2.830J / 6.780J / ESD.63J Control of Manufacturing Processes (SMA 6303)
Spring 2008

# Exponentially Weighted Moving Average: (EWMA)

$$A_i = rx_i + (1-r)A_{i-1}$$ Recursive EWMA

time

$$\sigma_A = \sqrt{\left(\frac{\sigma_x^{\ 2}}{n}\right)\left(\frac{r}{2-r}\right)\left[1-(1-r)^{2t}\right]}$$

$$UCL, LCL = \bar{\bar{x}} \pm 3\sigma_A$$

$$\sigma_A = \sqrt{\frac{\sigma_x^{\ 2}}{n}\left(\frac{r}{2-r}\right)}$$

for large t

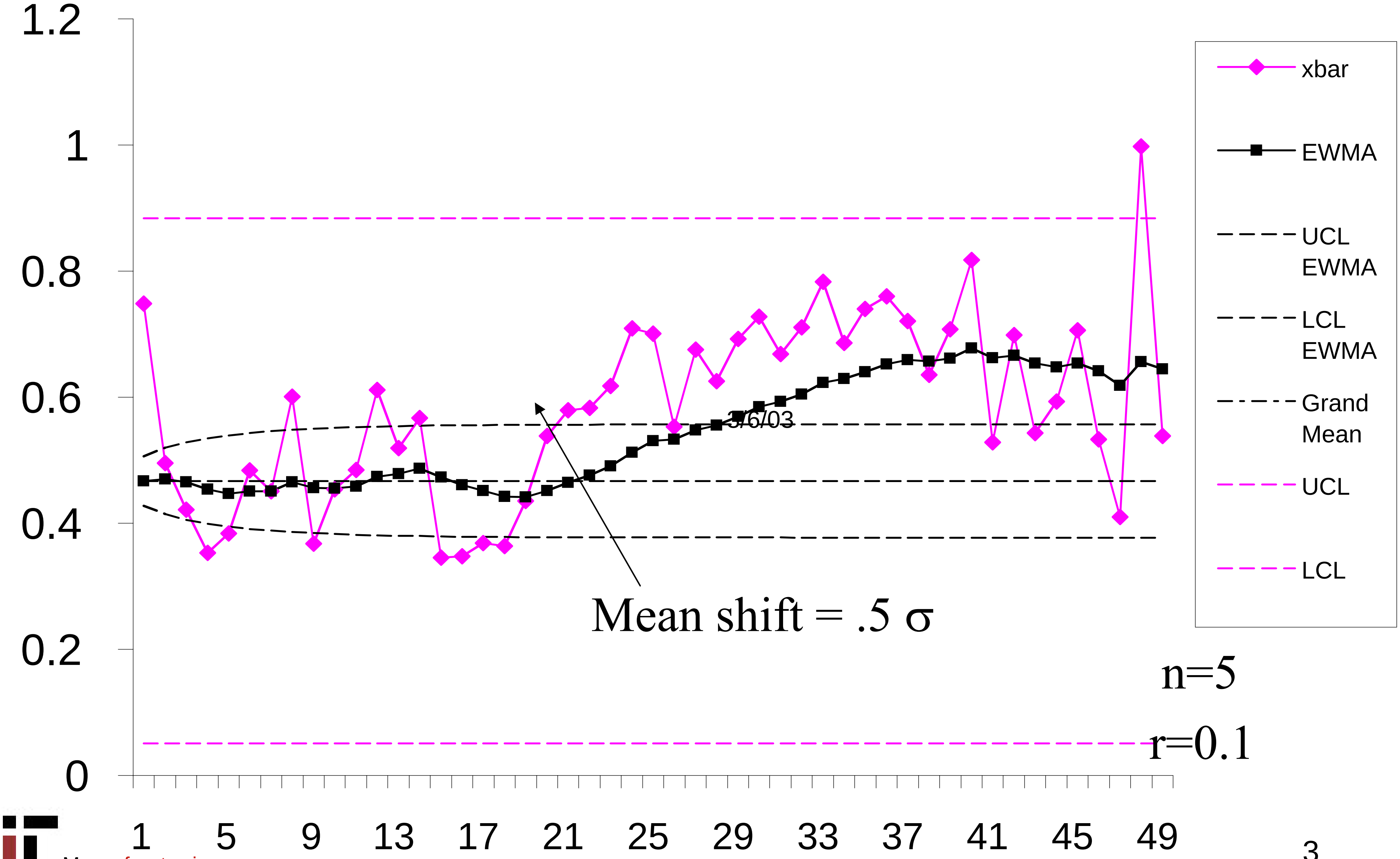
# SO WHAT?

- The variance will be less than with xbar,

$$\sigma_A = \frac{\sigma_x}{\sqrt{n}}\sqrt{\left(\frac{r}{2-r}\right)} \qquad = \sigma_{\bar{x}}\ \sqrt{\left(\frac{r}{2-r}\right)}$$

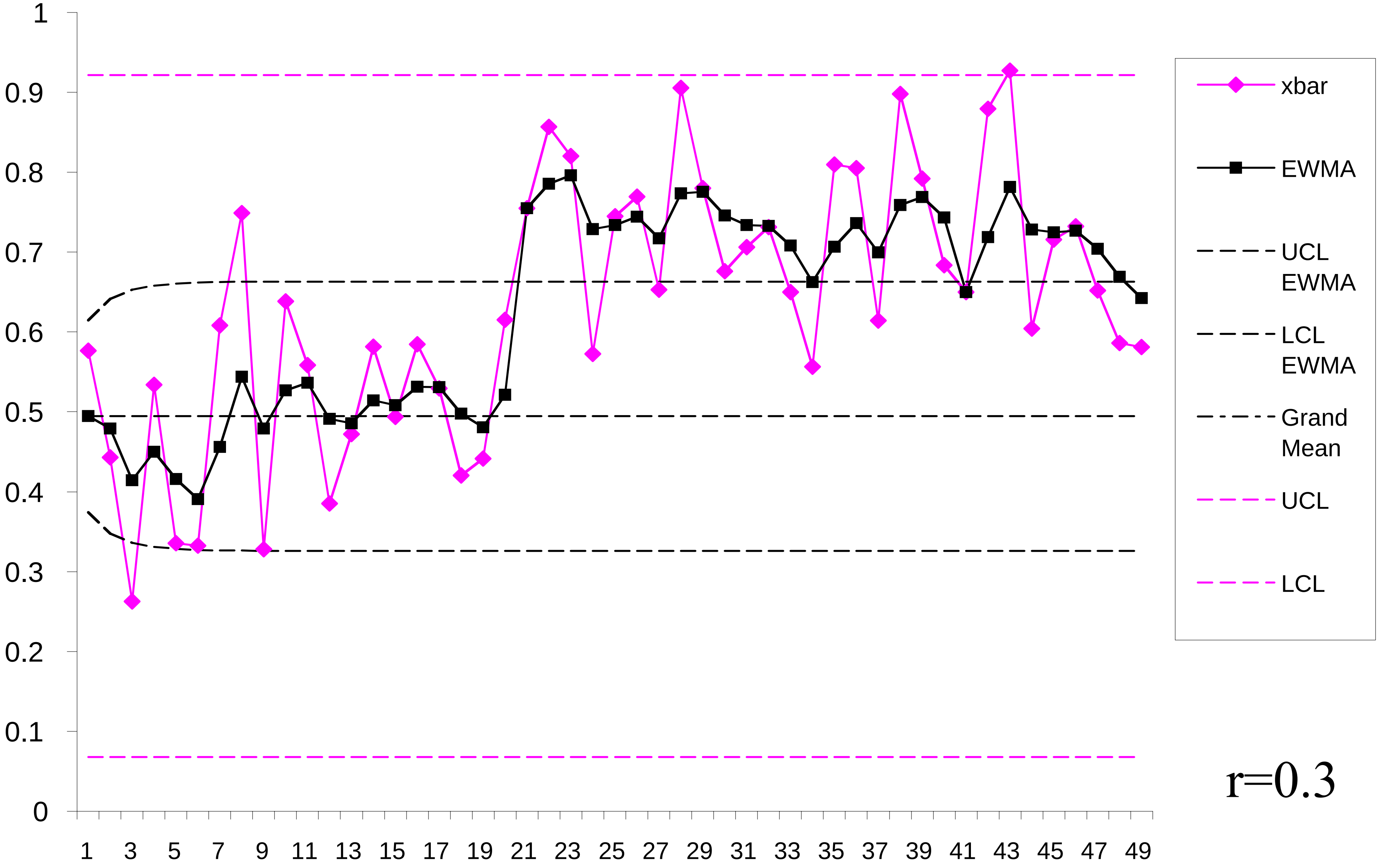
- n=1 case is valid
- If r=1 we have “unfiltered” data
  - Run data stays run data
  - Sequential averages remain
- If r<<1 we get long weighting and long delays
  - “Stronger” filter; longer response time

