Chapter 3. Experiments with a Single Factor: The Analysis of Variance

3.1 An Example

An engineer is interested in investigating the relationship between the RF power setting and the etch rate for this tool. The objective of an experiment like this is to model the relationship between etch rate and RF power, and to specify the power setting that will give a desired target etch rate. She is interested in a particular gas (C_2F_6) and gap (0.80 cm) and wants to test four levels of RF power: 160, 180, 200, 220 W. She decided to test five wafers at each level of RF power.

This is an example of a single-factor experiment with $a = 4$ **levels** of the factor and $n = 5$ **replicates**. The $4 \times 5 = 20$ runs should be made in random order.

Table 3.1 Etch Rate Data (in Å/min) from the Plasma Etching Experiment

Power	observations						
(W)	1	2	3	4	5	Totals	Averages
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707.0

The appropriate procedure for testing the equality of several means is the **analysis of variance**.

3.2 The Analysis of Variance

Suppose we have a **treatments** or different **levels** of **a single factor**.

Table 3.2 Typical Data for a single-Factor Experiment

Treatment (level)	Observations				Total	Average
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a.}$	$\bar{y}_{a.}$
					$y_{..}$	$\bar{y}_{..}$

Models for the Data

One way to write the model is

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, 2, ..., a; \quad j = 1, 2, ..., n. \tag{3.1}$$

where y_{ij} is the ijth observation, μ_i is the mean of the ith factor level or treatment,and ϵ_{ij} is a **random error**. $E(y_{ij}) = \mu_i$. This model is called the means model. An alternative model is

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, 2, ..., a; \quad j = 1, 2, ..., n. \tag{3.2}$$

where μ is a parameter common to all treatments called the overall mean and τ_i is a parameter unique to the ith treatment called the ith treatment effect. This model is called the effect model. Both are linear statistical models.

(3.2) (or (3.1)) is also called the **one-way** or **single-factor analysis of variance** model. Further the experiment is required to be performed in random order so that the environment is as uniform as possible. Thus, the experiment design is a **completely randomized design**. For hypothesis testing, we assume that ϵ_{ij} are i.i.d with $N(0, \sigma^2)$ and thus

$$y_{ij} \sim N(\mu + \tau_i, \sigma^2)$$

and the observations are mutually independent.

Fixed Effects Model: a treatments can be choose by experimenter.
Random Effects Model or **Components of Variance Model**: the a treatments could be a random sample.

3.3 Analysis of the Fixed Effects Model

This section will focus on the single-factor analysis of variance for the fixed effects model. Towards this end, we first introduce notation.

$$y_{i\cdot} = \sum_{j=1}^{n} y_{ij}, \quad \bar{y}_{i\cdot} = y_{i\cdot}/n, \quad i = 1, 2, ..., n$$

$$y_{\cdot\cdot} = \sum_{i=1}^{a}\sum_{j=1}^{n}, \quad \bar{y}_{\cdot\cdot} = y_{\cdot\cdot}/N$$

where $N = an$ is the total number of observations.

The interest is to test the equality of the a treatment means; that is, $E(y_{ij}) = \mu + \tau_i = \mu_i, i = 1, 2, ..., a$. The appropriate hypotheses are

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a \quad vs \quad H_1 : \mu_i \neq \mu_j \text{ for at least one pair } (i, j)$$

In the effects model, we think of μ as an overall mean so that $\sum_{i=1}^{a} \mu_i/a = \mu$, which implies that $\sum_{i=1}^{a} \tau_i = 0$. Thus, an equivalent way to write the above hypotheses is

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0, \quad vs \quad H_1 : \tau_i \neq 0 \text{ for at least one } i$$

3.3.1 Decomposition of the Total Sum of Squares

The name analysis of variance is derived from a partitioning of total variability into its component parts. The total corrected sum of squares

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

is used as a measure of overall variability in the data. If we divide it by the degrees of freedom, we obtain the sample variance for y, which is the standard measure of variability. Note that,

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 + n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2.$$

It is the fundamental ANOVA identity.It states that the total variability in the data can be partitioned into a sum of squares of the differences between the treatment averages and the grand average plus a sum of squares of the differences of observations within treatments from the treatment average.

$$\underbrace{SS_T}_{N-1} = \underbrace{SS_{Treatment}}_{a-1} + \underbrace{SS_E}_{N-a}$$

where $SS_{treatment}$ is called the sum of squares due to treatments, SS_E is called the sum of squares due to the error and $N = an$.

Next we examine the expression of SS_E.

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 = \sum_{i=1}^{a}\left[\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2\right] = \sum_{i=1}^{a}(n-1)S_i^2$$

where

$$S_i = \sum_{i=1}^{n}(y_{ij} - \bar{y}_{i.})^2/(n-1), \qquad i = 1, 2, ..., a.$$

Therefore $SS_E/(N-a)$ is a pooled estimate of the common variance within each of the a treatment. Similarly, if there is no difference between the a treatment means, the variation of the treatment average from the grand average can be used to estimate σ^2. That is,

$$\frac{SS_{Treatment}}{a-1} = \frac{n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2}{a-1}$$

is an estimate of σ^2 if the treatment means are equal. Therefore ANOVA identity provide two estimates of σ^2 – one is based on the inherent variability with in treatments and other based on the variability between treatments. The quantities

$$MS_{treatment} = \frac{SS_{Treatment}}{a-1}$$

and

$$MS_E = \frac{SS_E}{N-1}$$

are called mean squares.

