Faculty: BioScienze e Tecnologie Agro-Alimentari e Ambientali
MASTER DEGREE IN FOOD SCIENCE AND TECHNOLOGY
I YEAR

Course:
EXPERIMENTAL DESIGN AND
CHEMOMETRICS IN FOOD
(5 credits – 38 hours)

Teacher: Marcello Mascini

(mmascini@unite.it)

The Teacher is available to answer questions at the
end of the lesson, or on request by mail

The course is split in 4 units

UNIT 1: Univariate analysis

Data, information, models, data types, analytical representation of data

Calibration and regression, Introduction to Statistics

Average & Variance

The Normal distribution, theory of measurement errors, the central limit theorem and the theorem of Gauss

Maximum likelihood, method of least squares, Generalization of the method of least squares

Polynomial regression, non-linear regression, the χ^2 method, Validation of the model

UNIT 2: Multivariate analysis

Correlation

Multiple linear regression

Principal component analysis (PCA)

Principal component regression (PCR) and Partial least squares regression - (PLS)

UNIT 3: Design of Experiments

Basic design of experiments and analysis of the resulting data

Analysis of variance, blocking and nuisance variables

Factorial designs

Fractional factorial designs

Overview of other types of experimental designs (Plackett–Burman designs, D-optimal designs, Supersaturated designs, Asymmetrical designs)

Response surface methods and designs

Applications of designed experiments from various fields of food science

UNIT 4: Elements of Pattern recognition

cluster analysis

Normalization

The space representation (PCA) Examples of PCA

Discriminant analysis (DA) PLS-DA

Examples of PLS-DA

UNIT 3: Design of Experiments

Basic design of experiments and analysis of the resulting data

Analysis of variance, blocking and nuisance variables

Factorial designs

Fractional factorial designs

Overview of other types of experimental designs (Plackett–Burman designs, D-optimal designs, Supersaturated designs, Asymmetrical designs)

Response surface methods and designs

Applications of designed experiments from various fields of food science

Factors

Responses

Silver laydown,
Finish time…

**Film
Building**

Speed,
Contrast

Time,
Catalyst…

**Chemical
Process**

Yield,
Purity

Transport speed,
Capture lens...

**Digital
Imaging**

Image resolution,
Banding

Factors

Compensation plan,
Sales training

Method of shipping,
Order entry method

Product positioning,
Price

Responses

Factor	Response
Sales	Sales revenue, Volume of new sales
Supply Chain	Shipping cost, Inventory level
Product Develop.	Trial purchase, Share of market

Topics

- Review of Error Analysis

- Theory & Experimentation in Engineering

- Some Considerations in Planning Experiments

- Review of Statistical formulas and theory

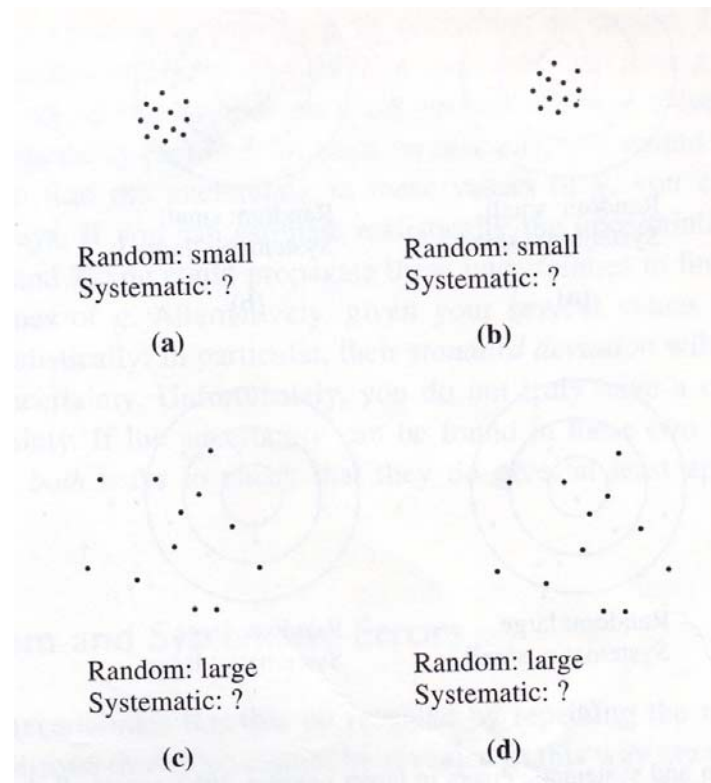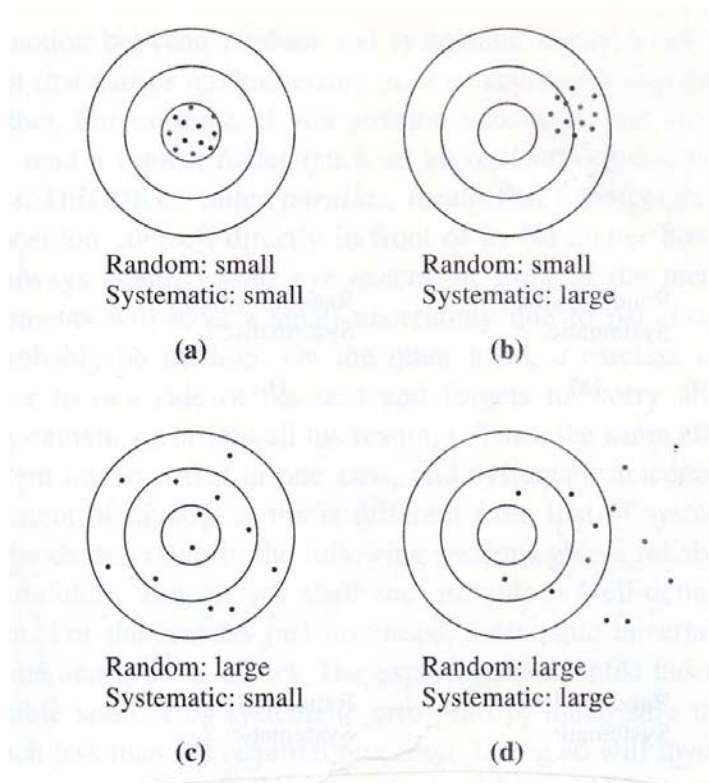- Begin Statistical Design of experiments ("DOE" or "DOX")

Review of Error Analysis

- Uncertainty or "random error" is inherent in all measurements
  - Statistical basis
  - Unavoidable- seek to *estimate* and take into account
  - Can minimize with better instruments, measurement techniques, etc.

Review of Error Analysis

- *Systematic* errors (or "method errors") are mistakes in assumptions, techniques etc. that lead to non-random bias
  - Careful experimental planning and execution can minimize
  - Difficult to characterize; can only look at evidence after the fact, troubleshoot process to find source and eliminate

Graphical Description of Random and Systematic Error

Random: small
Systematic: small
(a)

Random: small
Systematic: large
(b)

Random: large
Systematic: small
(c)

Random: large
Systematic: large
(d)

Random: small
Systematic: ?
(a)

Random: small
Systematic: ?
(b)

Random: large
Systematic: ?
(c)

Random: large
Systematic: ?
(d)

Why do we need to estimate uncertainty and include in stated experimental values?
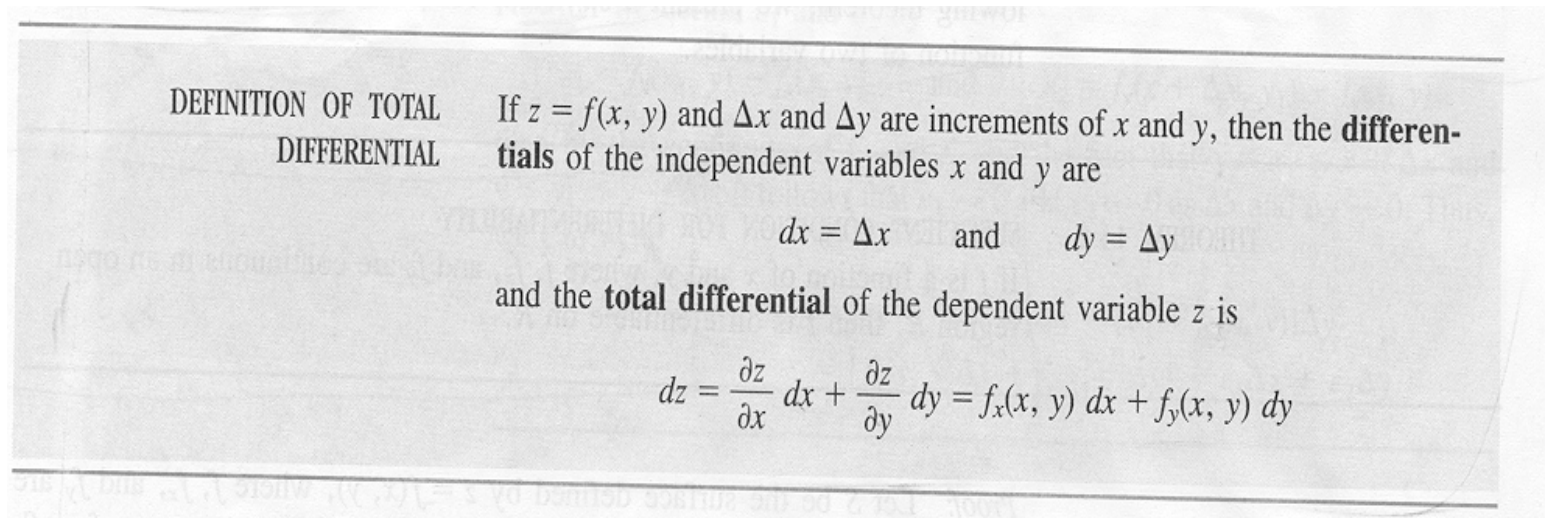
- Probability of being wrong will influence process and/or financial decisions
  - Cost / benefit of accepting result as "fact"?
  - What would be the effect downstream as the uncertainty propagates through the process?
- When comparing two values and determining if they are different
  - Overlap of uncertainty?
  - What is the probability that the difference is *significant*?

Stating Results +/- Uncertainty

- Rule for Stating Uncertainties
  - Experimental uncertainties should almost always be rounded to one significant figure.

- Rule for Stating Answers
  - The last significant figure in any stated answer should usually be of the same order of magnitude (in the same decimal position) as the uncertainty.
  - Express Uncertainty as error bars and confidence interval for graphical data and curve fits (regressions) respectively

Determining *Propagated* Error: Non-statistical Method

- Compute from total differential

DEFINITION OF TOTAL DIFFERENTIAL	If $z = f(x, y)$ and Δx and Δy are increments of x and y, then the **differentials** of the independent variables x and y are

$$dx = \Delta x \quad \text{and} \quad dy = \Delta y$$

and the **total differential** of the dependent variable z is

$$dz = \frac{\partial z}{\partial x}\, dx + \frac{\partial z}{\partial y}\, dy = f_x(x, y)\, dx + f_y(x, y)\, dy$$

Propagated error

- OR Can do *sensitivity analysis* in spreadsheet of other software program
  - Compute possible uncertainty in calculated result based on varying values of inputs according to the uncertainty of each input
  - Example: Use "Solver" optimization tool in Excel to find maximum and minimum values of computed value in a cell by varying the value of each input cell
    - Set constraint that the input values lie in the range of uncertainty of that value

Or Can Use *repeat measurements* to estimate uncertainty in a result using *probability and statistics* for *random* errors*:*

- mean
- standard deviation of each measurement
- standard deviation of the mean of the measurements

- Confidence intervals on dependant variable
- Confidence intervals on regression parameters

Statistical Formulas from chapter 4 of *Taylor*

THE STANDARD DEVIATION

The average uncertainty of the individual measurements $x_1, x_2, \ldots, x_N$ is given by the standard deviation, or SD:

$$\sigma_x = \sqrt{\frac{1}{N-1} \Sigma (x_i - \bar{x})^2}. \qquad \text{[See (4.9)]}$$

This definition of the SD, often called the *sample* standard deviation, is the most appropriate for our purposes. The *population* standard deviation is obtained by replacing the factor $(N-1)$ in the denominator by N. You will usually want to calculate standard deviations using the built-in function on your calculator; be sure you know which definition it uses.

The detailed significance of the standard deviation σ_x is that approximately 68% of the measurements of x (using the same method) should lie within a distance σ_x of the true value. (This claim is justified in Section 5.4.) This result is what allows us to identify σ_x as the *uncertainty* in any one measurement of x,

$$\delta x = \sigma_x,$$

and, with this choice, we can be 68% confident that any one measurement will fall within σ_x of the correct answer.

THE STANDARD DEVIATION OF THE MEAN

As long as systematic uncertainties are negligible, the uncertainty in our best estimate for x (namely $\bar{x}$) is the standard deviation of the mean, or SDOM,

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}. \qquad \text{[See (4.14)]}$$
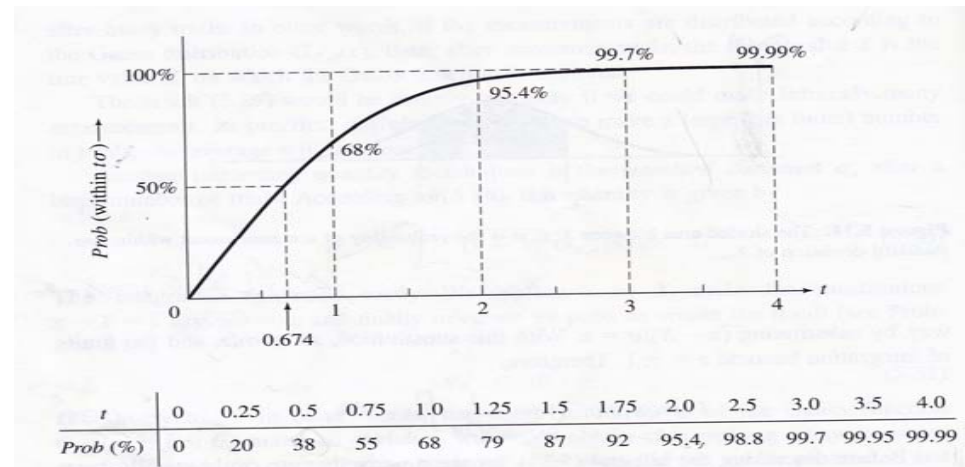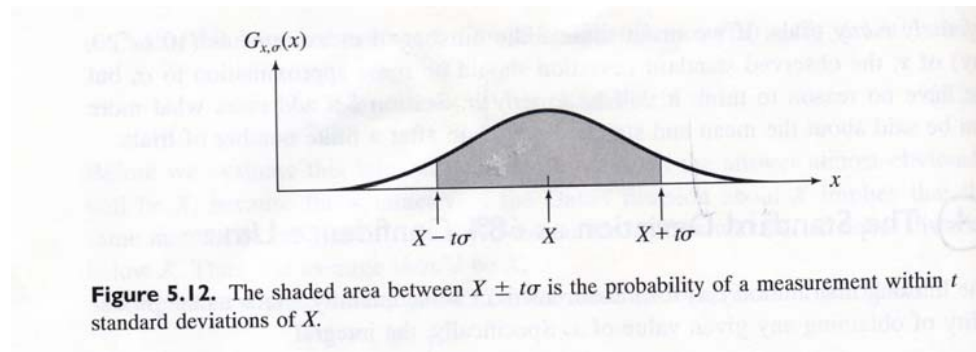
If there *are* appreciable systematic errors, then $\sigma_{\bar{x}}$ gives the *random component* of the uncertainty in our best estimate for x:

$$\delta x_{\text{ran}} = \sigma_{\bar{x}}.$$

If you have some way to estimate the systematic component δx_{sys}, a reasonable (but not rigorously justified) expression for the total uncertainty is the quadratic sum of δx_{ran} and δx_{sys}:

$$\delta x_{\text{tot}} = \sqrt{(\delta x_{\text{ran}})^2 + (\delta x_{\text{sys}})^2}. \qquad \text{[See (4.26)]}$$

Relationship of standard deviation to confidence intervals

Figure 5.12. The shaded area between $X \pm t\sigma$ is the probability of a measurement within t standard deviations of X.

| t	0	0.25	0.5	0.75	1.0	1.25	1.5	1.75	2.0	2.5	3.0	3.5	4.0
Prob (%)	0	20	38	55	68	79	87	92	95.4	98.8	99.7	99.95	99.99

Confidence intervals on non-linear regression coefficients

- Can be complex- use software but understand theory of how calculated for linear case

