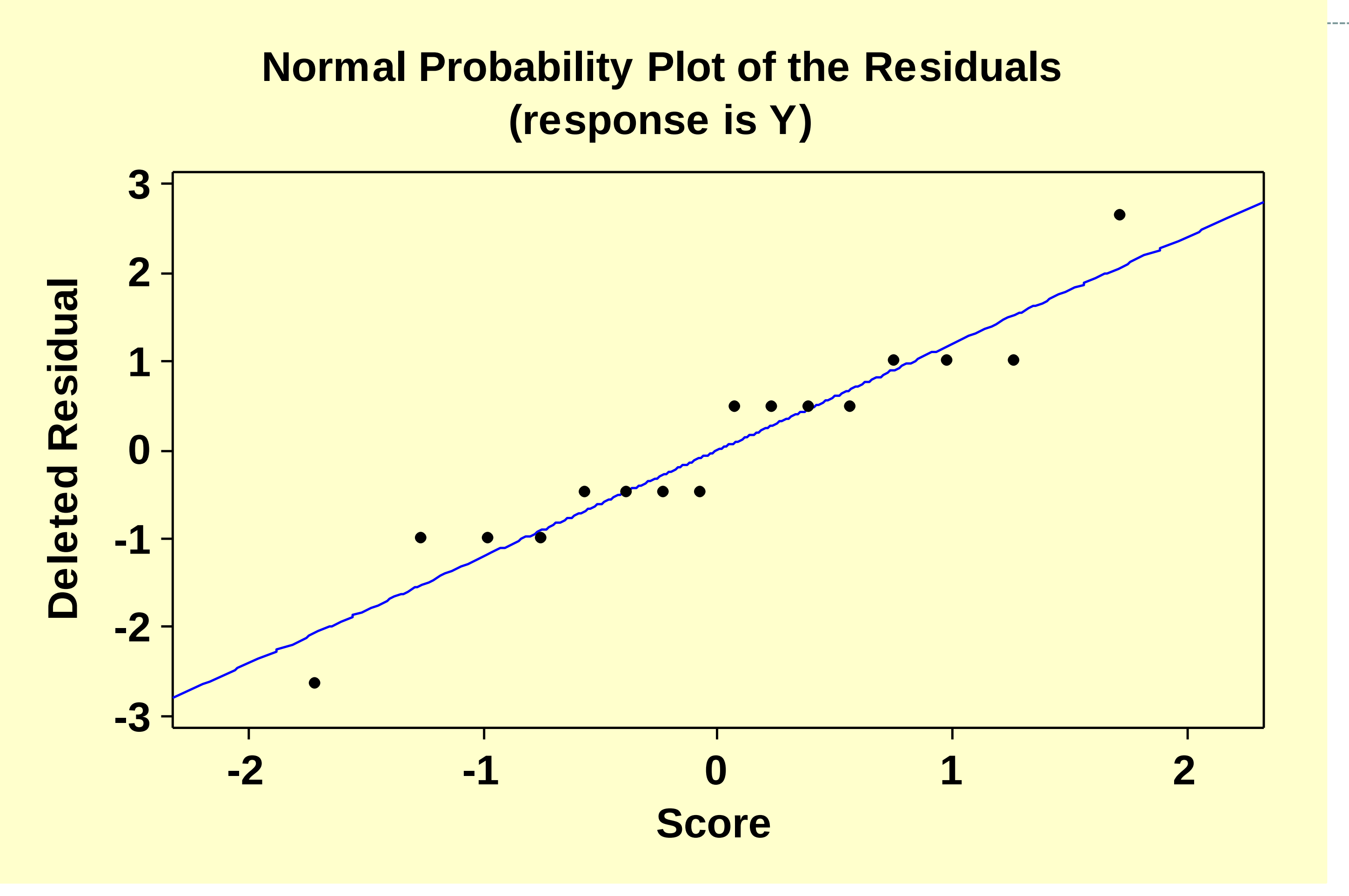
Minitab analysis

Lecture 2.2

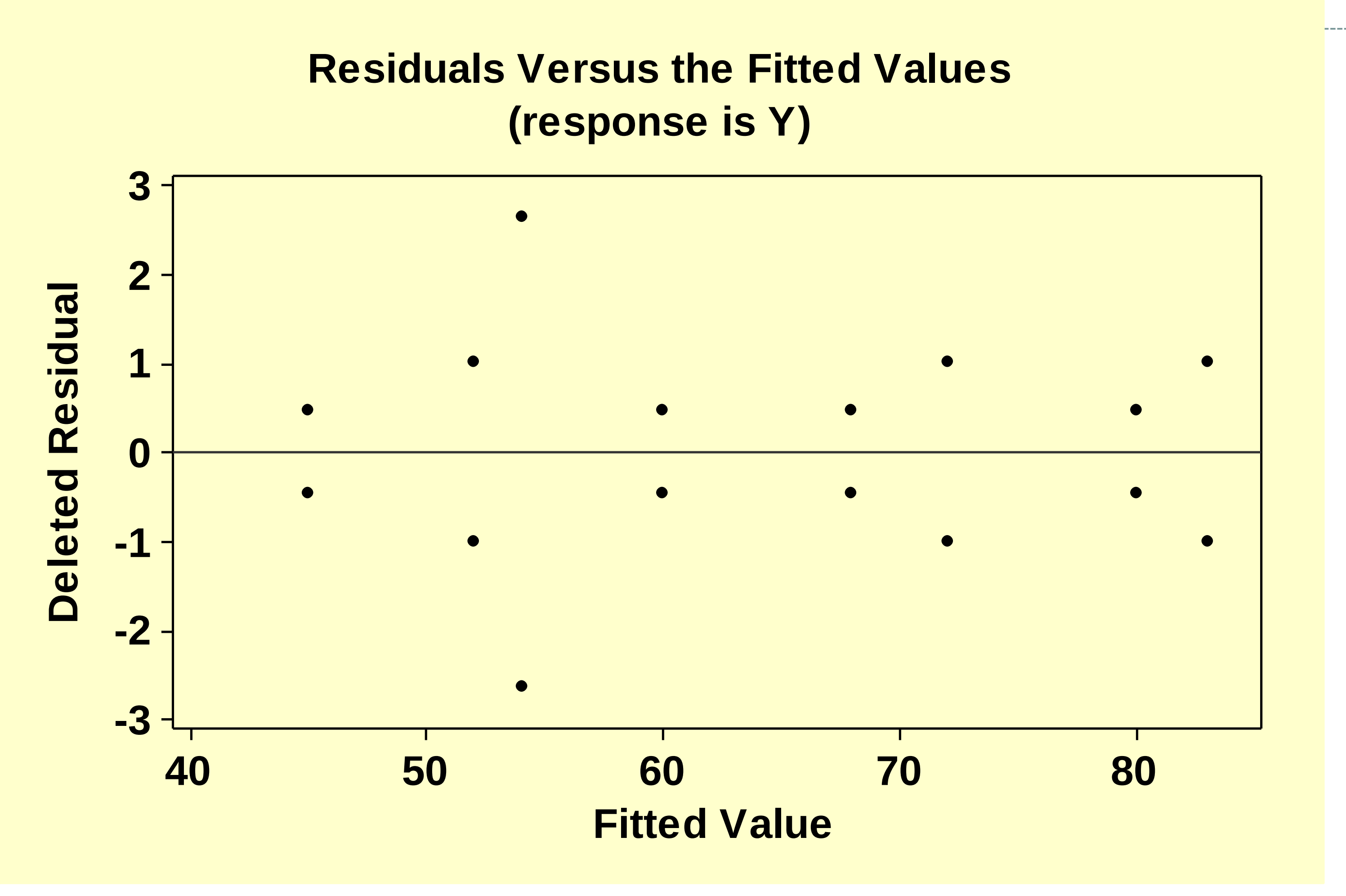




Minitab diagnostic analysis



Minitab diagnostic analysis



Homework 3.1.1

An experiment was run to assess the effects of three factors on the life of a cutting tool

A: Cutting speed

B: Tool geometry

C: Cutting angle.

The full 2^3 design was replicated three times. The results are shown in the next slide and are available in Excel file Tool Life.xls.

Carry out a full analysis and report.

Homework

Cutting Speed	Tool Geometry	Cutting Angle	Tool Life		
-	-	-	22	31	25
+	-	-	32	43	29
-	+	-	35	34	50
+	+	-	55	47	46
-	-	+	44	45	38
+	-	+	40	37	36
-	+	+	60	50	54
+	+	+	39	41	47

Part 4 2^4 in 16 runs, no replicates

A process development study
with four factors each at two levels

	Low (–)	High (+)
A: Catalyst Charge (lbs)	10	15
B: Temperature (°C)	220	240
C: Concentration (%)	10	12
D: Pressure (bar)	50	80

2^4 in 16 runs, no replicates

Process yields resulting from varying levels of four two-level factors arranged in a 2^4 design run in completely random order

Design Point	*Catalyst Charge*	*Temperature*	*Concentration*	*Pressure*	*Y*	*RunOrder*
1	10	220	10	50	70	8
2	15	220	10	50	60	2
3	10	240	10	50	89	10
4	15	240	10	50	81	4
5	10	220	12	50	60	16
6	15	220	12	50	49	5
7	10	240	12	50	88	11
8	15	240	12	50	82	14
9	10	220	10	80	69	15
10	15	220	10	80	62	9
11	10	240	10	80	88	1
12	15	240	10	80	81	13
13	10	220	12	80	60	3
14	15	220	12	80	52	12
15	10	240	12	80	86	6
16	15	240	12	80	79	7

