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

Control of Manufacturing Processes

Subject 2.830/6.780/ESD.63

Spring 2008

Lecture #17

Nested Variance Components

April 15, 2008

Readings/References

- D. Drain, *Statistical Methods for Industrial Process Control*, Chapter 3: Variance Components and Process Sampling Design, Chapman & Hall, New York, 1997.

Agenda

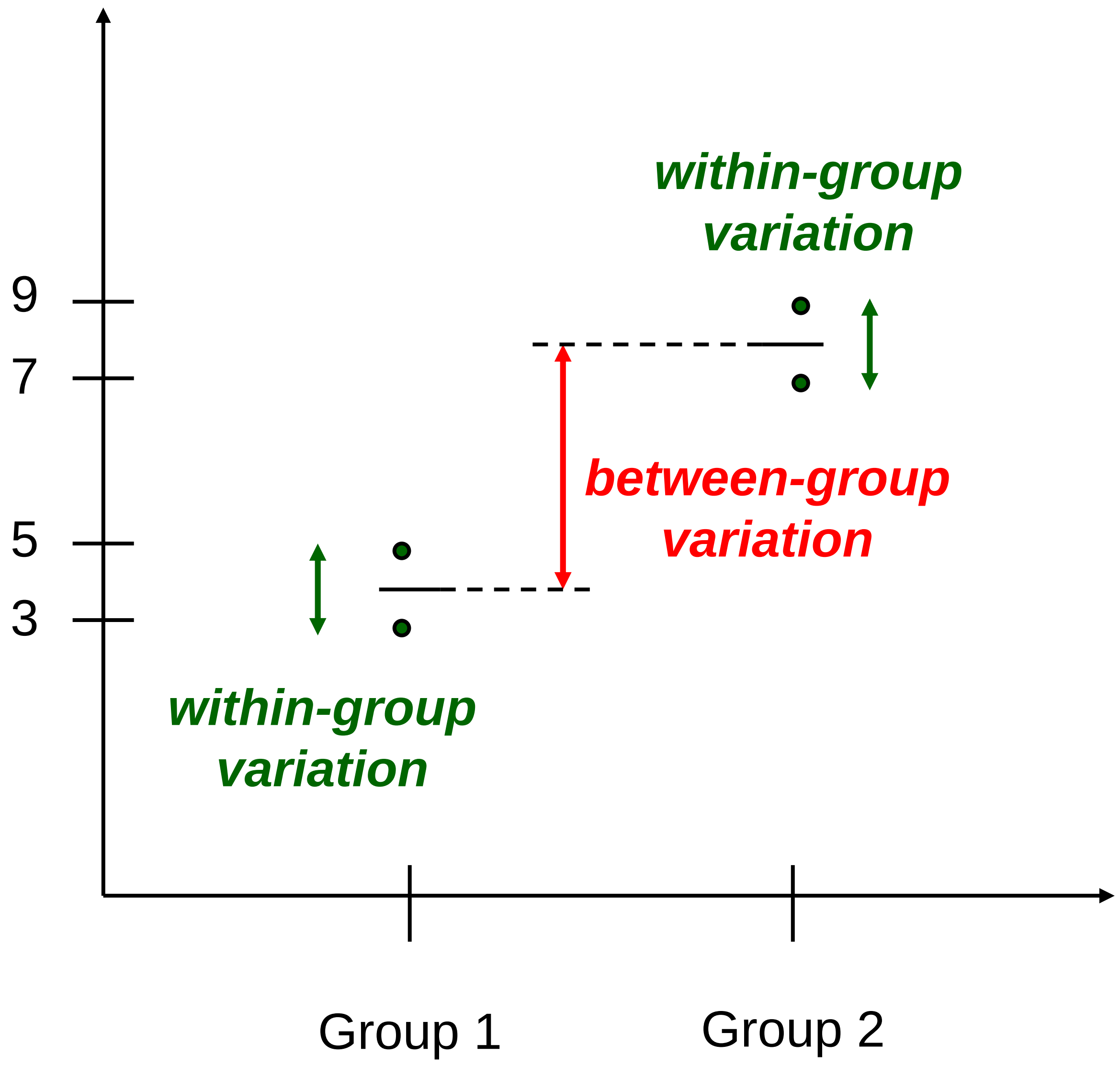
- Standard ANOVA
  - Looking for fixed effect vs. chance/sampling
- Nested variance structures
  - More than one zero-mean variance at work
  - Want to estimate these variances
- Examples
  - Based on simple ANOVA
  - Two-level example (from Drain)
  - Three-level example (from Drain)
- Implications for design of sampling and experimental plans

