DOE –Design of Experiments

Yossi Levy

Why use DOE?

- Decisions should be based on information
- Data is not information
- We need collect data that can be transformed into information
- We also need to know how to do the transformation
- DOE has the two capabilities:
  - Decide what data to collect
  - Analyze the data to extract useful information

DOE and QbD

- Design Of Experiments is an essential component in QbD.
- It leads to detailed verification of how product and process definition affect key quality characteristics.
- It establishes a basis for defining the design space.
- Experiments enable us to map the process.
- A reliable map lets us drive safely.
- A reliable process map permits changes without adverse effects on product performance.

Basic principles

- Randomization
- Blocking
- Replication
- Factorial Structure

Randomization

- Make arbitrary decisions randomly.
  - Who gets each treatment?
  - Where are materials placed?
  - What gets done first?

Randomization helps to guarantee that the experiment is fair; it is not biased in favor of one of the treatment

Blocking

- Often units can be grouped by factors that are not of direct interest, but will affect results.
  - Source of raw materials.
  - Day/time of preparation or measurement.
  - Location in shaker or plate

Replication

- We need to know something about the variability of outcomes when conditions are *not* changed.
- That gives a basis to know if differences between conditions are “just noise”.

Factorial Structure

- Most processes are affected by many factors
- An experiment can:
  - Modify factors systematically.
  - Hold factors constant.
  - Ignore factors.

The standard solution

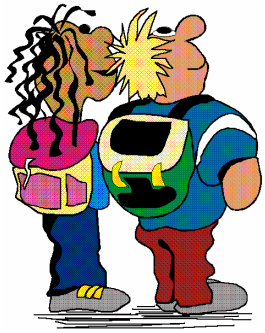
- One-Factor-at-a-Time Experiments – OFAT
- *Modify one factor; hold all other factors constant*

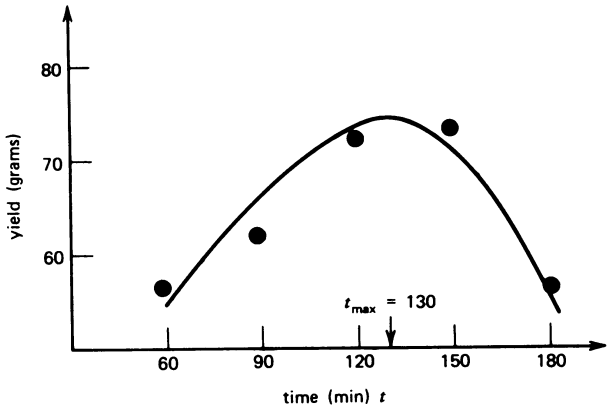
Example: optimizing yield

- Two factors affect the yield of a process:
  - Time of reaction
  - Temperature of reaction
- Feasibility experiments showed that the ranges for these factors are:
  - 60-180 minutes
  - 21-25^{o}C

One Factor at a Time

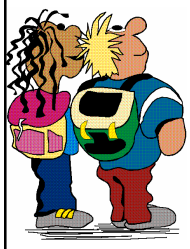
OFAT

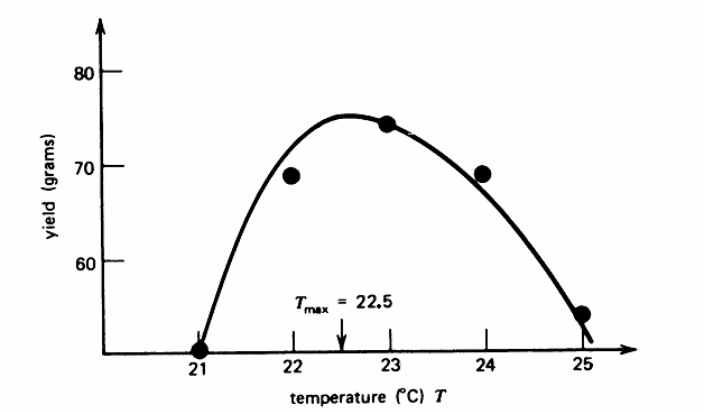




Experiment #1: Study effects of reaction time on yield
(reaction temperature held fixed at 22.5° C)

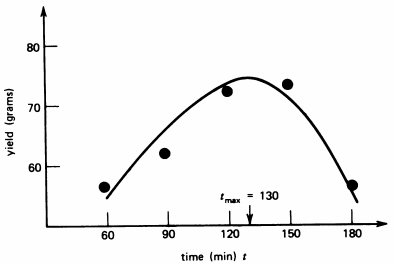
One Factor at a Time





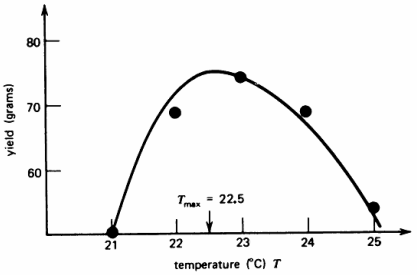
Experiment #2: Study effects of reaction temperature on yield
(reaction time held fixed at 130 minutes)

