Design of Experiments in Semiconductor Manufacturing

Costas J. Spanos

Department of Electrical Engineering
and Computer Sciences
University of California Berkeley, CA 94720,
U.S.A.

tel (510) 643 6776, fax (510) 642 2739
email spanos@eecs.berkeley.edu
http://bcam.eecs.berkeley.edu

Design of Experiments

- Comparison of Treatments
  - which recipe works the best?
- Simple Factorial Experiments
  - to explore impact of few variables
- Fractional Factorial Experiments
  - to explore impact of many variables
- Regression Analysis
  - to create analytical expressions that "model" process behavior
- Response Surface Methods
  - to visualize process performance over a range of input parameter values

Design of Experiments

- Objectives:
  - Compare Methods
  - Deduce Dependence
  - Create Models to Predict Effects
- Problems:
  - Experimental Error
  - Confusion of Correlation with Causation
  - Complexity of the Effects we study

Problems Solved

- Compare Recipes
  - Choose the recipe that gives the best results
  - Organize experiments to facilitate the analysis
  - Use experimental results to build process models
  - Use models to optimize the process

Comparison of Treatments

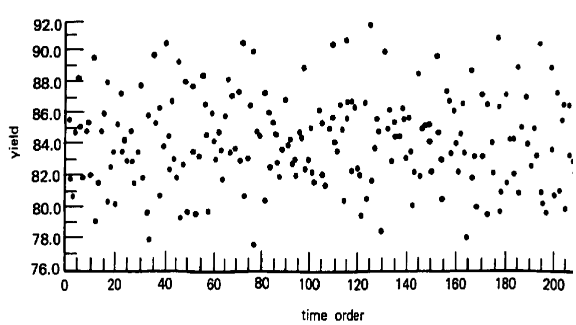
- Internal and External References
- The Importance of Independence
- Blocking and Randomization
- Analysis of Variance

The BIG Question in comparison of treatments:

- How does a process compare with other processes?
  - Is it the same?
  - Is it different?
  - How can we tell?

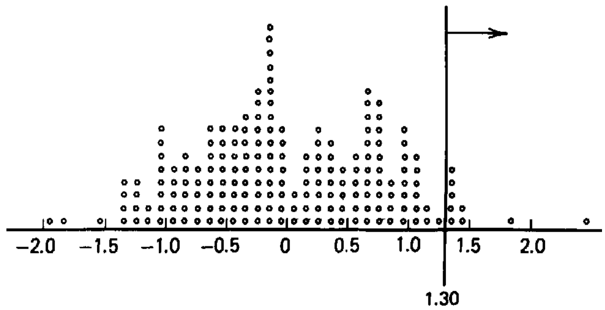
Using an External Reference to make a Decision

- An external reference can be used to decide whether a new observation is different than a group of old observations.
- Example: Create a comparison procedure for lot yield monitoring. Do it without "statistics".
- Use “external reference data“ (historical data from the same process, but *not* from the same experiment):



Example: Using an External Reference

To compare the difference between the average of successive groups of ten lots, I build the histogram from the reference data:



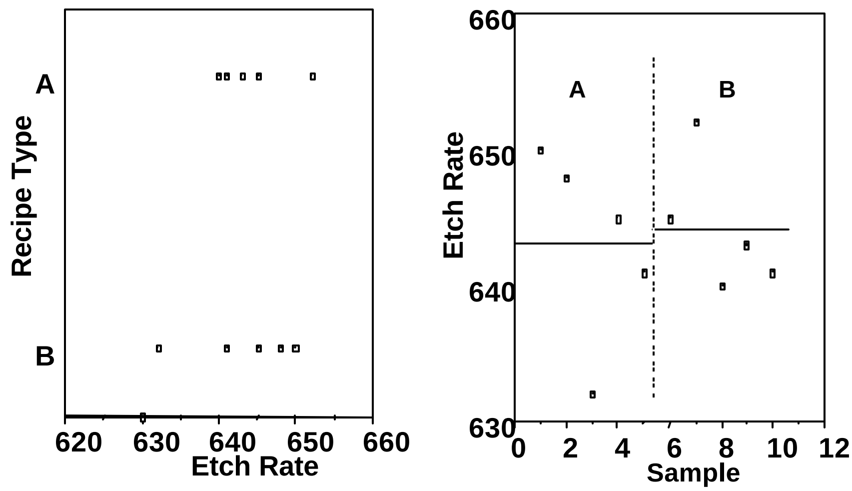
- Each *new* point can then be judged on the basis of the reference data.
- The only assumption here is that the reference data is relevant to my test!