Note that

$$E(MS_E) = E\left(\frac{SS_E}{N-a}\right) = \frac{1}{N-a}E\left[\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2\right]$$

$$= \frac{1}{N-a}E\left[\sum_{i=1}^{a}\sum_{j=1}^{n}y_{ij}^2 - \frac{1}{n}\sum_{j=1}^{a}\bar{y}_{i.}^2\right]$$

$$= \frac{1}{N-a}E\left[\sum_{i=1}^{a}\sum_{j=1}^{n}(\mu+\tau_i+\epsilon_{ij})^2 - \frac{1}{n}\sum_{i=1}^{a}\left(\sum_{j=1}^{n}(\mu+\tau_i+\epsilon_{ij})^2\right)\right]$$

After algebra, we have $E(MS_E) = \sigma^2$. By a similar approach we have that

$$E(MS_{Treatment}) = \frac{n}{a-1}E\left(\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2\right)$$

$$= \frac{n}{a-1}E\left[\sum_{i=1}^{a}\left(\frac{1}{n}\sum_{i=1}^{n}(\mu+\tau_i+\epsilon_{ij}) - \frac{1}{N}\sum_{i=1}^{a}\sum_{j=1}^{n}(\mu+\tau_i+\epsilon_{ij})\right)^2\right]$$

$$= \frac{n}{a-1}E\left[\sum_{i=1}^{a}(\tau_i+\bar{\epsilon}_{i.}-\bar{\epsilon}_{..})^2\right]$$

$$= \sigma^2 + \frac{n\sum_{i=1}^{a}\tau_i^2}{a-1}$$

3.3.2 Statistical Analysis

We now investigate how a formal test of the hypothesis of no differences in treatment means

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a \text{ or } H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$$

can be performed. Note that the errors ϵ_{ij} are assumed to be normally and independently distributed with mean zero and variance σ^2. Therefore the sums of squares SS_T/σ^2, $SS_{Treatment}/\sigma^2$ and SS_E/σ^2 in normally distributed random variables can be proved to have the chi-squared distributions. The following theorem can be used to establish the independence of SS_E and $SS_{Treatment}$.

Theorem

THEOREM 3-1: Cochran's Theorem

Let Z_i be NID(0,1) for $i = 1, 2, ..., \nu$ and

$$\sum_{i=1}^{\nu} Z_i^2 = Q_1 + Q_2 + \cdots + Q_s$$

where $s \leq \nu$ and Q_i has ν_i degrees of freedom $(i = 1, 2, ..., s)$. Then $Q_1, Q_2, ..., Q_s$ are independent chi-square random variables with $\nu_i, \nu_2, ..., \nu_s$ degrees of freedom, respectively. if and only if

$$\nu = \nu_1 + \nu_2 + \cdots + \nu_s.$$

Based on this theorem, we have that $SS_{Treatment}/\sigma^2$ and SS_E/σ^2 are independent with chi-square distributions. Therefore, if the null hypothesis of no difference in treatment means are true, the ratio

$$F_0 = \frac{SS_{Treatment}/(a-1)}{SS_E/(N-a)} = \frac{MS_{Treatment}}{MS_E}$$

is distributed as F with $a-1$ and $N-a$ degrees of freedom. The above expression is the test statistic for the hypothesis of no differences in treatments. The null hypothesis H_0 is rejected if $F_0 > F_{\alpha,a-1,N-a}$.

Now we give the alternative ways to computet the sums of squares SS_T, $SS_{Treatment}$ and SS_E.

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$SS_{Treatment} = \frac{1}{n}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N}$$

$$SS_E = SS_T - SS_{Treatment}$$

Table 3-3 The Analysis of Variance Table for the Single-Factor, Fixed Effects Model

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F_0
Between Treatment	$SS_{Treatment} = n\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	$a-1$	$MS_{Treatment}$	$F_0 = \frac{MS_{Treatment}}{MS_E}$
Error (within treatment)	$SS_E = SS_T - SS_{Treatment}$	$N-a$	MS_E	
Total	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot\cdot})^2$	$N-1$		

This is called an **analysis of variance (ANOVA)** table.

Example 3.1 The plasma Etching Experiment

Source of Variation Variation	Sum of Squares	Degree of Freedom	Mean Squares	F_0	P-Value
RF Power	66,870.55	3	22,290.18	66.80	< 0.01
Error	5339.20	16	333.70		
Total	72,209.75	19			

Coding the observations

3.3.3 Estimation of the Model Parameters

We now present estimators for the parameters in the single-factor model

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

and have that

$$\hat{\mu} = \bar{y}_{..}, \quad \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}, \quad i = 1, 2, ..., a.$$

A $100(1-\alpha)\%$ confidence interval estimate of the ith treatment mean $\mu_i = \mu + \tau_i$ can be obtained as follows:

$$\bar{y}_{i.} - t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}} \leq \mu_i \leq \bar{y}_{i.} + t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}}$$

and a $100(1-\alpha)\%$ confidence interval for the difference in any two treatments means, $\mu_i - \mu_j$ is

$$\bar{y}_{i.} - \bar{y}_{j.} - t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}} \leq \mu_i - \mu_j \leq \bar{y}_{i.} - \bar{y}_{j.} + t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}}.$$