One Factor at a Time



Experiment #1: Study effects of reaction time on yield (reaction temperature held fixed at 22.5° C)

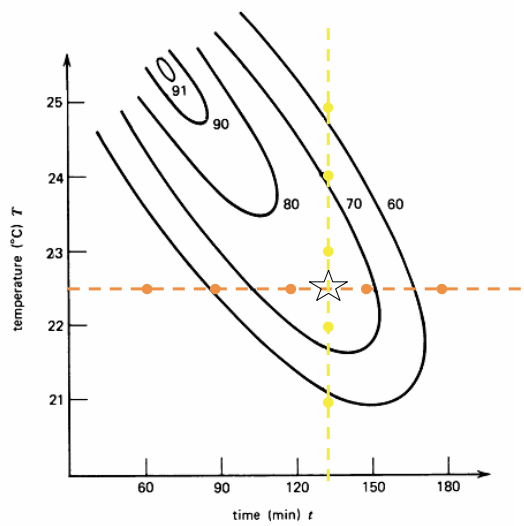
Experiment #2: Study effects of reaction temperature on yield (reaction time held fixed at 130 minutes)



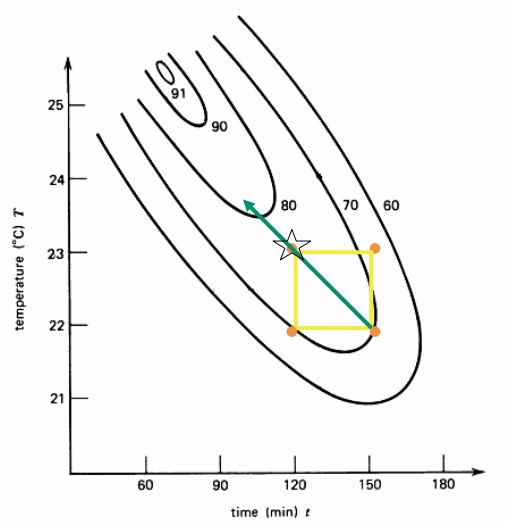
Conclusions

- Optimal conditions are 130 minutes, 22.5° C.
- Optimal yield is about 75 grams

What really happened



A better approach - DOE

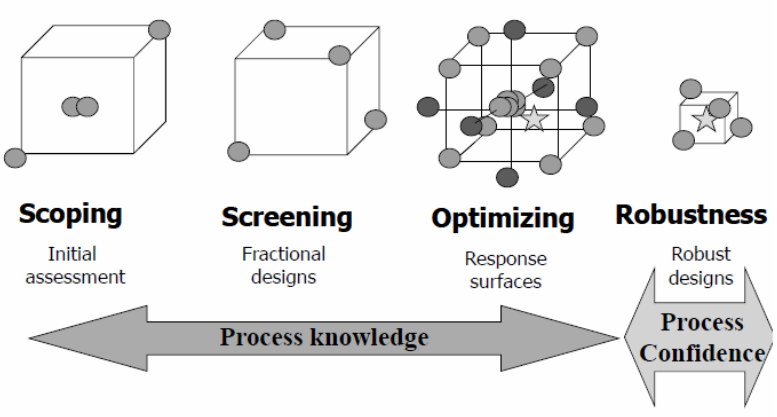


No aphorism is more frequently repeated in connection with field trials, than that we must ask Nature few questions, or, ideally, one question, at a time.

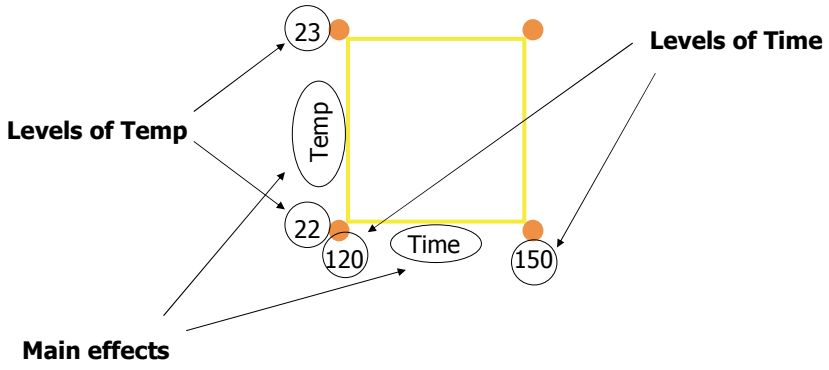
The writer is convinced that this view is wholly mistaken. Nature, he suggests, will best respond to a logical and carefully thought out questionnaire.