Using an Internal Reference...

- We could generate an "internal" reference distribution from the very data we are comparing.
- Sampling must be random, so that the data is independently distributed.
- Independence would allow us to use statistics such as the arithmetic average or the sum of squares.
- Internal references are based on *Randomization*.

Randomization Example

- Is recipe A different than recipe B?



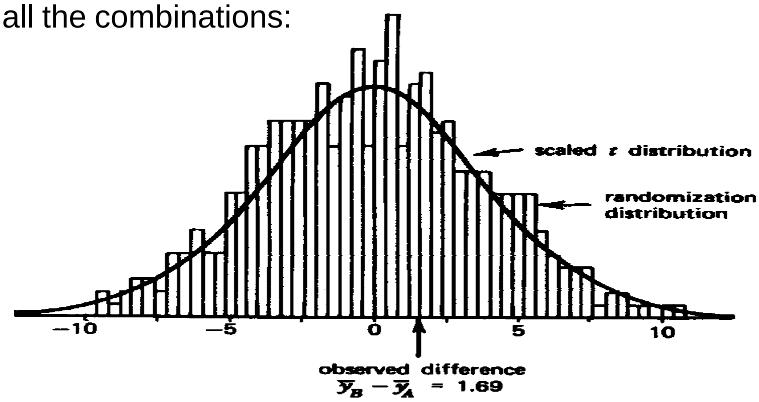
Randomization Example - cont.

- There are many ways to decide this...
  1. External reference distribution (based on old data.)
  2. Assumed, approximate external reference distr. (such as student-t, normal, etc).
  3. Internal reference distribution.
  4. "Distribution free" tests.

- Options 2, 3 and 4 depend on the assumption that the samples are independently distributed.

Randomization Example - cont.

- If there was no difference between A and B, then let me assume that I just have one out of the 10!/5!5! (252) possible arrangements of labels A and B.
- I use the data to calculate the differences in means for all the combinations:



The Origin of the student-t Distribution

The student-t distribution was, in fact, defined to approximate such randomized distributions, when the "parent" distribution is normal!

$$t_0 = \frac{(\bar{y}_B - \bar{y}_A) - (\mu_A - \mu_B)}{s\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

- For the etch example, $t_0 = 0.44$ and $\Pr(t > t_0) = 0.34$
- Randomized Distribution = 0.33

Example in Blocking

- Compare recipes A and B on five machines.
- If there are inherent differences from one machine to the other, what scheme would you use?

Random	Blocked
A A	A B
A B A	B A
B A	B A
B B	A B
B	B A

Example in Blocking - cont.

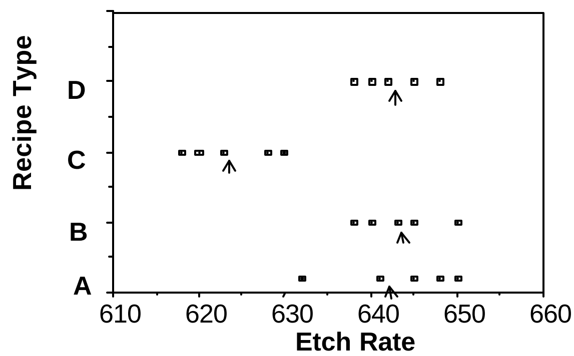
- With the blocked scheme, we could calculate the A-B difference for each machine.
- The machine-to-machine average of these differences could be randomized.

$$\bar{d} = \frac{\pm d_1 \pm d_2 \pm d_3 \pm d_4 \pm d_5}{5}$$

$$\frac{\bar{d} - \delta}{s_d/\sqrt{n}} \sim t_{n-1}$$

In general, *randomize* what you don't know and *block* what you do know.

