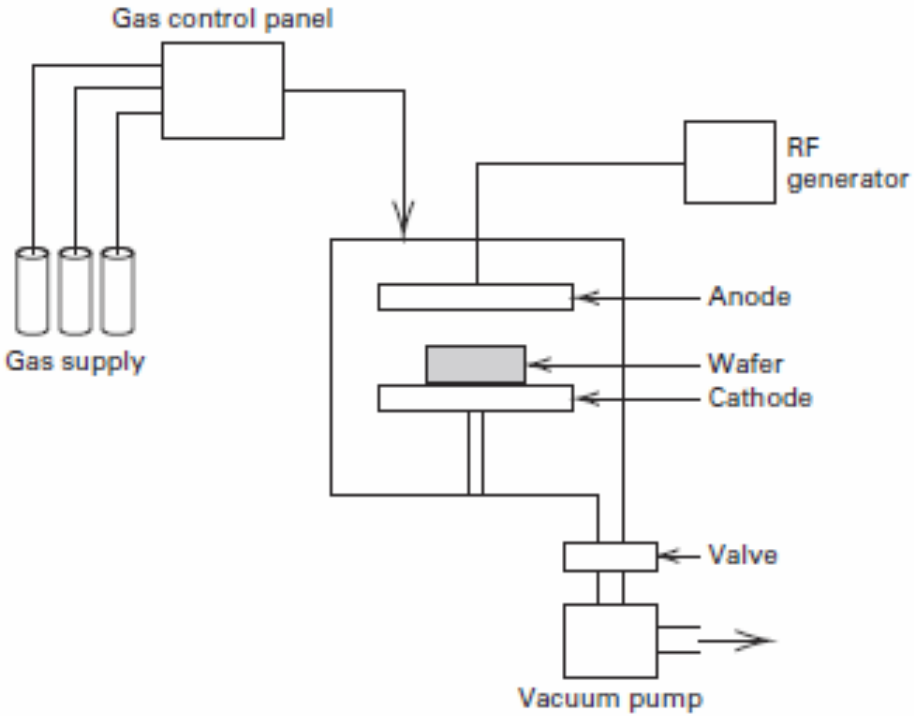
Chapter-3
ANALYSIS OF VARIANCE

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



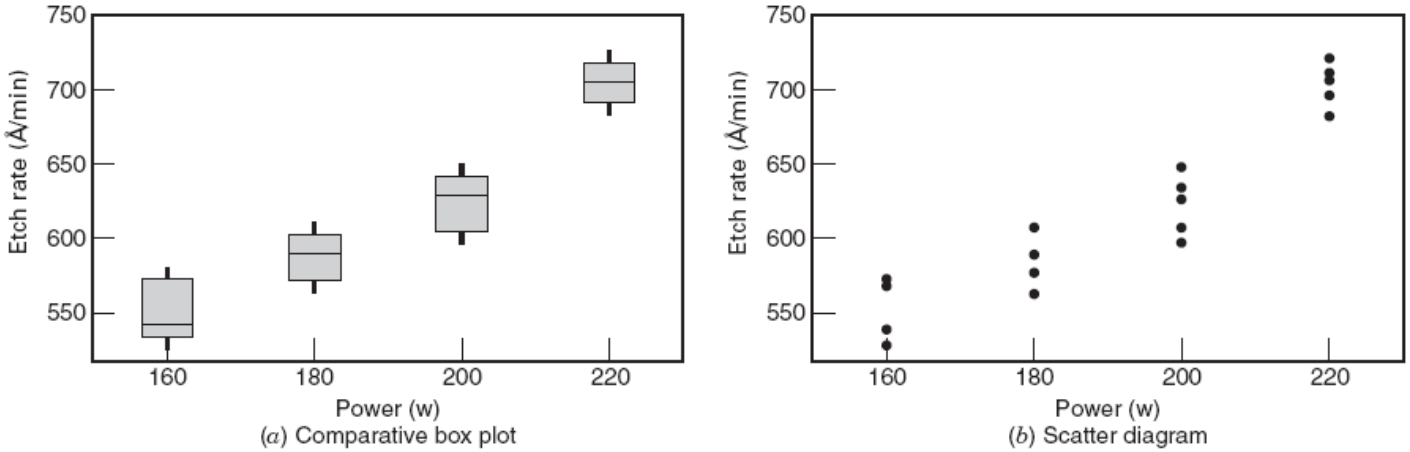
■ **FIGURE 3.1** **A single-wafer plasma etching tool**

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

- 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			

3-3.4 Unbalanced Data

In some single-factor experiments the number of observations taken within each treatment may be different. We then say that the design is **unbalanced**. The analysis of variance described above may still be used, but slight modifications must be made in the

sum of squares formulas. Let n_i observations be taken under treatment i ($i = 1, 2, \ldots, a$) and $N = \sum_{i=1}^{a} n_i$. The manual computational formulas for SS_T and $SS_{\text{Treatments}}$ become

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{3-14}$$

and

$$SS_{\text{Treatments}} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} \tag{3-15}$$

No other changes are required in the analysis of variance.

