ANALYSIS OF VARIANCE

What If There Are More Than Two Factor Levels?

- The *t*-test does not directly apply
- There are lots of practical situations where there are either more than two levels of interest, or there are several factors of simultaneous interest
- The **analysis of variance** (ANOVA) is the appropriate analysis "engine" for these types of experiments
- The ANOVA was developed by Fisher in the early 1920s, and initially applied to agricultural experiments
- Used extensively today for industrial experiments

An Example (See pg. 66)

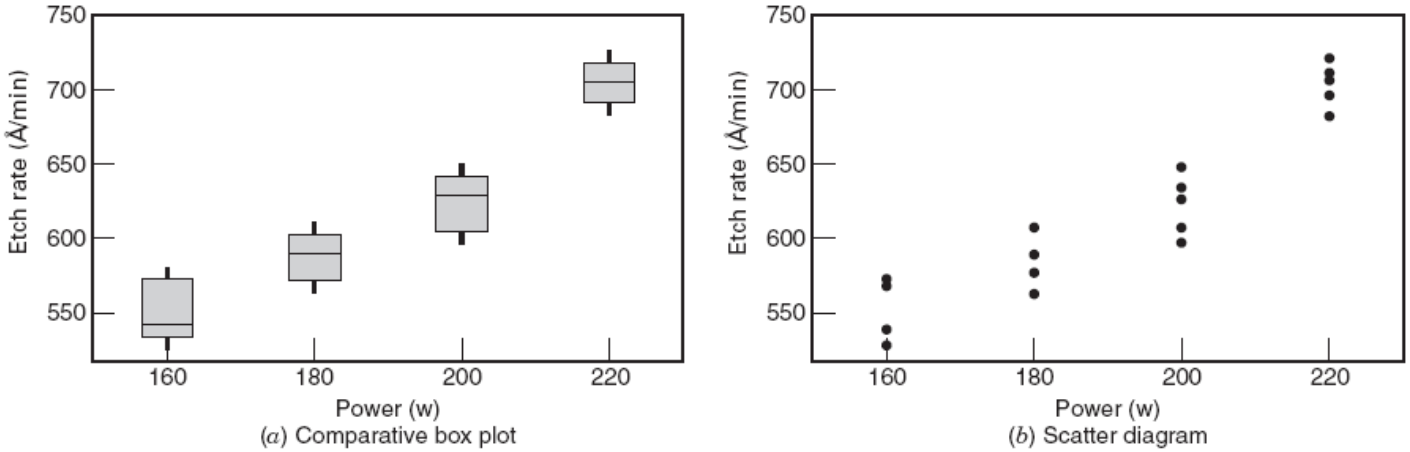
- An engineer is interested in investigating the relationship between the RF power setting and the etch rate for this tool. The objective of an experiment like this is to model the relationship between etch rate and RF power, and to specify the power setting that will give a desired target etch rate.
- The response variable is etch rate.
- She is interested in a particular gas (C2F6) and gap (0.80 cm), and wants to test four levels of RF power: 160W, 180W, 200W, and 220W. She decided to test five wafers at each level of RF power.
- The experimenter chooses 4 **levels** of RF power 160W, 180W, 200W, and 220W
- The experiment is **replicated** 5 times – runs made in random order

An Example (See pg. 66)

■ TABLE 3.1

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

Power (W)	Observations 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



■ FIGURE 3.2 Box plots and scatter diagram of the etch rate data

- Does **changing** the power change the mean etch rate?
- Is there an **optimum** level for power?
- We would like to have an objective way to answer these questions
- The *t*-test really doesn't apply here – more than two factor levels

The Analysis of Variance (Sec. 3.2, pg. 68)

■ TABLE 3.2
Typical Data for a Single-Factor Experiment

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

- In general, there will be *a* **levels** of the factor, or *a* **treatments, and *n* replicates** of the experiment, run in **random order…**a completely randomized design **(CRD)**
- $N = an$ total runs
- We consider the **fixed effects** case…the **random effects** case will be discussed later
- Objective is to test hypotheses about the equality of the *a* treatment means