# Mean Shift Sensitivity
# EWMA and Xbar comparison




Manufacturing

# Effect of r




r=0.3

# Small Mean Shifts

- What if $\Delta\mu_x$ is small with respect to $\sigma_x$ ?
- But it is "persistent"
- How could we detect?
  - ARL for xbar would be too large

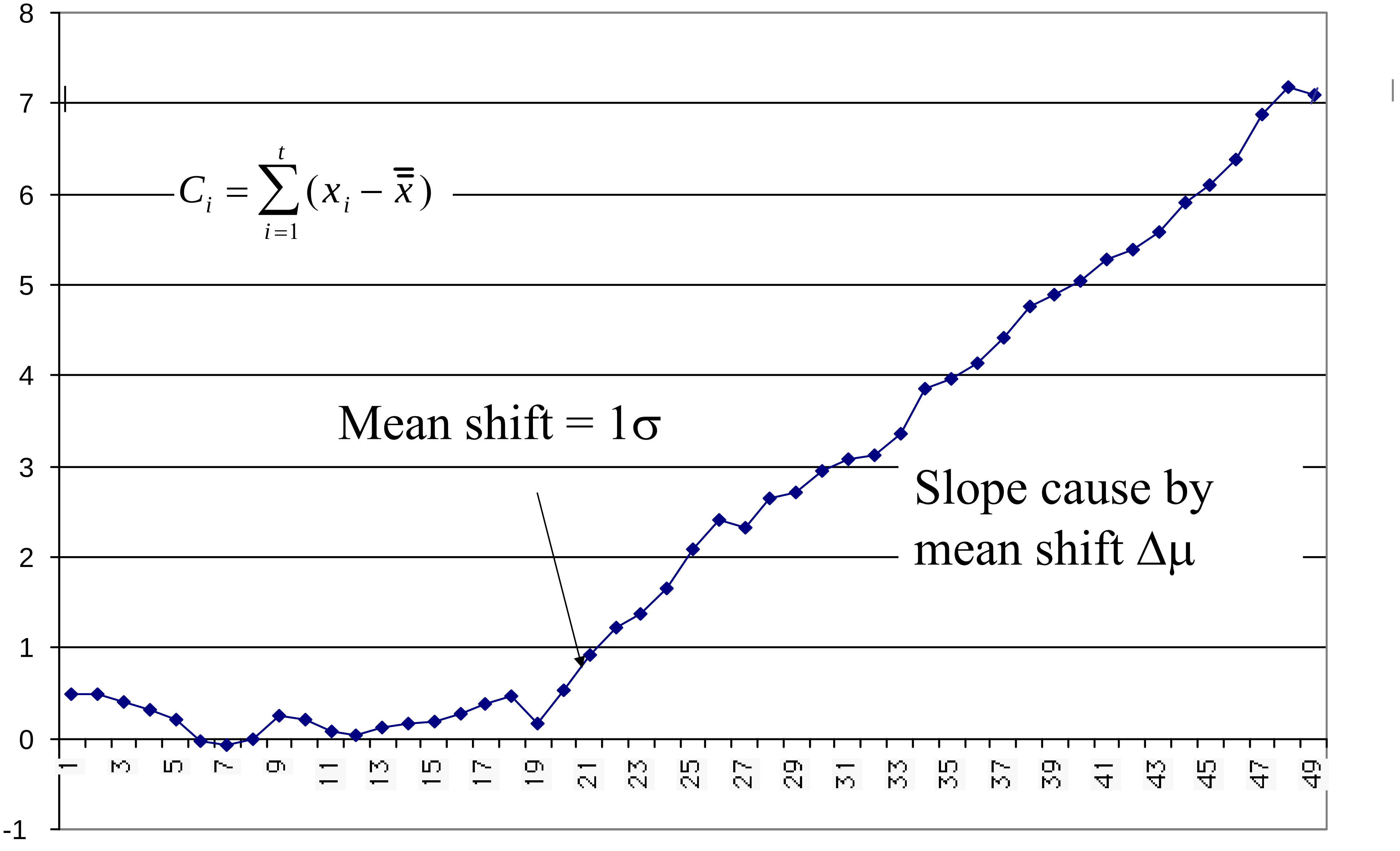
# Another Approach: Cumulative Sums

- Add up deviations from mean
  - A Discrete Time Integrator

$$C_j = \sum_{i=1}^{j} (x_i - \overline{\overline{x}})$$

- Since E{$x$-$\mu$}=0 this sum should stay near zero when in control
- Any bias (mean shift) in *x* will show as a trend

# Mean Shift Sensitivity: CUSUM



# An Alternative

- Define the Normalized Statistic

$$Z_i = \frac{X_i - \mu_x}{\sigma_x}$$

Which has an expected mean of 0 and variance of 1

- And the CUSUM statistic

$$S_i = \frac{\sum_{i=1}^{t} Z_i}{\sqrt{t}}$$

Which has an expected mean of 0 and variance of 1

Chart with Centerline =0 and Limits = ±3

# Tabular CUSUM

- Create Threshold Variables:

$$C_i^+ = \max[0, x_i - (\mu_0 + K) + C_{i-1}^+]$$
$$C_i^- = \max[0, (\mu_0 - K) - x_i + C_{i-1}^-]$$