Error bars that represent uncertainty in the dependant variable
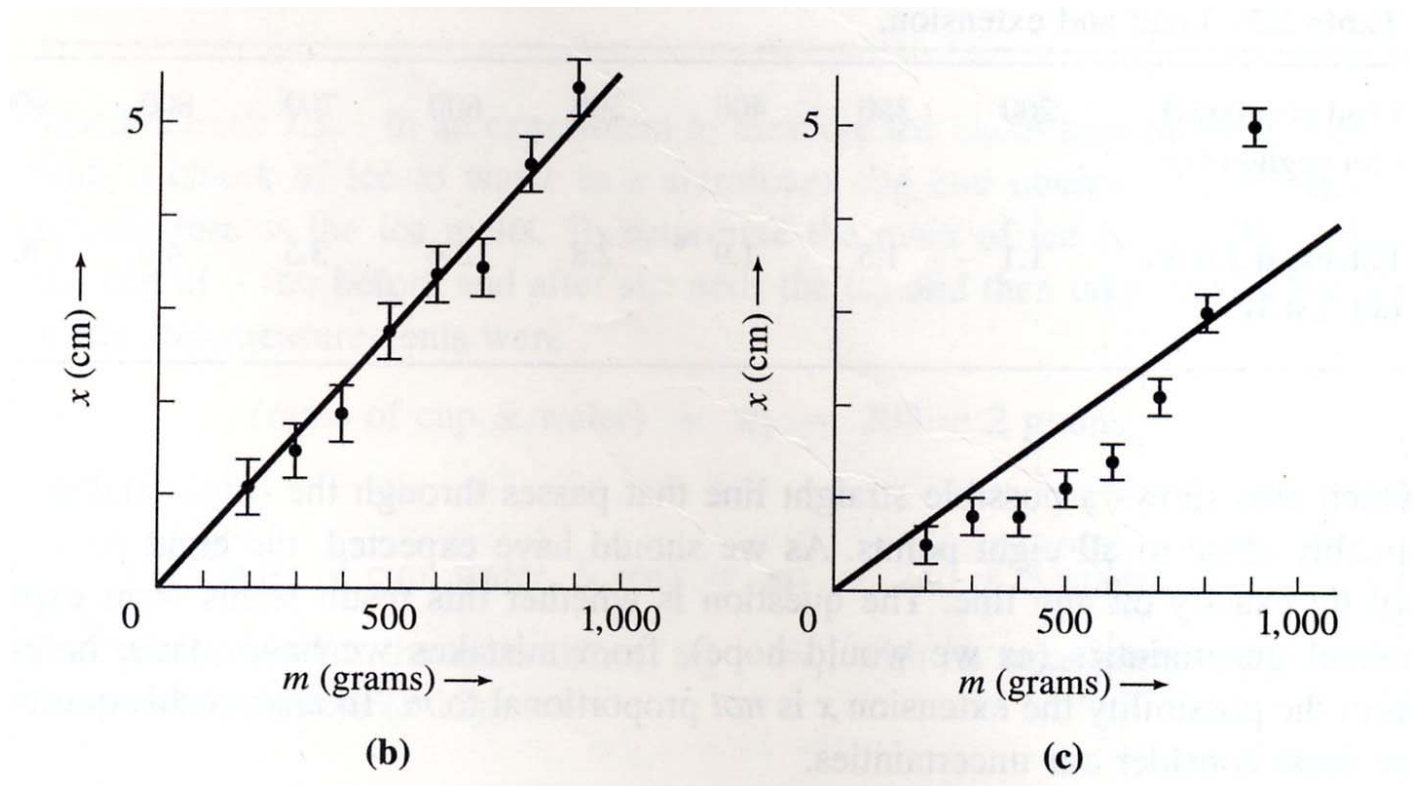
How measurements at a given x,y would be distributed for multiple measurements

Figure 43   Regression curve. The locus of the mean values of the y-distributions
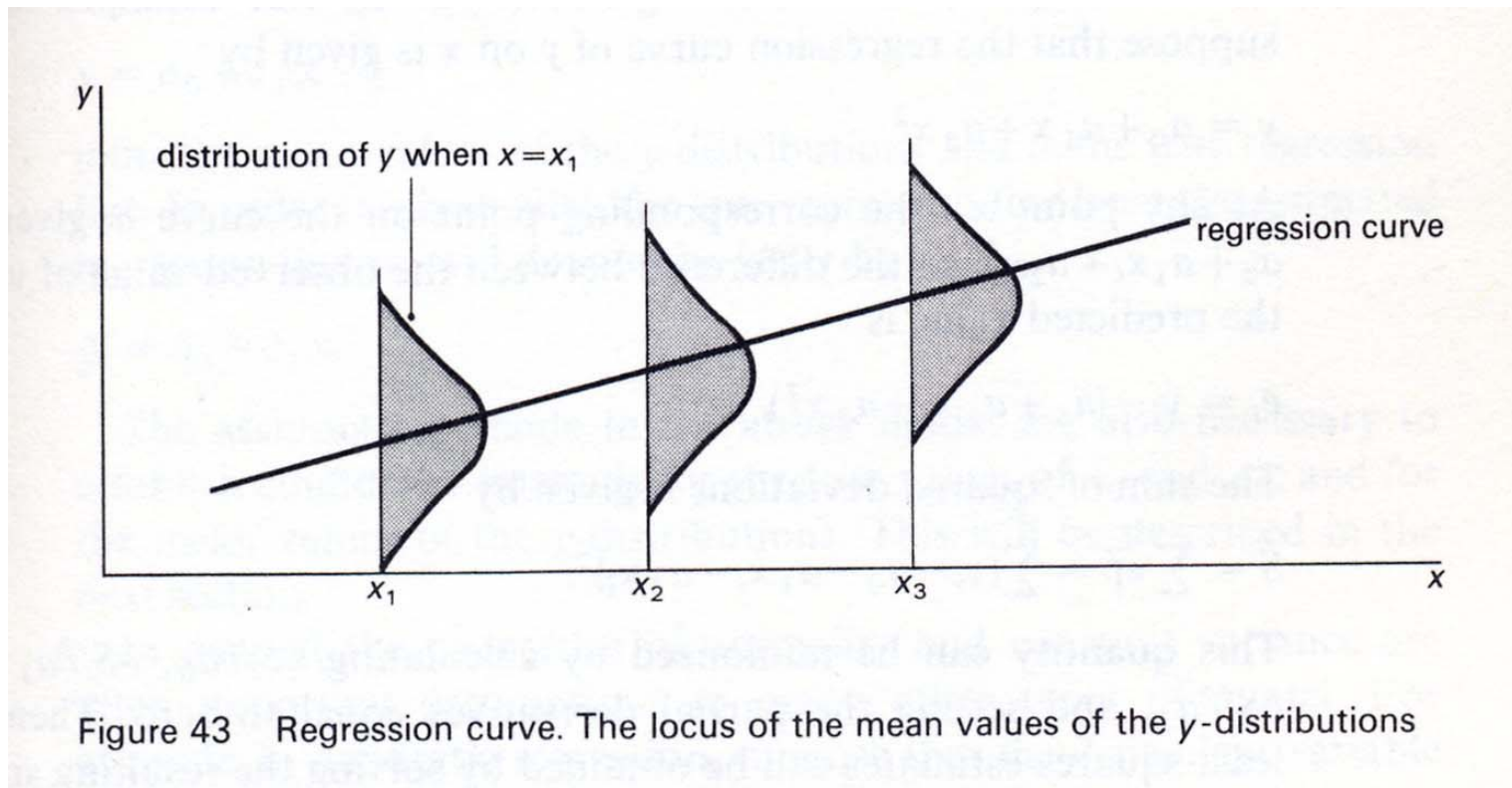
Determining Slope and Intercept In Linear Regression

A STRAIGHT LINE, $y = A + Bx$; EQUAL WEIGHTS

If y is expected to lie on a straight line $y = A + Bx$, and if the measurements of y all have the same uncertainties, then the best estimates for the constants A and B are:

$$A = \frac{\sum x^2 \sum y - \sum x \sum xy}{\Delta}$$

and

$$B = \frac{N \sum xy - \sum x \sum y}{\Delta},$$

where the denominator, Δ, is

$$\Delta = N \sum x^2 - (\sum x)^2. \qquad \text{[See (8.10) to (8.12)]}$$

Based on the observed points, the best estimate for the uncertainty in the measurements of y is

$$\sigma_y = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N} (y_i - A - Bx_i)^2}. \qquad \text{[See (8.15)]}$$

Confidence intervals (SD) on slope B and Intercept A

Chapter 8: Least-Squares Fitting
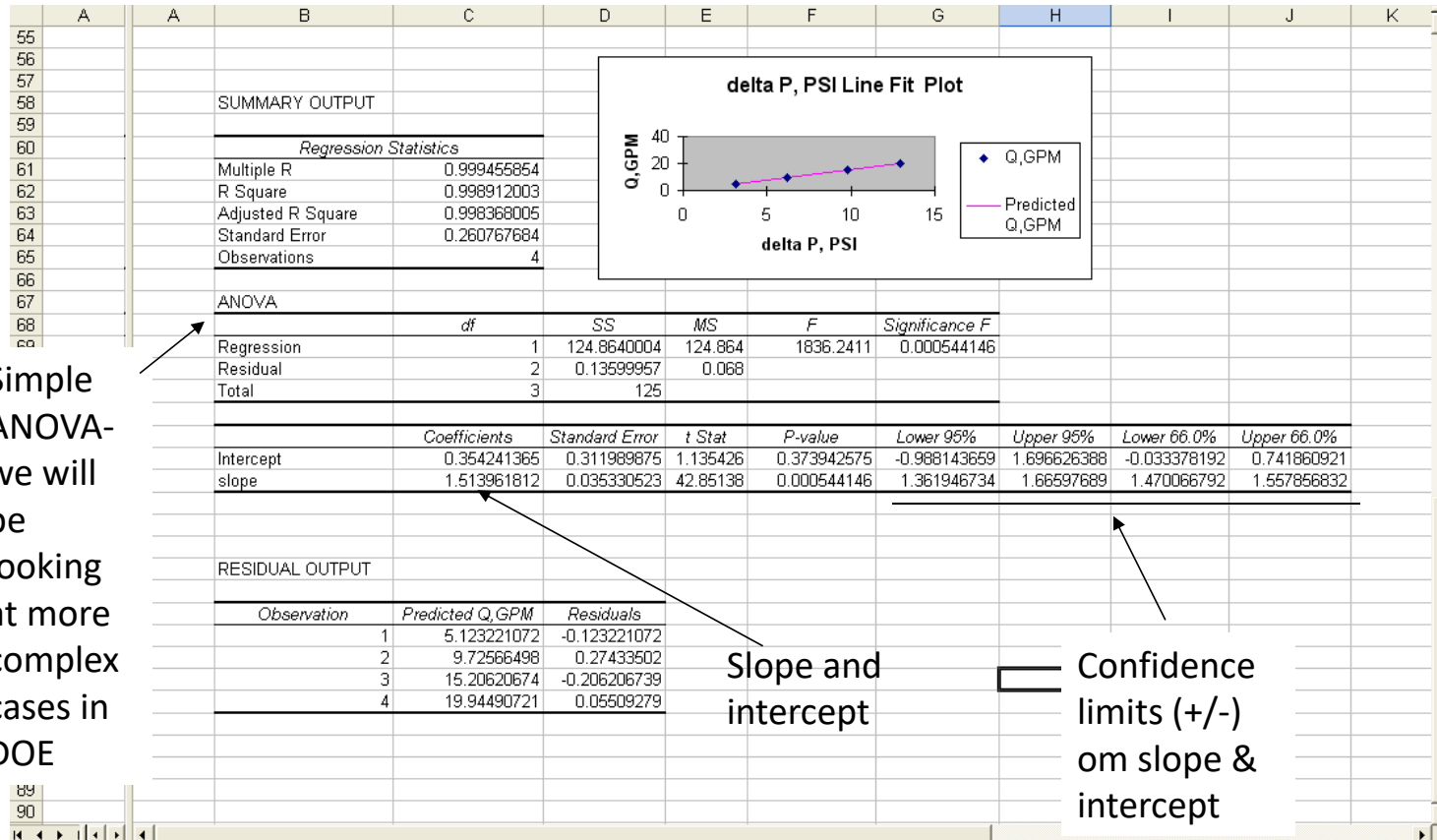
The uncertainties in A and B are:

$$\sigma_A = \sigma_y \sqrt{\frac{\sum x^2}{\Delta}}$$

and

$$\sigma_B = \sigma_y \sqrt{\frac{N}{\Delta}}.$$

[See (8.16) & (8.17)]

Regression Output in Excel

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.999455854
R Square	0.998912003
Adjusted R Square	0.998368005
Standard Error	0.260767684
Observations	4

delta P, PSI Line Fit Plot

Q,GPM Predicted Q,GPM

ANOVA

	df	SS	MS	F	Significance F
Regression	1	124.8640004	124.864	1836.2411	0.000544146
Residual	2	0.13599957	0.068		
Total	3	125			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 66.0%	Upper 66.0%
Intercept	0.354241365	0.311989875	1.135426	0.373942575	-0.988143659	1.696626388	-0.033378192	0.741860921
slope	1.513961812	0.035330523	42.85138	0.000544146	1.361946734	1.66597689	1.470066792	1.557856832

RESIDUAL OUTPUT

Observation	Predicted Q,GPM	Residuals
1	5.123221072	-0.123221072
2	9.72566498	0.27433502
3	15.20620674	-0.206206739
4	19.94490721	0.05509279

Simple ANOVA- we will be looking at more complex cases in DOE
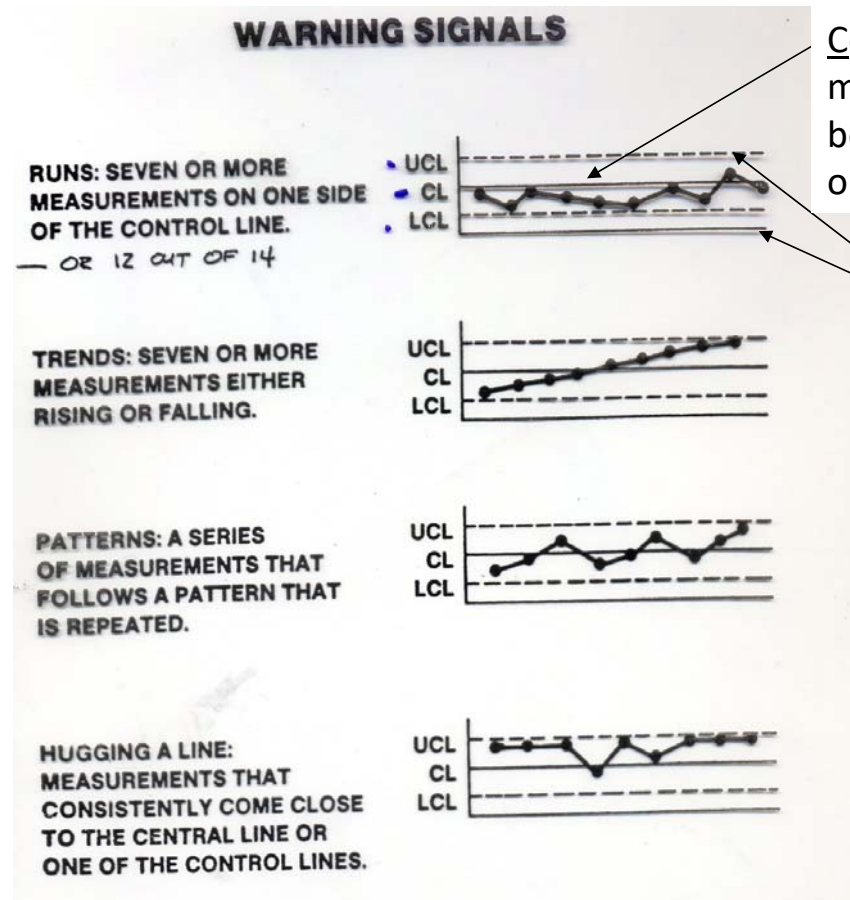
Slope and intercept

Confidence limits (+/-) om slope & intercept

Statistical Process Control

- Very Widely Used
- Used for quality control and **in conjunction with DOE** for process improvement
- *Control Charts* provide statistical evidence
  - That a process is behaving normally or if something wrong
  - Serve as data output (dependant variable)from process in designed statistical experiments

Variation from expected behavior in control charts- similar to regression and point statistics

WARNING SIGNALS

Control Limit is the mean of a well behaved process output (i.e. product)

RUNS: SEVEN OR MORE MEASUREMENTS ON ONE SIDE OF THE CONTROL LINE.
— OR 12 OUT OF 14

UCL
CL
LCL

Upper and lower Control Limits represent confidence limit on mean of "well behaved" process ouptut

Expect random deviations form mean just like in regression

TRENDS: SEVEN OR MORE MEASUREMENTS EITHER RISING OR FALLING.

UCL
CL
LCL

PATTERNS: A SERIES OF MEASUREMENTS THAT FOLLOWS A PATTERN THAT IS REPEATED.

UCL
CL
LCL

HUGGING A LINE: MEASUREMENTS THAT CONSISTENTLY COME CLOSE TO THE CENTRAL LINE OR ONE OF THE CONTROL LINES.