Analysis of Variance



<u>Your Question</u>: Are the four treatments the *same* or not?

The Statistician's Question: Are the discrepancies *between* the groups greater than the variation *within* each group?

Calculations for our Example

	i=1	i=2	i=3	i=4	i=5	Avg	s_t^2	ν_t	$(\bar{y}_t - \bar{y})^2$
1:	650	648	632	645	641	643.20	202.80	4	25.00
2:	645	650	638	643	640	643.20	86.80	4	25.00
3:	623	628	630	620	618	623.80	104.80	4	207.36
4:	645	640	648	642	638	642.60	63.20	4	19.36

$s_R^2 =$

$s_T^2 =$

$$\frac{s_T^2}{s_R^2} =$$

Variation Within Treatment Groups

First, lets assume that all groups have the same spread. Lets also assume that each group is normally distributed. The following is used to estimate their common σ:

$$S_t = \sum_{j=1}^{n_t} (y_{tj} - \bar{y}_t)^2 \qquad s_t^2 = \frac{S_t}{n_t - 1}$$

$$s_R^2 = \frac{\nu_1 s_1^2 + \nu_2 s_2^2 + \ldots + \nu_k s_k^2}{\nu_1 + \nu_2 + \ldots + \nu_k} = \frac{S_R}{N - k} = \frac{S_R}{\nu_R}$$

- This is an estimate of the unknown, within group s - square.
- It is called the within treatment mean square

Variation Between Treatment Groups

- Let us now form Ho by assuming that all the groups have the same mean.
- Assuming that there are no real differences between groups, a second estimate of s_T^2 would be:

$$s_T^2 = \frac{\sum_{t=1}^{k} n_t(\bar{y}_t - \bar{y})^2}{k - 1} = \frac{S_T}{\nu_T}$$

This is the <u>between treatment mean square</u>

> If all the treatments are the same, then the ***within*** and ***between*** treatment mean squares are estimating the same number!

What if the Treatments *are* different?

If the treatments are different then:

$$s_T^2 \text{ estimates } \sigma^2 + \left[\sum_{t=1}^{k} n_t \tau_t^2 / (k-1) \right]$$

where $\tau_t \equiv \mu_t - \mu$

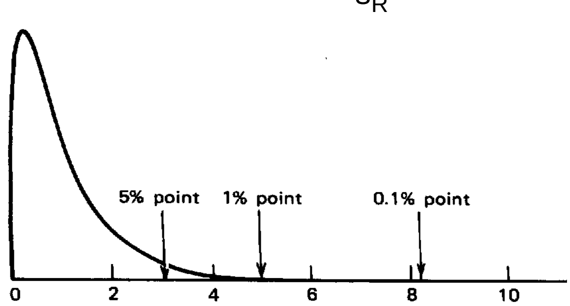
In other words, the between treatment mean square is *inflated* by a factor proportional to the *spread* among the treatments!