The Analysis of Variance

- The name “analysis of variance” stems from a **partitioning** of the total variability in the response variable into components that are consistent with a **model** for the experiment
- The basic single-factor ANOVA model is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., n \end{cases}$$

$\mu =$ an overall mean, $\tau_i = ith$ treatment effect,

$\varepsilon_{ij} =$ experimental error, $NID(0, \sigma^2)$

Models for the Data

There are several ways to write a model for the data:

$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ is called the effects model

Let $\mu_i = \mu + \tau_i$, then

$y_{ij} = \mu_i + \varepsilon_{ij}$ is called the means model

Regression models can also be employed

The Analysis of Variance

- **Total variability** is measured by the total sum of squares:

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

- The basic ANOVA partitioning is:

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

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

$$SS_T = SS_{Treatments} + SS_E$$

The Analysis of Variance

$$SS_T = SS_{Treatments} + SS_E$$

- A large value of $SS_{Treatments}$ reflects large differences in treatment means
- A small value of $SS_{Treatments}$ likely indicates no differences in treatment means
- Formal statistical hypotheses are:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a$$

$$H_1 : \text{ At least one mean is different}$$

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. Intuitively, this is reasonable because if we were to divide SS_T by the appropriate number of degrees of freedom (in this case, $an - 1 = N - 1$), we would have the **sample variance** of the y's. The sample variance is, of course, a standard measure of variability.

Note that the total corrected sum of squares SS_T may be written as

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}[(\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})]^2 \tag{3.5}$$

or

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

However, the cross-product term in this last equation is zero, because

$$\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.}) = y_{i.} - n\bar{y}_{i.} = y_{i.} - n(y_{i.}/n) = 0$$

Therefore, we have

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 \tag{3.6}$$

The Analysis of Variance

- While sums of squares cannot be directly compared to test the hypothesis of equal means, **mean squares** can be compared.
- A mean square is a sum of squares divided by its degrees of freedom:

$$df_{Total} = df_{Treatments} + df_{Error}$$

$$an - 1 = a - 1 + a(n-1)$$

$$MS_{Treatments} = \frac{SS_{Treatments}}{a-1}, MS_E = \frac{SS_E}{a(n-1)}$$

- If the treatment means are equal, the treatment and error mean squares will be (theoretically) equal.
- If treatment means differ, the treatment mean square will be larger than the error mean square.

The Analysis of Variance is Summarized in a Table

■ TABLE 3.3

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

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

- Computing…see text, pp 69
- The **reference distribution** for F_0 is the $F_{a\text{-}1,\, a(n\text{-}1)}$ distribution
- **Reject** the null hypothesis (equal treatment means) if

$$F_0 > F_{\alpha, a-1, a(n-1)}$$

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{3.8}$$

$$SS_{\text{Treatments}} = \frac{1}{n} \sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N} \tag{3.9}$$

$$SS_E = SS_T - SS_{\text{Treatments}} \tag{3.10}$$

ANOVA Table
Example 3-1

$$SS_T = \sum_{i=1}^{4}\sum_{j=1}^{5} y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$= (575)^2 + (542)^2 + \cdots + (710)^2 - \frac{(12{,}355)^2}{20}$$

$$= 72{,}209.75$$

$$SS_{\text{Treatments}} = \frac{1}{n}\sum_{i=1}^{4} y_{i.}^2 - \frac{y_{..}^2}{N}$$

$$= \frac{1}{5}[(2756)^2 + \cdots + (3535)^2] - \frac{(12{,}355)^2}{20}$$

$$= 66{,}870.55$$

$$SS_E = SS_T - SS_{\text{Treatments}}$$

$$= 72{,}209.75 - 66{,}870.55 = 5339.20$$

Usually, these calculations would be performed on a computer, using a software package with the capability to analyze data from designed experiments.