No replication: alternative analyses

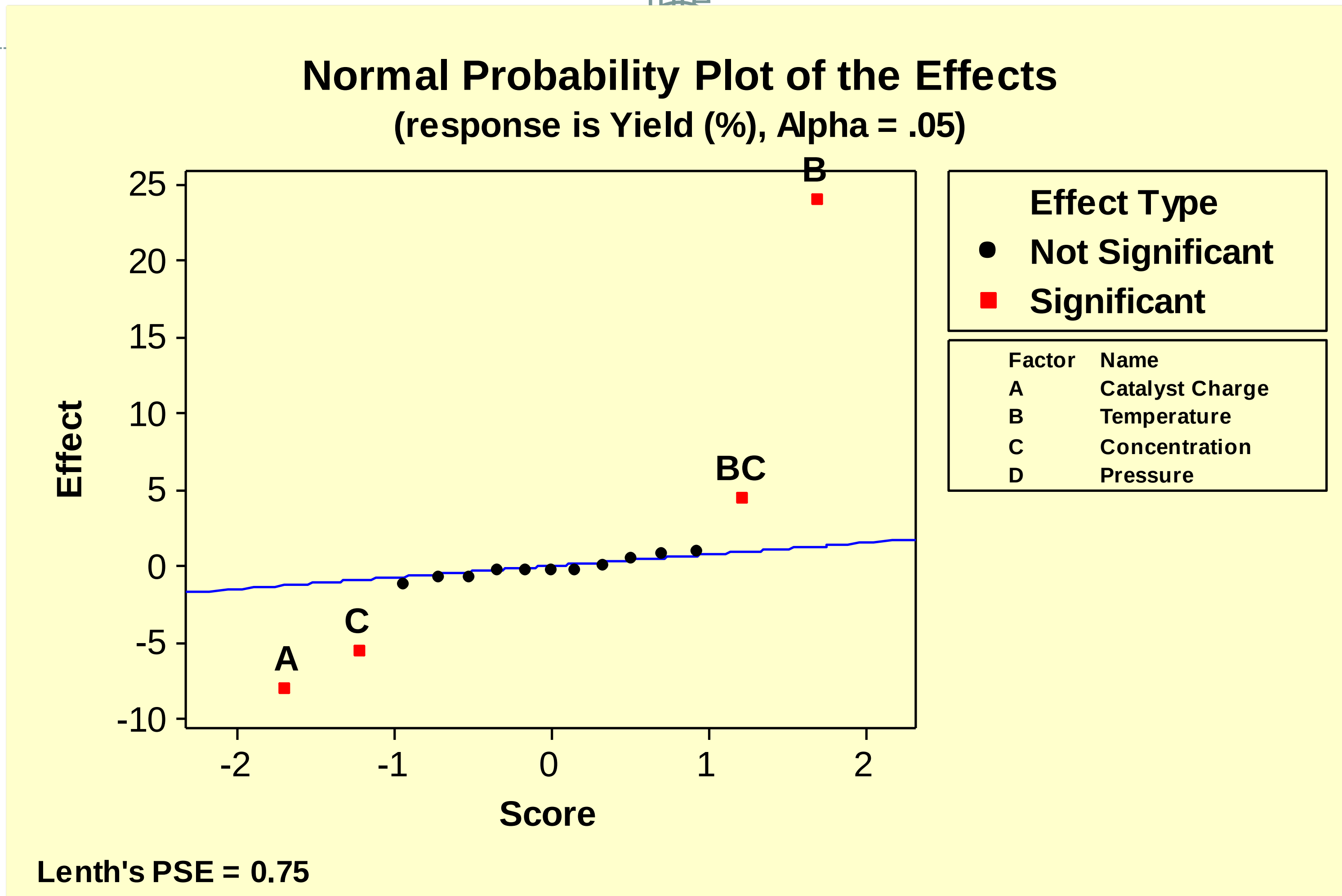
- Normal plots of effects
  - if no effects present, estimated effects reflect chance variation, follow Normal model
  - a few real effects will appear as exceptions in a Normal plot
- Lenth method
  - alternative estimate of s, given a few real effects
- Best approach: combine both!

