Design and Analysis of Experiments

04 - Interpretation of results

Ahmed A.(Msc)

DIRE DAWA UNIVERSITY
COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES
DEPARTMENT OF SATATISTICS

October 15, 2024

Outline

1. **Interpretation of results**
   - Comparison among treatment means
   - Graphical comparison of means
   - Contrasts
   - Comparing pairs of treatment means
   - Comparing treatment means with control

Interpretation of results

- After conducting the experiment, performing the statistical analysis, and investigating the underlying assumptions, the experimenter is ready to draw practical conclusions about the problem under study.
- Often this is relatively easy, particularly in simple experiments, where conclusions might be drawn informally, perhaps by inspecting graphical displays such as box plots and scatter diagrams.
- However, in some cases, more formal techniques need to be applied.

Regression Model

- Factors in an experiment can be either quantitative or qualitative.
- Quantitative factors: levels associated with points on a numerical scale (e.g., temperature, time).
- Qualitative factors: levels without meaningful numerical order (e.g., operators, batches).
- Both types are initially treated identically in design and analysis.

Prediction and Interpolation

1. When dealing with quantitative factors, the experimenter is often interested in predicting responses at intermediate levels.
2. For instance, with levels of 1.0, 2.0, and 3.0 hours, predicting the response at 2.5 hours may be desired.
3. This involves developing an empirical model, often using regression analysis.

Regression Analysis

- Regression analysis is used to fit empirical models to the data.
- The method of least squares is commonly applied to estimate parameters.

Scatter Diagrams and Model Fitting

Figure 3.10 presents scatter diagrams of etch rate y versus power x for the experiment in Example 3.1. A strong relationship is observed. We can first attempt to fit a linear model:

$$y = \beta_0 + \beta_1 x + \epsilon$$

where β_0 and β_1 are unknown parameters and ϵ is a random error term. The least squares fit yields:

$$y = 137.62 + 2.527x$$

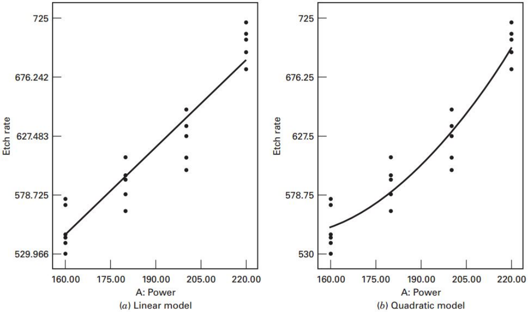
Quadratic Model

At higher power settings, the linear model may not be satisfactory. A quadratic term can be added:

$$y = 1147.77 - 8.2555x + 0.028375x^2$$

This quadratic model provides a better fit for the data at higher power levels.

Scatter Diagrams and Model Fitting



■ **FIGURE 3.10** Scatter diagrams and regression models for the etch rate data of Example 3.1

Comparison among treatment means

- Analysis of variance for fixed effects model
- Rejection of null hypothesis implies differences between treatment means
- Further comparisons and analysis may be useful

Definitions

- ith treatment mean: μ_i
- Estimate of μ_i: $\bar{y}_i$
- Comparisons made using treatment totals $\{y_{i.}\}$ or averages $\{\bar{y}_i\}$

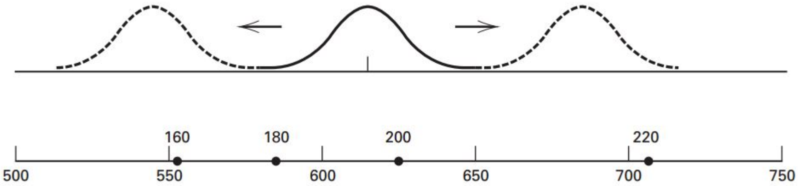
Multiple Comparison Methods

- Used for comparing individual treatment means or groups of means
- Various methods available (to be discussed in following sections)

Graphical Comparisons of Means

- Easy to develop graphical procedure
- Consider factor with a levels and treatment averages $\{\bar{y}_i\}$
- If all means are identical, $\{\bar{y}_i\}$ should behave like random samples from $N(\mu, \sigma/\sqrt{n})$
- Replace unknown σ with $\sqrt{MS_E}$ from ANOVA
- Use t-distribution with scale factor $\sqrt{MS_E/n}$
- Plot treatment means and overlay scaled t-distribution

Graphical Comparisons of Means



■ **FIGURE 3.11** Etch rate averages from Example 3.1 in relation to a t distribution with scale factor $\sqrt{MS_E/n} = \sqrt{330.70/5} = 8.13$

Example: Etch Rate Data

- Refer to Figure 3.11 (not included in this presentation)
- t-distribution origin at 615 Å/min
- Scale factor: $\sqrt{MS_E/n} = \sqrt{330.70/5} = 8.13$

Interpretation of Graphical Method

- No single position where all means fit the distribution
- Implies not all means are equal (confirms ANOVA results)
- Suggests all four power levels (160, 180, 200, 220 W) produce different mean etch rates

Conclusion

- Graphical method is simple but effective
- Useful for many multiple comparison problems
- More formal methods are available (to be discussed next)