The ANOVA is summarized in Table 3.4. Note that the RF power or between-treatment mean square (22,290.18) is many times larger than the within-treatment or error mean square (333.70). This indicates that it is unlikely that the treatment means are equal. More formally, we can compute

■ **TABLE 3.4**
ANOVA for the Plasma Etching Experiment

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

Model Adequacy Checking in the ANOVA Text reference, Section 3.4, pg. 80

- **Checking assumptions** is important
- Normality
- Constant variance
- Independence
- Have we fit the right model?
- Later we will talk about what to do if some of these assumptions are **violated**

Model Adequacy Checking in the ANOVA

- Examination of **residuals** (see text, Sec. 3-4, pg. 80)

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

$$= y_{ij} - \bar{y}_{i.}$$

- Computer software generates the residuals
- **Residual plots** are very useful
- **Normal probability plot** of residuals

■ FIGURE 3.4
Normal probability plot of residuals for Example 3.1

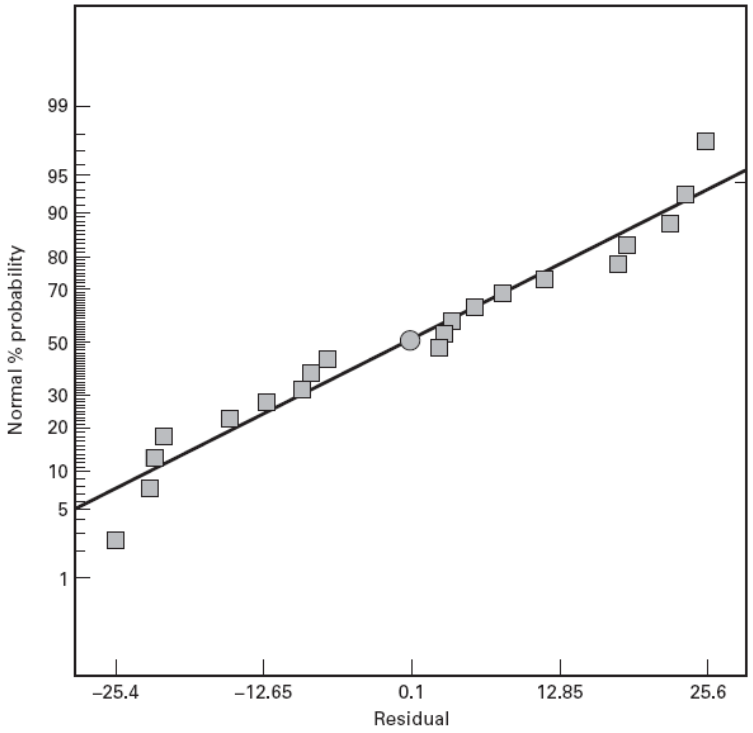


Table 4-1 Data and Residuals from Example 3-1[a]

Percentage of Cotton	Observations (j) 1	2	3	4	5	$\hat{y}_{ij} = \bar{y}_{i.}$
15	−2.8 7 (15)	−2.8 7 (19)	5.2 15 (25)	1.2 11 (12)	−0.8 9 (6)	9.8
20	−3.4 12 (8)	1.6 17 (14)	−3.4 12 (1)	2.6 18 (11)	2.6 19 (3)	15.4
25	−3.6 14 (18)	0.4 18 (13)	0.4 18 (20)	1.4 19 (7)	1.4 19 (9)	17.6
30	−2.6 19 (22)	3.4 25 (5)	0.4 22 (2)	−2.6 19 (24)	1.4 23 (10)	21.6
35	−3.8 7 (17)	−0.8 10 (21)	0.2 11 (4)	4.2 15 (16)	0.2 11 (23)	10.8

[a] The residuals are shown in the box in each cell. The numbers in parentheses indicate the order of data collection.

bution. Since the F test is only slightly affected, we say that the analysis of variance (and related procedures such as multiple comparisons) is *robust* to the normality assumption. Departures from normality usually cause both the true significance level and the power to differ slightly from the advertised values, with the power generally being lower. The random effects model is more severely impacted by nonnormality. In particular, the true confidence levels on interval estimates of variance components may differ greatly from the advertised values.