UCL
CL
LCL

Theory and Experimentation

- Two fundamental approaches to problem solving problems in the discovery of knowledge:

  1. Theoretical (physical/mathematical modeling)
  2. Experimental measurement

  (Most often a combination is used)

Example of combination of theory and experimentation to get semi-empirical correlation

Fig. 7.8.3 Heat transfer to gases flowing in packed beds.
○ Wilke and Hougen, $\epsilon = 0.34$–0.44, cylinders with $L/D \sim 1$. △ Taecker and Hougen, $\epsilon = 0.60$–0.74, commercial packings. □ McConnachie and Thodos, $\epsilon = 0.42$, spheres. ● Glaser and Thodos, $\epsilon = 0.44$–0.48, spheres and cylinders with $L/D \sim 1$. ▲ Gamson, Thodos, and Hougen, $\epsilon = 0.36$–0.43, spheres and cylinders with $L/D \sim 1$. ■ Gliddon and Cranfield, $\epsilon = 0.43$. ▲ Littman, Barile, and Pulsifer, $\epsilon \sim 0.4$, spheres.
Curve is given by

$$N_{Nu}/N_{Pr}^{0.4} = 0.4N_{Re}^{1/2} + 0.2N_{Re}^{2/3}$$

Features of alternative methods

- <u>Theoretical Models</u>
  - Simplifying assumptions needed
  - General results
  - Less facilities usually needed
  - Can start study immediately

- <u>Experimental approach</u>
  - Study the "real world"-no simplifying assumptions needed
  - Results specific to apparatus studied
  - High accuracy measurements need complex instruments
  - Extensive lab facilities maybe needed
  - Time delays from building apparatus, debugging

Functional Types of Engineering Experiments

1. Determine material properties
2. Determine component or system performance indices
3. Evaluate/improve theoretical models
4. Product/process improvement by testing
5. Exploratory experimentation
6. Acceptance testing
7. Teaching/learning through experimentation

Some important classes of Experiments

1. Estimation of parameter mean value
2. Estimate of parameter variability
3. Comparison of mean values
4. Comparison of variability
5. Modeling the dependence of dependant Variable on several quantitative and/or qualitative variables

Practical Experimental Planning

Experimental design:
- Consider *goals*
- Consider what data can be collected.
- Difficulty of obtaining data
- What data is *most* important
- What measurements can be ignored
- Type of data: Categorical? Quantitative?
- Test to make sure that measurements/apparatus are reliable
- Collect data carefully and document fully in ink using bound notebooks. Make copies and keep separately

Preview of Uses for DOE

- Lab experiments for research

- Industrial process experiments

Four engineering problem classes to which DOE is applied in manufacturing

1. Comparison

2. Screening/ characterization

3. Modeling

4. Optimization

Comparison

- Compares to see if a change in a single "factor" (variable) has resulted in a process change (ideally an improvement)

Screening/Characterization

- Used when you want to see the effect of a whole range of factors so as to know which one(s) are most important.

- Two factorial experiments usually used

Modeling

- Used when you want to be able to construct a mathematical model that will predict the effect on a process of manipulating a variables or multiple variables

Optimization

- When you want to determine the optimal settings for all factors to give an optimal process response.

Introduction to experimental design

Contents

- planning experiments
- regression analysis
- types of experiments
- software $A = \pi r^2$
- literature

Example of Experiment : synthesis of T8-POSS

- context: development of new synthesis route for polymer additive
- goal: optimize yield of reaction
- synthesis route consists of elements that are not uniquely determined (control variables):
  - time to let reaction run
  - concentration water
  - concentration silane
  - temperature
  - …

Issues in example T8-POSS synthesis

- ## how to measure yield
  - what to measure (begin/end weight,…)
  - when to measure (reaction requires at least one day)
- ## how to vary control variables
  - which values of pH, concentrations, … (*levels*)
  - which combinations of values
  - equipment only allows 6 simultaneous reactions, all with the same temperature
- ## how many combinations can be tested
  - reaction requires at least one day
  - only 4 experimentation days are available

Necessity of careful planning of experiment

- limited resources
  - time to carry out experiment
  - costs of required materials/equipment
- avoid reaching suboptimal settings
- avoid missing interesting parts of experimental region
- protection against external uncontrollable/undetectable influences
- getting precise estimates

Traditional approach to experimentation:
T8-POSS example

• set T = 40 °C, H_2O concentration = 10%; try c_{Si}=0.1, 0.2, 0.3, 0.8,0.9,1.0 M

• set T = 60 °C, c_{Si}=0.5M, H_2O concentration = 5, 10, 12.5, 15, 17.5, 20%

•…

This is called a One-Factor-At-a-Time (OFAT) or Change-One-Separate-factor-at-a-Time (COST) strategy. Disadvantages:

• may lead to suboptimal settings (see next slide)

• requires too many runs to obtain good coverage of experimental region (see later)

30

40

50

60

The *real* maximum

factor B has been optimised

The *apparent* maximum

factor A has been optimised
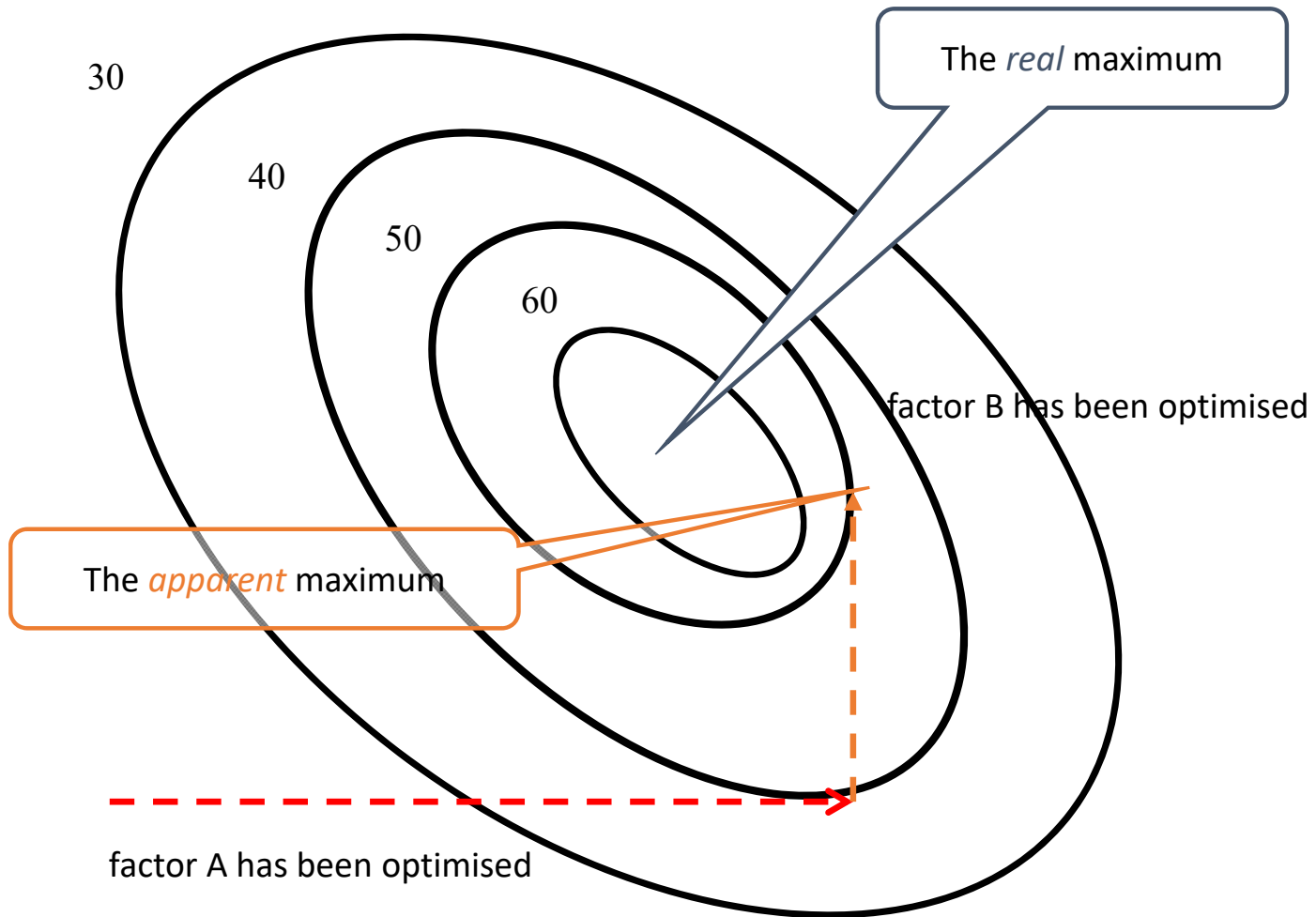
Statistical terminology for experiments:
illustrated by T8-POSS example

- response variable: yield
- factors: time, temperature, c_{Si}, H_2O concentration
- levels: actual values of factors (e.g., T=30 °C, 40 °C ,50 °C)
- runs: one combination of factor settings (e.g., T=30 °C, c_{Si}=0.5M, H_2O concentration = 15%)
- block: 6 simultaneous runs with same temperature in reaction station

Modern approach: DOE

- DOE = Design of Experiments
- key ideas:
  - change several factors simultaneously
  - carefully choose which runs to perform
  - use regression analysis to obtain effect estimates
- statistical software (Statgraphics, JMP, SAS,…) allows to
  - choose or construct designs
  - analyse experimental results

Example of analysis

simple experiment:
- response is conversion
- goal is screening (are time and temperature influencing conversion?)
- 2 factors (time and temperature), each at two levels
- 5 centre points (both time and temperature at intermediate values)

Statgraphics demo with conversion.sfx. (choose Special -> Experimental Design etc. from menu)

More advanced (5 factors, not all 2^5 combinations): colour.sfx

Example of construction: T8-POSS example

- 36 runs
  - 2 reactors available each day (each reactor 6 places)
  - 3 experimental days
- factors:
  - H_2O concentration
  - temperature
  - c_{Si}
- goal is optimization of response
- choose in Statgraphics: Special -> Experimental Design -> Create Design -> Response Surface

Goals in experimentation

- there may be more than one goal, e.g.:
  - yield
  - required reaction time until equilibrium
  - costs of required chemical substances
  - impact on environment (waste)
- these goals may contradict each other
- goals must be converted to explicitly measurable quantities

Types of experimental designs

- "screening designs"

These designs are used to investigate *which* factors are important ("significant").

- "response surface designs"

These designs are used to determine the *optimal* settings of the significant factors.

Interactions

Factors may *influence each other*. E.g, the optimal setting of a factor may depend on the settings of the other factors.

When factors are optimised *separately*, the overall result (as function of all factors) may be *suboptimal* ...

Interaction effects

Cross terms in linear regression models cause interaction effects:

$$Y = 3 + 2\,x_A + 4\,x_B + 7\,x_A\,x_B$$

$$x_A \to x_A + 1 \implies Y \to Y + 2 + 7\,x_B,$$

so increase depends on x_B. Likewise for $x_B \to x_B + 1$

This explains the notation AB for the interaction of factors A and B.

No interaction

Output

55

50

B low

20

25

B high

low

high

Factor A
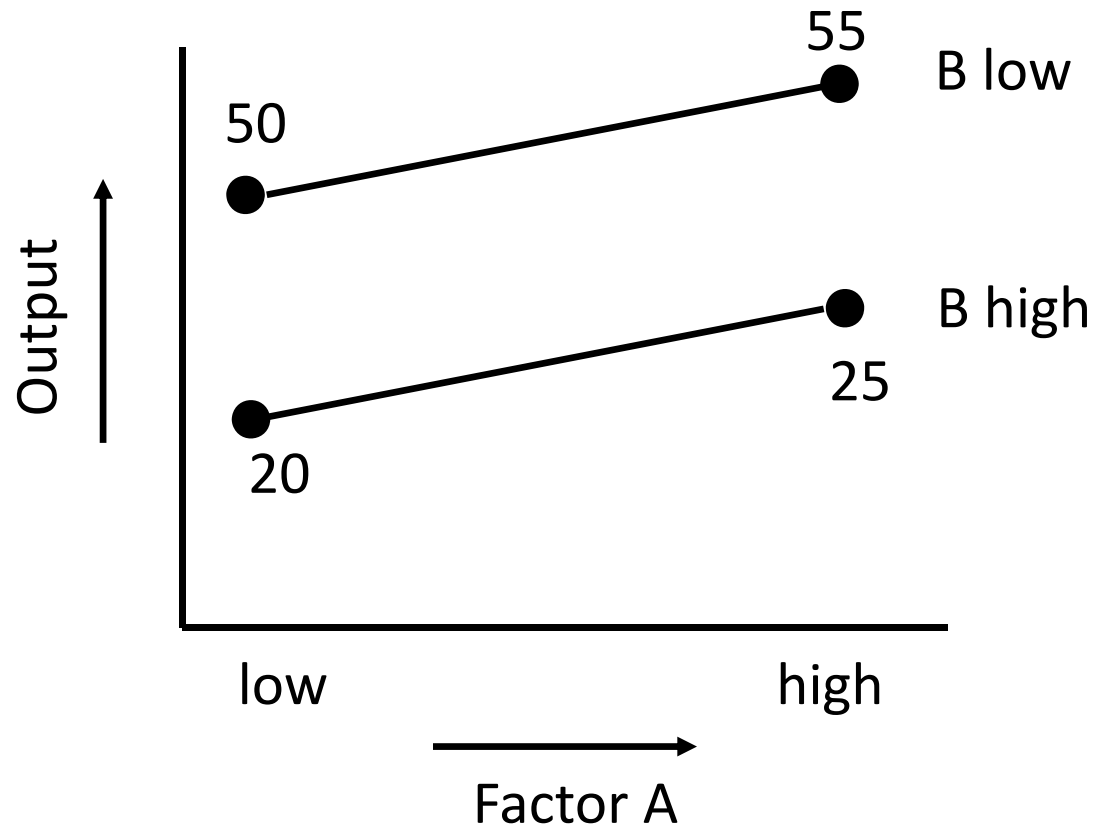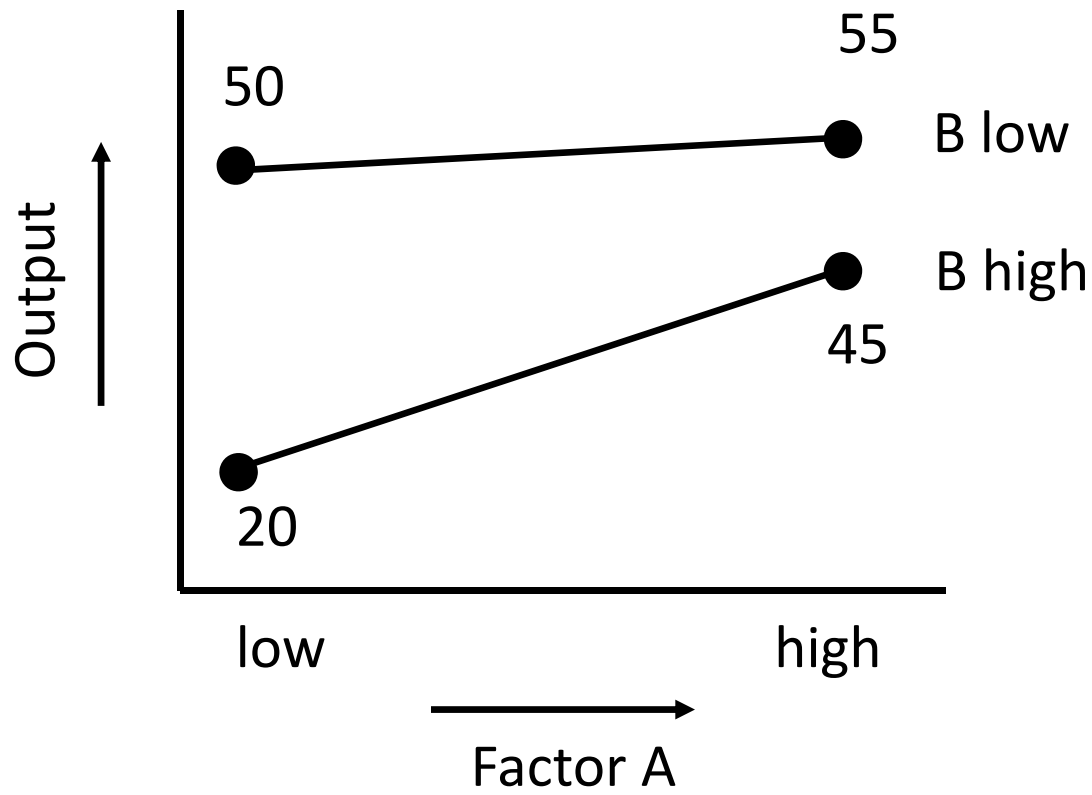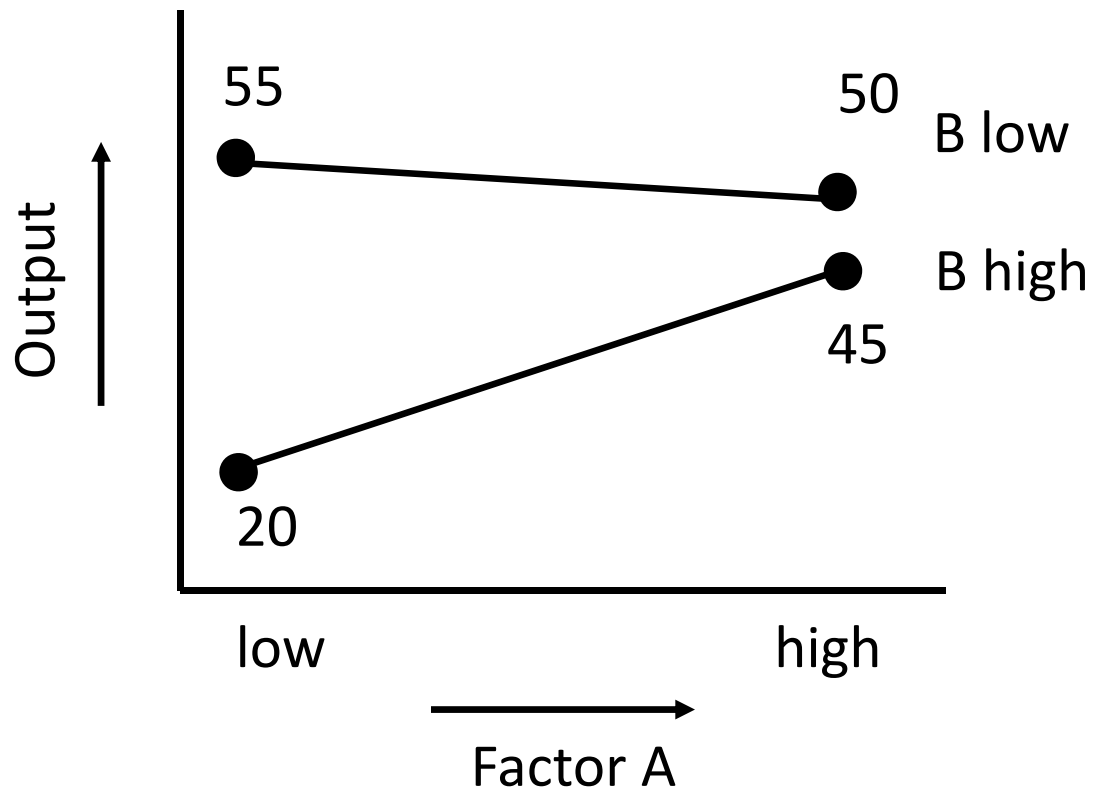
Interaction I
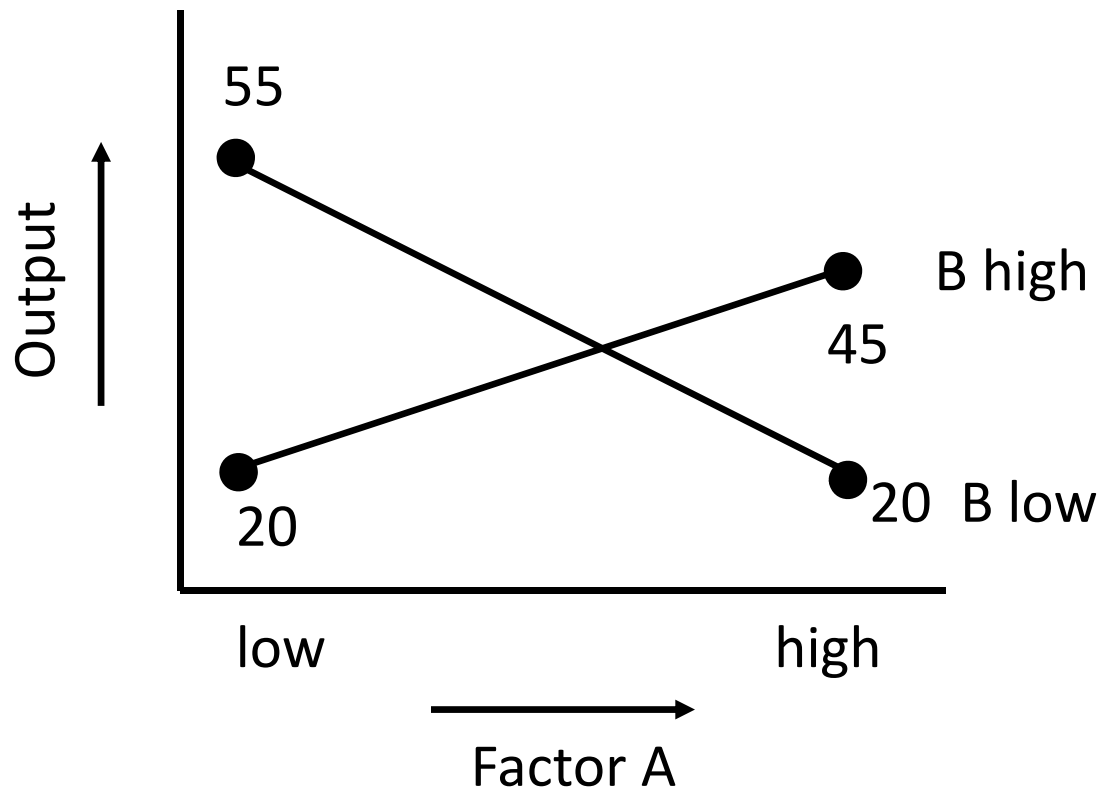
Centre points and Replications