Introduction to Contrasts

- Many multiple comparison methods use the idea of a **contrast**
- Example: Plasma etching experiment from Example 3.1
- Goal: Identify which power settings cause differences in etch rates

Hypothesis Testing with Contrasts

- Hypothesis 1: 200 W and 220 W produce the same etch rate

$$H_0 : \mu_3 = \mu_4$$
$$H_1 : \mu_3 \neq \mu_4$$

- Hypothesis 2: Average of lowest levels = Average of highest levels

$$H_0 : \mu_1 + \mu_2 = \mu_3 + \mu_4$$
$$H_1 : \mu_1 + \mu_2 \neq \mu_3 + \mu_4$$

Definition of a Contrast

- A contrast is a linear combination of parameters:

$$\sum_{i=1}^{a} c_i \mu_i$$

- Contrast constants $c_1, c_2, \ldots, c_a$ sum to zero: $\sum_{i=1}^{a} c_i = 0$

Expressing Hypotheses as Contrasts

$$H_0 : \sum_{i=1}^{a} c_i \mu_i = 0 \quad \text{vs} \quad H_1 : \sum_{i=1}^{a} c_i \mu_i \neq 0$$

- For $H_0 : \mu_3 = \mu_4$

$$c_1 = c_2 = 0, c_3 = +1, c_4 = -1$$

- For $H_0 : \mu_1 + \mu_2 = \mu_3 + \mu_4$

$$c_1 = c_2 = +1, c_3 = c_4 = -1$$

Testing Contrasts: Method 1 (t-test)

- Express contrast in terms of treatment averages:

$$C = \sum_{i=1}^{a} c_i \bar{y}_{i.}$$

- Variance of C (for equal sample sizes):

$$V(C) = \sigma^2 \frac{\sum_{i=1}^{a} c_i^2}{n}$$

Testing Contrasts: t-statistic

- If H_0 is true, the ratio follows a t-distribution:

$$t_0 = \frac{C}{\sqrt{MS_E \frac{\sum_{i=1}^{a} c_i^2}{n}}}$$

- MS_E is the mean square error from ANOVA
- Degrees of freedom: $N - a$

Conclusion

- Contrasts are powerful tools for multiple comparisons
- They allow testing specific hypotheses about differences among treatments
- t-test method is one way to test contrasts
- Other methods exist (to be discussed in future lectures)

Testing Contrasts: t-test Method (continued)

- Replace unknown variance σ^2 with MSE
- Use the test statistic:

$$t_0 = \frac{\sum_{i=1}^{a} c_i \bar{y}_{i.}}{\sqrt{MSE \frac{\sum_{i=1}^{a} c_i^2}{n}}}$$

- Reject H_0 if $|t_0| > t_{\alpha/2, N-a}$

Testing Contrasts: F-test Method

- Square of a t random variable with ν df is an F random variable with 1 numerator df and ν denominator df
- F-statistic:

$$F_0 = t_0^2 = \frac{\left(\sum_{i=1}^{a} c_i \bar{y}_{i.}\right)^2}{MSE \frac{\sum_{i=1}^{a} c_i^2}{n}}$$

- Reject H_0 if $F_0 > F_{\alpha,1,N-a}$

Contrast Sum of Squares

- F-statistic can be written as:

$$F_0 = \frac{MSC}{MSE}$$

- Where the contrast sum of squares is:

$$SSC = \frac{\left(\sum_{i=1}^{a} c_i \bar{y}_{i.}\right)^2}{\frac{1}{n}\sum_{i=1}^{a} c_i^2}$$

Graphical Comparisons of Means

■ **TABLE 3.11**
Analysis of Variance for the Plasma Etching Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	*P*-Value
Power setting	66,870.55	3	22,290.18	66.80	<0.001
Orthogonal contrasts					
$C_1 : \mu_1 = \mu_2$	(3276.10)	1	3276.10	9.82	<0.01
$C_2 : \mu_1 + \mu_3 = \mu_3 + \mu_4$	(46,948.05)	1	46,948.05	140.69	<0.001
$C_3 : \mu_3 = \mu_4$	(16,646.40)	1	16,646.40	49.88	<0.001
Error	5,339.20	16	333.70		
Total	72,209.75	19			

Confidence Interval for a Contrast

- Consider the contrast: $\sum_{i=1}^{a} c_i \mu_i$
- Estimate using sample means: $\sum_{i=1}^{a} c_i \bar{y}_{i.}$
- Variance of the estimate:

$$V(C) = \frac{\sigma^2}{n} \sum_{i=1}^{a} c_i^2$$

Confidence Interval Formula

- 100(1-α)% confidence interval:

$$\sum_{i=1}^{a} c_i \bar{y}_{i.} - t_{\alpha/2,N-a}\sqrt{\frac{MSE}{n}\sum_{i=1}^{a} c_i^2} \leq \sum_{i=1}^{a} c_i \mu_i$$

$$\leq \sum_{i=1}^{a} c_i \bar{y}_{i.} + t_{\alpha/2,N-a}\sqrt{\frac{MSE}{n}\sum_{i=1}^{a} c_i^2}$$