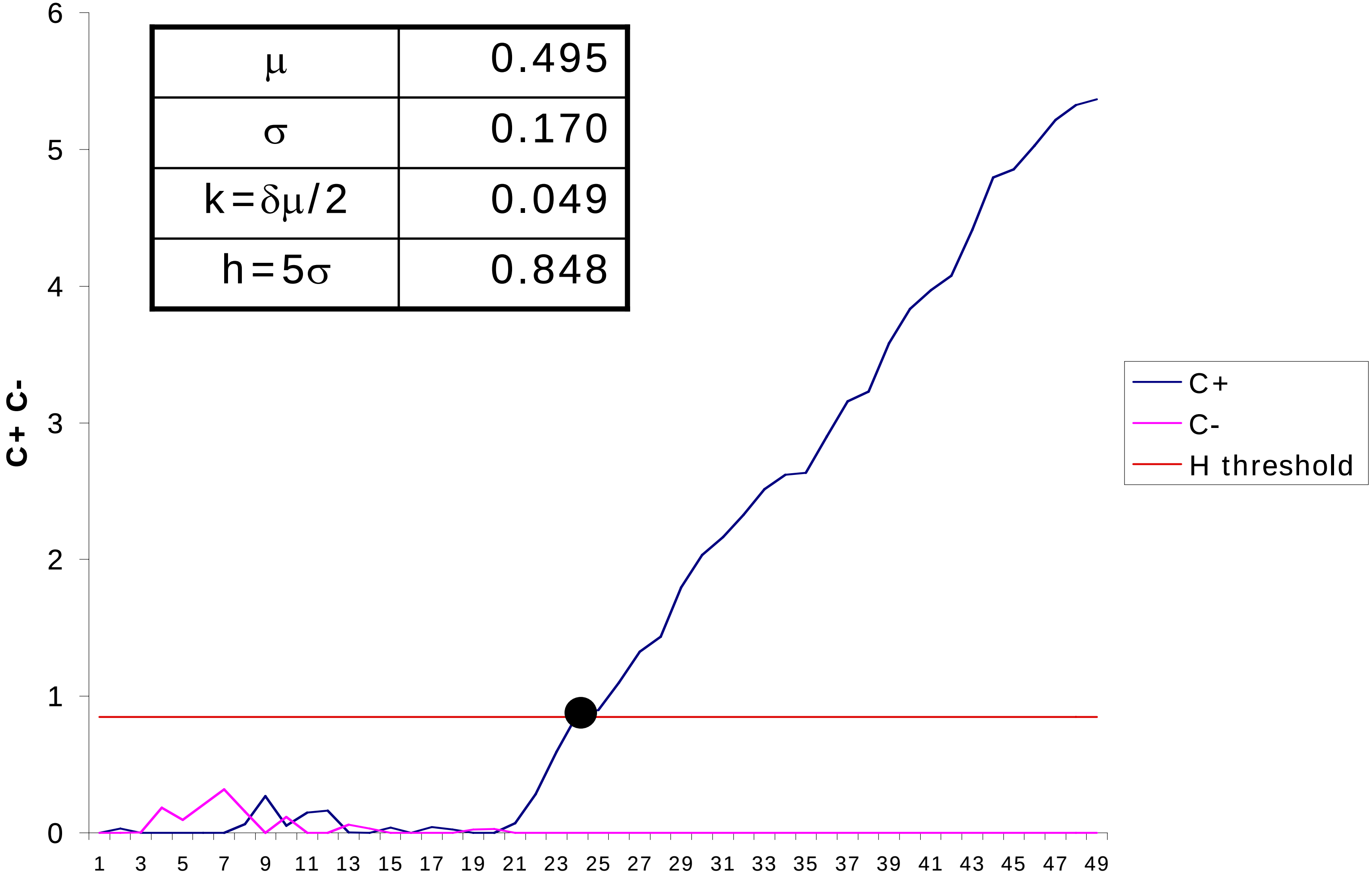
Accumulates deviations from the mean

$K$= threshold or slack value for accumulation

$$\underset{\boxed{\text{typical}}}{K} = \left|\frac{\Delta\mu}{2}\right| \qquad \Delta\mu = \text{mean shift to detect}$$

$H:$ alarm level (typically $5\sigma$)

# Threshold Plot




| | |
|---|---|
| $\mu$ | 0.495 |
| $\sigma$ | 0.170 |
| $k=\delta\mu/2$ | 0.049 |
| $h=5\sigma$ | 0.848 |

# Univariate vs. $\chi^2$ Chart

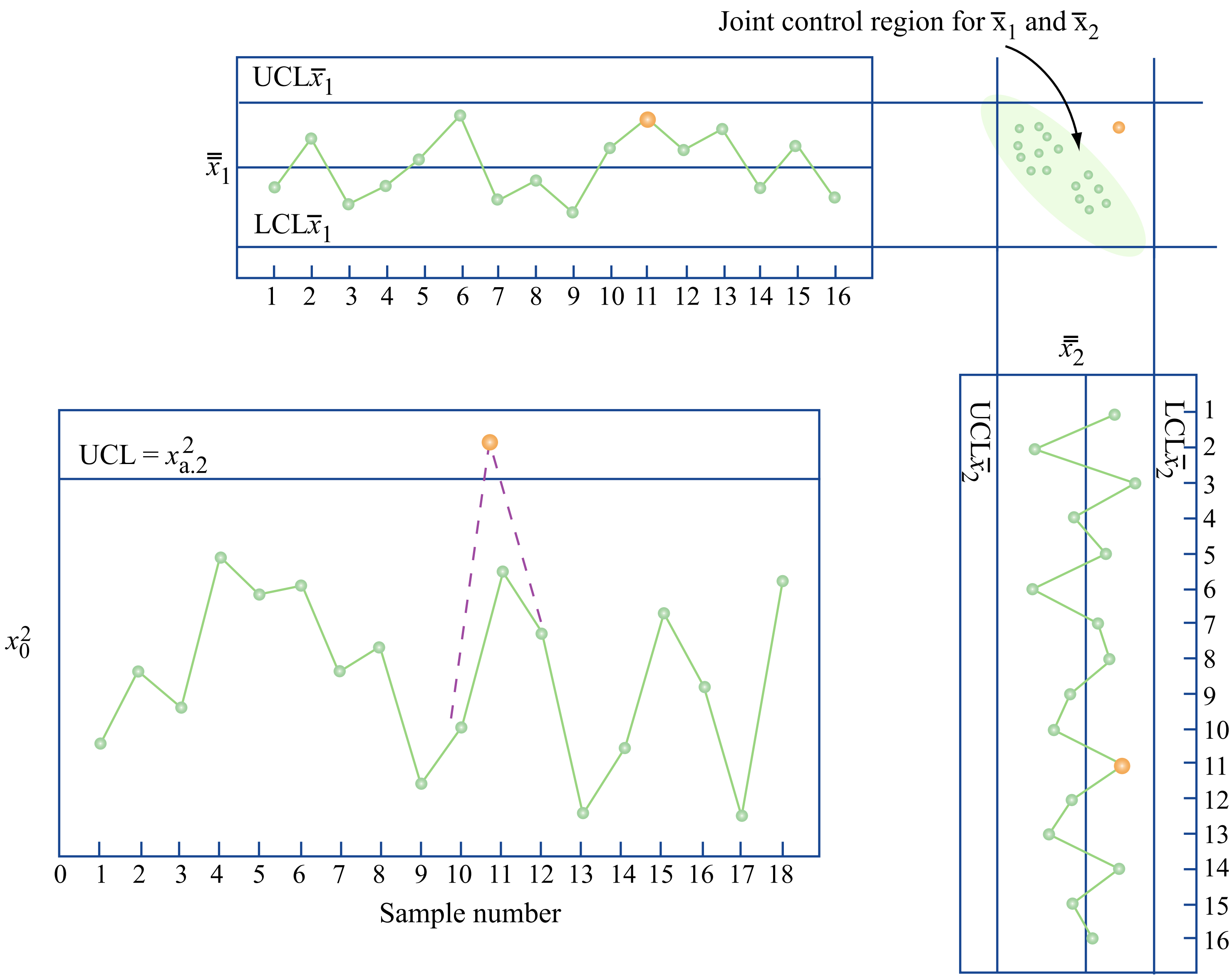



Figure by MIT OpenCourseWare.