Example 3.3 Using the data in Example 3.1, we have the following results for estimates of parameters $\mu, \tau_i, \mu_i, i = 1, 2, 3, 4$.

$$\hat{\mu} = 12355/12 = 617.75,$$

$$\tau_1 = \bar{y}_{1.} - \bar{y}_{..} = 551.20 - 617.75 = -66.55, \quad \tau_2 = \bar{y}_{2.} - \bar{y}_{..} = 587.40 - 617.75 = -30.35,$$

$$\tau_3 = \bar{y}_{3.} - \bar{y}_{..} = 625.40 - 617.75 = 7.65, \quad \tau_4 = \bar{y}_{4.} - \bar{y}_{..} = 707.00 - 617.75 = 89.25,$$

and a 95% confidence interval on the mean of treatment 4 is

$$707.00 - 2.120\sqrt{\frac{333.70}{5}} \leq \mu_4 \leq 707.00 + 2.120\sqrt{\frac{333.70}{5}}$$

and thus the desired 95% confidence interval is $689.68 \leq \mu_4 \leq 724.32$.

Simultaneous Confidence Intervals If there r such $100(1-\alpha)$ percent confidence intervals of interests, the probability that the r intervals will simultaneously be correct is at lease $1 - r\alpha$.

Bonferroni method:

3.3.4 Unbalanced Data

The single-factor design with the different numbers of observations for different treatments is caned the unbalanced design. In this case, $N = \sum_{i=1}^{a} n_i$ and

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{ij}^2}{N}, \quad \text{and} \quad SS_{Treatment} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}.$$

3.4 Model Adequacy Checking

Note that the use of the partitioning to test for no differences in treatment means requires that certain assumptions be satisfied. Specially, these assumptions are that the observations are adequately described by the model

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

and that the errors are normally and independently distributed with mean zero and constant but unknown variance σ^2. If these assumptions are valid, the analysis of variance procedure is an exact test of the hypothesis of no difference in treatment means. However these assumptions will usually not hold exactly. Violations of the basic assumptions and model adequacy can be easily investigated by the examination of residuals. We define the residual for observation j in treatment i as

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$

where $\hat{y}_{ij}$ is an estimate of the corresponding observation y_{ij} obtained as follows:

$$\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i = \bar{y}_{..} + \bar{y}_{i.} - \bar{y}_{..} = \bar{y}_{i.}$$

If the model is adequate, the residuals should be structureless; that is, they should contain no obvious patterns.

3.4.1 The Normality Assumption

A check of the normality assumption could be made by plotting a histogram of the residuals. An extremely useful procedure is to construct a normal probability plot of the residuals. In the analysis of variance, it is usually more effective to do this with the residuals. If the underlying error distribution is normal, this plot will resemble a straight line.

A very common defect that often shows up on normal probability plots is one residual that very much larger than any of the others. Such a residual is often called an **outlier**.

Several formal statistical procedures may be used for detecting outlier. A rough check for outliers may be made by examining the standardized residuals

$$d_{ij} = \frac{e_{ij}}{\sqrt{MS_E}}$$

For the tensile strength data of Example 3.1, the normal probability plot gives no indication of outliers. Furthermore, the largest standardized residual is

$$d_1 = \frac{e_i}{\sqrt{MS_E}} = \frac{25.6}{\sqrt{33.70}} = \frac{25.6}{18.27} = 1.40.$$

3.4.2 Plot of Residual in Time Sequence

Plotting the residuals in time order of data collection is helpful in detecting **correlation** between the residual. A tendency to have runs of positive and negative residuals indicates positive correlation. This would imply that the **independence assumption** on the errors has been violated.

3.4.3 Plot of Residual versus Fitted Values

If the model is correct and the assumptions are satisfied, the residuals should be structureless; in particular, they should be unrelated to any other variables including the predicted response. A simple check is to plot the residuals versus the fitted values $\hat{y}_{ij}$. his plot should not reveal any obvious pattern.

A defect that occasionally shows up on this plot is nonconstant variance. If this were the case, the residuals would get larger as y_{ij} gets larger, and the plot of residuals versus $\hat{y}_{ij}$ would look like an outward-opening funnel or megaphone.

The usual approach to dealing with nonconstant variance is to apply a variance-stabilizing transformation and then to run the analysis of variance on the transformed data. If the experimenters know the theoretical distribution of the observation, they may utilize this information in choosing a transformation.

- The square root transformation: $y_{ij}^* = \sqrt{y_{ij}}$ or $y_{ij}^* = \sqrt{1 + y_{ij}}$ (Poisson)
- The Logarithm transformation: $y_{ij}^* = \log y_{ij}$ (Lognormal)
- The arcsin transformation: $y_{ij}^* = \arcsin \sqrt{y_{ij}}$ (binomial)

Statistical Tests for Equality of Variances Although residual plots are frequently used to diagnose inequality of variance, several statistical tests have also been proposed. The hypotheses are

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_a^2, \quad H_1 : \text{ above not true for at least one } \sigma_i^2$$

A widely used procedure is **Bartlett's** test. The test statistic is

$$\chi_0^2 = 2.3026 \frac{q}{c}$$

where

$$q = (N-a)\log_{10} S_P^2 - \sum_{i=1}^{a}(n_i - 1)\log_{10} S_i^2$$

$$c = 1 + \frac{1}{3(a-1)}\left(\sum_{i=1}^{n}(n_i-1)^{-1} - (N-a)^{-1}\right)$$

$$S_p^2 = \frac{\sum_{i=1}^{a}(n_i-1)S_i^2}{N-a}$$

and S_i^2 is the sample variance of the ith population.