Sir Ronald A. Fischer

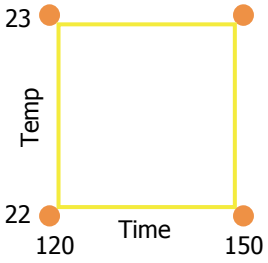
Strategy for Factorial Experiments



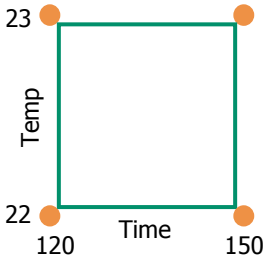
DOE terminology



Adding another main effect

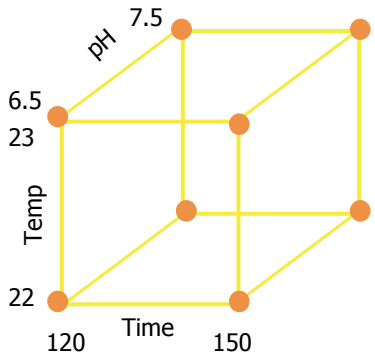


pH=6.5



pH=7.5

Three-way full design (2^3)



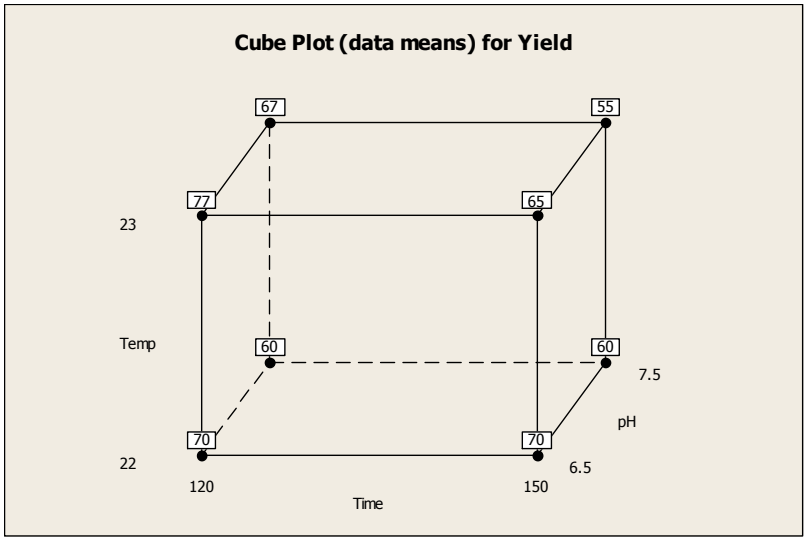
The data

Minitab - DOE.MPJ - [Worksheet 1 ***]

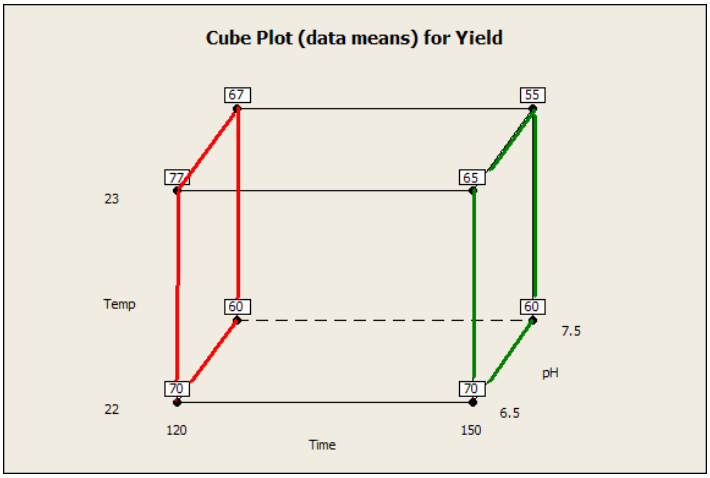
File Edit Data Calc Stat Graph Editor Tools Window Help

	C1	C2	C3	C4	C5	C6	C7	C8
	StdOrder	RunOrder	CenterPt	Blocks	Time	Temp	pH	Yield
1	8	1	1	1	150	23	7.5	55
2	6	2	1	1	150	22	7.5	60
3	2	3	1	1	150	22	6.5	70
4	7	4	1	1	120	23	7.5	67
5	5	5	1	1	120	22	7.5	60
6	1	6	1	1	120	22	6.5	70
7	4	7	1	1	150	23	6.5	65
8	3	8	1	1	120	23	6.5	77
9								