No replication: alternative analyses

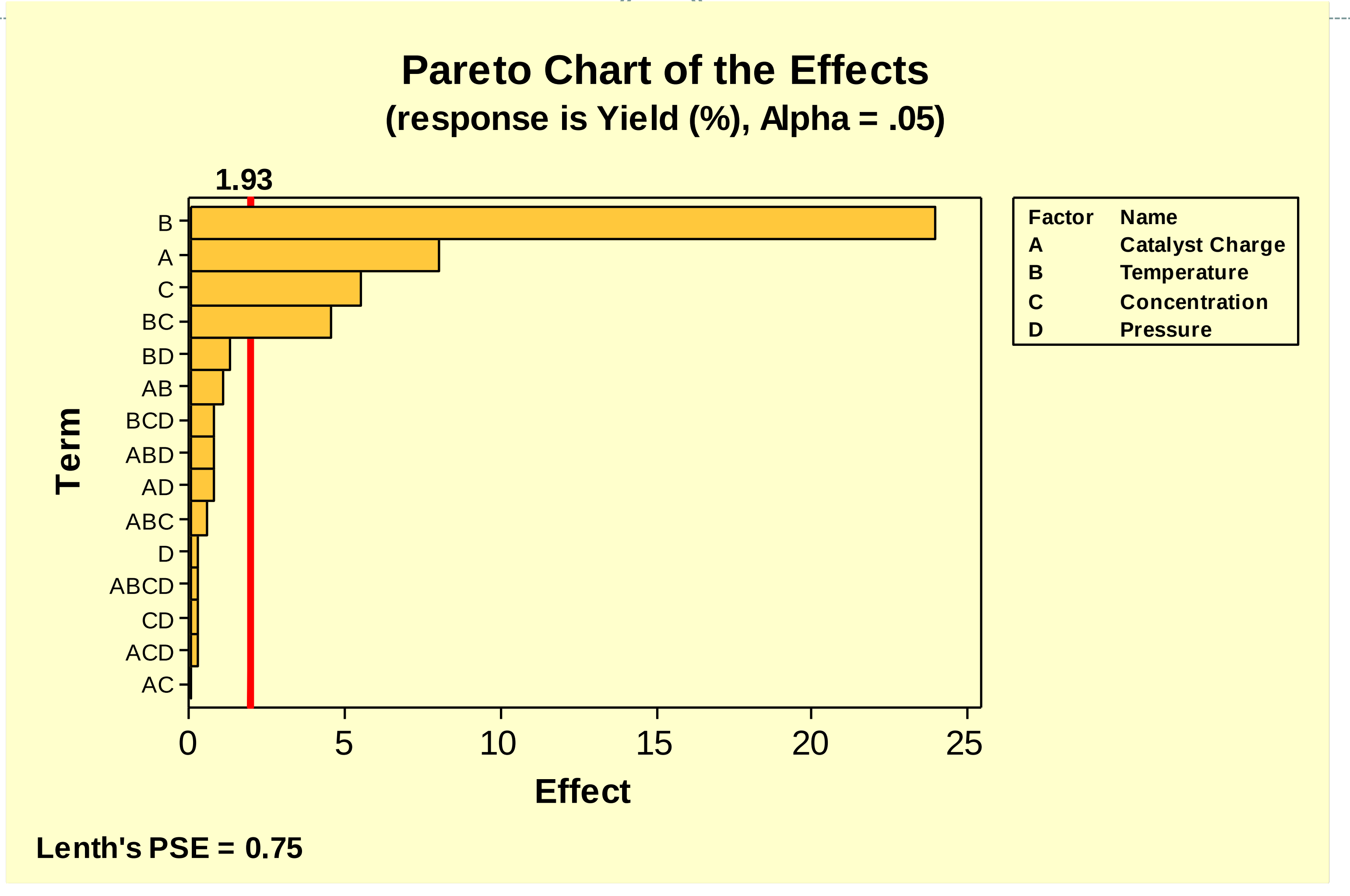
Estimated Effects for Yield (%) (use design matrix columns)

Term	Effect
Catalyst Charge	-8.000
Temperature	24.000
Concentration	-5.500
Pressure	-0.250
Catalyst Charge*Temperature	1.000
Catalyst Charge*Concentration	-0.000
Catalyst Charge*Pressure	0.750
Temperature*Concentration	4.500
Temperature*Pressure	-1.250
Concentration*Pressure	-0.250

Normal Effects Plot



Pareto Chart, vital few versus trivial many (Juran)



Lenth's method

Given several Normal values with mean 0

and given their absolute values (magnitudes, or values without signs), then it may be shown that

SD(Normal values) ≈ 1.5 × median(Absolute values).

Given a small number of effects with mean ≠ 0, then

SD(Normal values) is a small bit bigger.

Refinement: PSE ≈ 1.5 × median(Absolute values < 2.5)

Lenth's method illustrated

Example

Values	-41	14	-23	-1	-38	-5	-27	-34	-9	-32	29	-18	-48	-25	-37
Magnitudes	41	14	23	1	38	5	27	34	9	32	29	18	48	25	37
Sorted	1	5	9	14	18	23	25	27	29	32	34	37	38	41	48

Add 50 to 3 values, to represent 3 active effects; median will be 27, 29, 32 or 34; not much bigger, so s will be not much bigger,