Table 4-2 Ordered Residuals and Probability Points for the Tensile Strength Data

Order k	Residual e_{ij}	$P_k = (k - \frac{1}{2})/25$	Order k	Residual e_{ij}	$P_k = (k - \frac{1}{2})/25$
1	−3.8	.0200	14	0.4	.5400
2	−3.6	.0600	15	0.4	.5800
3	−3.4	.1000	16	1.2	.6200
4	−3.4	.1400	17	1.4	.6600
5	−2.8	.1800	18	1.4	.7000
6	−2.8	.2200	19	1.4	.7400
7	−2.8	.2600	20	1.6	.7800
8	−2.6	.3000	21	2.6	.8200
9	−0.8	.3400	22	2.6	.8600
10	−0.8	.3800	23	3.4	.9000
11	0.2	.4200	24	4.2	.9400
12	0.2	.4600	25	5.2	.9800
13	0.4	.5000			

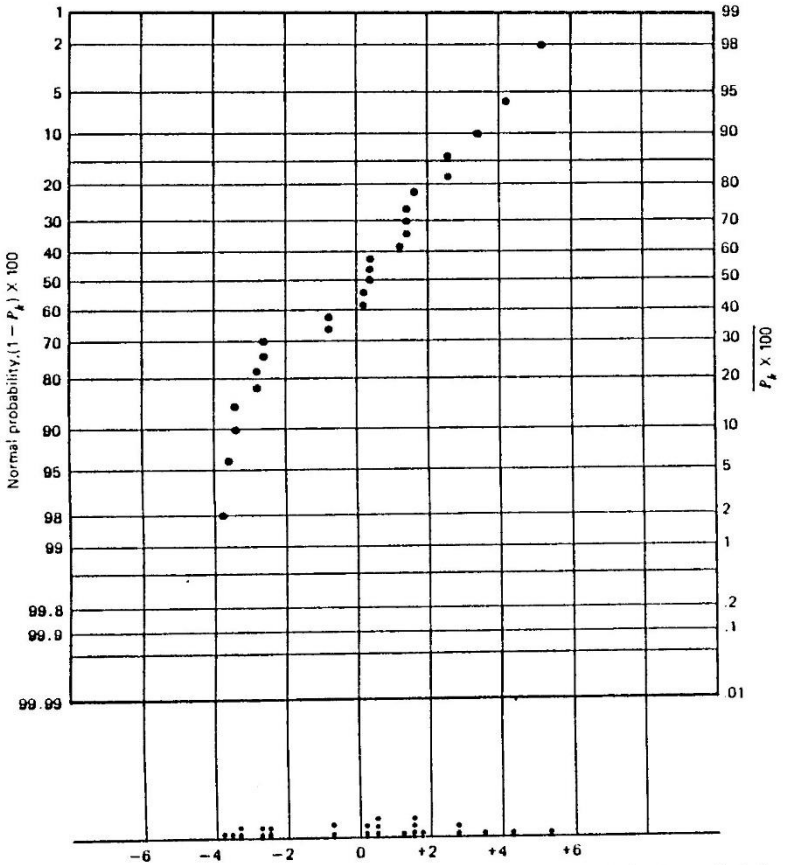
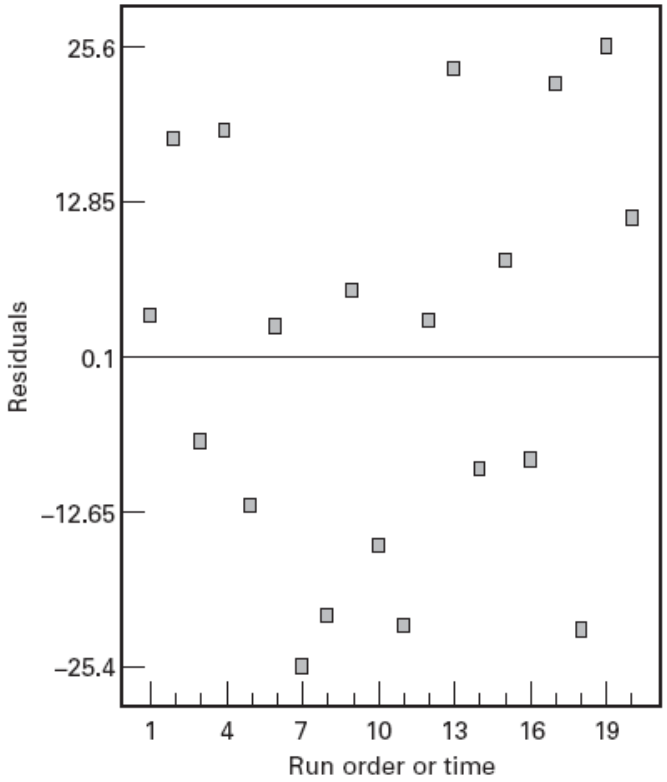