Reject H_0 when

$$\chi_0^2 > \chi_{\alpha,a-1}^2$$

Note that Bartlett's test is very sensitive to the normality assumption.

Example 3.4 In the plasma etch experiment, the normality assumption is not in question, so we can apply Bartlett's test to the etch rate data. We first compute the sample variances in each treatment and find that $S_1^2 = 400.7, S_2^2 = 280.3, S_3^2 = 421.3$ and $S_4^2 = 232.5$. Then,

$$S_p^2 = \frac{4(400.7) + 4(280.3) + 4(421.3) + 4(232.5)}{16} = 333.7$$

$$q = 16\log_{10}(333.7) - 4[\log_{10} 400.7 + \log_{10} 280.3 + \log_{10} 421.3 + \log_{10} 232.5] = 0.21,$$

$$c = 1 + \frac{1}{3(3)}\left(\frac{4}{4} - \frac{1}{16}\right) = 0.1$$

and the trest statistic is

$$\chi_0^2 = 2.3026\frac{0.21}{1.10} = 0.43$$

Note that $\chi_{0.05,3}^2 = 7.81$ and the p-value is 0.934.

Because Bartlett's test is sensitive to the normality assumption, there may be situations where an alternative procedure would be useful. The modified Levene test is a very nice procedure that is robust to departure from normality. The modified Levene test uses the absolute deviation of the observations y_{ij} in each treatment from the treatment median, say $\tilde{y}_i$. Denote these deviations by

$$d_{ij} = |y_{ij} - \tilde{y}_i|, \quad i = 1, 2, ..., a; j = 1, 2, ..., n.$$

The test statistic for Levene's test is simply the usual ANOVA F statistic for testing equality of means applied to the absolute deviations.

Example 3.5 A civic engineer is interested in determining whether four different methods of estimating flood flow frequency produce equivalent estimates of peak discharging when applied to the same watershed.

Empirical Selection of a Transformation

Let $E(y) = \mu$ be the mean of y, and suppose that the standard deviation of y is proportional to the power of the mean of y such that

$$\sigma_y \propto \mu^{\alpha}$$

We want to find a transformation on y that yields a constant variance. Suppose that $y^* = y^{\lambda}$, It can be shown that

$$\sigma_{y^*} \propto \mu^{\lambda+\alpha-1}$$

By setting $\lambda = 1 - \alpha$, the transformed data y^* has the constant variance.

Several of the common transformations are summarized in Table 3.9. Note that $\lambda = 0$ implies the log transformation. In many experimental design situation where there is replication, we can empirically estimate α from the data. In the case that $\sigma_{y_i} \propto \mu_i^{\alpha} = \theta\mu_i^{\alpha}$ where θ is a constant of proportionality, we may have

$$\log \sigma_i = \log \theta + \alpha \log \mu_i$$

Therefore a plot of $\log \sigma_i$ versus $\log \mu_i$ would be a straight line with slope α. Since we do not know σ_i and μ_i, we may plot the points $(\log \bar{y}_{i.}, \log S_i)$.

Table 3.9 Variance-stabilizing Transformations

Relationship Between σ_y and μ	α	$\lambda = 1 - \alpha$	Transformation	Comment
$\sigma_y \propto$constant	0	1	No transformation	
$\sigma_y \propto \mu^{1/2}$	1/2	1/2	Square root	Poisson data
$\sigma_y \propto \mu$	1	0	Log	
$\sigma_y \propto \mu^{3/2}$	3/2	-1/2	Reciprocal square root	
$\sigma_y \propto \mu^2$	2	-1	reciprocal	

To investigate the possibility of using a variance-stabilizing transformation on the peak discharge data from Example 3.5, we plot $\log S_i$ versus $\log \bar{y}_{i\cdot}$ and find that the slope is close to 1/2. Therefore we use the square root transformation.

Table 3.10 Analysis of Variance for Transformed Peak Discharge Data, $y^* = \sqrt{y}$

Source of Variation	Sum of Squares	Degree of Freedom	Mean Squares	F_0	p-value
Methods	32.6842	3	10.8947	76.99	¡0.001
Error	2.6884	19	0.1415		
Total	35.3726	22			

3.4.4 Plots of Residual Versus Other Variables

If data have been collected on any other variables that might possibly affect the response, the residuals should be plotted against these variables. For example, in the tensile strength experiment of Example 3.2, strength may be significantly affected by the thickness of the fiber, so the residual should be plotted versus fiber thickness. If different testing machine were used to collect the data, the residual should be plotted against machines. Patterns in such residual plots imply that the variable affect the response. This suggests that the variable should be either controlled more carefully in future experiments or included in the analysis.

3.5 Practical Interpretation of Results

3.5.1 A Regression Model

The factor involved in an experiment can be either **quantitative** or **qualitative**. Insofar as the initial design and analysis of the experiment are concerned, both types of factors are treated identically.In fact ANOVA treats the design factor as if it were qualitative or categorical. If the factor is really qualitative, it is meaningless to consider the response for a subsequent run at an intermediate level of the factor. However, with the quantitative factor, the experimenter is usually interested in the entire range of values. Thus, development of an interpolation equation for the response variable is of interest. This equation is an **empirical model** of the process.