If there are not enough measurements to obtain a good estimate of the variance, then one can perform replications. Another possibility is to add *centre points* .

Adding centre points serves two purposes:
 better variance estimate
 allow to test curvature using
 a lack-of-fit test

Centre point

b   ab
+1

B

-1 (1)   a

-1   +1

A
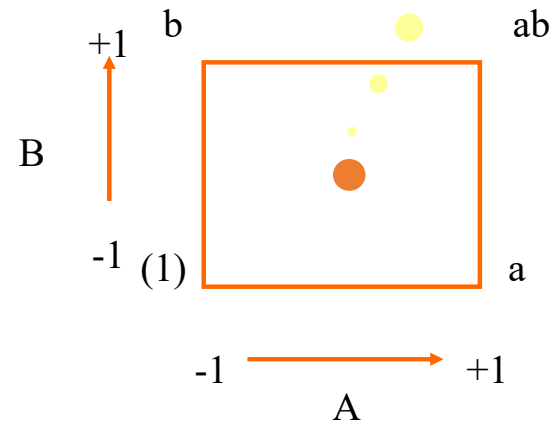
Multi-layered experiments

Experiments are not one-shot adventures. Ideally one performs:

- an initial experiment
  - check-out experimental equipment
  - get initial values for quantities of interest

- main experiment
  - obtain results that support the goal of the experiment

- confirmation experiment
  - verify results from main experiment
  - use information from main experiment to conduct more focussed experiment (e.g., near computed optimum)

Example

- *testing method for material hardness* :

force

pressure pin/tip

strip testing material
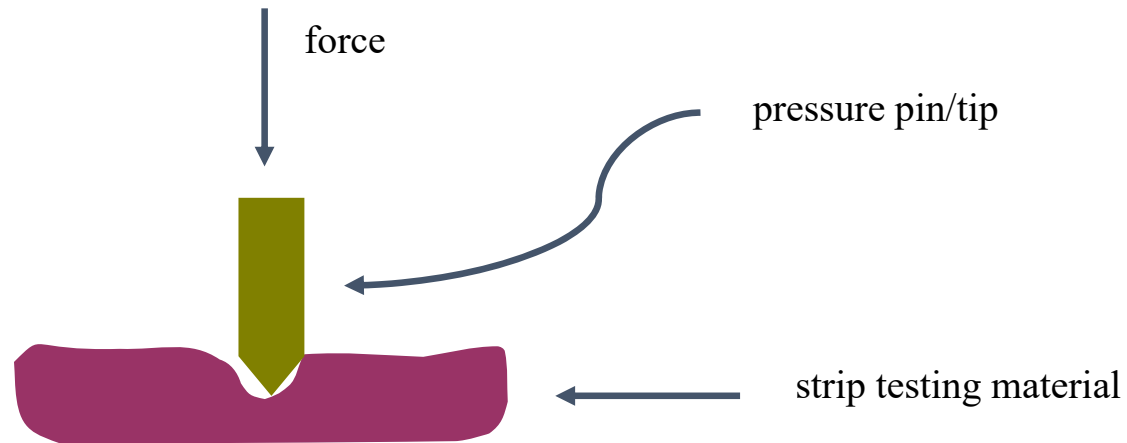
practical problem: 4 types of pressure pins
$\Rightarrow$ do these yield the same results?

Experimental design 1
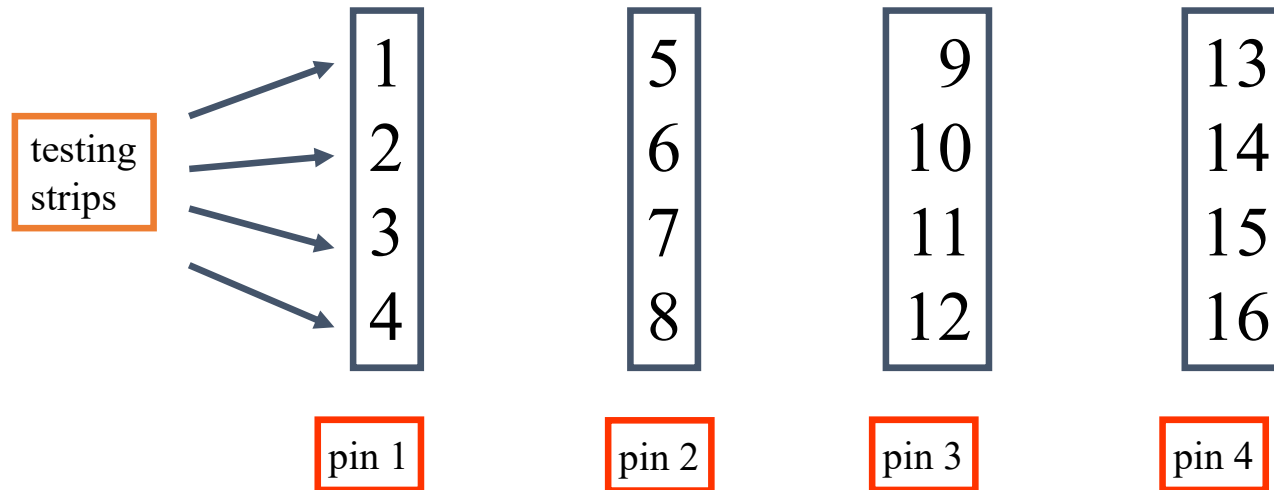
Problem: if the measurements of strips 5 through 8 differ, is this caused by the strips or by pin 2?

Experimental design 2

•Take 4 strips on which you measure (in random order)

each pressure pin once :
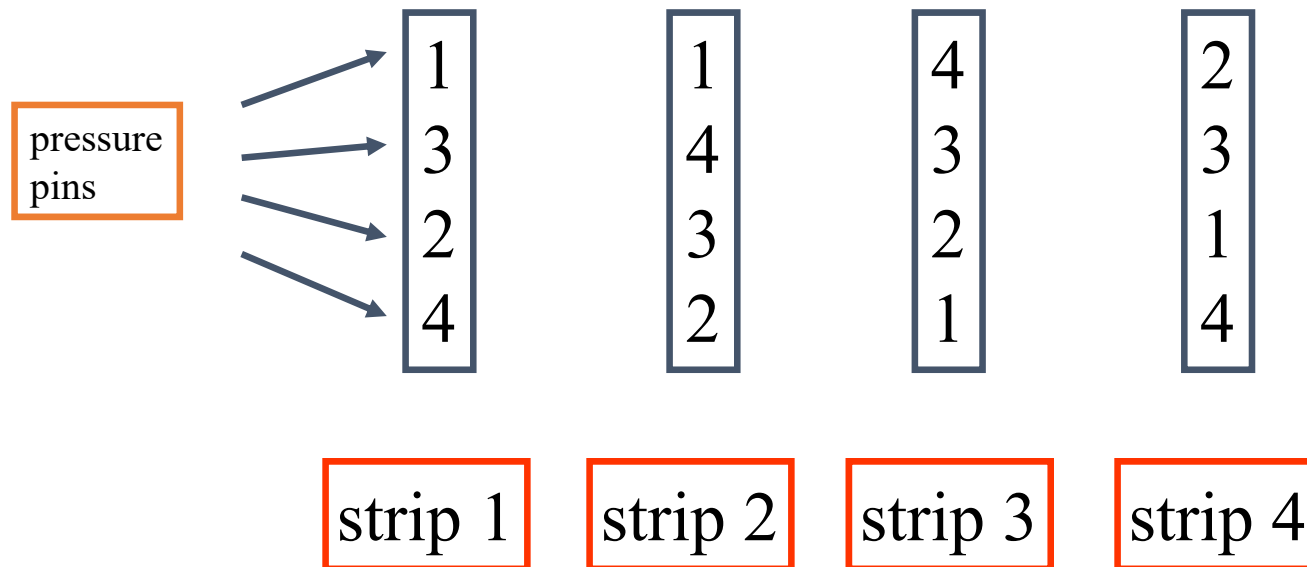
Blocking

- *Advantage* of *blocked* experimental design 2:

  differences between strips are filtered out

  - *Model*: $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$

    factor
    pressure pin

    block effect
    strip

    error term

  Primary goal: reduction error term

Short checklist for DOE (see protocol)

- clearly state objective of experiment
- check constraints on experiment
  - constraints on factor combinations and/or changes
  - constraints on size of experiment
- make sure that measurements are obtained under constant external conditions (if not, apply blocking!)
- include centre points to validate model assumptions
  - check of constant variance
  - check of non-linearity
- make clear protocol of execution of experiment (including randomised order of measurements)

Introduction: What is meant by DOE?

- Experiment -
  - a test or a series of tests in which purposeful changes are made to the *input variables or factors* of a system so that we may observe and identify the reasons for changes in the *output* response(s).

- Question: 5 factors, and 2 response variables
  - Want to know the effect of each factor on the response and how the factors may interact with each other
  - Want to predict the responses for given levels of the factors
  - Want to find the levels of the factors that optimizes the responses - e.g. maximize Y_1 but minimize Y_2
  - Time and budget allocated for 30 test runs only.

Strategy of Experimentation

- Strategy of experimentation
  - Best guess approach (trial and error)
    - can continue indefinitely
    - cannot guarantee best solution has been found
  - One-factor-at-a-time (OFAT) approach
    - inefficient (requires many test runs)
    - fails to consider any possible interaction between factors
  - Factorial approach (invented in the 1920's)
    - Factors varied together
    - Correct, modern, and most efficient approach
    - Can determine how factors interact
    - Used extensively in industrial R and D, and for process improvement.

- This course will focus on three very useful and important classes of factorial designs:
  - 2-level full factorial (2^k)
  - fractional factorial (2^{k-p}), and
  - response surface methodology (RSM)
- I will also cover split plot designs, and design and analysis of computer experiments if time permits.
- Dimensional analysis and how it can be combined with DOE will also be briefly covered.
- All DOE are based on the same statistical principles and method of analysis - ANOVA and regression analysis.
- *Answer to question: use a 2^{5-1} fractional factorial in a central composite design = 27 runs (min)*

Statistical Design of Experiments

- All experiments should be designed experiments

- Unfortunately, some experiments are poorly designed - valuable resources are used ineffectively and results inconclusive

- Statistically designed experiments permit efficiency and economy, and the use of statistical methods in examining the data result in scientific objectivity when drawing conclusions.
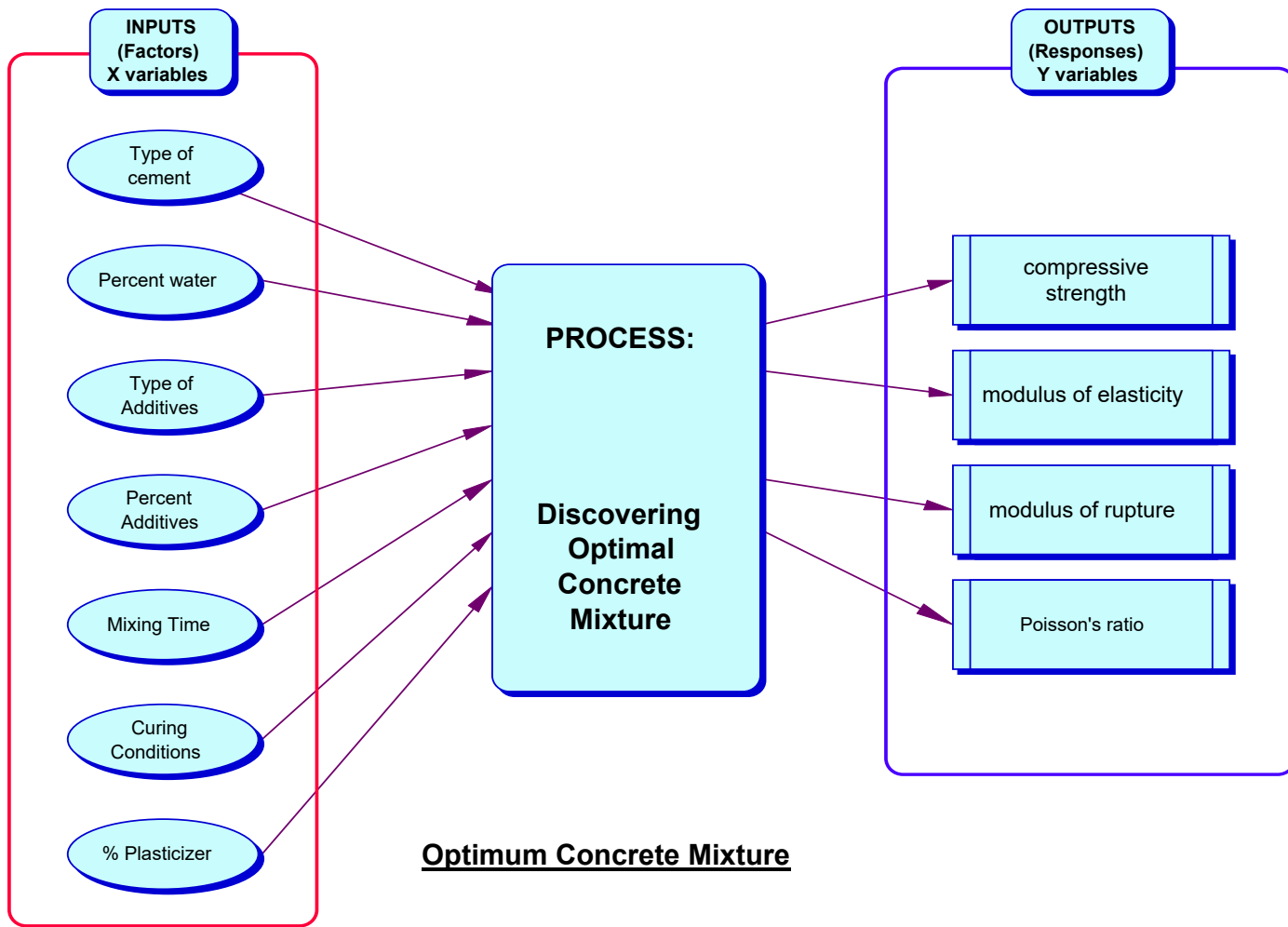
- DOE is a methodology for systematically applying statistics to experimentation.