# Multivariate Chart with No Prior Statistics: $T^2$

- If we must use data to get $\underline{\bar{x}}$ and $S$
- Define a new statistic, Hotelling $T^2$

$$T^2 = n(\underline{\bar{x}} - \underline{\bar{\bar{x}}})^T S^{-1} (\underline{\bar{x}} - \underline{\bar{\bar{x}}})$$

- Where $\underline{\bar{\bar{x}}}$ is the vector of the averages for each variable over all measurements
- $S$ is the matrix of sample *covariance* over all data

# Similarity of $T^2$ and $t^2$

$$T^2 = n(\underline{\bar{x}} - \underline{\bar{\bar{x}}})^T S^{-1} (\underline{\bar{x}} - \underline{\bar{\bar{x}}})$$

vs.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$t^2 = \frac{n(\bar{x} - \mu)^2}{s^2}$$

$$t^2 = n(\bar{x} - \mu) s^{-2} (\bar{x} - \mu)$$

# Yield – Negative Binomial Model

- Gamma probability distribution for *f(D)*
  - proposed by Ogabe, Nagata, and Shimada; popularized by Stapper

$$f(D) = \frac{D^{\alpha-1}e^{-D/\beta}}{\Gamma(\alpha)\beta^{\alpha}}$$

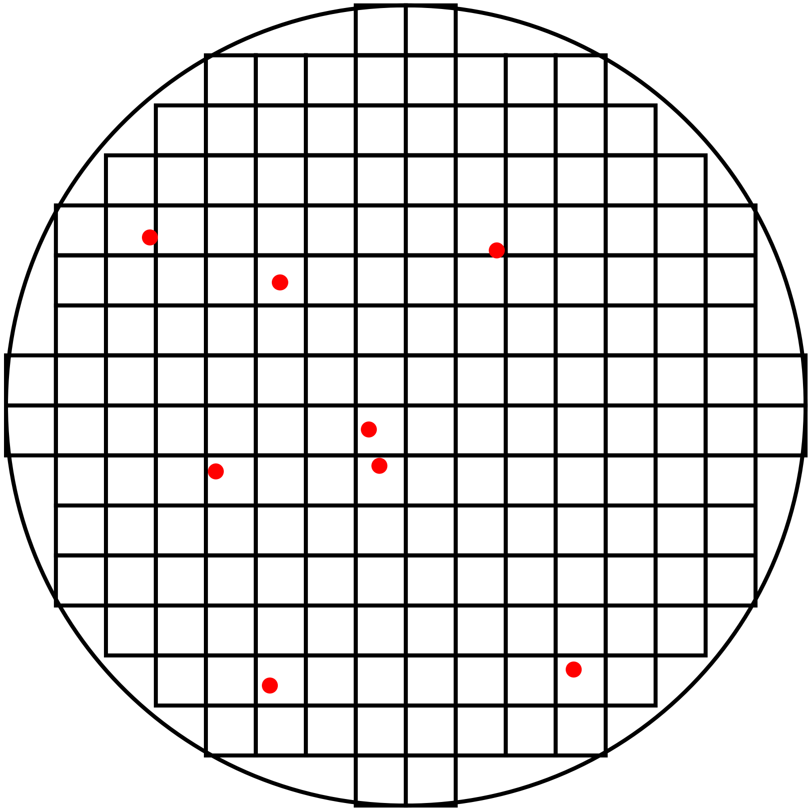


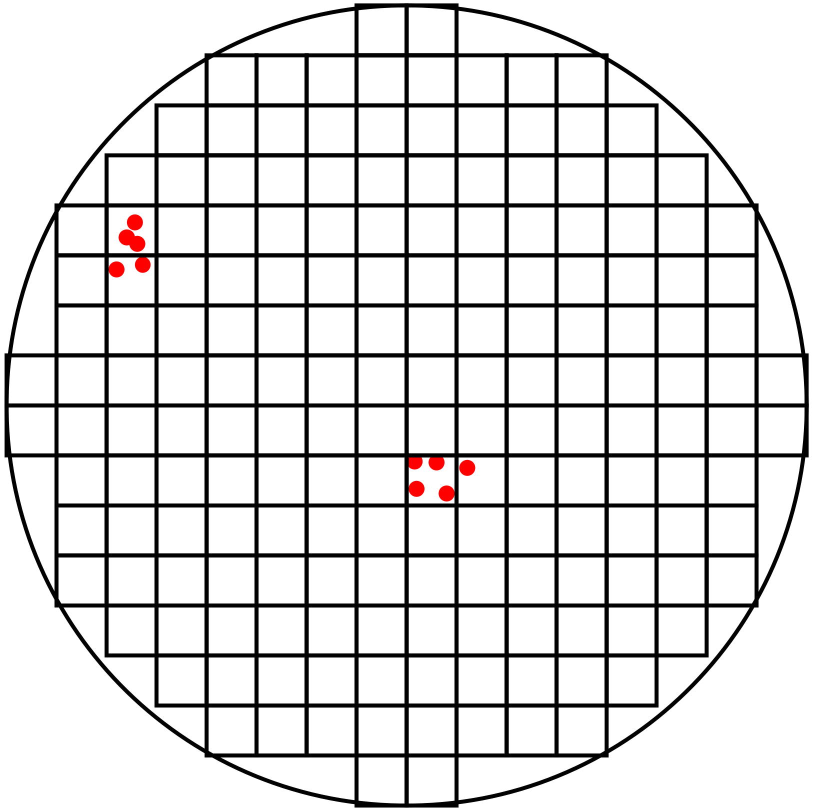
- $\alpha$ is a "cluster" parameter
  - High $\alpha$ means low variability of defects (little clustering)
- Resulting yield:

$$Y_{\text{gamma}} = \left(1 + \frac{A_0 D_0}{\alpha}\right)^{-\alpha}$$

*May & Spanos*

# Spatial Defects





- Random distribution
- Spatially uncorrelated
- Each defect “kills” one chip

- Spatially clustered
- Multiple defects within one chip (can’t already kill a dead chip!)

# Negative Binomial Model, p. 2

- Large $\alpha$ limit (little clustering)
  - gamma density approaches a delta function, and yield approaches the Poisson model:

$$Y = \lim_{\alpha \to \infty} \left(1 + \frac{A_0 D_0}{\alpha}\right)^{-\alpha} = \exp(-A_0 D_0)$$

- Small $\alpha$ limit (strong clustering)
  - yield approaches the Seeds model:

$$Y = \lim_{\alpha \to 0} \left(1 + \frac{A_0 D_0}{\alpha}\right)^{-\alpha} = \frac{1}{1 + A_0 D_0}$$

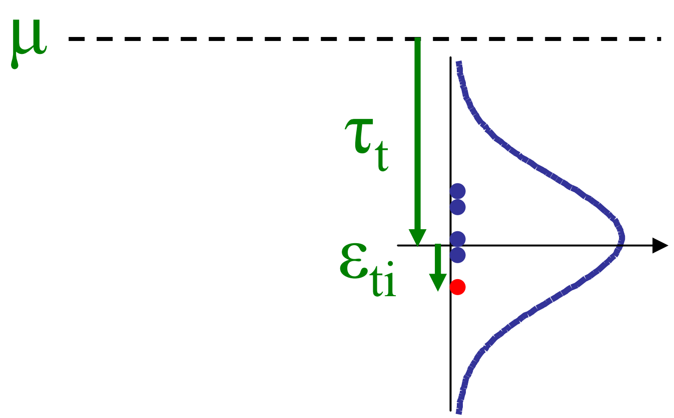
- Must empirically determine $\alpha$
  - typical memory and microprocessors: $\alpha$ = 1.5 to 2

# ANOVA – Fixed effects model

- The ANOVA approach assumes a simple mathematical model:

$$\begin{aligned} y_{ti} &= \mu + \tau_t + \epsilon_{ti} \\ &= \mu_t + \epsilon_{ti} \end{aligned}$$