The general approach to fitting empirical models is called **regression analysis**. As a first approximation, we fit the linear model to the data:

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

where β_0 and β_1 are unknown parameters and ϵ is a random error term. The parameters can be estimated by the **method of least square**.

Quadratic model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon.$$

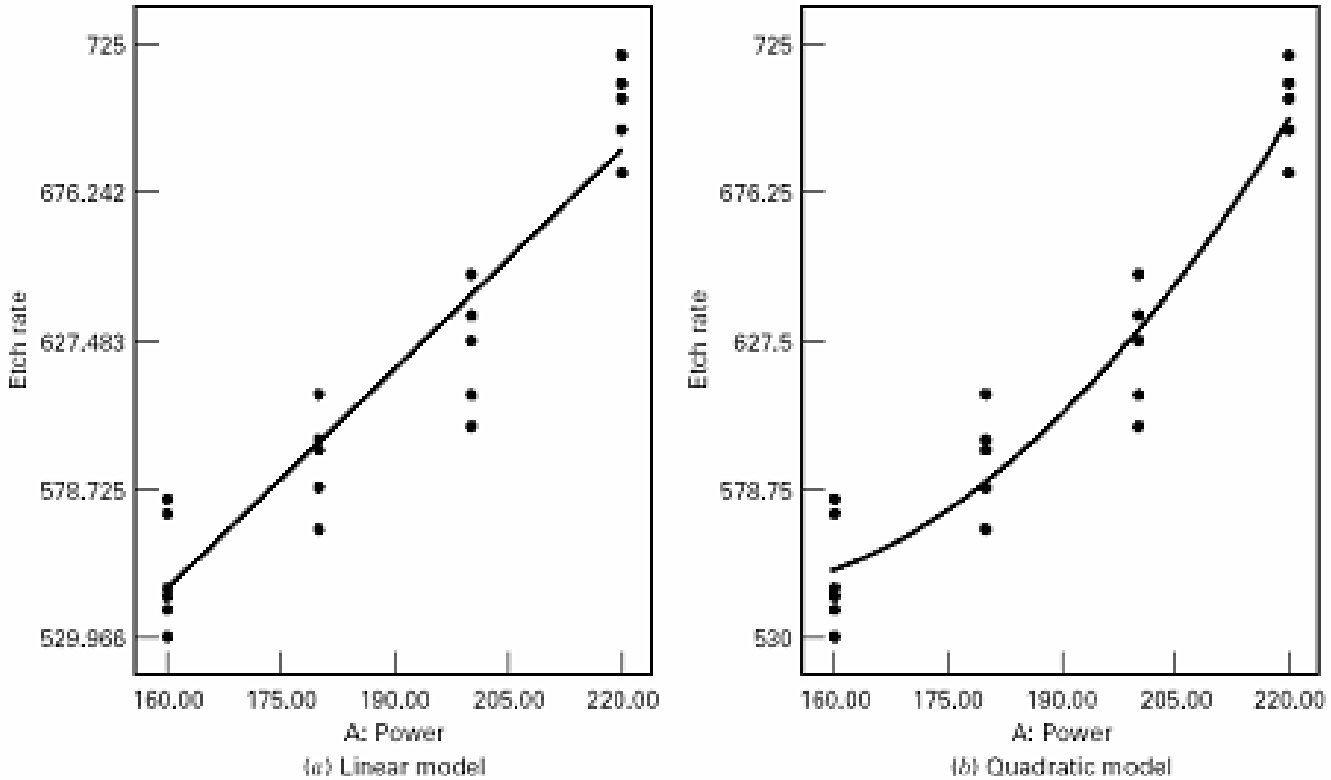

Figure 3-10 Scatter diagrams and regression models for the etch rate data of Example 3-1.

3.5.2 Comparisons Among Treatment Means

Suppose that in conducting an analysis of variance for the fixed effects model the null hypothesis is rejected. Thus, there are differences between the treatment means but exactly which means differ is not specified. Comparisons between treatment means are made in terms of either the treatment totals $\{y_{i.}\}$ or the treatment averages $\{\bar{y}_{i.}\}$. The procedure for making these comparisons are usually called **multiple comparison methods.**

3.5.3 Graphical Comparisons of Means

Suppose that the factor of interest has a levels and that $\bar{y}_{1.}, \bar{y}_{2.}, ..., \bar{y}_{a.}$ are the treatment average. If σ is known, any treatment average would have a standard deviation $\sigma/\sqrt{n}$. If the treatment means are equal, there should be some position for the normal distribution that makes it obvious that the $\bar{y}_{i.}$ values were drawn from the same distribution. However, σ is unknown. We replace it with $\sqrt{MS_E}$ and use a t distribution with a scale factor $\sqrt{MS_E/n}$ instead of the normal.