There are two advantages in choosing a balanced design. First, the test statistic is relatively insensitive to small departures from the assumption of equal variances for the a treatments if the sample sizes are equal. This is not the case for unequal sample sizes. Second, the power of the test is maximized if the samples are of equal size.

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

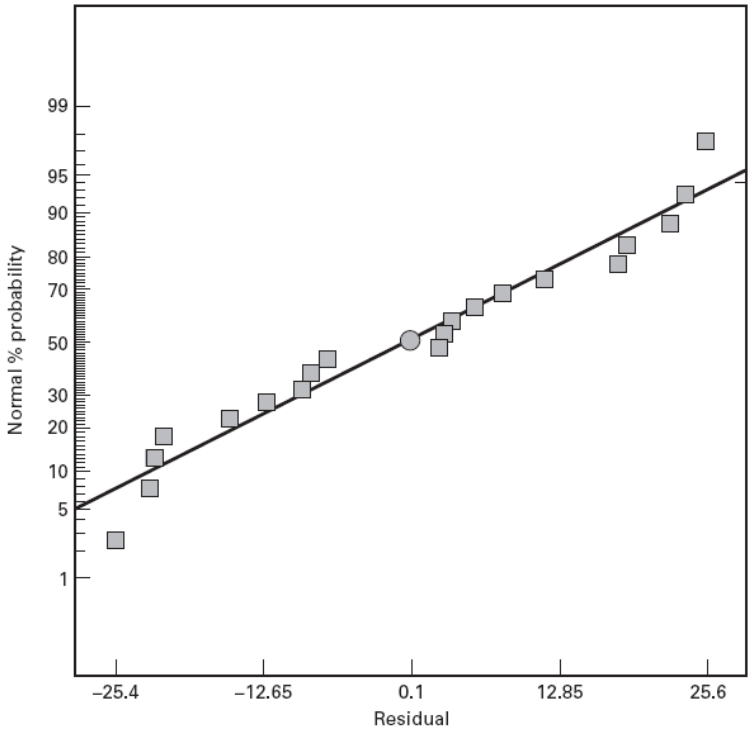


Table 4-1 Data and Residuals from Example 3-1[a]