vs.

Standard Analysis of Variance (ANOVA)

- Question in single variable ANOVA:
  – Are we seeing anything other than random sampling from a single (Normal) distribution?
- Approach:
  – Estimate variance of the natural variation from observed replication for each treatment level (i.e., estimate the within-group variance)
  – Estimate the between-group variance
    - Could be due to a *fixed effect*
    - Could be due to chance (random sampling)
  – Consider probability of a ratio of these two variances as large as what was observed, if only a single (Normal) distribution is at work

ANOVA Example



- Groups are different levels of some treatment
- Goal – determine if there is a non-zero fixed-effect or not

Hypotheses in ANOVA

- Null Hypothesis: Random Sampling from Single Distribution
  - E.g. we draw multiple samples of some size
  - What range of variance ratios among these samples would we expect to see purely by chance?
  - Assumed model:

$$x_i = \mu + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

- Fixed Effects Model
  - The alternative hypothesis is that there is a fixed effect between the treatment groups (where *i* indicates group, and *j* indicates replicate within group)
  - Assumed model:

$$x_{j(i)} = \mu + t_i + \epsilon_j$$

$$\epsilon_j \sim N(0, \sigma^2)$$

Some Definitions (for ANOVA Calculations)

- Deviations from grand mean
  - Individual data point from grand mean: $x_{ij} - \bar{\bar{x}}$
  - Squared deviation of point from grand mean: $S_D = (x_{ij} - \bar{\bar{x}})^2$
  - Sum of squared deviations from grand mean: $SS_D = \sum_{ij} (x_i - \bar{\bar{x}})^2$
- Deviations of group mean from grand mean
  - Deviation of group *i* mean from grand mean: $\bar{x}_i - \bar{\bar{x}}$
  - Squared dev of group mean from grand mean: $S_G = (\bar{x}_i - \bar{\bar{x}})^2$
  - Sum of squared deviations of group means: $SS_G = \sum_i \sum_{j(i)} (\bar{x}_i - \bar{\bar{x}})^2$
- Deviations from local group mean
  - Deviation of individual point *j* (within group *i*) from the group mean: $x_{j(i)} - \bar{x}_i$
  - Squared deviation from group mean: $S_E = (x_{j(i)} - \bar{x}_i)^2$
  - Sum of squared deviations from group mean: $SS_E = \sum_i \sum_{j(i)} (x_{j(i)} - \bar{x}_i)^2$

Simple ANOVA Example

Group	Value	Group Ave	Squared devs of point from grand ave S_D	Squared devs of group ave from grand ave S_G	Squared devs of point from group ave S_E
1	3	4	9	4	1
	5	4	1	4	1
2	7	8	1	4	1
	9	8	9	4	1
Grand Ave	6				
Grand Var	6.67				
		SS_D =	20		
			SS_G =	16	
				SS_E =	4

ANOVA

Source	d.o.f.	SS	MS	F	Pr > F
C TOTAL	3	20.00	6.67		
GROUP	1	16.00	16.00	8.00	0.11
ERROR	2	4.00	2.00		