Final Test for Treatment Significance

Therefore, the hypothesis of equivalence is rejected if:

$\frac{s_T^2}{s_R^2}$ is significantly greater than 1.0

This can be formalized since: $\frac{s_T^2}{s_R^2} \sim F_{k-1, N-k}$



More Sums of Squares

A measure of the overall variation:

$$S_D = \sum_{t=1}^{k} \sum_{j=1}^{n_t} (y_{tj} - \bar{y})^2 \qquad s_D^2 = \frac{S_D}{N-1} = \frac{S_D}{\nu_D}$$

Obviously (actually, this is not so obvious, but it can be proven):

$$S_D = S_T + S_R \text{ and } \nu_D = \nu_T + \nu_R$$

ANOVA Table

Source of Var	Sum of sq	DFs	Mean sq
between	S_T	ν_T (k-1)	s_T^2
within	S_R	ν_R (N-k)	s_R^2
total	S_D	ν_D (N-1)	s_D^2

ANOVA Table (full)

Source of Var	Sum of sq	DFs	Mean sq
average	S_A	ν_A (1)	s_A^2
between	S_T	ν_T (k-1)	s_T^2
within	S_R	ν_R (N-k)	s_R^2
total	S	ν (N)	

Anova for our example...

Data File: CompEtch

Source	Sum of Squares	Deg. of Freedom	Mean Squares	F-Ratio	Prob>F
Between Recipe	1.3836e+3	3	4.6120e+2	1.6126e+1	4.29e-5
Error	4.5760e+2	16	2.8600e+1		
Total	1.8412e+3	19			

Decomposition of Observations

$$\mathbf{Y} = \mathbf{A} + \mathbf{T} + \mathbf{R}$$

In Vector Form:

$$\begin{vmatrix} y_{ti} \\ \cdot \\ \cdot \\ \cdot \end{vmatrix} = \begin{vmatrix} \bar{y} \\ \cdot \\ \cdot \\ \cdot \end{vmatrix} + \begin{vmatrix} \bar{y}_t - \bar{y} \\ \cdot \\ \cdot \\ \cdot \end{vmatrix} + \begin{vmatrix} y_{ti} - \bar{y}_t \\ \cdot \\ \cdot \\ \cdot \end{vmatrix}$$

$N \quad 1 \quad k-1 \quad N-k$

The term *degrees of freedom* refers to the dimensionality of the space each vector is free to move into.

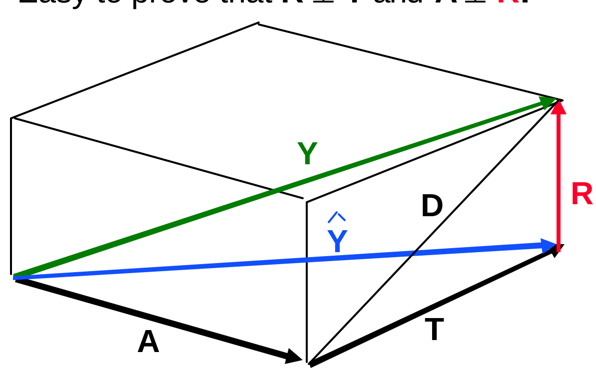
Geometric Interpretation of ANOVA

$\mathbf{Y} = \mathbf{A} + \mathbf{D}$

Easy to prove that $\mathbf{A} \perp \mathbf{D}$.

$\mathbf{D} = \mathbf{R} + \mathbf{T}$

Easy to prove that $\mathbf{R} \perp \mathbf{T}$ and $\mathbf{A} \perp \mathbf{R}$.



ANOVA "Model" and Diagnostics

$$y_{ti} = \mu_t + e_{ti} \qquad e_{ti} \sim N(0, \sigma^2)$$

So, the "sufficient statistics" are: $s_R^2,\ \bar{y}_1, \bar{y}_2, \ldots,\ \bar{y}_k$

as estimators of: $\sigma^2,\ \mu_1, \mu_2, \ldots,\ \mu_k$

For our example:

This "model" describes the data "sufficiently". Its values are the <u>sufficient statistics</u> of the dataset.