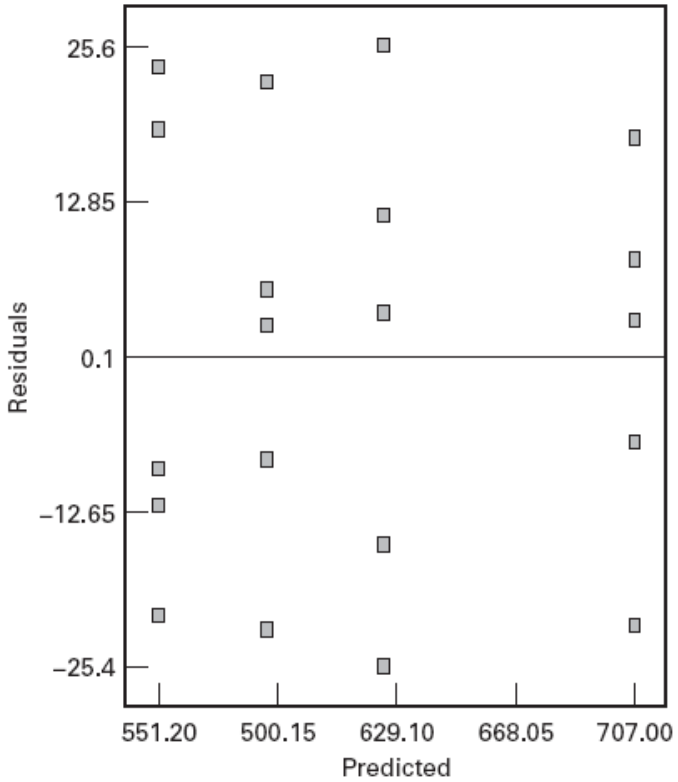
Figure 4-1. Normal probability plot and dot diagram of residuals for Example 3-1.

A very common defect that often shows up on normal probability plots is one residual that is very much larger than any of the others. Such a residual is often called an *outlier*. The presence of one or more outliers can seriously distort the analysis of variance, so when a potential outlier is located, careful investigation is called for. Frequently, the cause of the outlier is a mistake in calculations or a data coding or copying error. If this is not the cause, then the experimental circumstances surrounding this run must be carefully studied. If the outlying

Other Important Residual Plots



■ FIGURE 3.5 Plot of residuals versus run order or time



■ FIGURE 3.6 Plot of residuals versus fitted values

Post-ANOVA Comparison of Means

- The analysis of variance tests the hypothesis of equal treatment means
- Assume that residual analysis is satisfactory
- If that hypothesis is rejected, we don't know **which specific means** are different
- Determining which specific means differ following an ANOVA is called the **multiple comparisons problem**
- There are **lots** of ways to do this…see text, Section 3.5, pg. 89
- We will use pairwise *t*-tests on means…sometimes called Fisher's Least Significant Difference (or Fisher's **LSD**) Method and Tukey Method

Tukey's Test. Suppose that, following an ANOVA in which we have rejected the null hypothesis of equal treatment means, we wish to test all pairwise mean comparisons:

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

for all $i \neq j$. Tukey (1953) proposed a procedure for testing hypotheses for which the overall significance level is exactly α when the sample sizes are equal and at most α when the sample sizes are unequal. His procedure can also be used to construct confidence intervals on the differences in all pairs of means. For these intervals, the simultaneous confidence level is $100(1 - \alpha)$ percent when the sample sizes are equal and at least $100(1 - \alpha)$ percent when sample sizes are unequal. In other words, the Tukey procedure controls the **experimentwise** or "family" error rate at the selected level α. This is an excellent data snooping procedure when interest focuses on pairs of means.