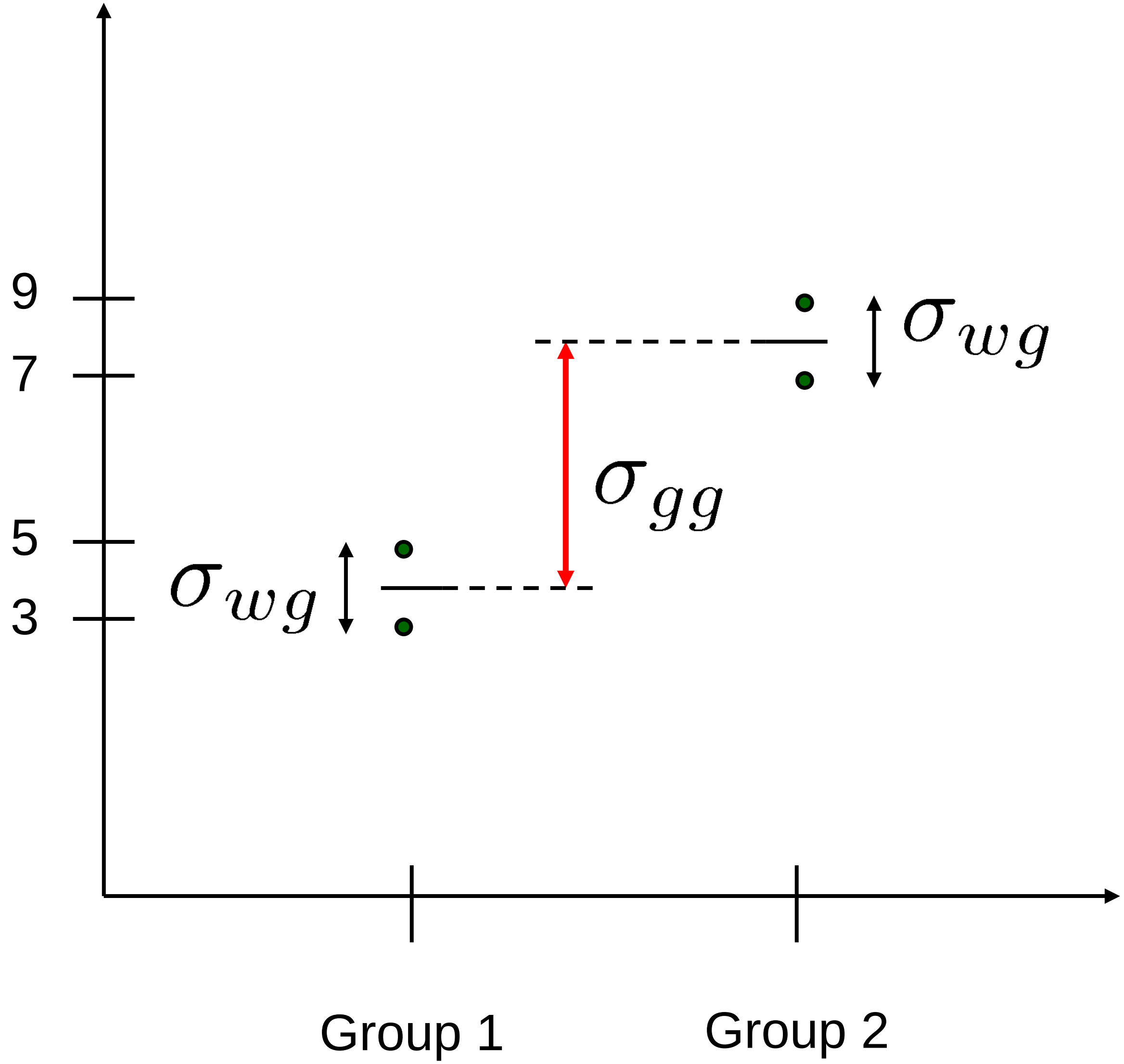
Nested Variance Structure

- Two different, independent sources of variation
  - "within group" variance (σ_{wg}^2)
  - "between group" variance (σ_{gg}^2)
- Assumed model: $x_{ij} = \mu + G_i + \epsilon_{j(i)}$

$$
\begin{aligned}
G_i &\sim N(0, \sigma_{gg}^2) \\
\epsilon_{j(i)} &\sim N(0, \sigma_{wg}^2)
\end{aligned}
$$