Summary

- Two methods for testing contrasts: t-test and F-test
- Both methods use the mean square error (MSE) from ANOVA
- Confidence intervals provide a range of plausible values for the contrast
- These methods allow for specific comparisons among treatment means

Comparing pairs of treatment means

- Often need to compare pairs of means after ANOVA
- Interest in contrasts of the form $\mu_i - \mu_j$ for all $i \neq j$
- Multiple methods available; we'll focus on Tukey's test

Pairwise Comparisons

- Null hypotheses: $H_0 : \mu_i = \mu_j$ for all $i \neq j$
- Alternative hypotheses: $H_1 : \mu_i \neq \mu_j$
- Goal: Control overall significance level at α

Tukey's Test

- Proposed by Tukey (1953)
- Controls experimentwise (family) error rate at α
- Exact when sample sizes are equal
- Conservative (at most α) when sample sizes are unequal

Studentized Range Statistic

- Tukey's test uses the studentized range statistic:

$$q = \frac{y_{\max} - y_{\min}}{\sqrt{MSE/n}}$$

- $y_{\max}$ and $y_{\min}$ are the largest and smallest sample means
- Distribution values found in Appendix Table VII

Tukey's Test: Equal Sample Sizes

- Declare means significantly different if:

$$|y_{i.} - y_{j.}| > T_\alpha = q_\alpha(a, f)\sqrt{\frac{MSE}{n}}$$

- Confidence intervals:

$$y_{i.} - y_{j.} - T_\alpha \leq \mu_i - \mu_j \leq y_{i.} - y_{j.} + T_\alpha$$

- $q_\alpha(a, f)$ is the upper α percentage point of q
- a is the number of means, f is degrees of freedom for MSE

Tukey's Test: Unequal Sample Sizes

- Known as Tukey-Kramer procedure
- Declare means significantly different if:

$$|y_{i.} - y_{j.}| > q_\alpha(a, f)\sqrt{\frac{MSE}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

- Confidence intervals:

$$y_{i.} - y_{j.} - T_\alpha \leq \mu_i - \mu_j \leq y_{i.} - y_{j.} + T_\alpha$$

where $T_\alpha = q_\alpha(a, f)\sqrt{\frac{MSE}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$

Advantages of Tukey's Test

- Controls experimentwise error rate
- Provides confidence intervals
- Suitable for all pairwise comparisons
- Relatively easy to compute and interpret
- Widely available in statistical software

Summary

- Tukey's test is useful for all pairwise comparisons
- Controls overall error rate at α
- Uses studentized range statistic
- Can be applied to equal or unequal sample sizes
- Provides both hypothesis tests and confidence intervals

Comparing treatment means with control

- In many experiments, one treatment is a control
- Goal: Compare each of the other $a - 1$ treatment means with the control
- Only $a - 1$ comparisons are needed
- Dunnett (1964) developed a procedure for this scenario

Dunnett's Procedure: Setup

- Assume treatment a is the control
- We want to test the hypotheses:

$$H_0 : \mu_i = \mu_a$$
$$H_1 : \mu_i \neq \mu_a$$

for $i = 1, 2, \ldots, a-1$

Dunnett's Procedure: Test Statistic

- Compute the observed differences in sample means:

$$|\bar{y}_{i.} - \bar{y}_{a.}| \quad \text{for } i = 1, 2, \ldots, a-1$$

- Reject $H_0 : \mu_i = \mu_a$ if:

$$|\bar{y}_{i.} - \bar{y}_{a.}| > d_\alpha(a-1, f)\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_a}\right)}$$

- $d_\alpha(a-1, f)$ is given in Appendix Table VIII
- f is the degrees of freedom associated with MSE

Dunnett's Procedure: Key Points

- Modification of the usual t-test
- α is the joint significance level for all $a - 1$ tests
- Controls the family-wise error rate
- Both two-sided and one-sided tests are possible
- More powerful than Tukey's test for this specific scenario

Confidence Intervals

- Dunnett's procedure can also be used to construct confidence intervals
- 100(1-α)% confidence intervals:

$$(\bar{y}_{i.} - \bar{y}_{a.}) \pm d_\alpha(a-1, f)\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_a}\right)}$$

- Simultaneous confidence level is 100(1-α)% for all $a-1$ intervals

Advantages of Dunnett's Procedure

- Specifically designed for comparisons with a control
- More powerful than general pairwise comparison methods
- Controls family-wise error rate
- Provides both hypothesis tests and confidence intervals
- Widely available in statistical software

When to Use Dunnett's Procedure

- Experiments with a clear control group
- When all comparisons are against the control
- No interest in comparisons between non-control treatments
- Examples:
  - Drug trials comparing new treatments to a placebo
  - Agricultural experiments comparing new varieties to a standard

Summary

- Dunnett's procedure is optimal for comparing treatments with a control
- Modifies t-test to account for multiple comparisons
- Controls family-wise error rate at α
- Provides both hypothesis tests and confidence intervals
- More powerful than general pairwise comparison methods for this specific scenario