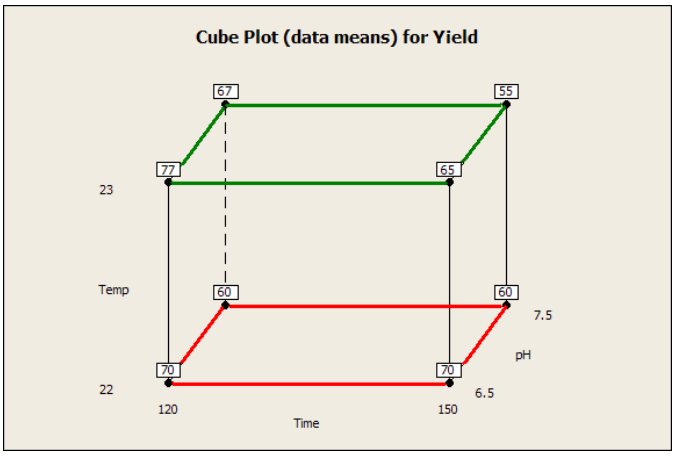
The data



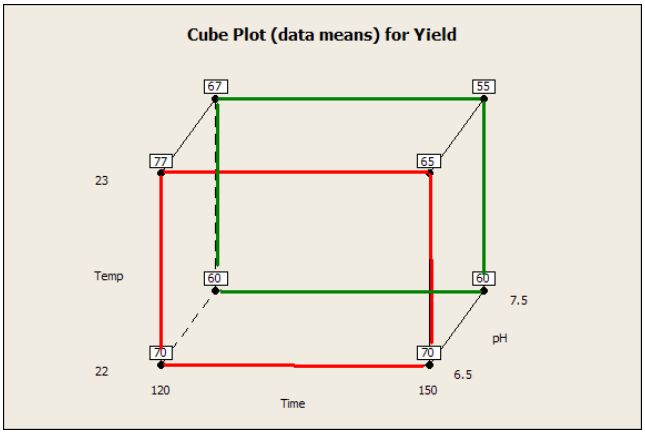
Main effect - Time



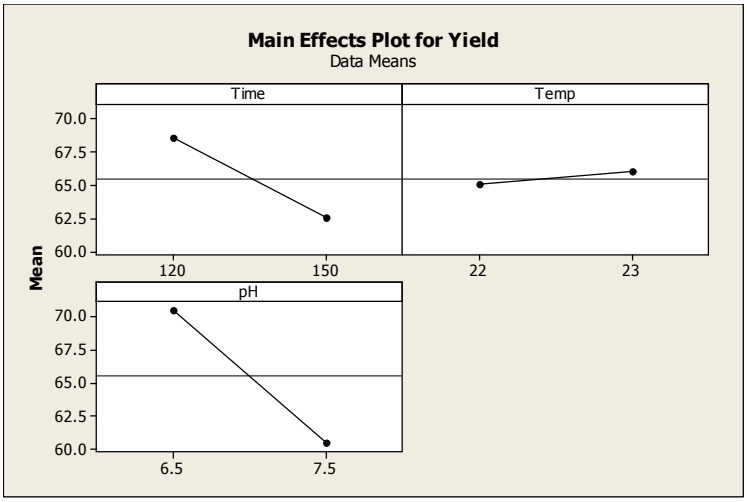
Main Effect - Temp



Main effect - pH



Main effects plot



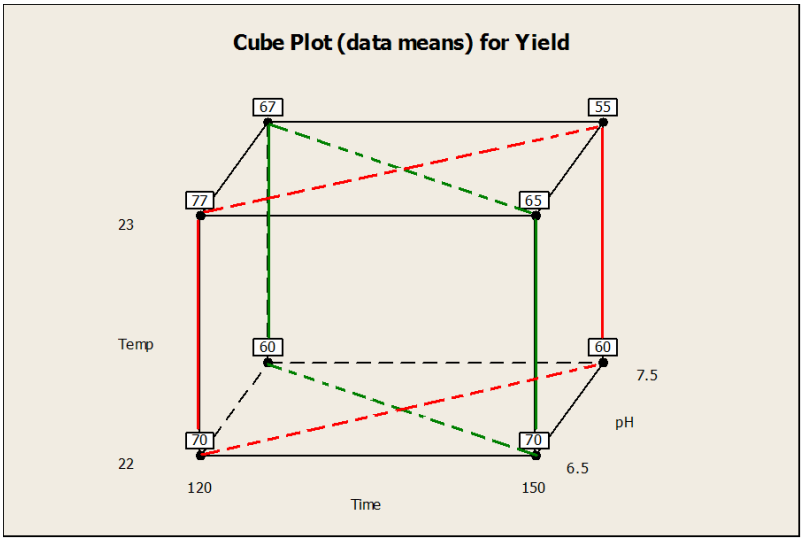
Analysis

- Regression model:

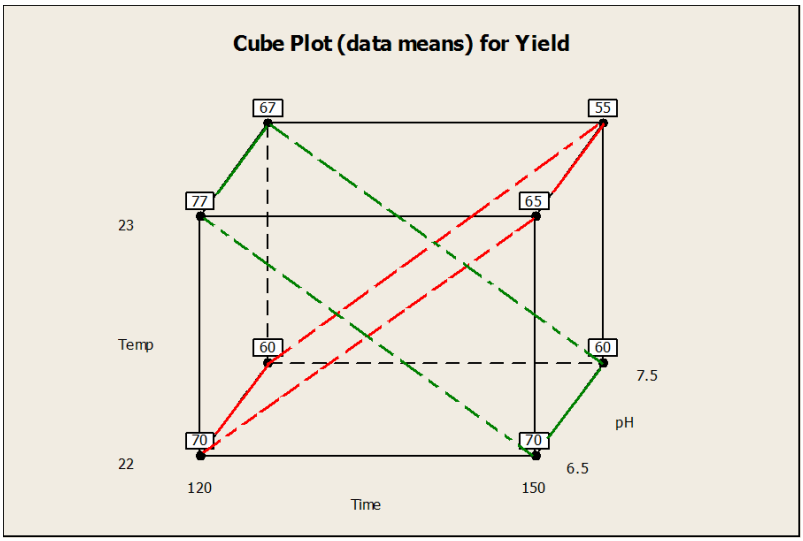
$$Yield_{ijk} = \beta_0 + \beta_1 \cdot Time_i + \beta_2 \cdot Temp_j + \beta_3 \cdot pH_k + \varepsilon_{ujk}$$

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

Term       Effect    Coef  SE Coef      T      P
Constant            65.500    1.500  43.67  0.000
Time       -6.000  -3.000    1.500  -2.00  0.116
Temp        1.000   0.500    1.500   0.33  0.756
pH        -10.000  -5.000    1.500  -3.33  0.029


S = 4.24264     PRESS = 288
R-Sq = 79.19%   R-Sq(pred) = 16.76%   R-Sq(adj) = 63.58%


Analysis of Variance for Yield (coded units)