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			

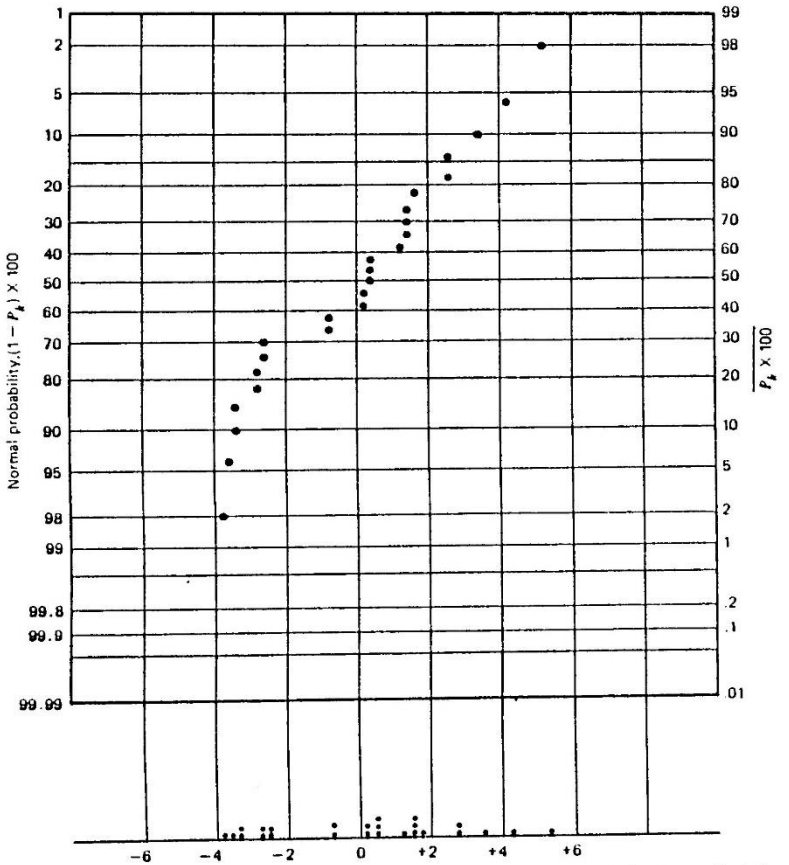
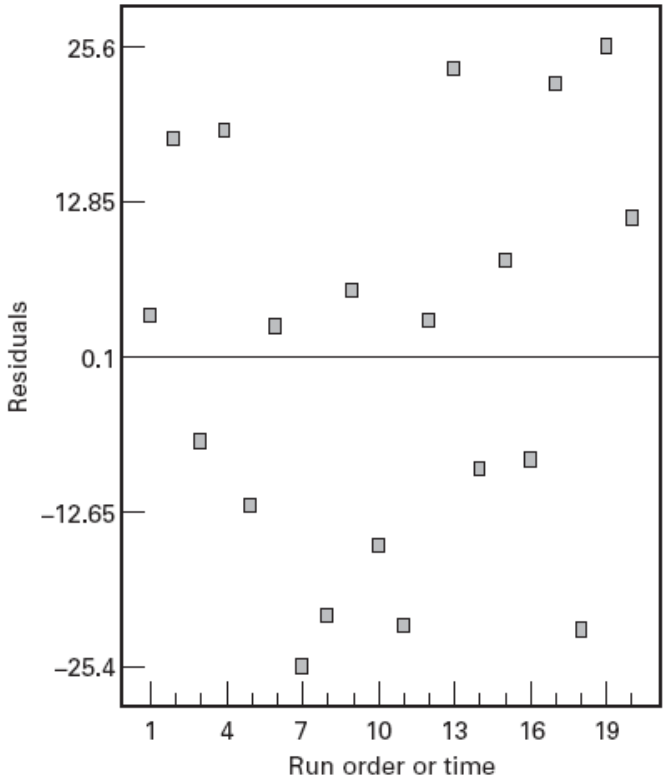


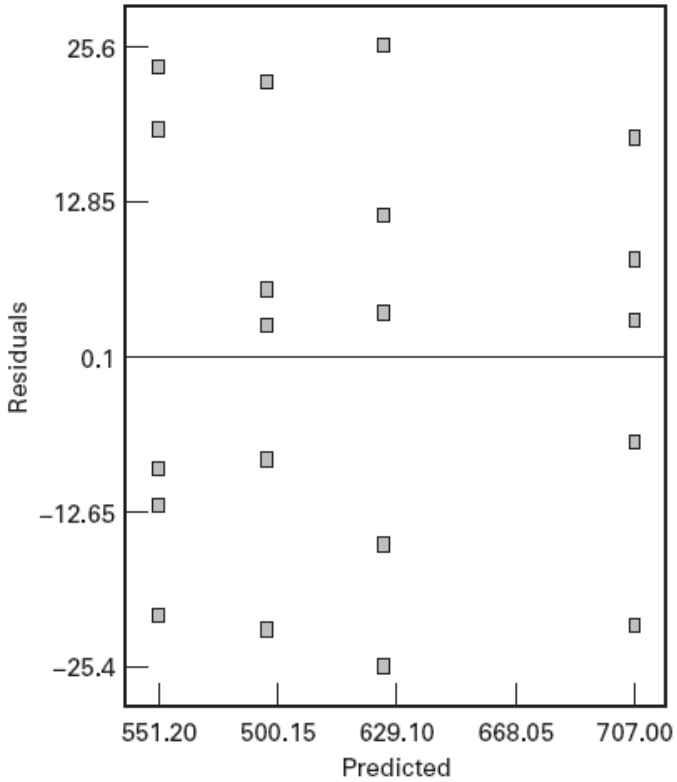
Figure 4-1. Normal probability plot and dot diagram of residuals for Example 3-1.