- DOE lets experimenters develop a mathematical model that predicts how input variables *interact* to create output variables or responses in a process or system.

- DOE can be used for a wide range of experiments for various purposes including nearly all fields of engineering and even in business marketing.

- Use of statistics is very important in DOE and the basics are covered in a first course in an engineering program.
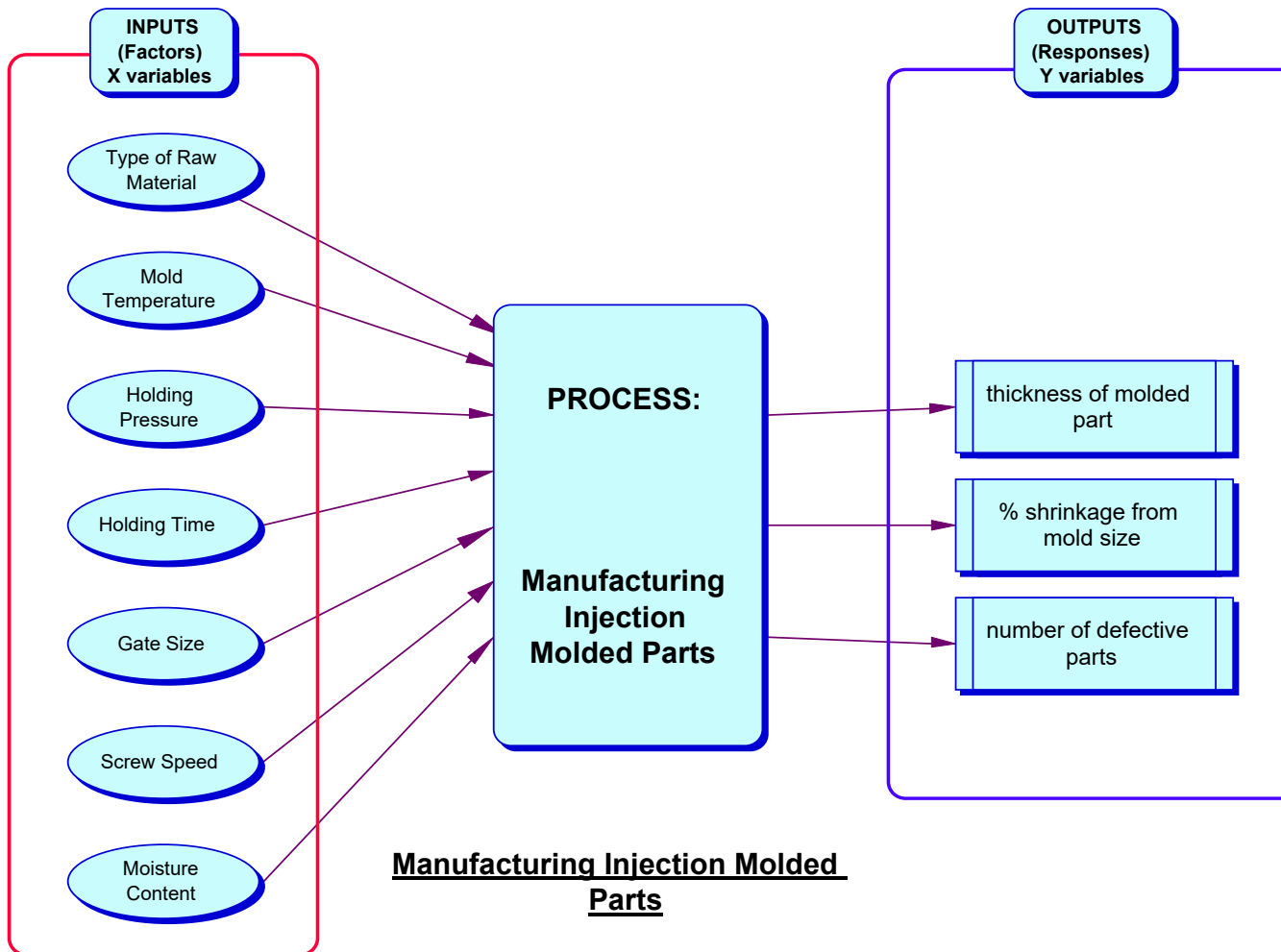
- In general, by using DOE, we can:
    - Learn about the process we are investigating
    - Screen important variables
    - Build a mathematical model
    - Obtain prediction equations
    - Optimize the response (if required)

- Statistical significance is tested using **ANOVA**, and the prediction model is obtained using **regression analysis.**
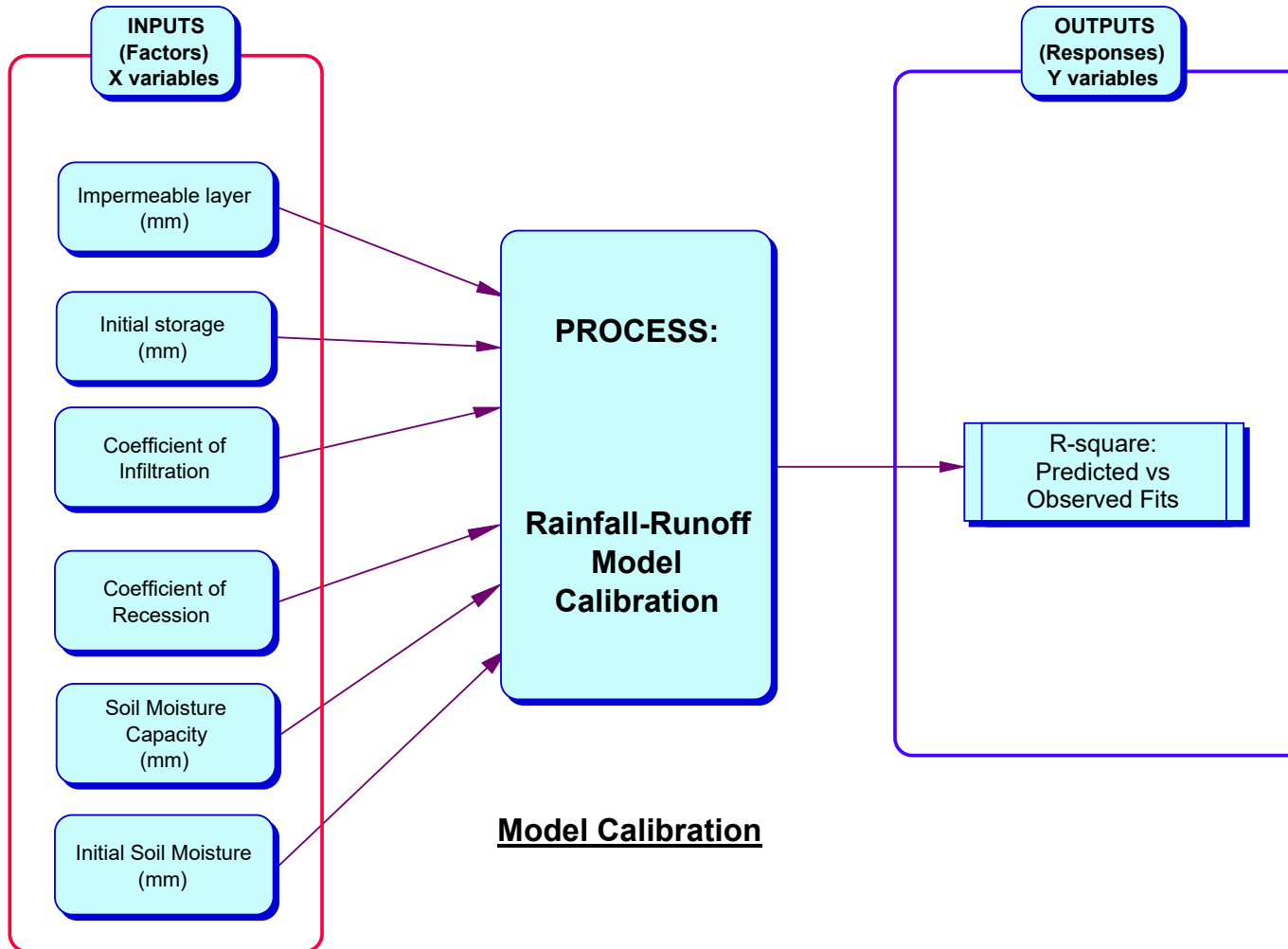
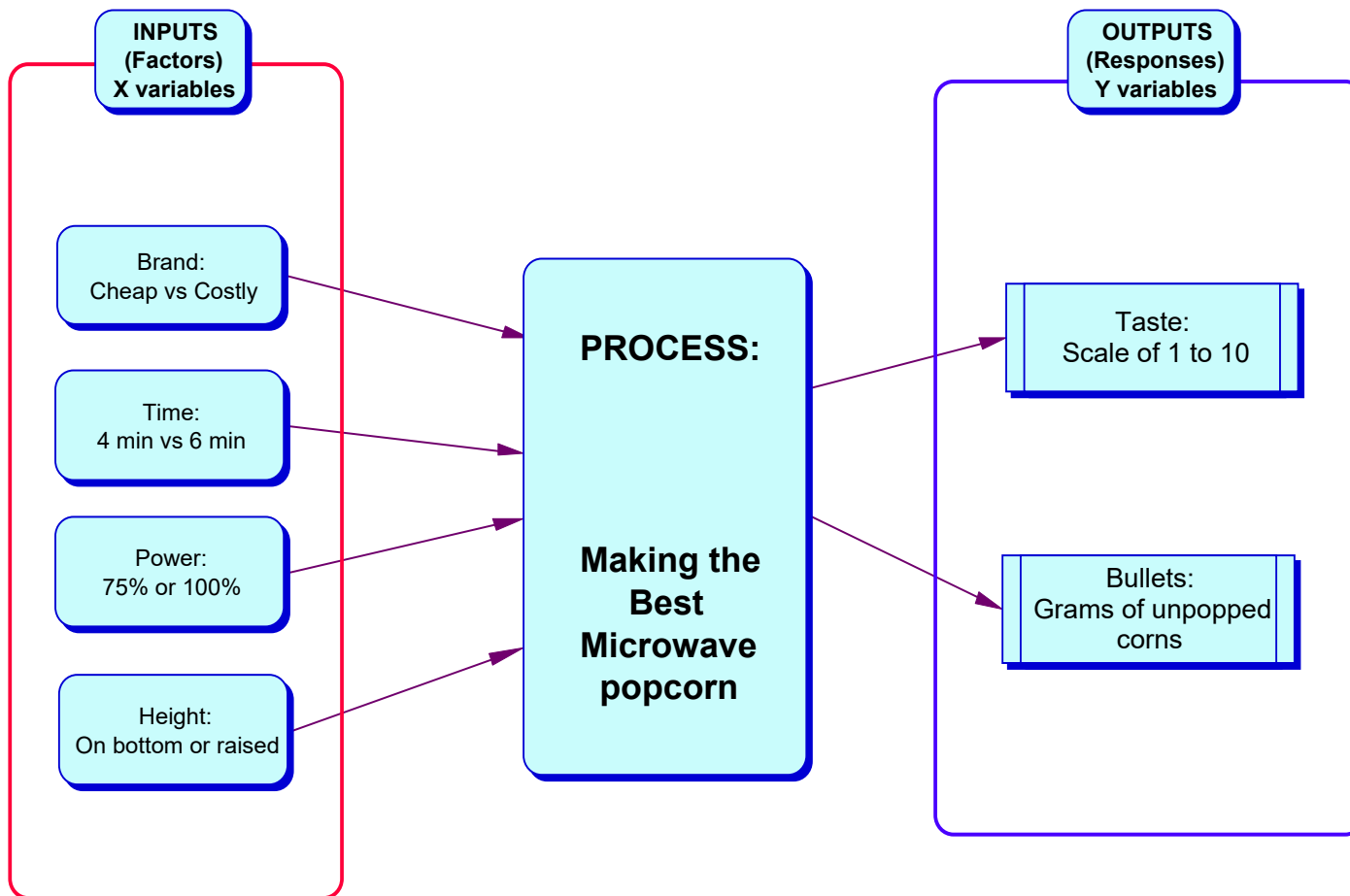# Applications of DOE in Engineering Design

- Experiments are conducted in the field of engineering to:
  - evaluate and compare basic design configurations
  - evaluate different materials
  - select design parameters so that the design will work well under a wide variety of field conditions (robust design)
  - determine key design parameters that impact performance

People

Materials

Equipment

Policies

Procedures

Methods

Environment

PROCESS:

A Blending of Inputs which Generates Corresponding Outputs

responses related to performing a service

responses related to producing a produce

responses related to completing a task

Illustration of a Process

INPUTS
(Factors)
X variables

- Type of cement
- Percent water
- Type of Additives
- Percent Additives
- Mixing Time
- Curing Conditions
- % Plasticizer

PROCESS:

Discovering Optimal Concrete Mixture

OUTPUTS
(Responses)
Y variables

- compressive strength
- modulus of elasticity
- modulus of rupture
- Poisson's ratio

Optimum Concrete Mixture

INPUTS
(Factors)
X variables

- Type of Raw Material
- Mold Temperature
- Holding Pressure
- Holding Time
- Gate Size
- Screw Speed
- Moisture Content

PROCESS:

Manufacturing Injection Molded Parts

OUTPUTS
(Responses)
Y variables

- thickness of molded part
- % shrinkage from mold size
- number of defective parts

Manufacturing Injection Molded Parts

INPUTS
(Factors)
X variables

Impermeable layer
(mm)

Initial storage
(mm)

Coefficient of
Infiltration

Coefficient of
Recession

Soil Moisture
Capacity
(mm)

Initial Soil Moisture
(mm)

PROCESS:

Rainfall-Runoff
Model
Calibration

Model Calibration

OUTPUTS
(Responses)
Y variables

R-square:
Predicted vs
Observed Fits

INPUTS
(Factors)
X variables

Brand:
Cheap vs Costly

Time:
4 min vs 6 min

Power:
75% or 100%

Height:
On bottom or raised

PROCESS:

**Making the
Best
Microwave
popcorn**

OUTPUTS
(Responses)
Y variables

Taste:
Scale of 1 to 10

Bullets:
Grams of unpopped
corns

Making microwave popcorn

Examples of experiments from daily life

- Photography
  - Factors: speed of film, lighting, shutter speed
  - Response: quality of slides made close up with flash attachment

- Boiling water
  - Factors: Pan type, burner size, cover
  - Response: Time to boil water

- D-day
  - Factors: Type of drink, number of drinks, rate of drinking, time after last meal
  - Response: Time to get a steel ball through a maze

- Mailing
  - Factors: stamp, area code, time of day when letter mailed
  - Response: Number of days required for letter to be delivered

More examples

- Cooking
  - Factors: amount of cooking wine, oyster sauce, sesame oil
  - Response: Taste of stewed chicken

- Sexual Pleasure
  - Factors: marijuana, screech, sauna
  - Response: Pleasure experienced in subsequent you know what

- Basketball
  - Factors: Distance from basket, type of shot, location on floor
  - Response: Number of shots made (out of 10) with basketball

- Skiing
  - Factors: Ski type, temperature, type of wax
  - Response: Time to go down ski slope

Basic Principles

- Statistical design of experiments (DOE)
  - the process of planning experiments so that appropriate data can be analyzed by statistical methods that results in valid, objective, and meaningful conclusions from the data
  - involves two aspects: design and statistical analysis

- Every experiment involves a sequence of activities:
  - Conjecture - hypothesis that motivates the experiment
  - Experiment - the test performed to investigate the conjecture
  - Analysis - the statistical analysis of the data from the experiment
  - Conclusion - what has been learned about the original conjecture from the experiment.

Three basic principles of Statistical DOE

- Replication
  - allows an estimate of experimental error
  - allows for a more precise estimate of the sample mean value

- Randomization
  - cornerstone of all statistical methods
  - "average out" effects of extraneous factors
  - reduce bias and systematic errors

- Blocking
  - increases precision of experiment
  - "factor out" variable not studied

Guidelines for Designing Experiments

- Recognition of and statement of the problem
  - need to develop all ideas about the objectives of the experiment - get input from everybody - use team approach.

- Choice of factors, levels, ranges, and response variables.
  - Need to use engineering judgment or prior test results.

- Choice of experimental design
  - sample size, replicates, run order, randomization, software to use, design of data collection forms.

- Performing the experiment
  - vital to monitor the process carefully. Easy to underestimate logistical and planning aspects in a complex R and D environment.
- Statistical analysis of data
  - provides objective conclusions - use simple graphics whenever possible.
- Conclusion and recommendations
  - follow-up test runs and confirmation testing to validate the conclusions from the experiment.
- Do we need to add or drop factors, change ranges, levels, new responses, etc.. ???

Using Statistical Techniques in Experimentation - things to keep in mind

- Use non-statistical knowledge of the problem
  - physical laws, background knowledge
- Keep the design and analysis as simple as possible
  - Don't use complex, sophisticated statistical techniques
  - If design is good, analysis is relatively straightforward
  - If design is bad - even the most complex and elegant statistics cannot save the situation
- Recognize the difference between practical and statistical significance
  - statistical significance $\neq$ practically significance

- Experiments are usually iterative
  - unwise to design a comprehensive experiment at the start of the study
  - may need modification of factor levels, factors, responses, etc.. - too early to know whether experiment would work
  - use a sequential or iterative approach
  - should not invest more than 25% of resources in the initial design.
  - Use initial design as learning experiences to accomplish the final objectives of the experiment.