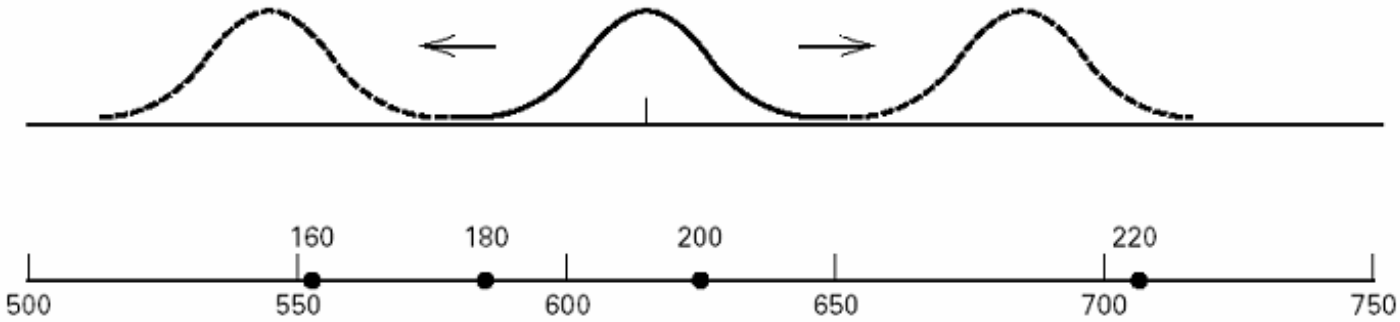

Figure 3-11 Etch rate averages from Example 3-1 in relation to a t distribution with scale factor $\sqrt{MS_E/n} = \sqrt{330.70/5} = 8.13$.

To sketch the t distribution, simply multiply the abscissa t value by the scale factor

$$\sqrt{MS_E/n} = \sqrt{330.70/5} = 8.13$$

and plot this against the ordinate of t at that point. Now visualize sliding the t distribution along the horizontal axis as indicated by the dashed lines and examine the four means plotted in the figure. There is no location for the distribution such that all four averages could be thought of as typical, randomly selected observations from the distribution. This implies that all four means are not equal.

3.5.4 Contrasts

Many multiple comparison methods use the ideal of a **contrast**. When the null hypothesis is rejected, we would like to know which ones actually cause this difference. We might suspect at the outset of the experiment that any two of a levels produce the same response value, implying that we would like to test the hypothesis

$$H_0 : \mu_i = \mu_j, \qquad H_1 : \mu_i \neq \mu_j$$

or equivalently

$$H_0 : \mu_i - \mu_j = 0, \qquad H_1 : \mu_i - \mu_j \neq 0$$

In general, a **contrast** is a linear combination of parameters of the form

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

where the contrast constants $c_1, c_2, \ldots, c_a$ sum to zero; that is $\sum_{i=1}^{a} c_i = 0$. The hypotheses can be expressed in terms of contrasts:

$$H_0 : \sum_{i=1}^{a} c_i \mu_i = 0, \qquad H_1 : \sum_{i=1}^{a} c_i \mu_i \neq 0$$

Testing hypotheses involving contrasts can be done in two basic ways. The first method uses a t-test. Write the contrast of interest in terms of the treatment average, giving

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

The variance of C is

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

when the sample sizes in each treatment are equal. If the null hypothesis is true, the ratio

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

has the $N(0,1)$ distribution. If σ^2 is unknown, the following t-test statistic can be used:

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

The null hypothesis would be rejected it $|t_0| > t_{\alpha/2, N-a}$.

The second approach uses an F test. We have

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

F_0 follows the F-distribution with 1 numerator degree of freedom and $N - a$ denominator degrees of freedom. The null hypothesis would be rejected if $F_0 > F_{\alpha,1,N-a}$. Moreover, we can write the test statistic as

$$F_0 = \frac{MS_C}{MS_E} = \frac{SS_C/1}{SS_E/(N-a)}$$

where the single degree of freedom contrast sum of square is

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

Confidence Interval for a Contrast Suppose that the contrast of interest is

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

Replace the treatment means with the treatment average yields

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

The $100(1-\alpha)\%$ confidence interval on the contrast $\sum_{i=1}^{a} c_i \mu_i$ is

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

Standardized Contrast: The standardized contrast is

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

where

$$c_i^* = \frac{c_i}{\sqrt{\frac{1}{n}\sum_{i=1}^{a} c_i^2}}$$

Unequal Sample Sizes: The definition of a contrast requires that $\sum_{i=1}^{a} n_i c_i = 0$.

$$t_0 = \frac{\sum_{i=1}^{a} c_i \bar{y}_{i.}}{\sqrt{MS_E \sum_{i=1}^{n} \frac{c_i^2}{n_i}}}, SS_c = \frac{\left(\sum_{i=1}^{a} c_i \bar{y}_{i.}\right)^2}{\sum_{i=1}^{n} \frac{c_i^2}{n_i}}$$

3.5.5. Orthogonal Contrasts

A useful special case of the procedure is that of **orthogonal contrasts**. Two contrasts with coefficients $\{c_i\}$ and $\{d_i\}$ are orthogonal if

$$\sum_{i=1}^{a} c_i d_i = 0 \text{ or } \sum_{i=1}^{a} n_i c_i d_i = 0 (\text{ for an unbalanced design})$$

For a treatments, the set of $a-1$ orthogonal contrasts partition the sum of squares due to treatments into $a-1$ independent single-degree-of-freedom components. Thus, tests performed on orthogonal contrasts are independent.

Example 3.6 Consider the plasma etching experiment in Example3.1. We consider the following set of comparisons among the four treatment means.