  - Key difference from standard ANOVA:
    - This does NOT postulate a fixed effect (mean offset) between groups
    - Rather, *random* group offset (still zero mean), the same for all members within that group *j*

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

Estimating Variances – A Naïve Attempt

- Within-group variance
  - Use Error Mean Square from ANOVA:
    - $\boldsymbol{\sigma_{wg}^2} = 2.0$
- Between-group variance
  - Use Group Mean Square from ANOVA:
    - $\boldsymbol{\sigma_{gg}^2} = 16.0$

WRONG!

- Total variance
  - Use CTOT Mean Square from ANOVA:
    - $\boldsymbol{\sigma_T^2} = 6.67$

WRONG!

Where Do Nested Variance Structures Arise?

- Typically occur in batch or parallel manufacturing processes
- Very common in semiconductor manufacturing
  - multiple chips or die within a wafer,
    multiple wafers within a lot,
    multiple lots within a batch
  - Physical causes of variation at each level are typically *different*
- Our Goal:
  - Point estimates for each source of variation
  - Confidence intervals for each variance

Nested Structures

- Items within a group tend to be more similar to each other:
  - Measure film thickness on a wafer
    - $T(x,y) \sim N(\mu, \sigma^2_{\text{within-wafer}})$ a reasonable model?
    - E.g. arise due to uniformity of temperature, gas flows within a particular deposition chamber
  - Measure film thickness averages on multiple wafers
    - $T_{ave} \sim N(\mu, \sigma^2_{\text{wafer-to-wafer}})$ a reasonable model?
    - E.g. may arise due to run-to-run repeatability of the tool as a whole

Single Level Variance Structure

- Multiple measurements on same wafer

$$X_i = \mu + M_i$$

$$M_i \quad \sim \quad N(0, \sigma_M^2)$$

- M_i indicates measurement
  - measurement location is randomly selected
  - each measurement is IIND
    - Independent & Identically Normally Distributed
    - zero mean, variance = σ^2_M

Two Level Variance Structure

- Multiple measurements on multiple wafers

$$X_{ij} = \mu + W_i + M_{j(i)}$$

$$\begin{aligned} W_i &\sim N(0, \sigma_W^2) \quad \text{for} \quad i = 1 \cdots W \\ M_{j(i)} &\sim N(0, \sigma_M^2) \quad \text{for} \quad j = 1 \cdots M \end{aligned}$$

- W_i indicates wafer
  - wafer selected at random from wafer group
  - each wafer mean is assumed to be IIND as above
- $M_{j(i)}$ indicates measurements within wafer i
  - measurement location is randomly selected
  - each measurement is IIND as above

Total Variance (for Individual Measurement)

- Variance components add

$$\mathrm{Var}[X_{ij}] = \mathrm{Var}[\mu] + \mathrm{Var}[W_i] + \mathrm{Var}[M_{j(i)}]$$

$$\sigma_T^2 = \sigma_W^2 + \sigma_M^2$$

  – Individual variances are assumed independent

Note: this relationship did not hold
in naïve attempt!