Source          DF  Seq SS  Adj SS  Adj MS     F      P
Main Effects     3  274.00  274.00   91.33  5.07  0.075
Residual Error   4   72.00   72.00   18.00
Total            7  346.00
```

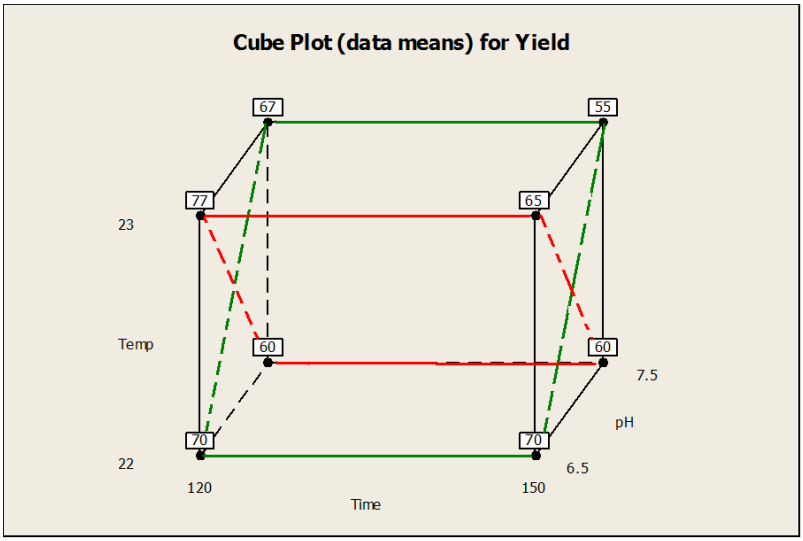

Time and pH interaction



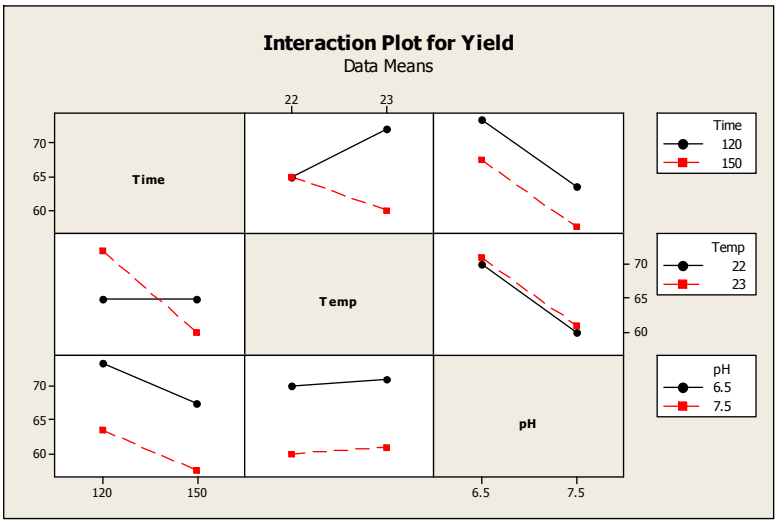
Time and Temp interaction



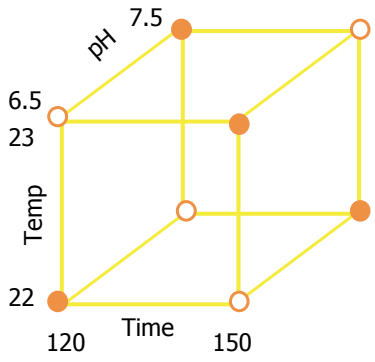
Temp and pH interaction



Interaction plots



Three-way factorial design (2^{3-1})



Development of cell culture

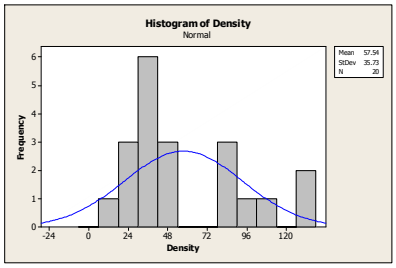
- Experiment done in a bioreactor
- Only one experiment can be set up in a day
- Response: cell density
- Factors:
  - Solution pH: low=3.25 high=4.75
  - Add carbohydrates: low=0, high=20 g/L
  - Bioreactor type – two types
  - Reaction time: low=4 hours, high=8 hours
  - Solution temperature: low=30^{o}C, high=50^{o}C

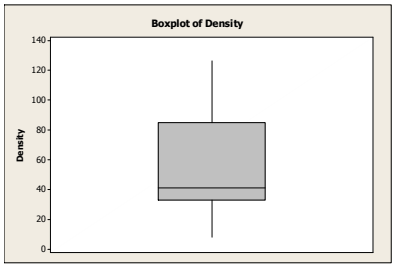
A total of 2^5=32 factor combinations

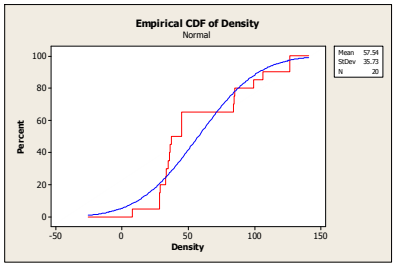
The data

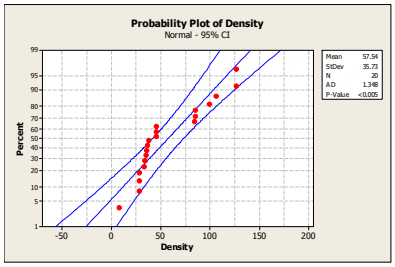
StdOrder	RunOrder	CenterPt	Blocks	pH	Carbo	Type	Time	Temp	Density
2	1	1	1	4.75	0	A	4	30	35
1	2	1	1	3.25	0	A	4	50	28
19	3	0	1	4	10	A	6	40	37.5
5	5	1	1	3.25	0	B	4	30	33.5
6	6	1	1	4.75	0	B	4	50	84