$\bar{y}_t$

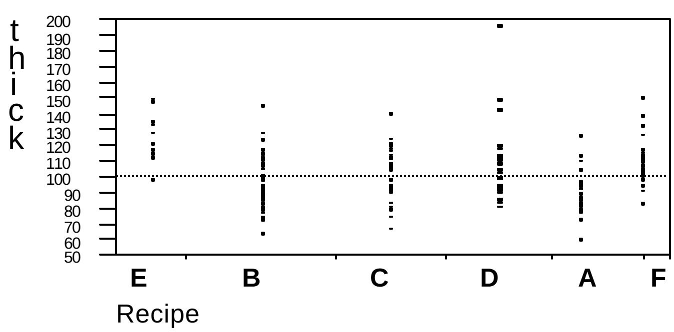
A: 643.20
B: 643.20
C: 623.80
D: 642.60

s_R^2 28.6

According to this model, the residuals are IIND.

How do you verify that?

ANOVA Example: Poly Deposition

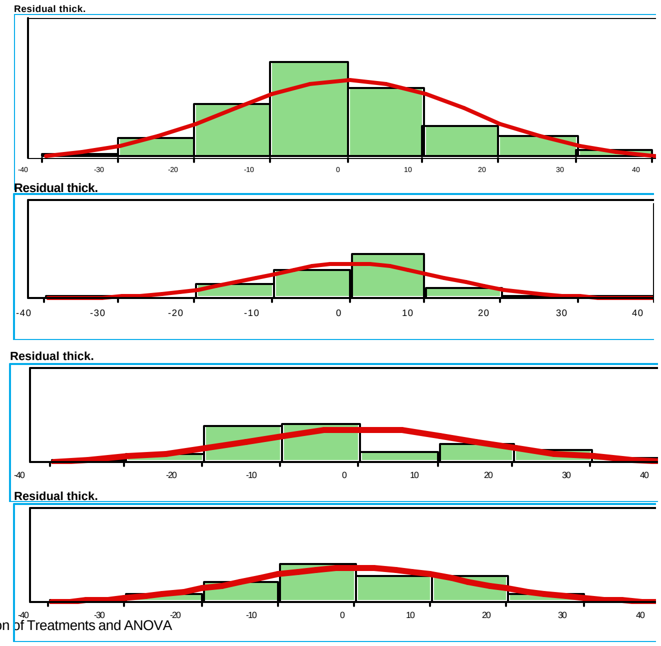


Are these recipes *significantly* different?

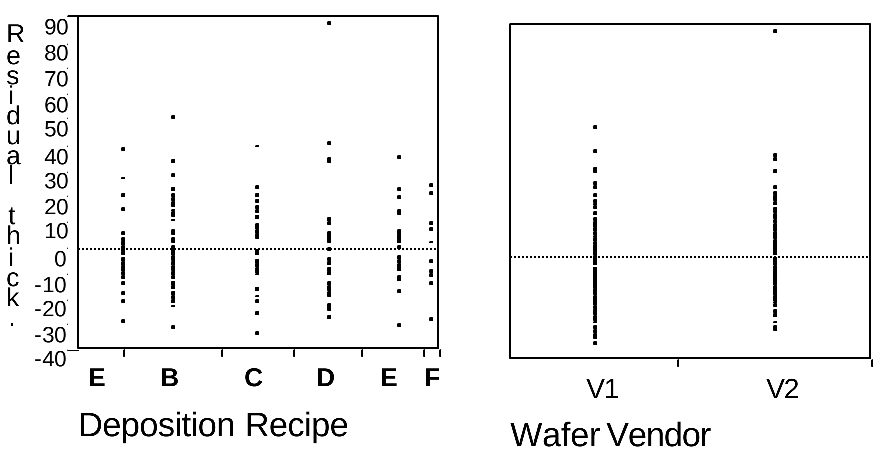
Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	5	26969.525	5393.91	20.9593
Error	227	58418.758	257.35	**Prob > F**
C Total	232	85388.283		0.0000

Residual Plots:



Residual Plots (cont):



ANOVA Summary

- Plot Originals
- Construct ANOVA table
- Are the treatment effects significant?
- Plot residuals versus:
  - treatment
  - group mean
  - time sequence
  - other?
- ANOVA is the basic tool behind most empirical modeling techniques.