MIT 2.830/6.780/ESD.63
Control of Manufacturing Processes

Introduction to Analysis of Variance: a tool for assessing input-output relationships
Have focused so far on interpreting output
## Review of tools for interpreting outputs

<table>
<thead>
<tr>
<th>Tool</th>
<th># outputs</th>
<th># samples</th>
<th># inputs</th>
<th>Levels per input</th>
</tr>
</thead>
<tbody>
<tr>
<td>t, F tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>control charts, cusum, EWMA etc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2$, $T^2$ charts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Also talked about yield modeling and process capability, but need ways of *modeling* and thus *improving* processes
Injection molding data

IM Run data

Levels changes:
- v - low
- t - low
- v - high
- t - low
- v - low
- t - high
- v - high
- t - high

Diameter (in)
Want to start relating input(s) to output(s)

\[ Y = \Phi(\alpha) \]

\[ \alpha \equiv \text{process parameters} \]

\[ \Delta Y = \frac{\partial Y}{\partial \alpha} \Delta \alpha + \frac{\partial Y}{\partial u} \Delta u \]
What is our goal?

- **Developing a process model**
  - Relating inputs and disturbances to outputs
  - Determining significance of the input effect
    - Does it really matter?

- **Process optimization**
  - Max (Cpk) or Min (QLF)
  - Models for mean shifting
  - Models for variance reduction
Empirical Modeling

• What is the objective?
• What is the output?
• What are the input(s)?
• What do we want to vary?
• What model form should we use?
  – $Y = \Phi(\alpha, u)$ is not specific!
• How many data can we take?
First step: determining which inputs matter

<table>
<thead>
<tr>
<th>Tool</th>
<th># inputs</th>
<th>Levels per input</th>
<th># samples</th>
<th># outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>t, F tests</td>
<td>?</td>
<td>? (2?)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>control charts, cusum</td>
<td>?</td>
<td>? (2?)</td>
<td>many</td>
<td>1</td>
</tr>
<tr>
<td>$\chi^2$, $T^2$ charts</td>
<td>?</td>
<td>? (2?)</td>
<td>many</td>
<td>2</td>
</tr>
<tr>
<td>Analysis of variance</td>
<td>$\geq 1$</td>
<td>$\geq 2$</td>
<td>$\geq 2$</td>
<td>1</td>
</tr>
</tbody>
</table>
Agenda

1. Comparison of treatments (one variable)
   - Fixed effects model
   - Analysis of Variance (ANOVA) technique
   - Example

2. Multivariate analysis of variance
   - Model forms
   - MANOVA technique
Comparison of Treatments

- Consider multiple conditions (treatments, settings for some variable)
  - There is an overall mean $\mu$ and real “effects” or deltas between conditions $\tau_i$.
  - We observe samples at each condition of interest

- Key question: are the observed differences in mean “significant”?
  - Typical assumption (should be checked): the underlying variances are all the same – usually an unknown value ($\sigma_0^2$)
The ANOVA approach assumes a simple mathematical model:

\[ y_{ti} = \mu + \tau_t + \epsilon_{ti} \]

\[ = \mu_t + \epsilon_{ti} \]

Where \( \mu_t \) is the treatment mean (for treatment type \( t \))
And \( \tau_t \) is the treatment effect
With \( \epsilon_{ti} \) being zero mean normal residuals \( \sim N(0,\sigma_0^2) \)
Steps/Issues in Analysis of Variance

1. Within-group variation
   - Estimate underlying population variance

2. Between-group variation
   - Estimate group to group variance

3. Compare the two estimates of variance
   - If there is a difference between the different treatments, then the between group variation estimate will be *inflated* compared to the within group estimate
   - We will be able to establish confidence in whether or not observed differences between treatments are significant

Hint: we’ll be using *F* tests to look at ratios of variances

(1) Within Group Variation

- Assume that each group is normally distributed and shares a common variance $\sigma_0^2$
- $SS_t = \text{sum of square deviations within } t^{th} \text{ group (there are } k \text{ groups)}$
  \[ SS_t = \sum_{i=1}^{n_t} (y_{ti} - \bar{y}_t)^2 \text{ where } n_t \text{ is number of samples in treatment } t \]
- Estimate of within group variance in $t^{th}$ group (just variance formula)
  \[ s_t^2 = \frac{SS_t}{\nu_t} = \frac{SS_t}{n_t - 1} \text{ where } \nu_t \text{ is d.o.f. in treatment } t \]
(2) Between Group Variation

• We will be testing hypothesis $\mu_1 = \mu_2 = \ldots = \mu_k$

• If all the means are in fact equal, then a $2^{nd}$ estimate of $\sigma^2$ could be formed based on the observed differences between group means:

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

where $n_t$ is number of samples in treatment $t$, and $k$ is the number of different treatments.