Hypothesis	Contrast
$H_0 : \mu_1 = \mu_2$	$C_1 = \bar{y}_{1.} - \bar{y}_{2.}$
$H_0 : \mu_1 + \mu_2 = \mu_3 + \mu_4$	$C_2 = \bar{y}_{1.} + \bar{y}_{2.} - \bar{y}_{3.} - \bar{y}_{4.}$
$H_0 : \mu_3 = \mu_4$	$C_3 = \bar{y}_{3.} - \bar{y}_{4.}$

Table 3.11 Analysis of Variance for the Plasma Etching Experiment

Source of Variation	Sum of Squares	Degree of Freedom	Mean Squares	F_0	p-value
Power setting	66870.55	3	22290.18	66.80	< 0.001
Orthogonal contrast					
$C_1 : \mu_1 = \mu_2$	3276.10	1	3276.10	9.82	< 0.001
$C_2 : \mu_1 + \mu_2 = \mu_3 + \mu_4$	46948.05	1	46948.05	140.69	< 0.001
$C_3 : \mu_3 = \mu_4$	16646.40	1	16646.40	48.88	< 0.001
Error	5339.20	16	333.70		
Total	72209.75	19			

3.5.6 Scheffeé's Method for Comparing All Contrasts Suppose that a set of m contrasts in the treatment means

$$\Gamma = c_{1u}\mu_1 + c_{2u}\mu_2 + \cdots + c_{au}\mu_a, \quad u = 1, 2, ..., m$$

of interest have been determined. The corresponding contrast in the treatment averages $\bar{y}_{i.}$ is

$$C_u = c_{1u}\bar{y}_{1.} + c_{2u}\bar{y}_{2.} + \cdots + c_{au}\bar{y}_{a.}, \quad u = 1, 2, ..., m$$

and the standard error of this contrast is

$$S_{C_u} = \sqrt{MS_E \sum_{i=1}^{a}(c_{iu}^2/n_i)}$$

The critical value against which C_u should be compared is

$$S_{\alpha,u} = S_{C_u}\sqrt{(a-1)F_{\alpha,a-1,N-a}}$$

To test the hypothesis that the contrast Γ_u differ significantly from zero, refer C_u to the critical value. $|C_u| > S_{\alpha,u}$, the hypothesis that the contrast Γ_u equals zero is rejected. The simultaneous confidence intervals are

$$C_u - S_{\alpha,u} \leq \Gamma \leq C_u + S_{\alpha,u}, u = 1, 2, ..., m$$

in that the probability that all of them are simultaneously true is at least $1 - \alpha$.

To illustrate the procedure, consider the data in example and suppose that the contrasts of interest are

$$\Gamma_1 = \mu_1 + \mu_2 - \mu_3 - \mu_4, \quad \text{and } \Gamma = \mu_1 - \mu_4$$

The numerical values of these contrasts are

$$C_1 = \bar{y}_{1.} + \bar{y}_{2.} - \bar{y}_{3.} - \bar{y}_{4.} = -193.80, \quad C_2 = \bar{y}_{1.} - \bar{y}_{4.} = -155.8$$

The standard error are

$$S_{C_1} = \sqrt{MS_E \sum_{i=1}^{a} (c_{i1}^2/n_i)} = 16.34, \quad S_{C_2} = \sqrt{MS_E \sum_{i=1}^{a} (c_{i2}^2/n_i)} = 11.55$$

Now the 1% critical values are

$$S_{0.01,1} = S_{C_1}\sqrt{(a-1)F_{0.01,a-1,N-a}} = 16.34\sqrt{3(5.29)} = 65.09$$

$$S_{0.01,2} = S_{C_2}\sqrt{(a-1)F_{0.01,a-1,N-a}} = 11.55\sqrt{3(5.29)} = 45.97$$

3.5.7 Comparing Pairs of Treatment Means Suppose that we are interested in comparing all pairs of a treatment means and that the null hypotheses are $H_0 : \mu_i = \mu_j$ for all $\imath \neq j$. Now we present two popular methods for making such comparisons.

Tukey's Test: Tukey's procedure makes use of the distribution of the studentized range statistic

$$q = \frac{\bar{y}_{\max} - \bar{y}_{\min}}{\sqrt{MS_E/n}}$$

Appendix Table VII contains values of $q_\alpha(p, f)$, the upper α points of q, where f is the number of degrees of freedom associated with the MS_E. For equal sample sizes, Tukey's test declares two means significantly different if the absolute value of their sample differences exceeds

$$T_\alpha = q_\alpha(a, f)\sqrt{\frac{MS_E}{n}}$$

A set of $100(1-\alpha)\%$ confidence intervals for all pairs of means are

$$\bar{y}_{i.} - \bar{y}_{j.} - q_\alpha(a, f)\sqrt{\frac{MS_E}{n}} \leq \mu_i - \mu_j \leq \bar{y}_{i.} - \bar{y}_{j.} + q_\alpha(a, f)\sqrt{\frac{MS_E}{n}}, \quad i \neq j$$

When sample sizes are not equal, we have

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

and

$$\bar{y}_{i.} - \bar{y}_{j.} - \frac{q_\alpha(a,f)}{\sqrt{2}}\sqrt{MS_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \le \mu_i - \mu_j \le \bar{y}_{i.} - \bar{y}_{j.} + \frac{q_\alpha(a,f)}{\sqrt{2}}\sqrt{MS_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

The unequal sample size version is sometimes called the **Tukey-Kramer Procedure**.