Factorial v.s. OFAT

- Factorial design - experimental trials or runs are performed at all possible combinations of factor levels in contrast to OFAT experiments.

- Factorial and fractional factorial experiments are among the most useful multi-factor experiments for engineering and scientific investigations.

- The ability to gain competitive advantage requires extreme care in the design and conduct of experiments. Special attention must be paid to joint effects and estimates of variability that are provided by factorial experiments.

- Full and fractional experiments can be conducted using a variety of statistical designs. The design selected can be chosen according to specific requirements and restrictions of the investigation.

Factorial Designs

- In a factorial experiment, **all possible combinations** of factor levels are tested
- The golf experiment:
  - Type of driver (over or regular)
  - Type of ball (balata or 3-piece)
  - Walking vs. riding a cart
  - Type of beverage (Beer vs water)
  - Time of round (am or pm)
  - Weather
  - Type of golf spike
  - Etc, etc, etc…
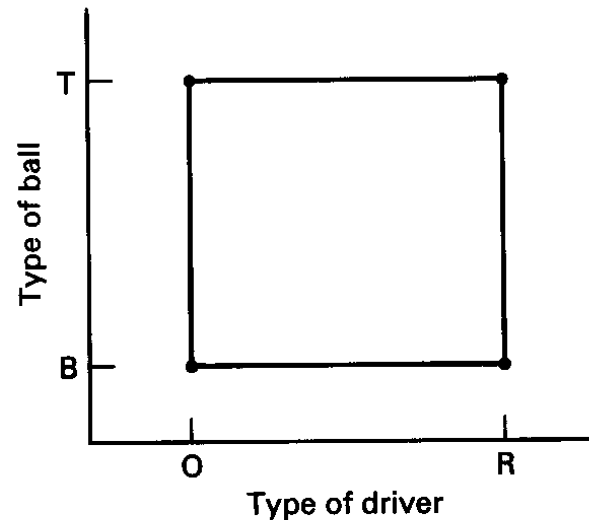
A two-factor factorial experiment involving type of driver and type of ball.

Factorial Design
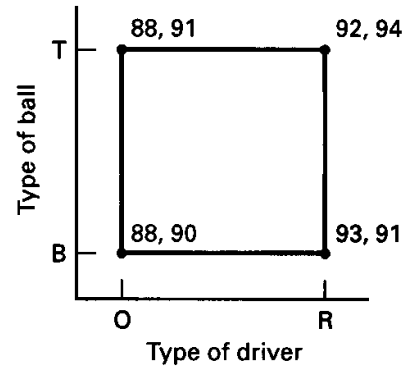
(a) Scores from the golf experiment
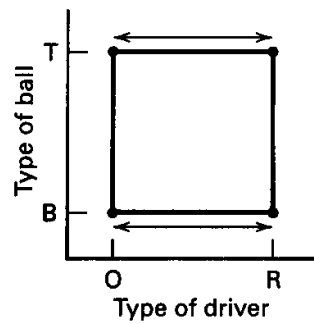
(b) Comparison of scores leading to the driver effect
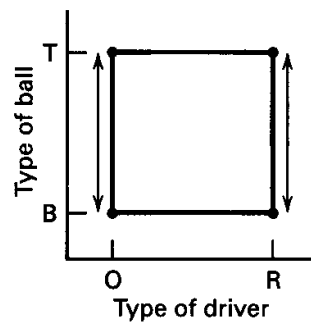
(c) Comparison of scores leading to the ball effect
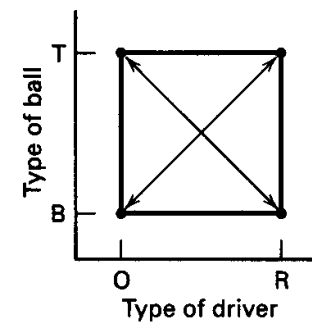
(d) Comparison of scores leading to the ball-driver interaction effect

Figure 1-5 Scores from the golf experiment in Figure 1-4 and calculation of the factor effects.
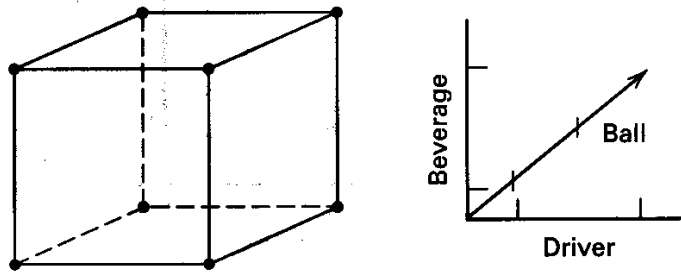
Factorial Designs with Several Factors

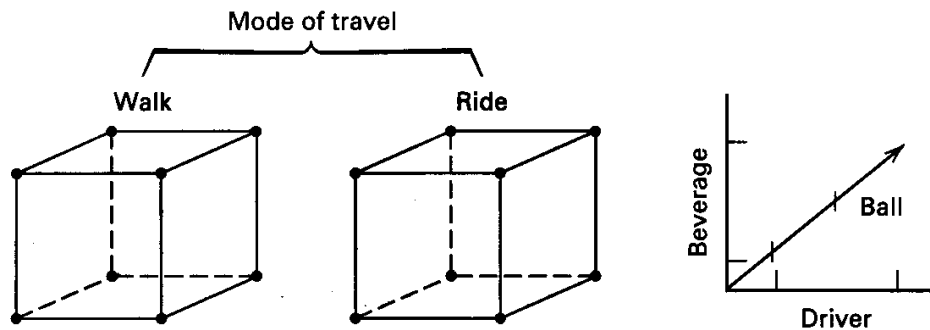**Figure 1-6** A three-factor factorial experiment involving type of driver, type of ball, and type of beverage.

Figure 1-7 A four-factor factorial experiment involving type of driver, type of ball, type of beverage, and mode of travel.

Erroneous Impressions About Factorial Experiments

- Wasteful and do not compensate the extra effort with additional useful information - this folklore presumes that one knows (not assumes) that factors independently influence the responses (i.e. there are no factor interactions) and that each factor has a linear effect on the response - almost any reasonable type of experimentation will identify optimum levels of the factors

- Information on the factor effects becomes available only after the entire experiment is completed. Takes too long. Actually, factorial experiments can be blocked and conducted sequentially so that data from each block can be analyzed as they are obtained.

One-factor-at-a-time experiments (OFAT)

- OFAT is a prevalent, but potentially disastrous type of experimentation commonly used by many engineers and scientists in both industry and academia.

- Tests are conducted by systematically changing the levels of one factor while holding the levels of all other factors fixed. The "optimal" level of the first factor is then selected.

- Subsequently, each factor in turn is varied and its "optimal" level selected while the other factors are held fixed.

One-factor-at-a-time experiments (OFAT)

- OFAT experiments are regarded as easier to implement, more easily understood, and more economical than factorial experiments. Better than trial and error.

- OFAT experiments are believed to provide the optimum combinations of the factor levels.

- Unfortunately, each of these presumptions can generally be shown to be false except under very special circumstances.

- The key reasons why OFAT should not be conducted except under very special circumstances are:
    - *Do not provide adequate information on interactions*
    - *Do not provide efficient estimates of the effects*

Factorial vs OFAT (2-levels only)
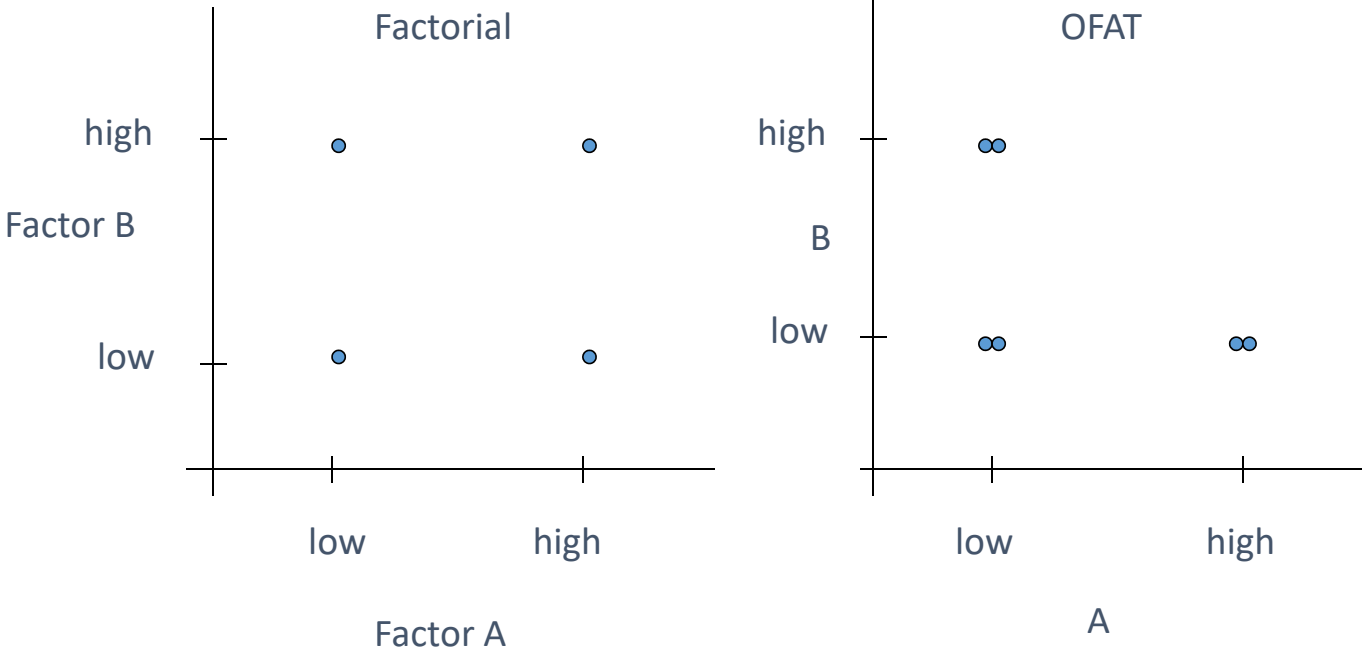
Factorial

- 2 factors: 4 runs
  - 3 effects
- 3 factors: 8 runs
  - 7 effects
- 5 factors: 32 or 16 runs
  - 31 or 15 effects
- 7 factors: 128 or 64 runs
  - 127 or 63 effects

OFAT

- 2 factors: 6 runs
  - 2 effects
- 3 factors: 16 runs
  - 3 effects
- 5 factors: 96 runs
  - 5 effects
- 7 factors: 512 runs
  - 7 effects

Example: Factorial vs OFAT

E.g. Factor A: Reynold's number, Factor B: k/D

Example: Effect of Re and k/D on friction factor f

- Consider a 2-level factorial design (2^2)
- Reynold's number = Factor A; k/D = Factor B
- Levels for A: 10^4 (low) 10^6 (high)
- Levels for B: 0.0001 (low)   0.001 (high)
- Responses: (1) = 0.0311,   a = 0.0135, b = 0.0327, ab = 0.0200
- Effect (A) = -0.66, Effect (B) = 0.22, Effect (AB) = 0.17
- % contribution: A = 84.85%, B = 9.48%, AB = 5.67%
- The presence of interactions implies that one cannot satisfactorily describe the effects of each factor using main effects.
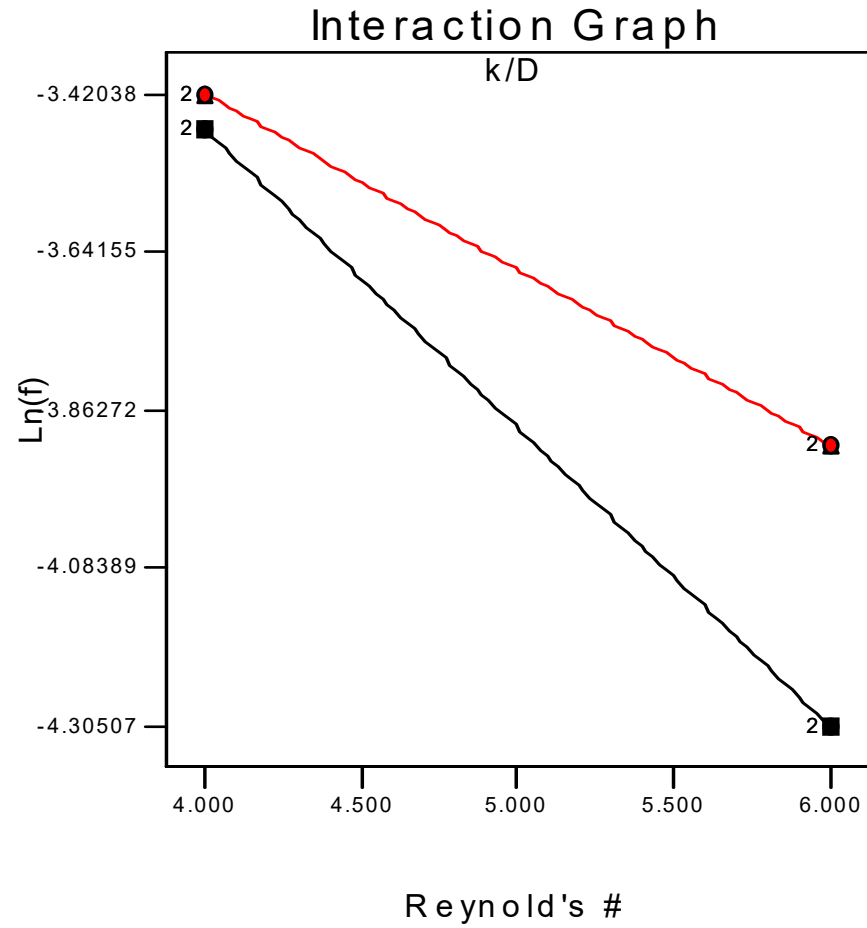
DESIGN-EASE Plot

Ln(f)

X = A: Reynold's #
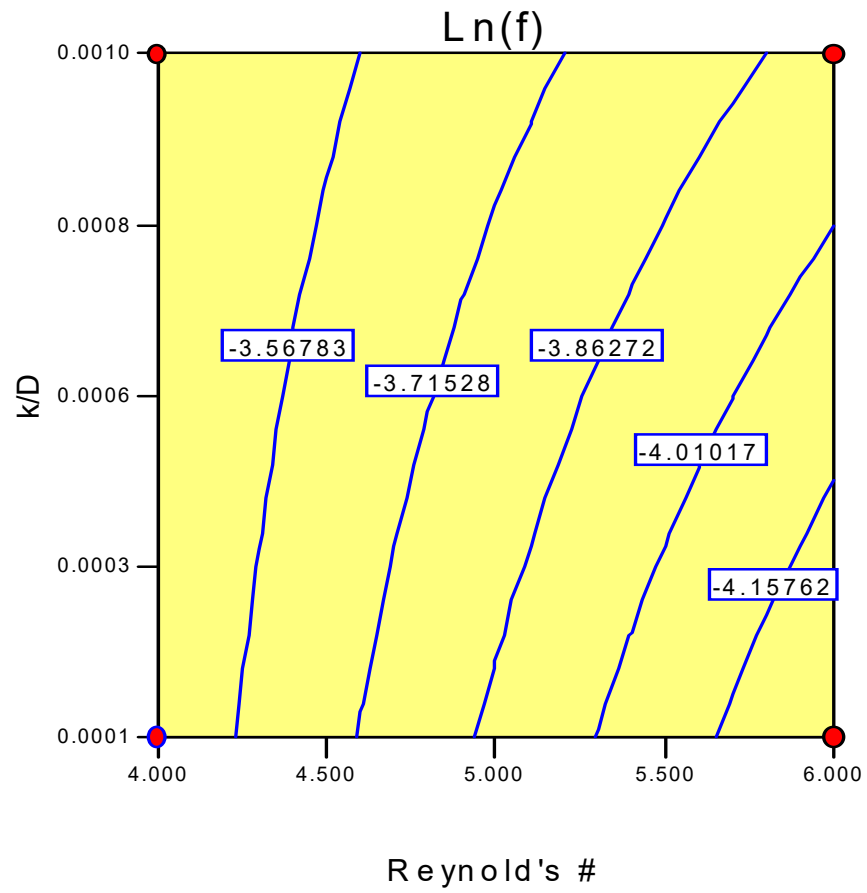Y = B: k/D

● Design Points

■ B- 0.000
▲ B+ 0.001

Interaction Graph
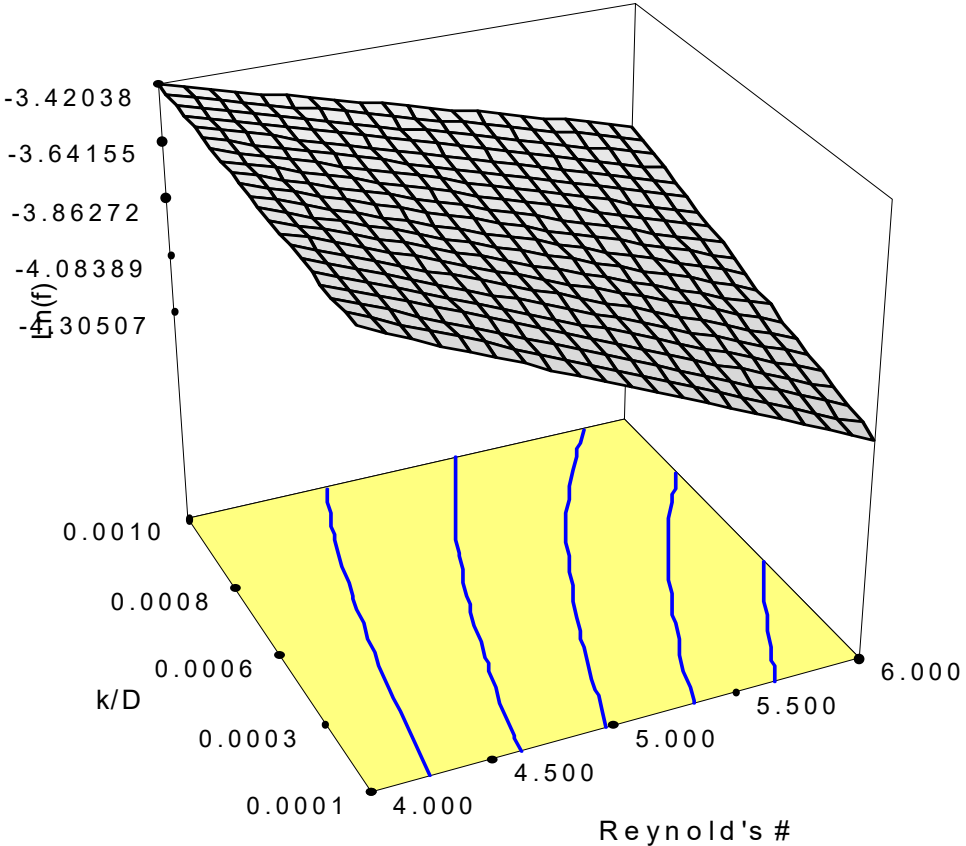
k/D

Ln(f)

Reynold's #

DESIGN-EASE Plot

Ln(f)
X = A: Reynold's #
Y = B: k/D

● Design Points

Ln(f)

k/D

-3.56783
-3.71528
-3.86272
-4.01017
-4.15762

0.0010
0.0008
0.0006
0.0003
0.0001

4.000   4.500   5.000   5.500   6.000

Reynold's #

DESIGN-EASE Plot

Ln(f)
X = A: Reynold's #
Y = B: k/D

With the addition of a few more points

- Augmenting the basic 2^2 design with a center point and 5 axial points we get a central composite design (CCD) and a 2nd order model can be fit.