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

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

where $\bar{y}_{\max}$ and $\bar{y}_{\min}$ are the largest and smallest sample means, respectively, out of a group of p sample means. Appendix Table VII contains values of $q_\alpha(p, f)$, the upper α percentage 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

$$T_\alpha = q_\alpha(a, f)\sqrt{\frac{MS_E}{n}} \tag{3.35}$$

Equivalently, we could construct a set of $100(1 - \alpha)$ percent confidence intervals for all pairs of means as follows:

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

$$\le \bar{y}_{i.} - \bar{y}_{j.} + q_\alpha(a, f)\sqrt{\frac{MS_E}{n}}, \quad i \ne j. \tag{3.36}$$

When sample sizes are not equal, Equations 3.35 and 3.36 become

$$T_\alpha = \frac{q_\alpha(a, f)}{\sqrt{2}}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \tag{3.37}$$

and

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

$$\le \bar{y}_{i.} - \bar{y}_{j.} + \frac{q_\alpha(a, f)}{\sqrt{2}}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}, \; i \ne j \tag{3.38}$$

respectively. The unequal sample size version is sometimes called the **Tukey–Kramer procedure**.

Tukey's Test. Suppose that, following an ANOVA in which we have rejected the null hypothesis of equal treatment means, we wish to test all pairwise mean comparisons:

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

for all $i \neq j$. Tukey (1953) proposed a procedure for testing hypotheses for which the overall significance level is exactly α when the sample sizes are equal and at most α when the sample sizes are unequal. His procedure can also be used to construct confidence intervals on the differences in all pairs of means. For these intervals, the simultaneous confidence level is $100(1-\alpha)$ percent when the sample sizes are equal and at least $100(1-\alpha)$ percent when sample sizes are unequal. In other words, the Tukey procedure controls the **experimentwise** or "family" error rate at the selected level α. This is an excellent data snooping procedure when interest focuses on pairs of means.

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

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

where $\bar{y}_{\max}$ and $\bar{y}_{\min}$ are the largest and smallest sample means, respectively, out of a group of p sample means. Appendix Table VII contains values of $q_\alpha(p, f)$, the upper α percentage 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

$$T_\alpha = q_\alpha(a, f)\sqrt{\frac{MS_E}{n}} \tag{3.35}$$

Equivalently, we could construct a set of $100(1 - \alpha)$ percent confidence intervals for all pairs of means as follows:

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

$$\le \bar{y}_{i.} - \bar{y}_{j.} + q_\alpha(a, f)\sqrt{\frac{MS_E}{n}}, \quad i \ne j. \tag{3.36}$$

When sample sizes are not equal, Equations 3.35 and 3.36 become

$$T_\alpha = \frac{q_\alpha(a, f)}{\sqrt{2}}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \tag{3.37}$$

and

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

$$\le \bar{y}_{i.} - \bar{y}_{j.} + \frac{q_\alpha(a, f)}{\sqrt{2}}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}, \; i \ne j \tag{3.38}$$

respectively. The unequal sample size version is sometimes called the **Tukey–Kramer procedure**.

EXAMPLE 3.7

To illustrate Tukey's test, we use the data from the plasma etching experiment in Example 3.1. With $\alpha = 0.05$ and $f =$ 16 degrees of freedom for error, Appendix Table VII gives $q_{0.05}(4, 16) = 4.05$. Therefore, from Equation 3.35,

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

Thus, any pairs of treatment averages that differ in absolute value by more than 33.09 would imply that the corresponding pair of population means are significantly different. The four treatment averages are

$$\bar{y}_{1.} = 551.2 \quad \bar{y}_{2.} = 587.4$$
$$\bar{y}_{3.} = 625.4 \quad \bar{y}_{4.} = 707.0$$

and the differences in averages are

$$\bar{y}_{1.} - \bar{y}_{2.} = 551.2 - 587.4 = -36.20^*$$
$$\bar{y}_{1.} - \bar{y}_{3.} = 551.2 - 625.4 = -74.20^*$$
$$\bar{y}_{1.} - \bar{y}_{4.} = 551.2 - 707.0 = -155.8^*$$
$$\bar{y}_{2.} - \bar{y}_{3.} = 587.4 - 625.4 = -38.0^*$$
$$\bar{y}_{2.} - \bar{y}_{4.} = 587.4 - 707.0 = -119.6^*$$
$$\bar{y}_{3.} - \bar{y}_{4.} = 625.4 - 707.0 = -81.60^*$$

The starred values indicate the pairs of means that are significantly different. Note that the Tukey procedure indicates that all pairs of means differ. Therefore, each power setting results in a mean etch rate that differs from the mean etch rate at any other power setting.