Variance in *Observed* Averages

- **Key Idea:** the variance observed for the wafer average will NOT be equal to the true wafer to wafer variance, due to additional measurement variance and sampling:

$$\sigma^2_{\bar{W}} = \sigma^2_W + \frac{\sigma^2_M}{M}$$

I.e., wafer average is inflated by the measurement variance

- Thus, if we want to estimate the actual wafer-to-wafer variance:

$$\sigma^2_W = \sigma^2_{\bar{W}} - \frac{\sigma^2_M}{M}$$

Derivation: Variance in *Observed* Averages

- Observed wafer average for wafer *i:*

$$
\begin{aligned}
\bar{W}_i &= \frac{1}{M}\sum_{j=1}^{M} X_{ij} = \frac{1}{M}\sum_{j=1}^{M}(\mu + W_i + M_{j(i)}) \\
&= \frac{1}{M}(M\mu + MW_i + \sum_{j=1}^{M} M_{j(i)}) \\
&= \mu + W_i + \frac{1}{M}\sum_{j=1}^{M} M_{j(i)}
\end{aligned}
$$

- So variance in observed wafer averages:

$$
\mathrm{Var}[\bar{W}_i] = \mathrm{Var}[\mu] + \mathrm{Var}[W_i] + \frac{1}{M^2} \cdot M \cdot \mathrm{Var}[M_{j(i)}]
$$

$$
\sigma_{\bar{W}}^2 = \sigma_W^2 + \frac{\sigma_M^2}{M}
$$

Note: *Observed* Total Variance is Always Smaller than *Estimated* Total Variance

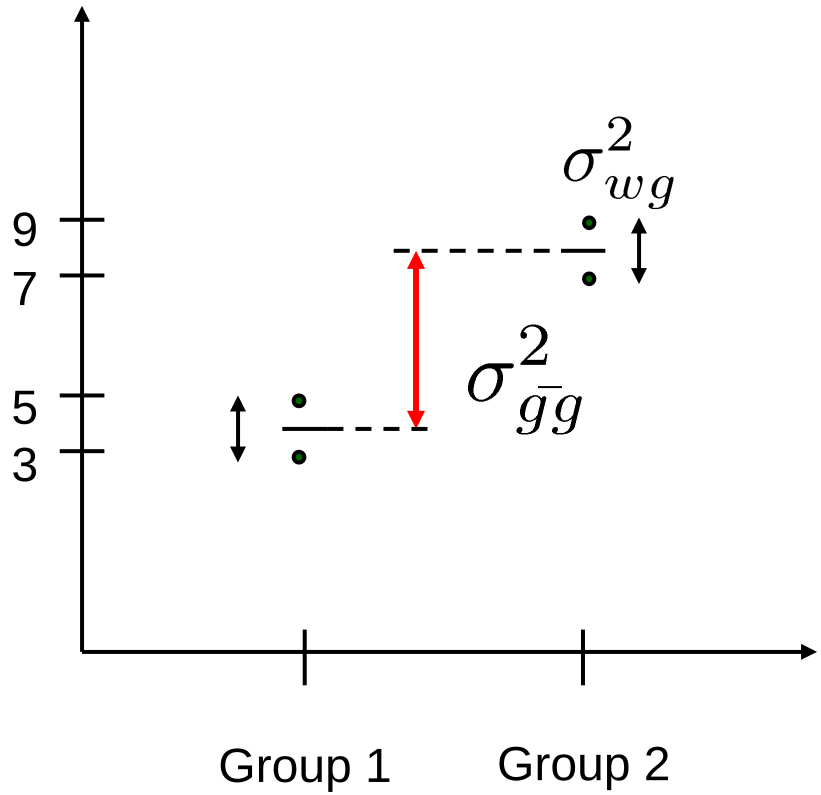
- We assume two independent sources of variation are at work, so estimated total variance is:

$$\sigma_T^2 = \sigma_W^2 + \sigma_M^2$$

- The "observed" total variance has sampling effects in it, making it smaller than actual total variance:

$$\begin{aligned}\sigma^2_{T_{\text{observed}}} &= \frac{(W-1)\cdot\sigma_W^2 + W(M-1)\cdot\sigma_M^2}{WM-1} = \frac{SS_E}{N-1} \\ &= \underbrace{\frac{W-1}{WM-1}}_{<1}\cdot\sigma_W^2 + \underbrace{\frac{WM-W}{WM-1}}_{<1}\cdot\sigma_M^2 \quad < \quad \sigma_W^2 + \sigma_M^2\end{aligned}$$