Example 3.7 By using data in Table 3.1, for $\alpha = 0.05$ and $f = 16$, we have $q_{0.05}(4,16) = 4.05$ and thus,

$$T_{0.05} = q_{0.05}(4,16)\sqrt{\frac{MS_E}{n}} = 4.05\sqrt{\frac{333.70}{5}} = 33.09$$

Note that

$$\bar{y}_{1.} - \bar{y}_{2.} = -36.20^*, \quad \bar{y}_{1.} - \bar{y}_{3.} = -74.20^*, \quad \bar{y}_{1.} - \bar{y}_{4.} = -155.80^*,$$

$$\bar{y}_{2.} - \bar{y}_{3.} = -38.00^*, \quad \bar{y}_{2.} - \bar{y}_{4.} = -119.60^*, \quad \bar{y}_{3.} - \bar{y}_{4.} = -81.60^*$$

The Fisher Least Significant Difference (LSD) Method This procedure uses the t statistic for testing $H : \mu_i = \mu_j$

$$t_0 = \frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{MA_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

Assuming a two-sided alternative, the pair of means μ_i and μ_j would be declared significantly different if $|\bar{y}_{i.} - \bar{y}_{j.}| > t_{\alpha/2,N-a}\sqrt{MS_E(1/n_i + 1/n_j)}$, the quantity

$$LSD = t_{\alpha/2,N-a}\sqrt{MS_E(1/n_i + 1/n_j)}$$

is called the least significant difference. If the design is balanced,

$$LSD = t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}}$$

Example 3.8 For the data in Table 3.1, the LSD at $\alpha = 0.05$ is

$$LSD = t_{0.025,16}\sqrt{\frac{2MS_E}{n}} = 2.120\sqrt{\frac{2(333.70)}{5}} = 24.49.$$

3.5.8 Comparing Treatment Means with a Control In many experiments, one of the treatments is a control, and the analysis is interested in comparing each of the other $a-1$ treatment means with the control. Thus, only $a-1$ comparisons are to be made. Suppose that treatment a is the control and we wish to test null hypothesis

$$H_0 : \mu_i = \mu_a, \qquad H_1 : \mu_i \neq \mu_a, i = 1, 2, ..., a-1$$

Dunnett's procedure is a modification of the usual t-test. For each hypothesis, the observed differences in the sample means are computed.

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

The hypothesis $H_0 : \mu_i = \mu_a$ is rejected at the level α if

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

where the constant $d_\alpha(a-1, f)$ is given in Appendix Table VIII (note that α is the joint significant level associated with all $a-1$ tests.

Example 3.9 Again we illustrate Dunnett's test using the data in Table 3.1. In this case, $a = 4, a-1 = 3, f = 16$, and $n_i = n = 5$. At the level $\alpha = 0.05, d_{0.05}(3, 16) = 2.59$. Thus, the critical difference become

$$d_{0.05}(3, 16)\sqrt{\frac{2MS_E}{n}} = 2.59\sqrt{\frac{2(333.70)}{5}} = 29.92.$$

We choose treatment 4 as the control. We have that the absolute values of all the

3.7 Determining Sample Size

Can choose the sample size to detect a specific difference in means and achieve desired values of type I and type II errors

3.7.1 Operating Characteristic Curves

Type I error α – reject H_0 when it is true

Type II error β – fail to reject H_0 when it is false .

Power $= 1 - \beta$

Operating characteristic curves plot β against a parameter Φ where

$$\Phi^2 = \frac{n\sum_{i=1}^{a}\tau_i^2}{a\sigma^2}$$

In using the OC curves, the parameter Φ and the value of σ^2 must be specified. One way to determine Φ is to choose the actual value of the treatment means for which we would like to reject the null hypothesis with high probability. The estimate of σ^2 may be available from prior experience.

A significant problem with this approach to use OC curves is that it is usually difficult to select a set of treatment means. An alternate approach is to select a sample size such that if the difference between any two treatment means exceeds a specified value, the null hypothesis should be rejected. Let D be the largest difference, then we can show that the minimum value of Φ is

$$\Phi^2 = \frac{nD^2}{2a\sigma^2} (= \frac{n(75)^2}{2(4)(25^2)} = 1.125n$$

Since it is a minimum value, the corresponding sample size obtained is a conservative value. $(D = 75)$

Example 3.11 Let

$$\mu_1 = 575, \mu_2 = 600, \mu_3 = 650, \mu_4 = 675$$

We are interested in rejecting H_0 with 0.9 probability $(\beta = 0.1)$. Set $\alpha = 0.01$. Note that $\bar{\mu} = 625$, $\tau_1 = -50, \tau_2 = -25, \tau_3 = 25, \tau_4 = 50$ and $\sigma = 25$, thus,

$$\Phi^2 = \frac{n\sum_{i=1}^{a}\tau_i^2}{a\sigma^2} = \frac{n(6250)}{4(25^2)} = 2.5n$$

n	Φ^2	Φ	$a(n-1)$	β	Power $(1-\beta)$
3	7.5	2.74	8	0.25	0.75
4	10.0	3.16	12	0.04	0.96
5	12.5	3.54	16	< 0.01	> 0.99