- The nonlinear nature of the relationship between Re, k/D and the friction factor f can be seen.

- If Nikuradse (1933) had used a factorial design in his pipe friction experiments, he would need far less experimental runs!!

- If the number of factors can be reduced by dimensional analysis, the problem can be made simpler for experimentation.
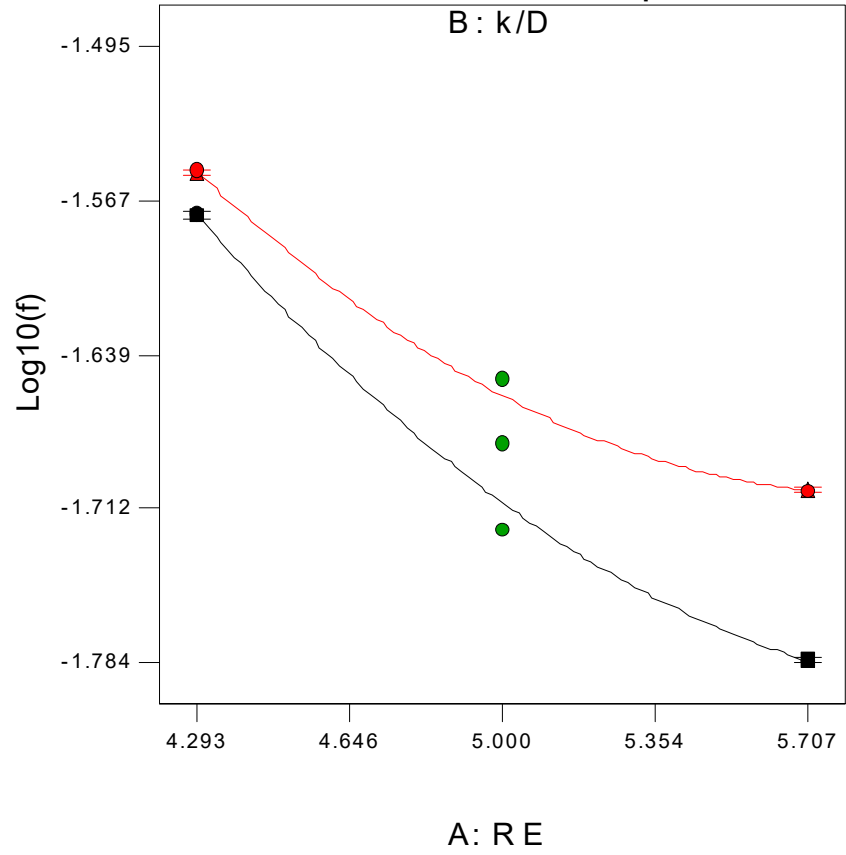
DESIGN-EXPERT Plot

Log10(f)

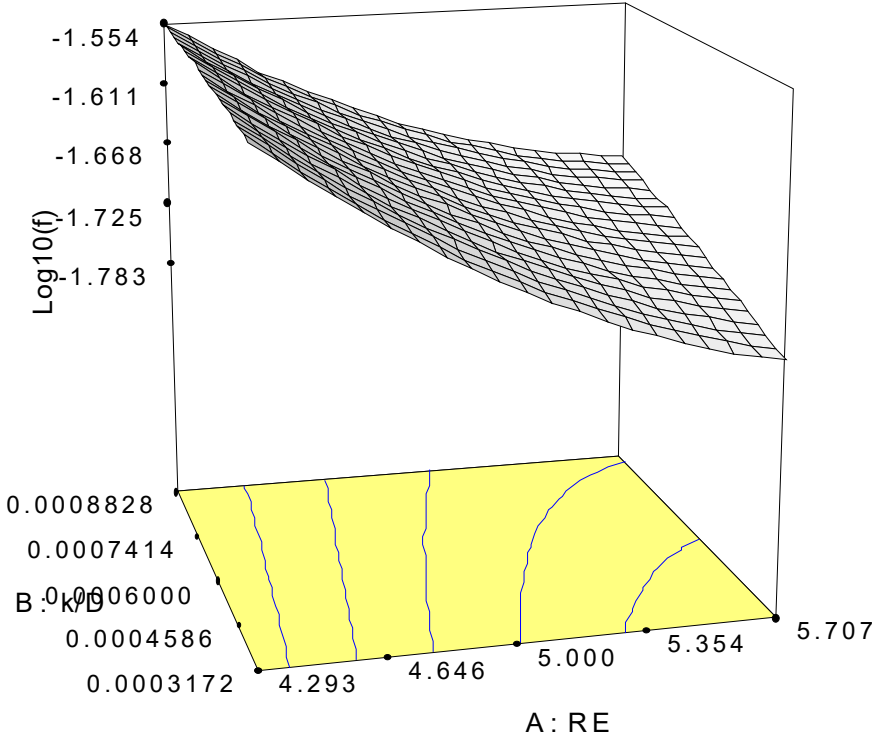X = A: RE
Y = B: k/D

● Design Points

■ B- 0.000
▲ B+ 0.001

Interaction Graph

B: k/D

Log10(f)

DESIGN-EXPERT Plot
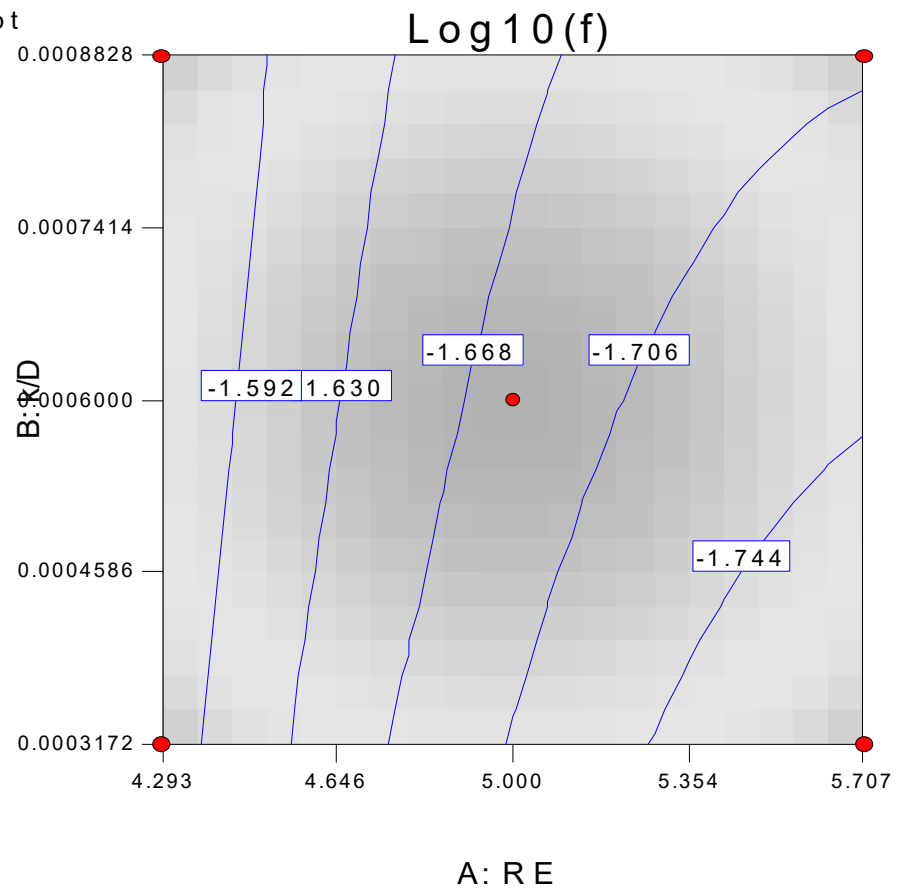
Log10(f)
X = A: RE
Y = B: k/D

Log10(f)

-1.554
-1.611
-1.668
-1.725
-1.783

0.0008828
0.0007414
0.0006000
0.0004586
0.0003172

B: k/D

4.293
4.646
5.000
5.354
5.707

A: RE

DESIGN-EXPERT Plot

Log10(f)
● Design Points

X = A: RE
Y = B: k/D

DESIGN-EXPERT Plot
Log10(f)

Predicted vs. Actual

Predicted (y-axis): -1.494, -1.566, -1.639, -1.711, -1.783

Actual (x-axis): -1.783, -1.711, -1.639, -1.566, -1.494
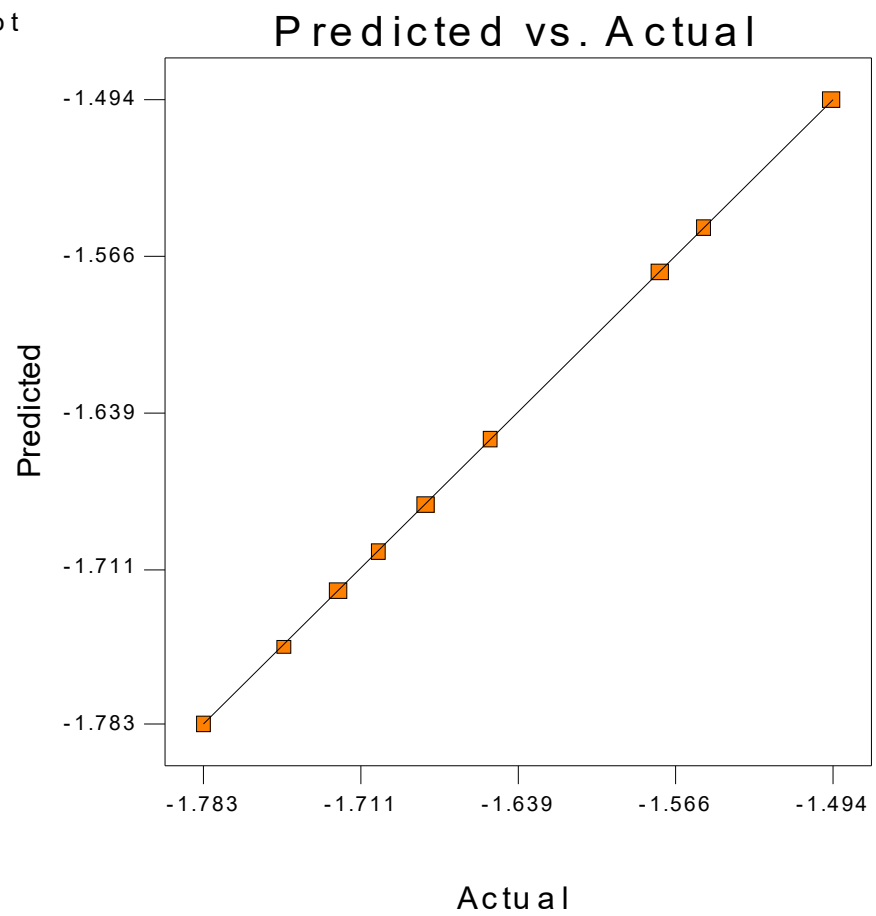
FACTORIAL (2^k) DESIGNS

- Experiments involving several factors (k = # of factors) where it is necessary to study the joint effect of these factors on a specific response.

- Each of the factors are set at two levels (a "low" level and a "high" level) which may be qualitative (machine A/machine B, fan on/fan off) or quantitative (temperature 80^0/temperature 90^0, line speed 4000 per hour/line speed 5000 per hour).

FACTORIAL (2^k) DESIGNS

- Factors are assumed to be fixed (fixed effects model)

- Designs are completely randomized (experimental trials are run in a random order, etc.)

- The usual normality assumptions are satisfied.

FACTORIAL (2^k) DESIGNS

- Particularly useful in the early stages of experimental work when you are likely to have many factors being investigated and you want to minimize the number of treatment combinations (sample size) but, at the same time, study all k factors in a complete factorial arrangement (the experiment collects data at all possible combinations of factor levels).

FACTORIAL (2^k) DESIGNS

- As k gets large, the sample size will increase exponentially. If experiment is replicated, the # runs again increases.

k	# of runs
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	1024

FACTORIAL (2^k) DESIGNS (k = 2)

- Two factors set at two levels (normally referred to as low and high) would result in the following design where each level of factor A is paired with each level of factor B.

Generalized Settings			
RUN	Factor A	Factor B	RESPONSE
1	low	low	y_1
2	high	low	y_2
3	low	high	y_3
4	high	high	y_4

| Orthogonal Settings			
RUN	Factor A	Factor B	RESPONSE
1	-1	-1	y_1
2	+1	-1	y_2
3	-1	+1	y_3
4	+1	+1	y_4

FACTORIAL (2^k) DESIGNS (k = 2)

- Estimating main effects associated with changing the level of each factor from low to high. This is the estimated effect on the response variable associated with changing factor A or B from their low to high values.

$$Factor\ A\ Effect = \frac{(y_2 + y_4)}{2} - \frac{(y_1 + y_3)}{2}$$

$$Factor\ B\ Effect = \frac{(y_3 + y_4)}{2} - \frac{(y_1 + y_2)}{2}$$

FACTORIAL (2^k) DESIGNS (k = 2): GRAPHICAL OUTPUT

- Neither factor A nor Factor B have an effect on the response variable.
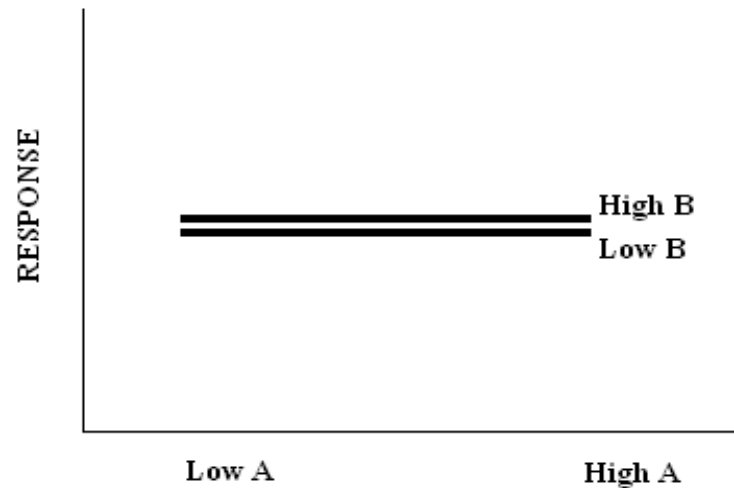
FACTORIAL (2^k) DESIGNS (k = 2): GRAPHICAL OUTPUT

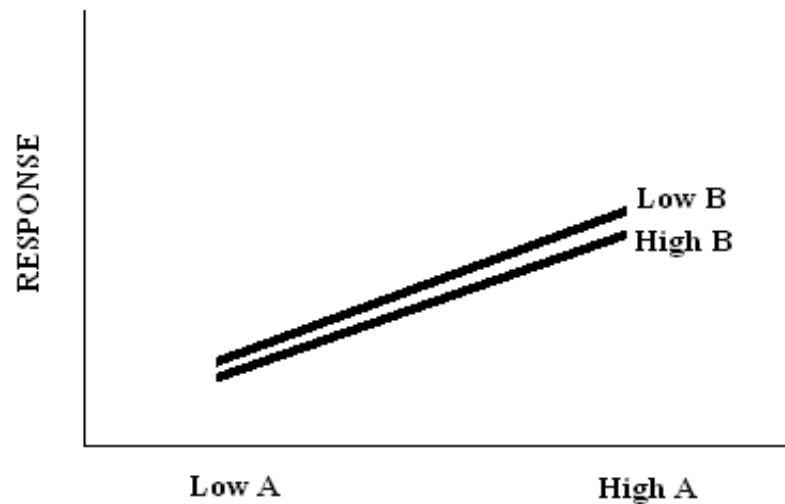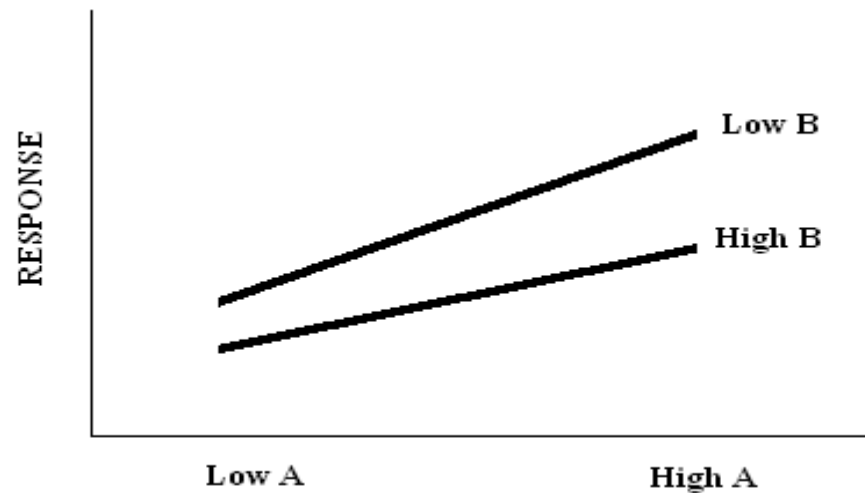- Factor A has an effect on the response variable, but Factor B does not.