Back to Simple Nested Variance Example



- Within-group variance

$$\sigma^2_{wg} = 2.0$$

- Observed group-group variance

$$\begin{aligned} \sigma^2_{\bar{g}g} &= ((8-6)^2 + (6-4)^2)/1 \\ &= 8.0 \end{aligned}$$

- Estimated actual group-group variance

$$\begin{aligned} \sigma^2_{gg} &= \sigma^2_{\bar{g}g} - \frac{\sigma^2_{wg}}{2} \\ &= 7.0 \end{aligned}$$

M = 2 measurements in each group

- Estimated total variance

$$\begin{aligned} \sigma^2_T &= \sigma^2_{gg} + \sigma^2_{wg} \\ &= 9.0 \end{aligned}$$

Example: Resistivity across Multiple Wafers

Wafer	Resistivity Measurement
1	47.85 46.48 47.68
2	55.97 55.67 56.26
3	48.43 50.39 50.86
4	47.45 49.49 45.81
5	47.12 47.43 48.73
6	51.09 49.04 47.72

Figure by MIT OpenCourseWare.

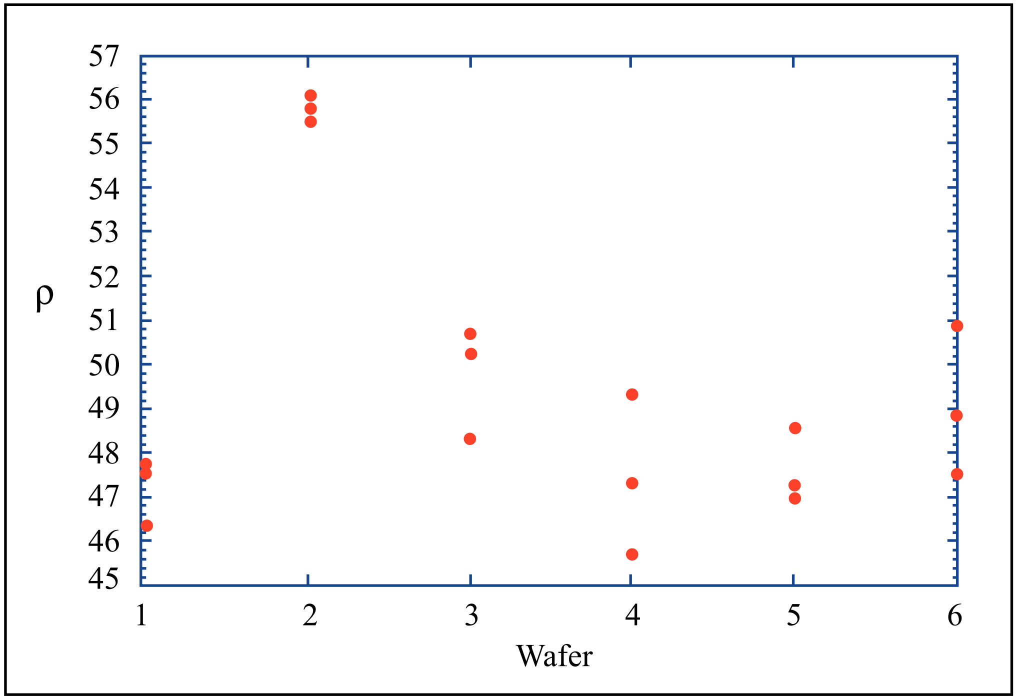


Figure by MIT OpenCourseWare.