- Where $\mu_t$ is the treatment mean (for treatment type t)
- And $\tau_t$ is the treatment effect
- With $\varepsilon_{ti}$ being zero mean normal residuals $\sim N(0,\sigma_0^2)$



# The ANOVA Table

| source of variation | sum of squares | degrees of freedom | mean square | $F_0$ | Pr($F_0$) |
|---|---|---|---|---|---|
| Between treatments | $SS_T$ | $k-1$ | $s_T^2 = \frac{SS_T}{k-1}$ | $\frac{s_T^2}{s_R^2}$ | table |
| Within treatments | $SS_R$ | $N-k$ | $s_R^2 = \frac{SS_R}{N-k}$ | | |
| Total about the grand average | $SS_D$ | $N-1$ | $s_D^2 = \frac{SS_D}{N-1}$ | | |

Also referred to as "residual" SS

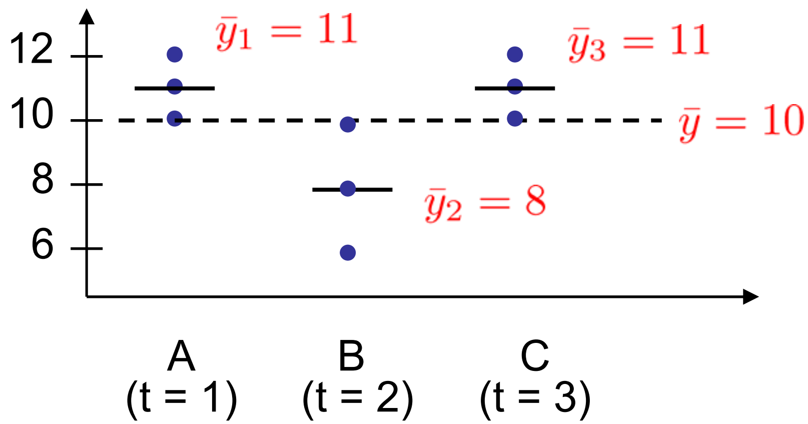
$SS_D = SS_T + SS_R$

$\nu_D = \nu_T + \nu_R$

Manufacturing

# Example: Anova

| A | B | C |
|---|---|---|
| 11 | 10 | 12 |
| 10 | 8 | 10 |
| 12 | 6 | 11 |




## Excel: Data Analysis, One-Variation Anova

| Anova: Single Factor | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | |
| SUMMARY | | | | | | |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* | | |
| A | 3 | 33 | 11 | 1 | | |
| B | 3 | 24 | 8 | 4 | | |
| C | 3 | 33 | 11 | 1 | | |
| | | | | | | |
| | | | | | | |
| ANOVA | | | | | | |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 18 | 2 | 9 | 4.5 | 0.064 | 5.14 |
| Within Groups | 12 | 6 | 2 | | | |
| | | | | | | |
| Total | 30 | 8 | | | | |

$$F = \frac{S_T^2}{S_R^2} = \frac{9}{2} = 4.5$$

$$F_{0.05,2,6} = 5.14$$

$$F_{0.10,2,6} = 3.46$$

$$\begin{aligned} SS_1 &= (12-11)^2 + (11-11)^2 + (10-11)^2 = 2 \\ SS_2 &= 2^2 + 0^2 + 2^2 = 8 \\ SS_3 &= 1^2 + 0^2 + 1^{=}2 \end{aligned}$$

$$\begin{aligned} s_1^2 &= MS_1 = SS_1/2 = 2/2 = 1 \\ s_2^2 &= MS_2 = 8/2 = 4 \\ s_3^2 &= MS_3 = 2/2 = 1 \end{aligned}$$

$$s_R^2 = \frac{SS_1 + SS_2 + SS_3}{N-k} = \frac{12}{6} = 2$$

$$\begin{aligned} s_T^2 &= \frac{3(11-10)^2 + 3(8-10)^2 + 3(11-10)^2}{3-1} \\ &= \frac{SS_T}{\nu_T} = \frac{18}{2} = 9 \end{aligned}$$

# Definition: Contrasts

$$A = \frac{1}{2n} \underbrace{[a + ab - b - (1)]}$$

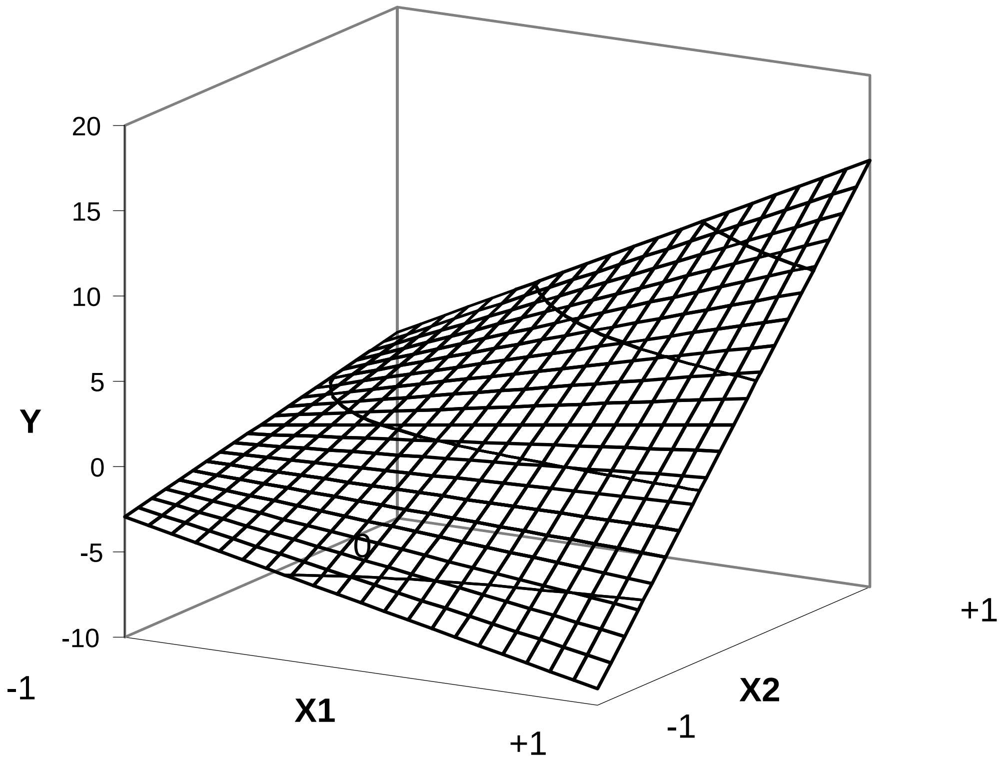
$$B = \frac{1}{2n} \underbrace{[b + ab - a - (1)]}$$

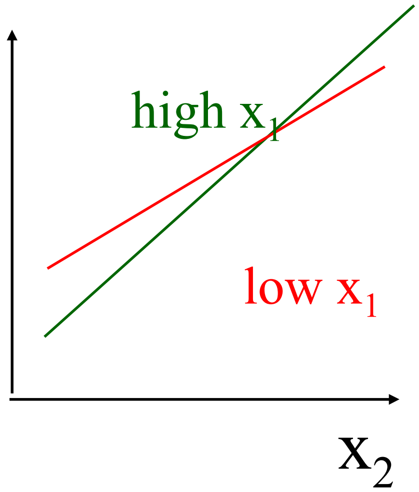
$$AB = \frac{1}{2n} \underbrace{[ab + (1) - a - b]}$$

[.....] = "Contrast"

$$\hat{y} = \bar{\bar{y}} + \frac{A}{2}x_1 + \frac{B}{2}x_2 + \frac{AB}{2}x_1x_2$$