16	7	1	1	4.75	20	B	8	50	126.5
18	8	0	1	4	10	B	6	40	99
17	9	0	1	4	10	A	6	40	45
15	10	1	1	3.25	20	B	8	30	45
11	11	1	1	3.25	20	A	8	50	45
14	12	1	1	4.75	0	B	8	30	126.5
20	13	0	1	4	10	B	6	40	84.5
10	14	1	1	4.75	0	A	8	50	85
12	16	1	1	4.75	20	A	8	30	36.5
13	17	1	1	3.25	0	B	8	50	106
9	19	1	1	3.25	0	A	8	30	33
8	20	1	1	4.75	20	B	4	30	28.5
4	21	1	1	4.75	20	A	4	50	28.25
3	22	1	1	3.25	20	A	4	30	8
7	23	1	1	3.25	20	B	4	50	36

Look at cell density distribution

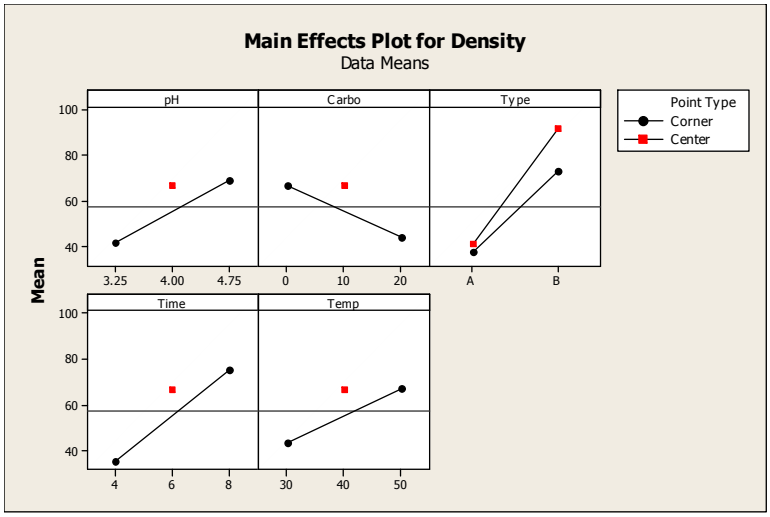




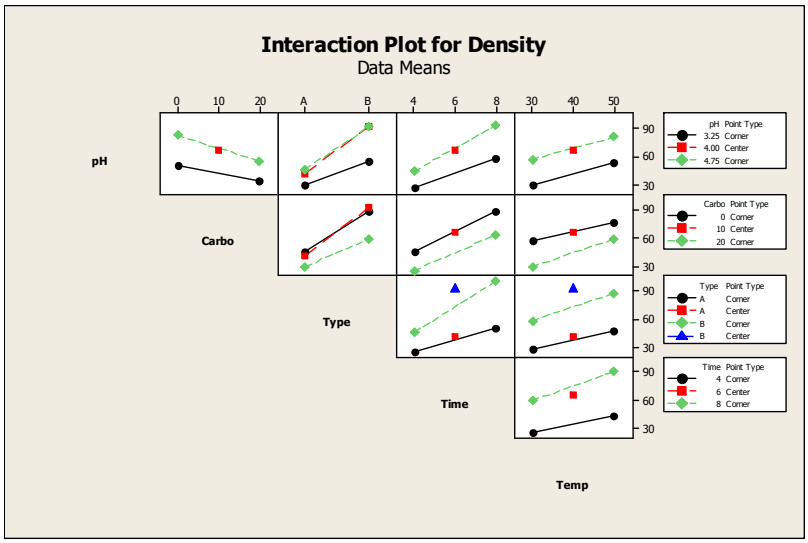




Main effects



Interactions



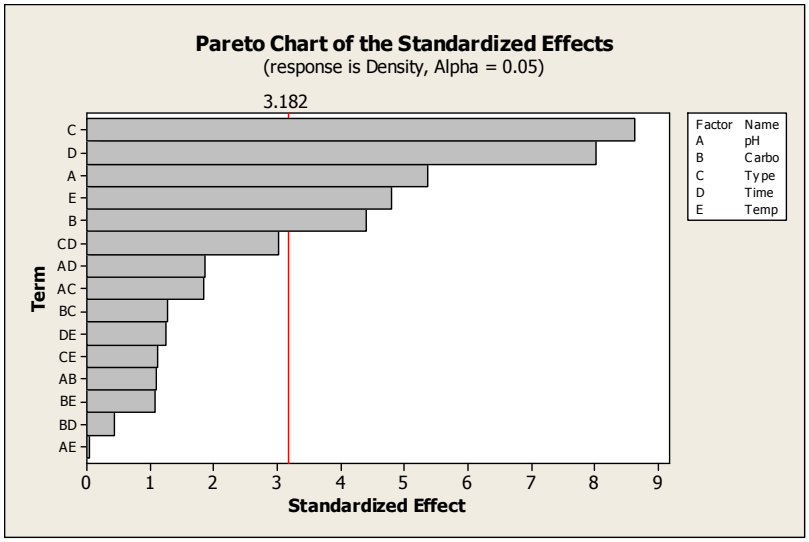
Factorial Fit for base model

Estimated Effects and Coefficients for Density (coded units)

Term	Effect	Coef	SE Coef	T	P	
Constant		55.30	2.515	21.99	0.000	
pH	26.97	13.48	2.515	5.36	0.013	*
Carbo	-22.16	-11.08	2.515	-4.40	0.022	*
Type	38.82	19.41	2.250	8.63	0.003	*
Time	40.28	20.14	2.515	8.01	0.004	*
Temp	24.09	12.05	2.515	4.79	0.017	*
pH*Carbo	-5.53	-2.77	2.515	-1.10	0.352	
pH*Type	9.28	4.64	2.515	1.85	0.162	
pH*Time	9.41	4.70	2.515	1.87	0.158	
pH*Temp	0.22	0.11	2.515	0.04	0.968	
Carbo*Type	-6.34	-3.17	2.515	-1.26	0.296	
Carbo*Time	-2.22	-1.11	2.515	-0.44	0.689	
Carbo*Temp	5.34	2.67	2.515	1.06	0.366	