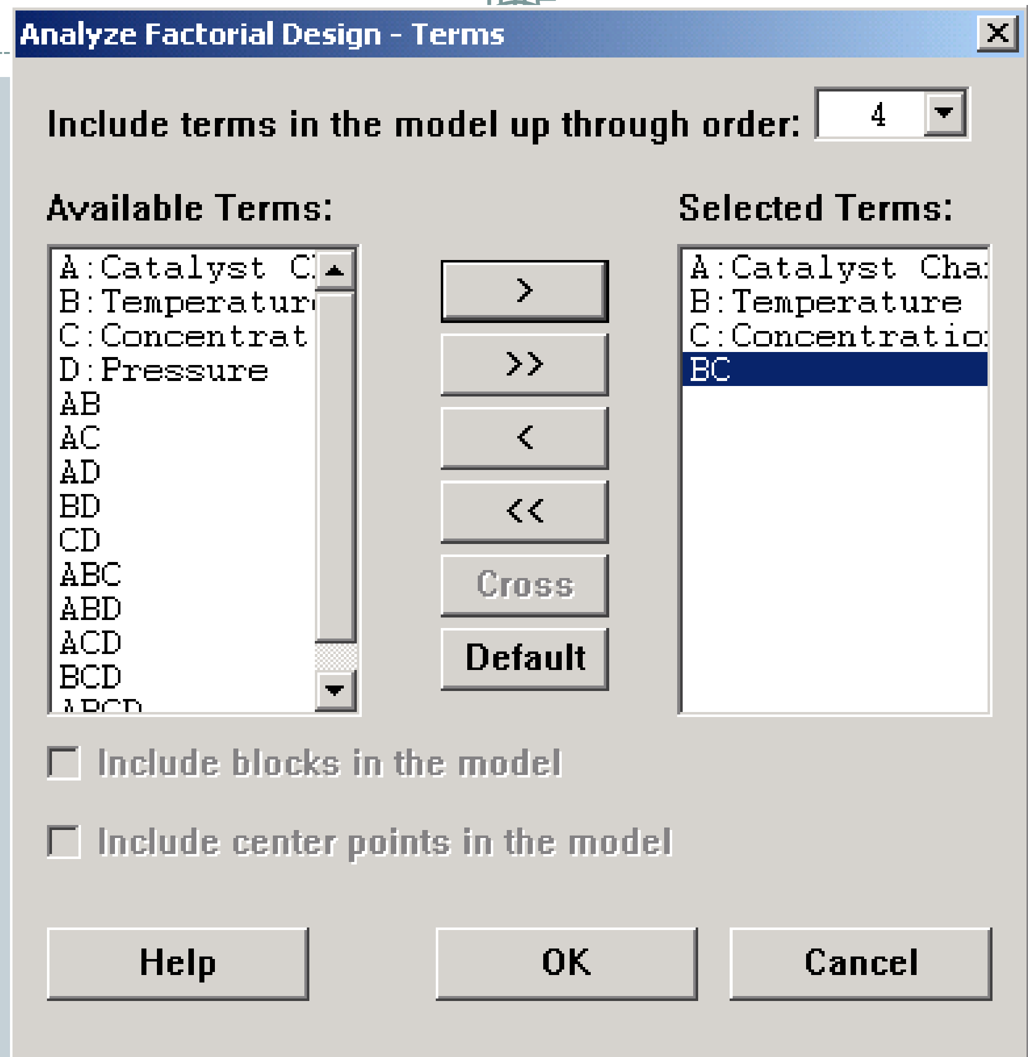
- provides a suitable basis for a "t"-test.