When using any procedure for pairwise testing of means, we occasionally find that the overall F test from the ANOVA is significant, but the pairwise comparison of means fails to reveal any significant differences. This situation occurs because the F test is simultaneously considering all possible contrasts involving the treatment means, not just pairwise comparisons. That is, in the data at hand, the significant contrasts may not be of the form $\mu_i - \mu_j$.

The derivation of the Tukey confidence interval of Equation 3.36 for equal sample sizes is straightforward. For the studentized range statistic q, we have

$$P\left(\frac{\max(\bar{y}_{i.} - \mu_i) - \min(\bar{y}_{i.} - \mu_i)}{\sqrt{MS_E/n}} \leq q_\alpha(a, f)\right) = 1 - \alpha$$

If $\max(\bar{y}_{i.} - \mu_i) - \min(\bar{y}_{i.} - \mu_i)$ is less than or equal to $q_\alpha(a, f)\sqrt{MS_E/n}$, it must be true that $|(\bar{y}_{i.} - \mu_i) - (\bar{y}_{j.} - \mu_j)| \leq q_\alpha(a, f)\sqrt{MS_E/n}$ for every pair of means. Therefore

$$P\left(-q_\alpha(a, f)\sqrt{\frac{MS_E}{n}} \leq \bar{y}_{i.} - \bar{y}_{j.} - (\mu_i - \mu_j) \leq q_\alpha(a, f)\sqrt{\frac{MS_E}{n}}\right) = 1 - \alpha$$

Rearranging this expression to isolate $\mu_i - \mu_j$ between the inequalities will lead to the set of $100(1 - \alpha)$ percent simultaneous confidence intervals given in Equation 3.38.

The Fisher Least Significant Difference (LSD) Method. The Fisher method for comparing all pairs of means controls the error rate α for each individual pairwise comparison but does not control the experimentwise or family error rate. This procedure uses the t statistic for testing $H_0: \mu_i = \mu_j$

$$t_0 = \frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \tag{3.39}$$

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

$$\text{LSD} = t_{\alpha/2,N-a}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \tag{3.40}$$

is called the **least significant difference**. If the design is balanced, $n_1 = n_2 = \cdots = n_a = n$, and

$$\text{LSD} = t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}} \tag{3.41}$$

To use the Fisher LSD procedure, we simply compare the observed difference between each pair of averages to the corresponding LSD. If $|\bar{y}_{i.} - \bar{y}_{j.}| > \text{LSD}$, we conclude that the population means μ_i and μ_j differ. The t statistic in Equation 3.39 could also be used.

EXAMPLE 3.8

To illustrate the procedure, if we use the data from the experiment in Example 3.1, the LSD at $\alpha = 0.05$ is

$$\text{LSD} = t_{.025,16}\sqrt{\frac{2MS_E}{n}} = 2.120\sqrt{\frac{2(333.70)}{5}} = 24.49$$

Thus, any pair of treatment averages that differ in absolute value by more than 24.49 would imply that the corresponding pair of population means are significantly different. The differences in averages are

$$\bar{y}_{1.} - \bar{y}_{2.} = 551.2 - 587.4 = -36.2^*$$

$$\bar{y}_{1.} - \bar{y}_{3.} = 551.2 - 625.4 = -74.2^*$$

$$\bar{y}_{1.} - \bar{y}_{4.} = 551.2 - 707.0 = -155.8^*$$

$$\bar{y}_{2.} - \bar{y}_{3.} = 587.4 - 625.4 = -38.0^*$$

$$\bar{y}_{2.} - \bar{y}_{4.} = 587.4 - 707.0 = -119.6^*$$

$$\bar{y}_{3.} - \bar{y}_{4.} = 625.4 - 707.0 = -81.6^*$$

The starred values indicate pairs of means that are significantly different. Clearly, all pairs of means differ significantly.