FACTORIAL (2^k) DESIGNS (k = 2): GRAPHICAL OUTPUT

- Factor A and Factor B have an effect on the response variable.

FACTORIAL (2^k) DESIGNS (k = 2): GRAPHICAL OUTPUT

- Factor B has an effect on the response variable, but only if factor A is set at the "High" level. This is **called interaction** and it basically means that the effect one factor has on a response is dependent on the level you set other factors at. Interactions can be major problems in a DOE if you fail to account for the interaction when designing your experiment.
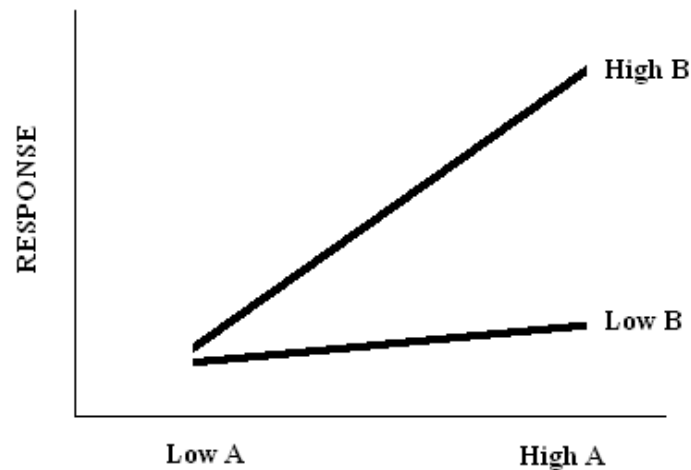
EXAMPLE:
FACTORIAL (2^k) DESIGNS (k = 2)

- A microbiologist is interested in the effect of two different culture mediums [medium 1 (low) and medium 2 (high)] and two different times [10 hours (low) and 20 hours (high)] on the growth rate of a particular CFU [Bugs].

EXAMPLE:
FACTORIAL (2^k) DESIGNS (k = 2)

- Since two factors are of interest, k =2, and we would need the following four runs resulting in

Generalized Settings			
RUN	Medium	Time	Growth Rate
1	low	low	17
2	high	low	15
3	low	high	38
4	high	high	39

EXAMPLE:
FACTORIAL (2^k) DESIGNS (k = 2)

- Estimates for the medium and time effects are

- Medium effect = [(15+39)/2] – [(17 + 38)/2] = -0.5

- Time effect = [(38+39)/2] – [(17 + 15)/2] = 22.5

EXAMPLE:
FACTORIAL (2^k) DESIGNS (k = 2)

Plot of Effects
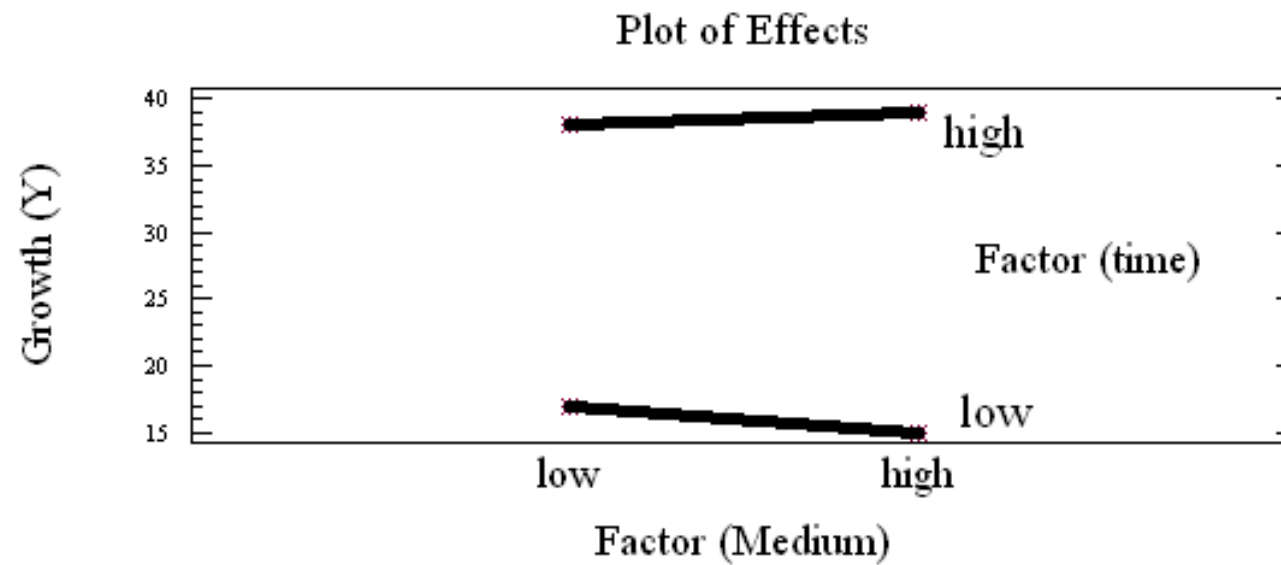
EXAMPLE:
FACTORIAL (2^k) DESIGNS (k = 2)

- A statistical analysis using the appropriate statistical model would result in the following information. Factor A (medium) and Factor B (time)

```
Type III Sums of Squares
--------------------------------------------------------------------------------
Source                          Sum of Squares    Df   Mean Square    F-Ratio    P-Value
--------------------------------------------------------------------------------
FACTOR A                                  0.25     1          0.25       0.11     0.7952
FACTOR B                                506.25     1        506.25     225.00     0.0424
Residual                                  2.25     1          2.25
--------------------------------------------------------------------------------
Total (corrected)                       508.75     3
All F-ratios are based on the residual mean square error.
```

EXAMPLE: CONCLUSIONS

- In statistical language, one would conclude that factor A (medium) is not statistically significant at a 5% level of significance since the p-value is greater than 5% (0.05), but factor B (time) is statistically significant at a 5 % level of significance since this p-value is less than 5%.

EXAMPLE: CONCLUSIONS

- In layman terms, this means that we have no evidence that would allow us to conclude that the medium used has an effect on the growth rate, although it may well have an effect (our conclusion was incorrect).

EXAMPLE: CONCLUSIONS

- Additionally, we have evidence that would allow us to conclude that time does have an effect on the growth rate, although it may well not have an effect (our conclusion was incorrect).

EXAMPLE: CONCLUSIONS

- In general we control the likelihood of reaching these incorrect conclusions by the selection of the level of significance for the test and the amount of data collected (sample size).

2^k DESIGNS (k $\geq$ 2)

- As the number of factors increase, the number of runs needed to complete a complete factorial experiment will increase dramatically. The following 2^k design layout depict the number of runs needed for values of k from 2 to 5. For example, when k = 5, it will take 2^5 = 32 experimental runs for the complete factorial experiment.

Interactions for 2k Designs (k = 3)

- Interactions between various factors can be estimated for different designs above by multiplying the appropriate columns together and then subtracting the average response for the lows from the average response for the highs.

Interactions for 2k Designs (k = 3)

a	b	c	ab	ac	bc	abc
-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

2^k DESIGNS (k $\geq$ 2)

- Once the effect for all factors and interactions are determined, you are able to develop a prediction model to estimate the response for specific values of the factors. In general, we will do this with statistical software, but for these designs, you can do it by hand calculations if you wish.
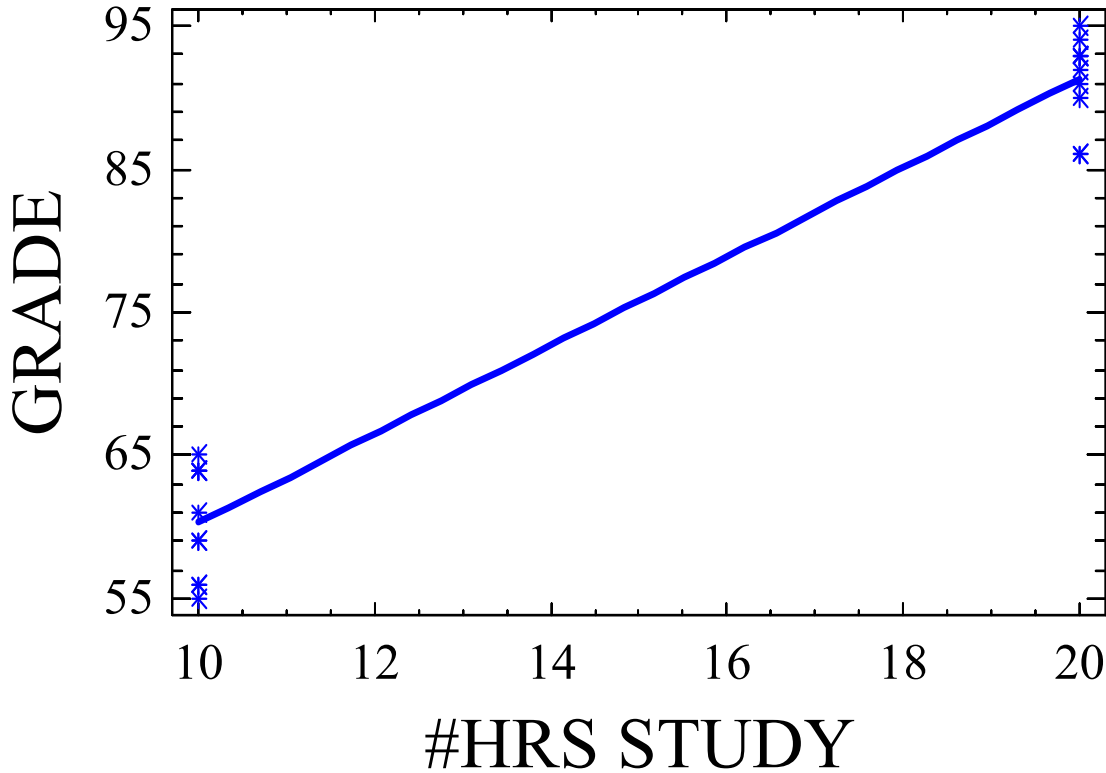
2^k DESIGNS (k $\geq$ 2)

- For example, if there are no significant interactions present, you can estimate a response by the following formula. (for quantitative factors only)

$$Y = (\text{average of all responses}) + \sum [(\frac{factorEFFECT}{2}) * (factorLEVEL)]$$

$$= \bar{\bar{Y}} + (\frac{\Delta_A}{2}) * A + (\frac{\Delta_B}{2}) * B$$

ONE FACTOR EXAMPLE

Plot of Fitted Model

ONE FACTOR EXAMPLE

- The output shows the results of fitting a general linear model to describe the relationship between GRADE and #HRS STUDY. The equation of the fitted general model is

- GRADE = 29.3 + 3.1* (#HRS STUDY)

- The fitted orthogonal model is

- GRADE = 75 + 15 * (SCALED # HRS)

Two Level Screening Designs

- Suppose that your brainstorming session resulted in 7 factors that various people think "might" have an effect on a response. A full factorial design would require $2^7 = 128$ experimental runs without replication. The purpose of screening designs is to reduce (identify) the number of factors down to the "major" role players with a minimal number of experimental runs. One way to do this is to use the 2^3 full factorial design and use interaction columns for factors.

Note that

* Any factor d effect is now confounded with the a*b interaction
* Any factor e effect is now confounded with the a*c interaction
* etc.
* What is the d*e interaction confounded with????????

| a	b	c	d = ab	e = ac	f = bc	g = abc
-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

Problems that Interactions Cause!

- Interactions – If interactions exist and you fail to account for this, you may reach erroneous conclusions. Suppose that you plan an experiment with four runs and three factors resulting in the following data:

Run	Factor A	Factor B	Results
1	+1	+1	10
2	+1	-1	5
3	-1	+1	5
4	-1	-1	10

Problems that Interactions Cause!

- Factor A Effect = 0

- Factor B Effect = 0

- In this example, if you were assuming that "smaller is better" then it appears to make no difference where you set factors A and B. If you were to set factor A at the low value and factor B at the low value, your response variable would be larger than desired. In this case there is a factor A interaction with factor B.

Problems that Interactions Cause!
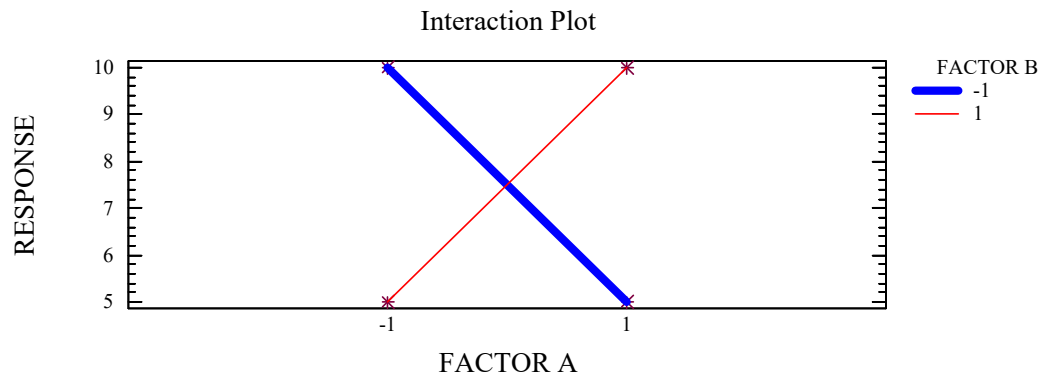
Interaction Plot

Resolution of a Design

- Resolution III Designs – No main effects are aliased with any other main effect BUT some (or all) main effects are aliased with two way interactions

- Resolution IV Designs – No main effects are aliased with any other main effect OR two factor interaction, BUT two factor interactions may be aliased with other two factor interactions

- Resolution V Designs – No main effect OR two factor interaction is aliased with any other main effect or two factor interaction, BUT two factor interactions are aliased with three factor interactions.

Common Screening Designs

- Fractional Factorial Designs – the total number of experimental runs must be a power of 2 (4, 8, 16, 32, 64, ...). If you believe first order interactions are small compared to main effects, then you could choose a resolution III design. Just remember that if you have major interactions, it can mess up your screening experiment.

Common Screening Designs

- Plackett-Burman Designs – Two level, resolution III designs used to study up to n-1 factors in n experimental runs, where n is a multiple of 4 (# of runs will be 4, 8, 12, 16, …).   Since n may be quite large, you can study a large number of factors with moderately small sample sizes. (n = 100 means you can study 99 factors with 100 runs)

Other Design Issues

- May want to collect data at center points to estimate non-linear responses

- More than two levels of a factor – no problem (multi-level factorial)

- What do you do if you want to build a non-linear model to "optimize" the response. (hit a target, maximize, or minimize) – called response surface modeling

Response Surface Designs – Box-Behnken

RUN	F1	F2	F3	Y_{100}
1	10	45	60	11825
2	30	45	40	8781
3	20	30	40	8413
4	10	30	50	9216
5	20	45	50	9288
6	30	60	50	8261
7	20	45	50	9329
8	30	45	60	10855
9	20	45	50	9205
10	20	60	40	8538
11	10	45	40	9718
12	30	30	50	11308
13	20	60	60	10316
14	10	60	50	12056
15	20	30	60	10378

Response Surface Designs – Box-Behnken

```
Regression coeffs. for Var_3
-----------------------------------------------------------
constant     = 2312.5
A:Factor_A   = 36.575
B:Factor_B   = 200.067
C:Factor_C   = 3.85
AA           = 9.09875
AB           = -9.81167
AC           = -0.0825
BB           = 0.117222
BC           = -0.311667
CC           = 1.10875
```

Response Surface Designs – Box-Behnken

Contours of Estimated Response Surface

Factor_C=60.0