(3) Compare Variance Estimates

• We now have two different possibilities for $s_T^2$, depending on whether the observed sample mean differences are “real” or are just occurring by chance (by sampling)

• Use $F$ statistic to see if the ratios of these variances are likely to have occurred by chance!

• Formal test for significance:

  \[
  \text{Reject } H_0 \quad (H_0 : \text{no mean difference})
  \]

  \[
  \text{if } \frac{s_T^2}{s_R^2} \text{ is significantly greater than 1.}
  \]
(4) Compute Significance Level

• Calculate observed $F$ ratio (with appropriate degrees of freedom in numerator and denominator)

• Use $F$ distribution to find how likely a ratio this large is to have occurred by chance alone
  – This is our “significance level”
  – Define observed ratio: $F_0 = \frac{s^2_T}{s^2_R}$
  – If $F_0 > F_{\alpha,k-1,N-k}$
    then we say that the mean differences or treatment effects are significant to $(1-\alpha)100\%$ confidence or better
(5) Variance Due to Treatment Effects

• We also want to estimate the sum of squared deviations from the grand mean among all samples:

\[ SS_D = \sum_{t=1}^{k} \sum_{i=1}^{n_t} (y_{ti} - \bar{y})^2 \]

\[ s_D^2 = SS_D / \nu_D = \frac{SS_D}{N - 1} = MS_D \]

where \( N \) is the total number of measurements
### (6) Results: The ANOVA Table

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>$F_0$</th>
<th>$Pr(F_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between treatments</td>
<td>$SS_T$</td>
<td>$k - 1$</td>
<td>$s^2_T = \frac{SS_T}{k-1}$</td>
<td>$\frac{s^2_T}{s^2_R}$</td>
<td>Table</td>
</tr>
<tr>
<td>Also referred to as “residual” SS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within treatments</td>
<td>$SS_R$</td>
<td>$N - k$</td>
<td>$s^2_R = \frac{SS_R}{N-k}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total about the grand average</td>
<td>$SS_D$</td>
<td>$N - 1$</td>
<td>$s^2_D = \frac{SS_D}{N-1}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$SS_D = SS_T + SS_R$

$\nu_D = \nu_T + \nu_R$
Example: Anova

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>6</td>
<td>11</td>
</tr>
</tbody>
</table>

Excel: Data Analysis, One-Variation Anova

### SUMMARY

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>33</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>24</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>33</td>
<td>11</td>
<td>1</td>
</tr>
</tbody>
</table>

### ANOVA

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>18</td>
<td>2</td>
<td>9</td>
<td>4.5</td>
<td>0.064</td>
<td>5.14</td>
</tr>
<tr>
<td>Within Groups</td>
<td>12</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
F = \frac{S_{R}^{2}}{S_{T}^{2}} = \frac{9}{2} = 4.5
\]

\[
F_{0.05.2.6} = 5.14
\]

\[
F_{0.10.2.6} = 3.46
\]

\[
SS_{1} = (12 - 11)^2 + (11 - 11)^2 + (10 - 11)^2 = 2
\]

\[
SS_{2} = 2^2 + 0^2 + 2^2 = 8
\]

\[
SS_{3} = 1^2 + 0^2 + 1^2 = 2
\]

\[
s_{1}^{2} = MS_{1} = SS_{1}/2 = 2/2 = 1
\]

\[
s_{2}^{2} = MS_{2} = 8/2 = 4
\]

\[
s_{3}^{2} = MS_{3} = 2/2 = 1
\]

\[
S_{R}^{2} = \frac{SS_{1} + SS_{2} + SS_{3}}{N - k} = \frac{12}{6} = 2
\]

\[
S_{T}^{2} = \frac{3(11-10)^2 + 3(8-10)^2 + 3(11-10)^2}{3-1} = \frac{18}{2} = 9
\]
ANOVA – residuals assumed $\sim N(0, \sigma_0^2)$ for every treatment

• Checks
  – Plot residuals against time order
  – Examine distribution of residuals: should be IID, Normal
  – Plot residuals vs. estimates
  – Plot residuals vs. other variables of interest
MANOVA – Two Dependencies

• Can extend to two (or more) variables of interest. MANOVA assumes a mathematical model, again simply capturing the means (or treatment offsets) for each discrete variable level:

\[ y_{tqi} = \mu + \tau_t + \beta_q + \epsilon_{tqi} \]

^ indicates estimates:

\[ \hat{y}_{tq} = \hat{\mu} + \hat{\tau}_t + \hat{\beta}_q \]

\[
\begin{align*}
\text{# model coeffs} & = 1 + k + n \\
\text{# independent model coeffs} & = 1 + (k - 1) + (n - 1)
\end{align*}
\]

Recall that our \( \hat{\tau}_t \) are not all independent model coefficients, because \( \sum \tau_t = 0 \). Thus we really only have \( k - 1 \) independent model coeffs, or \( \nu_t = k - 1 \).