- Same process for all wafers
  - Not introducing different 'treatments"
- Three measurements (randomly chosen on each wafer) of resistivity

Ref: Drain, p. 196

Example: Resistivity across Multiple Wafers (2)

- Nested variance ANOVA results from Drain

Variance Source	*Degrees of Freedom*	*Sum of Squares*	*F Value*	*Pr > F*	*Error Term*
Total	17	178.499361			
Wafer	05	159.863028	20.5873	0.000017	Error
Error	12	18.636333			

Variance Source	*Mean Square*	*Variance Component*	*Percent of Total*
Total	10.499962	11.692887	100.0000
Wafer	31.972606	10.139859	86.7182
Error	1.553028	1.553028	13.2818

Observed (Mean Square) **Estimated** (Variance Component)

Figure by MIT OpenCourseWare.

- Based on "SAS PROC NESTED"
  - What does it mean?
  - How did he do that? … See spreadsheet

Interval Estimates on Variance Components

Variance Source	*Lower Limit*	*Point Estimate*	*Upper Limit*
Total	5.83196	11.692887	47.5806
Wafer	4.27950	10.139859	45.9730
Error	0.88634	1.553028	3.56606

Figure by MIT OpenCourseWare.

- From Drain
  - Error (site-to-site variance): use Chi-square distribution
    - Claims to be 95% c.i., … but table shows 90% c.i.
  - Wafer (wafer-to-wafer variance):
    - Not sure what relationship used for c.i. calculation by Drain (SAS PROC NESTED). See spreadsheet for conservative approach.

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$$

Three Level Variance Structure

- Multiple measurements on multiple wafers in multiple lots

$$X_i = \mu + L_i + W_{j(i)} + M_{k(ij)}$$

$$\begin{aligned} L_i &\sim N(0, \sigma_L^2) \ \text{ for } \ i = 1 \cdots L \\ W_{j(i)} &\sim N(0, \sigma_W^2) \ \text{ for } \ j = 1 \cdots W \\ M_{k(ij)} &\sim N(0, \sigma_M^2) \ \text{ for } \ k = 1 \cdots M \end{aligned}$$

- L_i indicates lot
  - lot selected at random from set of lots
  - each lot mean is assumed to be IIND as above
- $W_{j(i)}$ indicates wafer *j* within lot *i*
- $M_{k(ij)}$ indicates measurement *k* within wafer *j* within lot *i*

Variance in *Observed* Averages, Three Levels

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

$$\sigma^2_{\bar{L}} = \sigma^2_L + \frac{\sigma^2_W}{W} + \frac{\sigma^2_M}{MW}$$