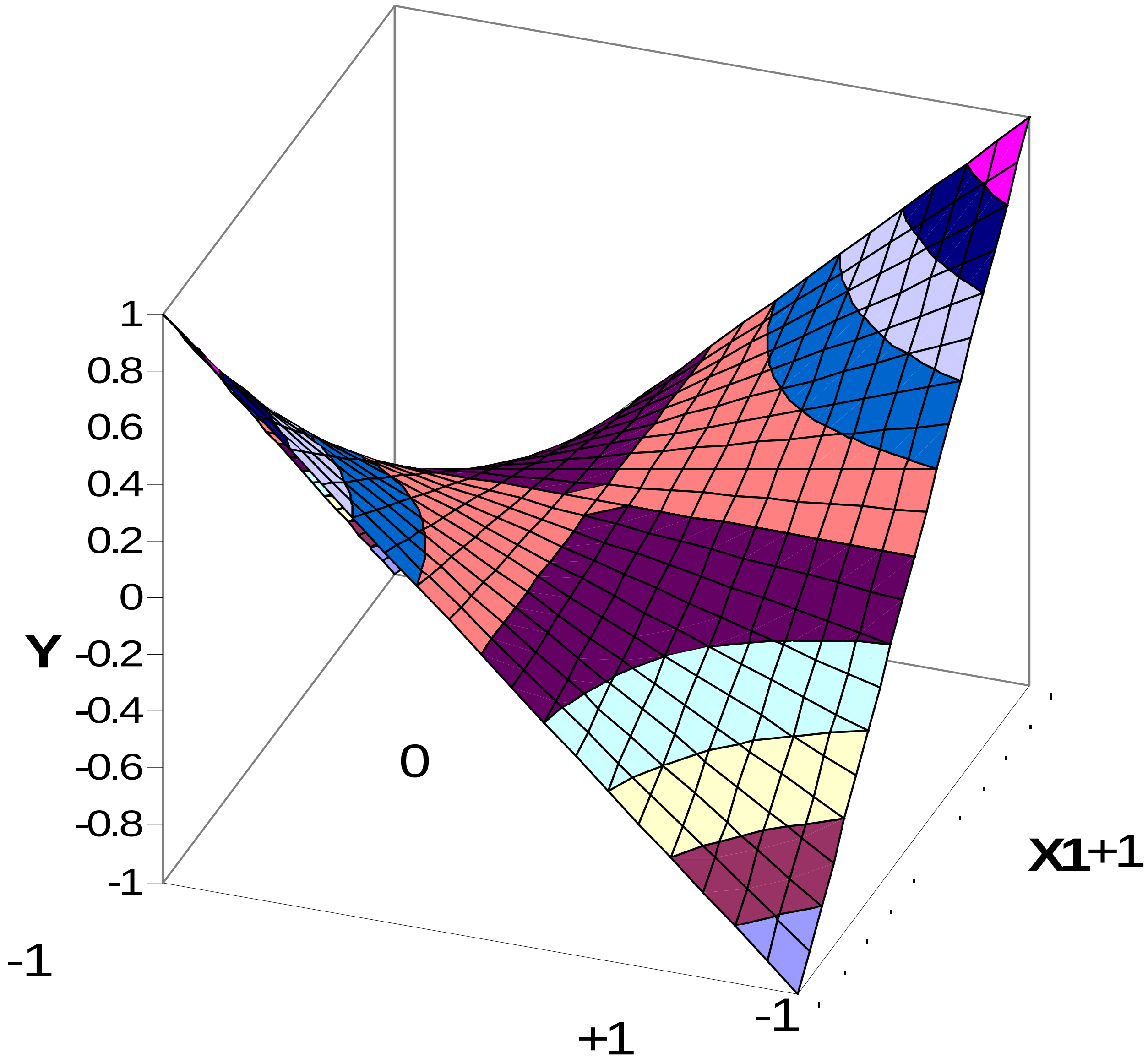
# Response Surface: Positive Interaction

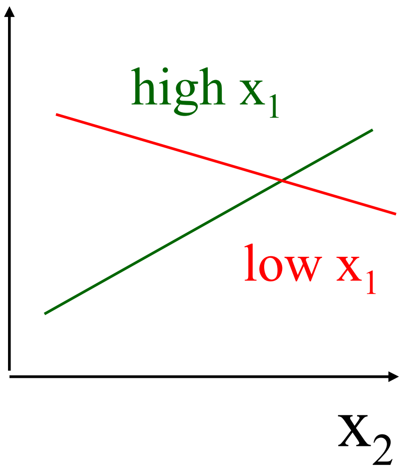



$$y = 1 + 7x_1 + 2x_2 + 5x_1x_2$$

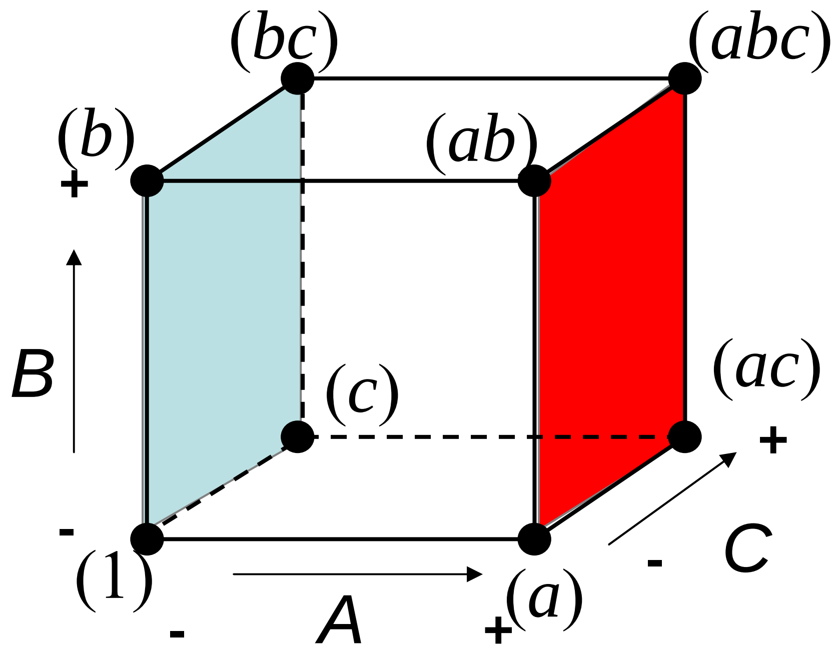
# Response Surface: Negative Interaction




$$y = 1 + 7x_1 + 2x_2 - 5x_1x_2$$

# "Surface" Averages




$$A = \frac{1}{4}\left[(abc)+(ab)+(ac)+(a)\right] - \frac{1}{4}\left[(b)+(c)+(bc)+(1)\right]$$

Courtesy of Dan Frey. Used with permission.