Type*Time	15.22	7.61	2.515	3.03	0.057	*
Type*Temp	5.66	2.83	2.515	1.12	0.343	
Time*Temp	6.28	3.14	2.515	1.25	0.300	
Ct Pt		11.20	5.624	1.99	0.140	

S = 10.0604 PRESS = 219069
R-Sq = 98.75% R-Sq(pred) = 0.00% R-Sq(adj) = 92.07%

Pareto chart for base model



Improved model

- Keep all main effects
- Keep only Type*Time interaction

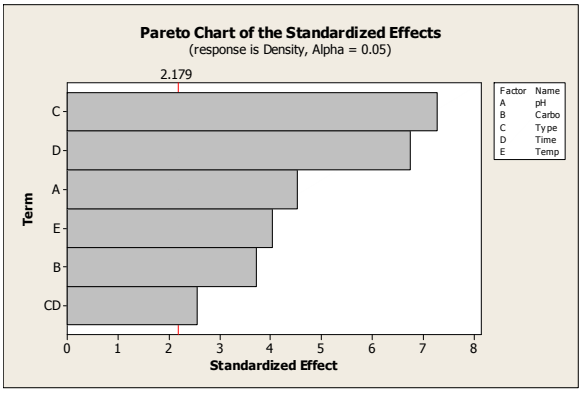
Factorial Fit for new model

Estimated Effects and Coefficients for Density (coded units)

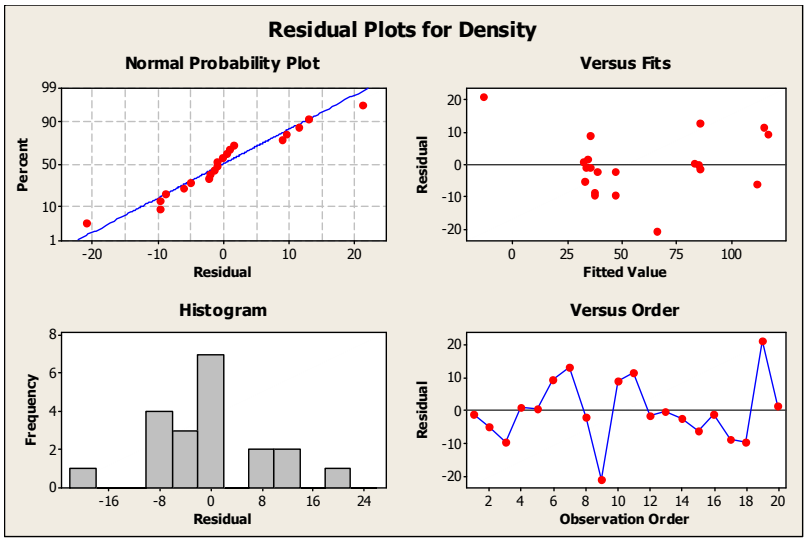
Term	Effect	Coef	SE Coef	T	P
Constant		55.30	2.980	18.55	0.000
pH	26.97	13.48	2.980	4.52	0.001
Carbo	-22.16	-11.08	2.980	-3.72	0.003
Type	38.82	19.41	2.666	7.28	0.000
Time	40.28	20.14	2.980	6.76	0.000
Temp	24.09	12.05	2.980	4.04	0.002
Type*Time	15.22	7.61	2.980	2.55	0.025
Ct Pt		11.20	6.664	1.68	0.119

S = 11.9211 PRESS = 5322.29
R-Sq = 92.97% R-Sq(pred) = 78.06% R-Sq(adj) = 88.87%

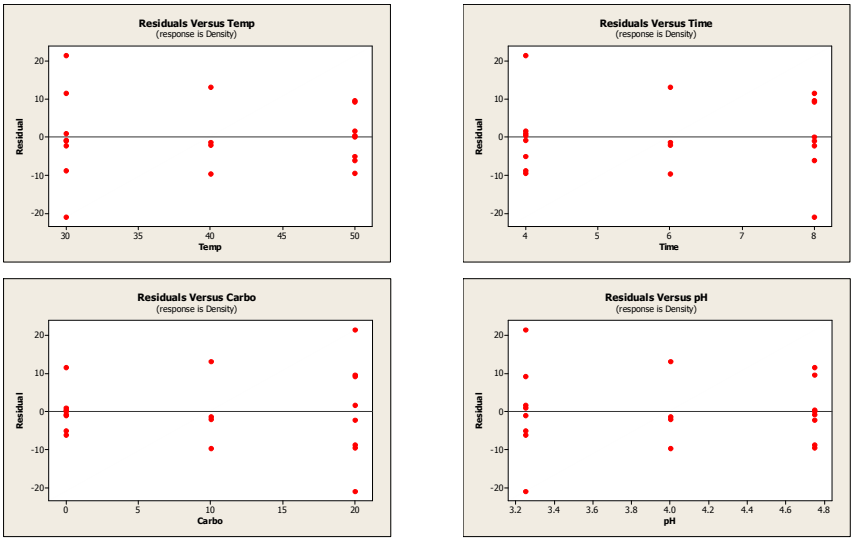
Pareto chart for new model



Check residuals



Residuals vs. main effects



How to design an experiment

1. Problem definition
2. Response variables
3. Controlled effects/factors
4. Effect levels
5. Noise factors
6. Experimental matrix
7. Number of runs
8. Protocol and SAP