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

Three Level Example

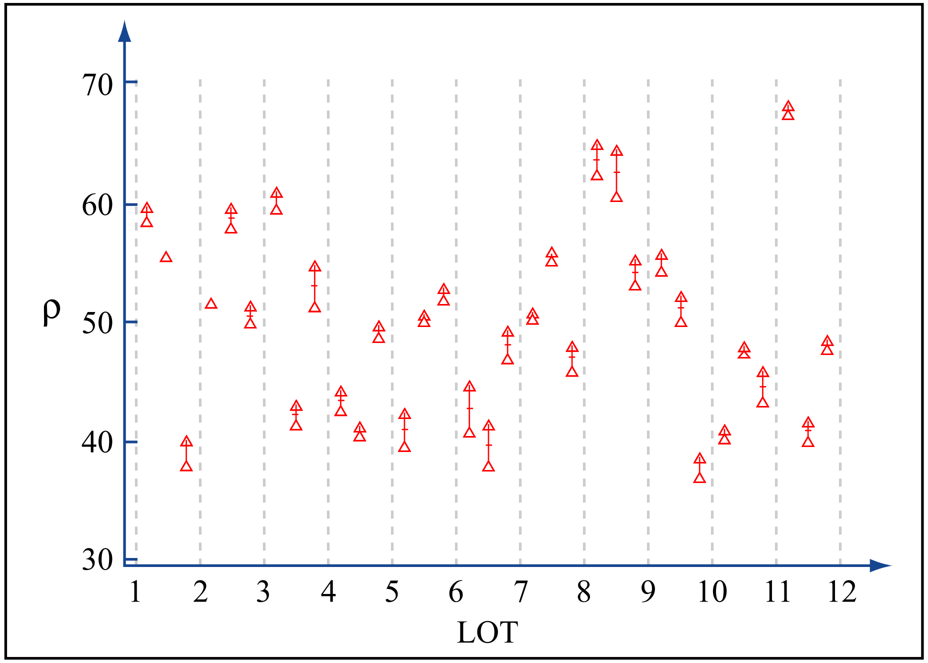


Figure by MIT OpenCourseWare.

- 11 Lots
- 3 Wafers within each lot
- 2 Measurements within each wafer

Ref: Drain, p. 198

Three Level Analysis – Point Estimates

Variance Source	*Degrees of Freedom*	*Sum of Squares*	*F Value*	*Pr > F*	*Error Term*
Total	65	4025.487062			
Lot	10	1453.333712	1.27299	0.303499	Wafer
Wafer	22	2511.673500	62.2936	0.000000	Error
Error	33	60.479850			

Variance Source	*Mean Square*	*Variance Component*	*Percent of Total*
Total	61.930570	63.194249	100.0000
Lot	145.333371	5.194399	8.2197
Wafer	114.166977	56.167127	88.8801
Error	1.832723	1.832723	2.9000

Figure by MIT OpenCourseWare.

- See spreadsheet example. Several tricky parts!

Three Level Analysis – Interval Estimates

Variance Source	*Lower Limit*	*Point Estimate*	*Upper Limit*
Total	44.4215	63.194249	127.324
Lot	-94.8509	5.194399	222.068
Wafer	33.2254	56.167127	113.423
Error	1.19231	1.832723	3.17535

Figure by MIT OpenCourseWare.

- "Negative" variance – set to lower bound of zero

Outer vs. Inner Levels of Variance

- When we observe/calculate an outer (higher) level average, what most strongly affects this?

$$\sigma^2_{\bar{\bar{L}}} = \sigma^2_L + \frac{\sigma^2_W}{W} + \frac{\sigma^2_M}{MW}$$

  – With appreciable number of wafers and measurements, the inner levels of variance are "averaged away"

Why worry about $\sigma_{\bar{x}}^2$?

- Often make decisions based on estimates for the true outer-level average
- One approach:
  - Calculate/observe multiple averages empirically
  - Use these to estimate variance in the mean
    - E.g. confidence interval on average

$$\bar{x} - z_{\alpha/2} \cdot \sigma_{\bar{x}} \;\leq\; \mu \;\leq\; \bar{x} + z_{\alpha/2} \cdot \sigma_{\bar{x}}$$

    - Or with small number of samples

$$\bar{x} - t_{\alpha/2} \cdot s_{\bar{x}} \;\leq\; \mu \;\leq\; \bar{x} + t_{\alpha/2} \cdot s_{\bar{x}}$$

- So… want sampling plans to minimize $\sigma_{\bar{x}}^2$

Implication: Sampling in Nested Cases

- Suppose we have a limited set of resources (e.g. lots, wafers, measurement), or given cost constraints
  - Use variance estimates to decide how to nest the measurements
  - If estimating an outer level value, e.g. lot average, we can often improve variance estimate by replicating at the outer rather than inner levels (i.e. increase *W* rather than *M*)

$$\sigma^2_{\bar{L}} = \sigma^2_L + \frac{\sigma^2_W}{W} + \frac{\sigma^2_M}{MW}$$

Summary

- **Nested Variance Structures**
  - When have sets of measurements “within” another spatial construct
  - Assumes independent sources of variance
- **Variance Components**
  - Unwrap variances from inside toward outside
  - Point and interval estimates possible
- **Implications in sampling plan design**
  - Allocate measurements, replications where most valuable for variance being estimated