# ANOVA for $2^k$

- Now have more than one "effect"
- We can derive:

$$SS_{Effect} = (Contrast)^2 / n2^k$$

- And it can be shown that:

$$SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{Error}$$

# Use of Central Data

- Determine Deviation from Linear Prediction
  - Quadratic Term, or Central Error Term
- Determine MS of that Error
  - SS/dof
- Compare to Replication Error

# Definitions

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x^2 + \beta_2 x^2$$

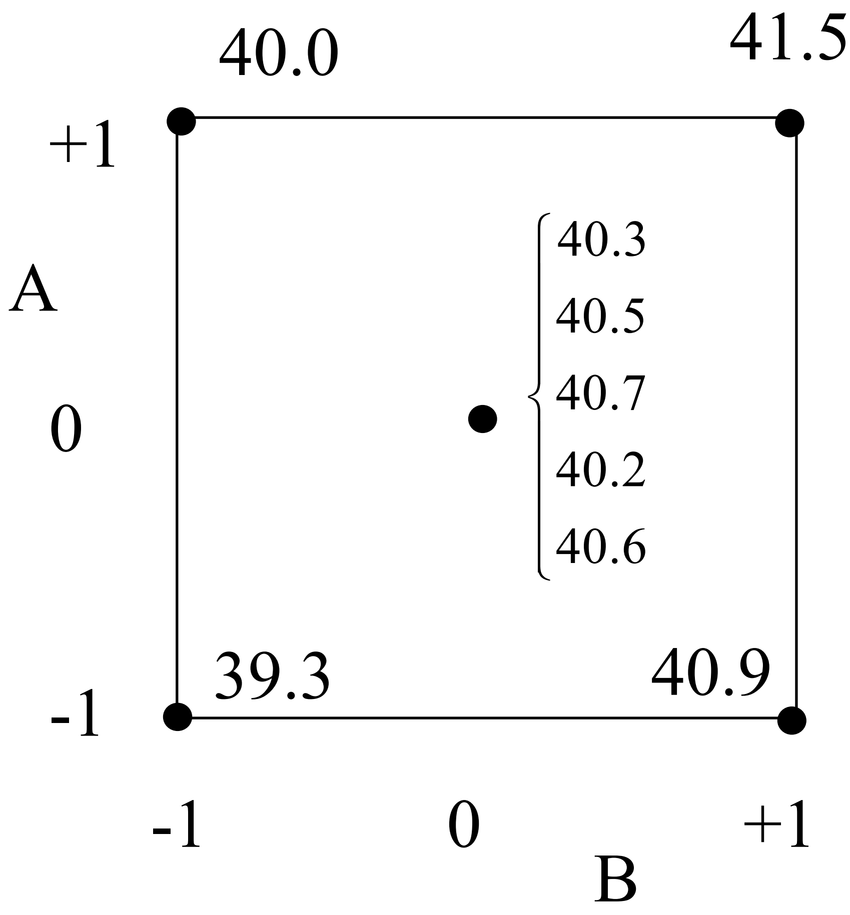
$\bar{y}_F$ = grand mean of all factorial runs

$\bar{y}_C$ = grand mean of all center point runs

$$SS_{Quadratic} = \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{n_F + n_C}$$

$$MS_{Quadratic} = \frac{SS_{Quadratic}}{n_c}$$

# Example: $2^2$ Without Replicates; Replicated Intermediate Points




| | | I | A | B | AB |
|---|---|---|---|---|---|
| (1) | 39.3 | 1 | -1 | -1 | 1 |
| a | 40.9 | 1 | 1 | -1 | -1 |
| b | 40 | 1 | -1 | 1 | -1 |
| ab | 41.5 | 1 | 1 | 1 | 1 |

| | | | | |
|---|---|---|---|---|
| Contrasts | 161.7 | 3.1 | 1.3 | -0.1 |
| Effect | 80.85 | 1.55 | 0.65 | -0.05 |
| Model Coefficients | 40.43 | 0.775 | 0.325 | -0.025 |

*Just using corner points:*

$$y = 40.43 + 0.775_1 x_1 + 0.325 x_2 - 0.025 x_1 x_2$$

# Measures of Model Goodness – $R^2$

- Goodness of fit – $R^2$
  - Question considered: how much better does the model do than just using the grand average?

$$R^2 = \frac{SS_T}{SS_D}$$

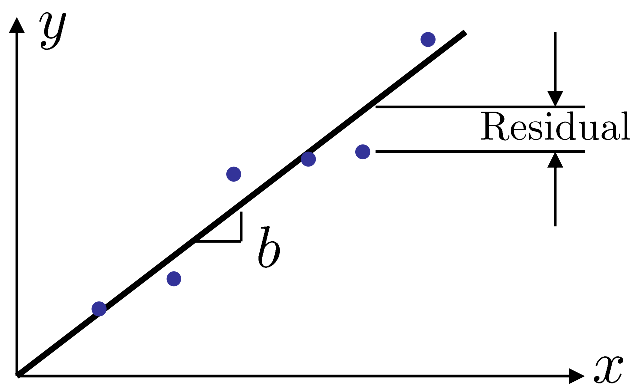
  - Think of this as the fraction of squared deviations (from the grand average) in the data which is captured by the model
- Adjusted $R^2$
  - For "fair" comparison between models with different numbers of coefficients, an alternative is often used

$$R^2_{\text{adj}} = 1 - \frac{SS_R/\nu_R}{SS_D/\nu_D} = 1 - \frac{s_R^2}{s_D^2}$$

  - Think of this as (1 – variance remaining in the residual). Recall $\nu_R = \nu_D - \nu_T$

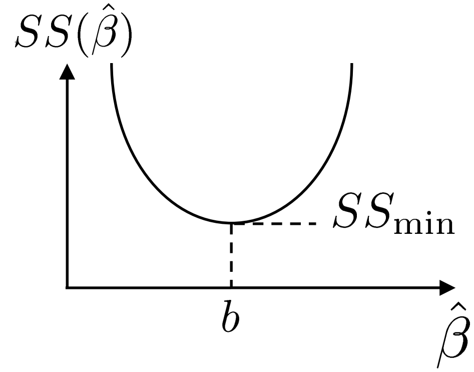
# Least Squares Regression

- We use ***least-squares*** to estimate coefficients in typical regression models
- $$\begin{aligned} y_i &= \beta x_i + \epsilon_i, \quad i = 1, 2, ..., n; \ \epsilon_i \sim N(0, \sigma^2) \\ \hat{y}_i &= b x_i \end{aligned}$$




- Goal is to estimate β with "best" *b*
- How define "best"?
  - That b which minimizes sum of squared error between prediction and data

$$SS(\hat{\beta}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta} x_i)^2$$




  - The residual sum of squares (for the best estimate) is

$$SS_{\min} = \sum_{i=1}^{n} (y_i - b x_i)^2 = SS_R$$

# Least Squares Regression, cont.

- Least squares estimation via normal equations
  - For linear problems, we need not calculate SS(β); rather, direct solution for *b* is possible
  - Recognize that vector of residuals will be normal to vector of *x* values at the least squares estimate

$$\begin{aligned} \sum (y-\hat{y})x &= 0 \\ \sum (y-bx)x &= 0 \\ \sum xy &= \sum bx^2 \end{aligned}$$

$$\Rightarrow \quad b = \frac{\sum xy}{\sum x^2}$$

- Estimate of experimental error
  - Assuming model structure is adequate, estimate $s^2$ of $\sigma^2$ can be obtained:

$$s^2 = \frac{SS_R}{n-1}$$

# Precision of Estimate: Variance in *b*

- We can calculate the variance in our estimate of the slope, *b*:

$$b = \frac{\sum xy}{\sum x^2} \quad \Rightarrow \quad \hat{V}(b) = \frac{s^2}{\sum x_i^2}$$

$$\text{s.e.}(b) = \sqrt{\hat{V}(b)}$$

$$b \pm \text{s.e.}(b)$$

- Why?