See Laboratory 1 for simulation

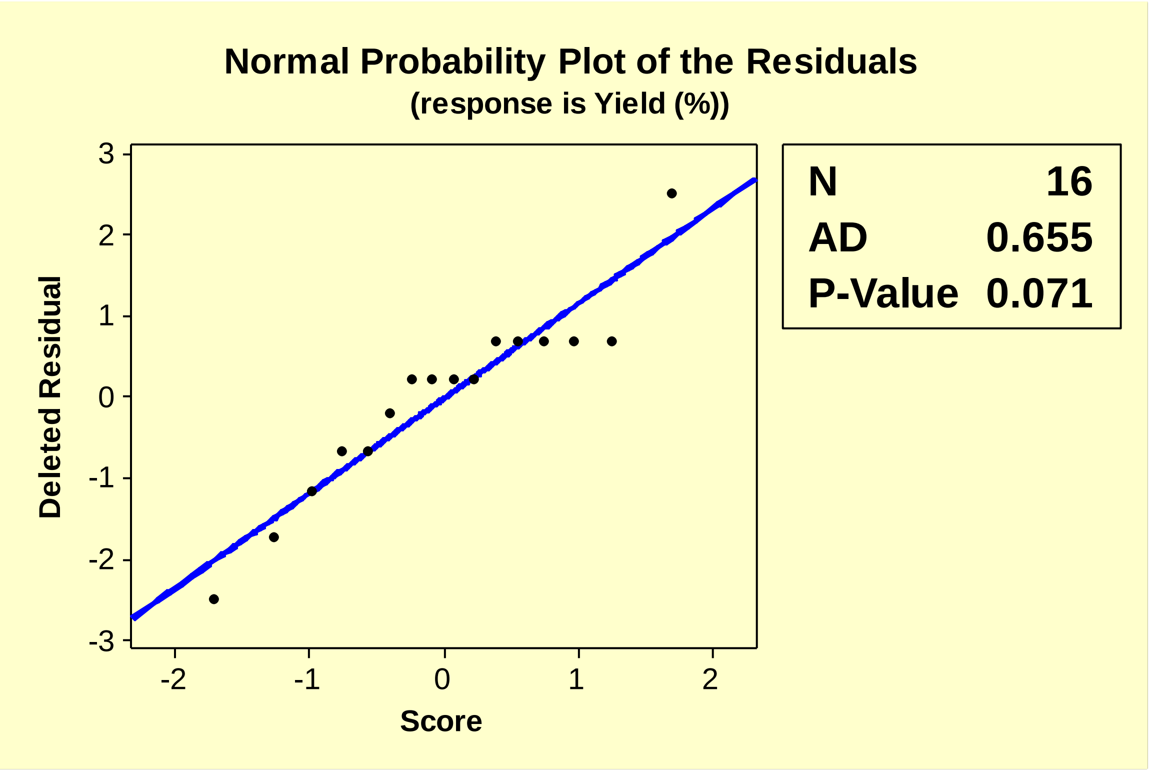
Reduced Model method

- Select identified terms for a fitted model
  - omitted terms provide basis for estimating σ
- Check diagnostics
- Estimate effects
  - ANOVA used to estimate s
- Identify optimal operating conditions

Reduced model

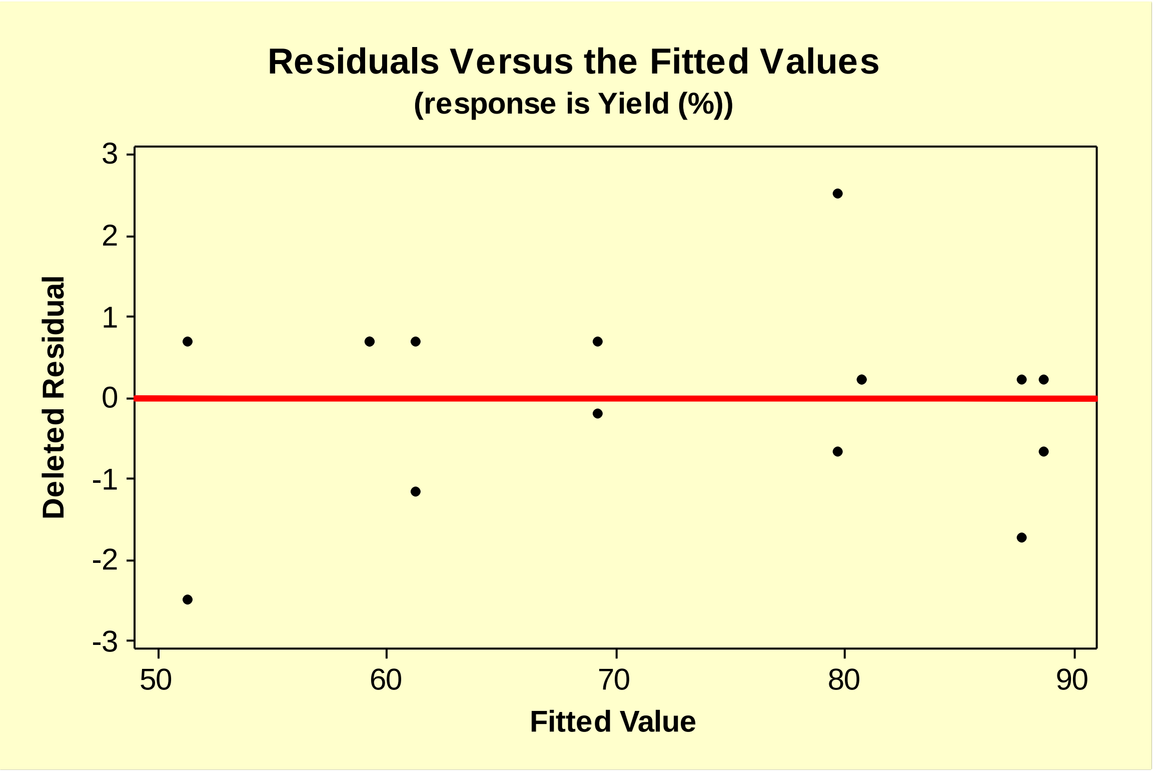


Diagnostics



Diagnostics

Lecture 2.2



Estimated effects

Estimated Effects for Yield (%)