• Assumes that the effects from the two variables are additive
MANOVA – Two Factors with Interactions

• May be interaction: not simply additive – effects may depend synergistically on both factors:

$$y_{tqi} = \mu_{tq} + \epsilon_{tqi}$$

IID, ~N(0, \sigma^2)

An effect that depends on both t & q factors simultaneously

t = first factor = 1, 2, … k  \quad (k = \# \text{ levels of first factor})
q = second factor = 1, 2, … n  \quad (n = \# \text{ levels of second factor})
i = replication = 1, 2, … m  \quad (m = \# \text{ replications at t, q^{th} combination of factor levels})

• Can split out the model more explicitly…

Estimate by:

$$\hat{y}_{tq} = \bar{y} + (\bar{y}_t - \bar{y}) + (\bar{y}_q - \bar{y}) + (\bar{y}_{tq} - \bar{y}_t - \bar{y}_q + \bar{y})$$

$$\omega_{tq} = \text{interaction effects} = (\bar{y}_{tq} - \bar{y}_t - \bar{y}_q + \bar{y})$$

$$\tau_t, \beta_q = \text{main effects}$$
MANOVA – Two Factors with Interactions

\[ S_T^2 = \frac{\sum_{t=1}^{k} m_t n_t (\bar{y}_t - \bar{y})^2}{k - 1} \]
\[ S_B^2 = \frac{\sum_{q=1}^{n} m_q k_q (\bar{y}_q - \bar{y})^2}{n - 1} \]
\[ S_I^2 = \frac{\sum_{q=1}^{n} \sum_{t=1}^{k} m_{tq} (\bar{y}_{tq} - \bar{y}_t - \bar{y}_q + \bar{y})^2}{(k - 1)(n - 1)} \]
\[ S_E^2 = \frac{\sum_{q=1}^{n} \sum_{i=1}^{m} \sum_{t=1}^{k} \sum_{t=1}^{k} (y_{tqi} - \bar{y}_{tq})^2}{nk(m - 1)} \]
## MANOVA Table – Two Way with Interactions

<table>
<thead>
<tr>
<th>source of variation</th>
<th>sum of squares</th>
<th>degrees of freedom</th>
<th>mean square</th>
<th>$F_0$</th>
<th>$Pr(F_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between levels of factor 1 (T)</td>
<td>$SS_T$</td>
<td>$k - 1$</td>
<td>$s^2_T$</td>
<td>$s^2_T/s^2_E$</td>
<td>table</td>
</tr>
<tr>
<td>Between levels of factor 2 (B)</td>
<td>$SS_B$</td>
<td>$n - 1$</td>
<td>$s^2_B$</td>
<td>$s^2_B/s^2_E$</td>
<td>table</td>
</tr>
<tr>
<td>Interaction</td>
<td>$SS_I$</td>
<td>$(k - 1)(n - 1)$</td>
<td>$s^2_I$</td>
<td>$s^2_I/s^2_E$</td>
<td>table</td>
</tr>
<tr>
<td>Within Groups (Error)</td>
<td>$SS_E$</td>
<td>$nk(m - 1)$</td>
<td>$s^2_E$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total about the grand average</td>
<td>$SS_D$</td>
<td>$nkm - 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example: plasma metal etch nonuniformity (lateral etch)

200-mm wafer

Average resistances of all structures in a ‘flash’ (Ohms)

Each flash has many copies of a feature set, with different ‘padding’ densities

Each feature set has a range of line/space widths; can be electrically probed
Relevant factors

- **Geometry**
  - Position on wafer
  - Locally averaged pattern density
  - Feature size and pitch

- **Physical perspective**
  - Reactant fluxes in etch chamber
  - ...

Pattern density dependency

Resistances averaged over all features up to 1 micron linewidth, and all flashes

Average resistance of a snake feature (Ohms)

Local metal pattern density (%)
Wafer-scale nonuniformity

Resistance of a particular snake feature (Ohms)

Flash index starts

Local pattern density = 85%

Local pattern density = 5%

‘Flash’ location on wafer
Next time

• Building models based on effects