$$\begin{aligned} b &= \frac{x_1}{\sum x^2} \cdot y_1 + \frac{x_2}{\sum x^2} \cdot y_2 + \cdots \frac{x_n}{\sum x^2} \cdot y_n \\ &= a_1 y_1 + a_2 y_2 + \cdots + a_n y_n \end{aligned}$$

$$\begin{aligned} V(b) &= (a_1^2 + a_2^2 + \cdots + a_n^2)\sigma^2 \\ &= \left[ \left(\frac{x_1}{\sum x^2}\right)^2 + \cdots + \left(\frac{x_n}{\sum x^2}\right)^2 \right] \sigma^2 \\ &= \frac{\sum x^2}{(\sum x^2)^2} \sigma^2 \\ &= \frac{\sigma^2}{\sum x^2} \end{aligned}$$

# Confidence Interval for β

- Once we have the standard error in *b*, we can calculate confidence intervals to some desired
  (1-α)100% level of confidence

$$\frac{b-\beta}{\text{s.e.}(b)} \sim t \qquad \Rightarrow \qquad \beta = b \pm t_{\alpha/2} \cdot \text{s.e.}(b)$$

- Analysis of variance
  - Test hypothesis: $H_0 : \beta = b = 0$
  - If confidence interval for β includes 0, then β not significant
  - Degrees of freedom (need in order to use t distribution)

$$\begin{array}{ccccc} \sum y_i^2 & = & \sum \hat{y}_i^2 & + & \sum (y_i - \hat{y}_i)^2 \\ n & = & p & + & n-p \end{array}$$

***p* = # parameters estimated by least squares**

# Lack of Fit Error vs. Pure Error

- Sometimes we have replicated data
  - E.g. multiple runs at same *x* values in a designed experiment

- We can decompose the residual error contributions

$$SS_R = SS_L + SS_E$$

Where
$SS_R$ = residual sum of squares error
$SS_L$ = lack of fit squared error
$SS_E$ = pure replicate error

- This allows us to TEST for lack of fit
  - By "lack of fit" we mean evidence that the linear model form is inadequate

$$\frac{s_L^2}{s_E^2} \sim F_{\nu_L,\nu_E}$$

# Regression: Mean Centered Models

- Model form $y = \alpha + \beta(x - \bar{x})$
- Estimate by $\hat{y} = a + b(x - \bar{x}), \quad (y_i - \hat{y}_i) \sim \mathrm{N}(0, \sigma^2)$

$$\text{Minimize } SS_R = \sum (y_i - \hat{y}_i)^2 \text{ to estimate } \alpha \text{ and } \beta$$

$$a = \bar{y}$$

$$\mathrm{E}(a) = \alpha$$

$$\mathrm{Var}(a) = \mathrm{Var}\left[\frac{\sum y_i}{n}\right] = \frac{\sigma^2}{n}$$

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\mathrm{E}(b) = \beta$$

$$\mathrm{Var}(b) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

# Regression: Mean Centered Models

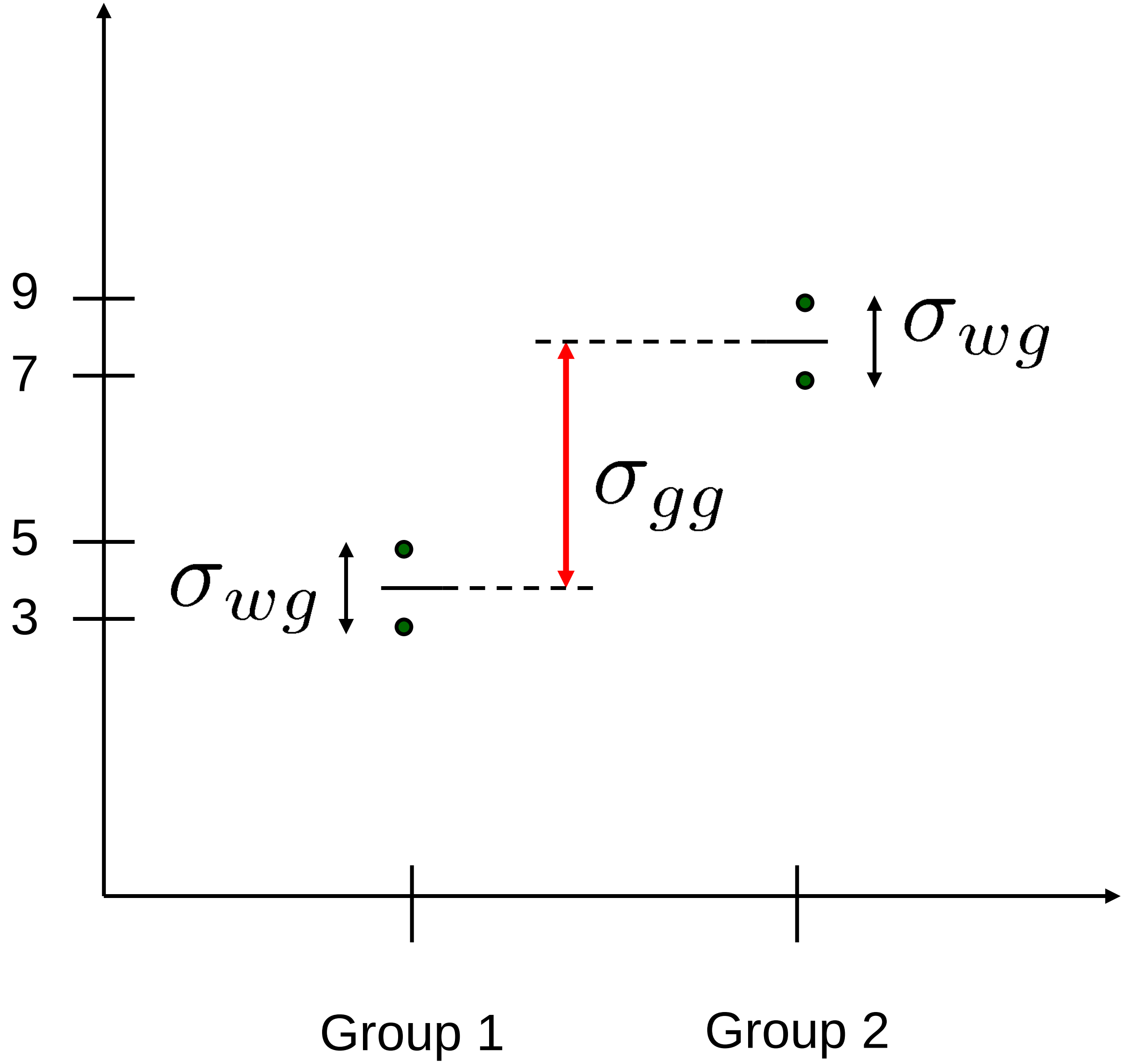
- Confidence Intervals

$$\begin{aligned}\hat{y}_i &= \bar{y} + b(x_i - \bar{x}) \\ \\ \mathrm{Var}(\hat{y}_i) &= \mathrm{Var}(\bar{y}) + (x_i - \bar{x})^2 \mathrm{Var}(b) \\ &= \frac{s^2}{n} + \frac{s^2 (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} = s^2_{\hat{y}_i}\end{aligned}$$

- Our confidence interval on output *y* widens as we get further from the center of our data!

$$\hat{y}_i \pm t_{\alpha/2} \cdot s_{\hat{y}_i}$$

# Nested Variance Example (Same Data)




- Now – groups are simply replicates (not changing treatment)
- But… assume there are two different sources of ***zero mean*** variances
- Goal – estimate these two variances

# Variance in *Observed* Averages, Three Levels

- As in the two level case, the observed averages include lower level variances, reduced by number of samples

$$\sigma_{\bar{L}}^2 = \sigma_L^2 + \frac{\sigma_W^2}{W} + \frac{\sigma_M^2}{MW}$$

  – Above is for a balanced sampling plan, with equal number of wafers and measurements for each lot