Term	Effect	SE	T	P
Catalyst Charge	-8.000	0.657	-12.17	0.000
Temperature	24.000	0.657	36.52	0.000
Concentration	-5.500	0.657	-8.37	0.000
Temperature*Concentration	4.500	0.657	6.85	0.000

S = 1.31426

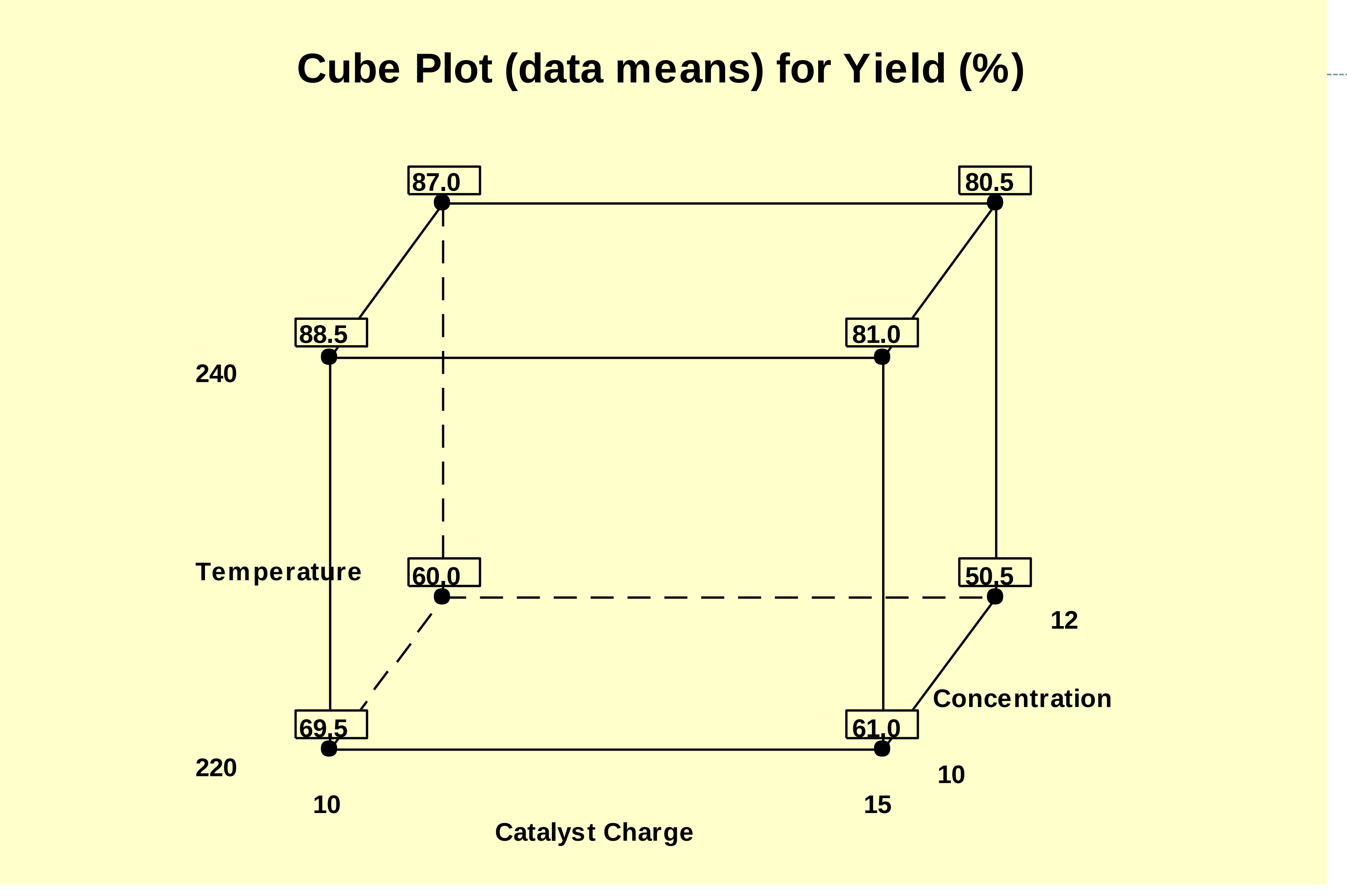
Analysis of Variance (basis for s)

Analysis of Variance for Yield (%)

Source	DF	SS	MS	F	P
Main Effects	3	2681.00	893.667	517.39	0.000
2-Way Interactions	1	81.00	81.000	46.89	0.000
Residual Error	11	19.00	1.727		
Total	15	2781.00			

s = sqrt(1.727) = 1.31

Identify optimal operating conditions



Identify optimal operating conditions

Confirm the calculation of the confidence interval for optimum yield.

$CI = 88.5 \pm 2.2 \times 0.93 = (86.45, 90.55)$

Exercise

Calculate a confidence interval for the 'next best' yield.

Homework

Test the statistical significance of the difference between best and next best yields.

Homework

Design Projection

Since Pressure is not statistically significant, it may be treated as an "inert" factor and the design may be treated as a 2^3 with duplicate observations.

Analyze these data accordingly.

Compare results with the Lenth method and the Reduced Model method.

5 Introduction to Fractional Factorial Designs

Design Point	A	B	C	D= ABC	Y
1	–	–	–	–	Y_1
2	+	–	–	+	Y_2
3	–	+	–	+	Y_3
4	+	+	–	–	Y_4
5	–	–	+	+	Y_5
6	+	–	+	–	Y_6
7	–	+	+	–	Y_7
8	+	+	+	+	Y_8

Fourth column estimates D main effect.

Fourth column also estimates ABC interaction effect in 2^3.

In fact, fourth column estimates D + ABC in 2^{4-1}.

Fractional factorial designs

Design Point	A= BCD	B= ACD	C= ABD	D= ABC	Y
1	–	–	–	–	Y_1
2	+	–	–	+	Y_2
3	–	+	–	+	Y_3
4	+	+	–	–	Y_4
5	–	–	+	+	Y_5
6	+	–	+	–	Y_6
7	–	+	+	–	Y_7
8	+	+	+	+	Y_8

Confirm "confounding" patterns shown.

Also, confirm AB = CD.

What other effects are confounded?

Fractional factorial designs

First half fraction

Design Point	A	B	C	D	Y
1	–	–	–	–	70
2	+	–	–	+	62
3	–	+	–	+	88
4	+	+	–	–	81
5	–	–	+	+	60
6	+	–	+	–	49
7	–	+	+	–	88
8	+	+	+	+	79

Identify corresponding design points

Full factorial design

Design Point	A	B	C	D	Y
1	–	–	–	–	70
2	+	–	–	–	60
3	–	+	–	–	89
4	+	+	–	–	81
5	–	–	+	–	60
6	+	–	+	–	49
7	–	+	+	–	88
8	+	+	+	–	82
9	–	–	–	+	69
10	+	–	–	+	62
11	–	+	–	+	88
12	+	+	–	+	81
13	–	–	+	+	60
14	+	–	+	+	52
15	–	+	+	+	86
16	+	+	+	+	79

Fractional factorial designs

First half fraction

Design Point	A	B	C	D	Y
1	–	–	–	–	70
10	+	–	–	+	62
11	–	+	–	+	88
4	+	+	–	–	81
13	–	–	+	+	60
6	+	–	+	–	49
7	–	+	+	–	88
16	+	+	+	+	79

Column A estimates A + BCD

Second half fraction

Design Point	A	B	C	D	Y
9	–	–	–	+	69
1	+	–	–	–	60
2	–	+	–	–	89
12	+	+	–	+	81
5	–	–	+	–	60
14	+	–	+	+	52
15	–	+	+	+	86
8	+	+	+	–	82

Column A estimates A – BCD

Full 2^4 design: Column A estimates ½[(A + BCD) + (A – BCD)] = A

6 "Coefficients" in Minitab output


```
Estimated Effects and Coefficients for Yield
  (coded units)

Term        Effect     Coef  SE Coef      T      P
Constant             64.250   0.7071  90.86  0.000
T           23.000   11.500   0.7071  16.26  0.000
C           -5.000   -2.500   0.7071  -3.54  0.008
K            1.500    0.750   0.7071   1.06  0.320
T*C          1.500    0.750   0.7071   1.06  0.320
T*K         10.000    5.000   0.7071   7.07  0.000
C*K          0.000    0.000   0.7071   0.00  1.000
T*C*K        0.500    0.250   0.7071   0.35  0.733

S = 2.82843
```

Recall Slide 28:

Estimated Effects for Yield

Term	Effect	SE	T	P
T	23.0	1.414	16.26	0.000
C	-5.0	1.414	-3.54	0.008
K	1.5	1.414	1.06	0.320
T*C	1.5	1.414	1.06	0.320
T*K	10.0	1.414	7.07	

Minitab Regression output


```
Regression Analysis: Yield versus T_160, C_20,
  K_A

The regression equation is
Yield = 64.3 - 11.5 T_160 + 2.50 C_20 - 0.75 K_A

Predictor       Coef   SE Coef       T      P
Constant      64.250     1.571   40.89  0.000
T_160        -11.500     1.571   -7.32  0.000
C_20           2.500     1.571    1.59  0.138
K_A           -0.750     1.571   -0.48  0.642

S = 6.28490

NB:  Values for T_160, C_20, K_A are +1 and -1
```

Reading

EM	§5.3, §5.4, §5.6, §5.7.1
DCM	§6-2, §6-3 to p.218, §6.5 to p. 235, §8.1, §8.2.1
BHH	§ 5.14 (Lenth plots) *and all of Ch. 5!*