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

Statistical Tests for Equality of Variance. Although residual plots are frequently used to diagnose inequality of variance, several statistical tests have also been proposed. These tests may be viewed as formal tests of the hypotheses

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

A widely used procedure is **Bartlett's test**. The procedure involves computing a statistic whose sampling distribution is closely approximated by the chi-square distribution with $a - 1$ degrees of freedom when the a random samples are from independent normal populations. The test statistic is

$$\chi_0^2 = 2.3026 \frac{q}{c} \tag{3.19}$$

where

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

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

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

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

The quantity q is large when the sample variances S_i^2 differ greatly and is equal to zero when all S_i^2 are equal. Therefore, we should reject H_0 on values of χ_0^2 that are too large; that is, we reject H_0 only when

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

where $\chi_{\alpha,a-1}^2$ is the upper α percentage point of the chi-square distribution with $a - 1$ degrees of freedom. The P-value approach to decision making could also be used.

Bartlett's test is very sensitive to the normality assumption. Consequently, when the validity of this assumption is doubtful, Bartlett's test should not be used.

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**
- 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.

Note that the overall α risk may be considerably inflated using this method. Specifically, as the number of treatments a gets larger, the experimentwise or family type I error rate (the ratio of the number of experiments in which at least one type I error is made to the total number of experiments) becomes large.

Which Pairwise Comparison Method Do I Use? Certainly, a logical question at this point is, Which one of these procedures should I use? Unfortunately, there is no clear-cut answer to this question, and professional statisticians often disagree over the utility of the various procedures. Carmer and Swanson (1973) have conducted Monte Carlo simulation studies of a number of multiple comparison procedures, including others not discussed here. They report that the least significant difference method is a very effective test for detecting true differences in means if it is applied *only after* the F test in the ANOVA is significant at 5 percent. However, this method does not contain the experimentwise error rate. Because the Tukey method does control the overall error rate, many statisticians prefer to use it.

As indicated above, there are several other multiple comparison procedures. For articles describing these methods, see O'Neill and Wetherill (1971), Miller (1977), and Nelson (1989). The books by Miller (1991) and Hsu (1996) are also recommended.

Why Does the ANOVA Work?

We are sampling from normal populations, so

$$\frac{SS_{Treatments}}{\sigma^2} \sim \chi^2_{a-1} \text{ if } H_0 \text{ is true, and } \frac{SS_E}{\sigma^2} \sim \chi^2_{a(n-1)}$$

Cochran's theorem gives the independence of these two chi-square random variables

$$\text{So } F_0 = \frac{SS_{Treatments}/(a-1)}{SS_E/[a(n-1)]} \sim \frac{\chi^2_{a-1}/(a-1)}{\chi^2_{a(n-1)}/[a(n-1)]} \sim F_{a-1,a(n-1)}$$

$$\text{Finally, } E(MS_{Treatments}) = \sigma^2 + \frac{n\sum_{i=1}^{n}\tau_i^2}{a-1} \text{ and } E(MS_E) = \sigma^2$$

Therefore an upper-tail F test is appropriate.

Sample Size Determination
Text, Section 3.7, pg. 105

- **FAQ** in designed experiments
- Answer depends on lots of things; including what type of experiment is being contemplated, how it will be conducted, resources, and desired **sensitivity**
- Sensitivity refers to the **difference in means** that the experimenter wishes to detect
- Generally, **increasing** the number of **replications increases** the **sensitivity** or it makes it easier to detect small differences in means

Sample Size Determination
Fixed Effects Case

- Can choose the sample size to detect a specific difference in means and achieve desired values of **type I and type II errors**
- Type I error – reject H_0 when it is true (α)
- Type II error – fail to reject H_0 when it is false (β)
- **Power** = 1 - β
- **Operating characteristic curves** plot β against a parameter Φ where

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

Sample Size Determination
Fixed Effects Case---use of OC Curves

- The **OC curves** for the fixed effects model are in the Appendix, Table V
- A very common way to use these charts is to define a difference in two means *D* of interest, then the minimum value of Φ^2 is

$$\Phi^2 = \frac{nD^2}{2a\sigma^2}$$

- Typically work in term of the ratio of D/σ and try values of *n* until the **desired power** is achieved
- Most statistics software packages will perform power and sample size calculations – see page 108
- There are some other methods discussed in the text

Example:

- **EXAMPLE:** Consider the tensile strength experiment described earlier. Suppose that the experimenter is interested in rejecting the null hypothesis with a probability of at least
- 0.90 if the five treatment means are
- $\mu 1=11$, $\mu 2=12$, $\mu 3=15$, $\mu 4=18$, and $\mu 5=19$
- She plans to use $\alpha = 0.01$. In this case, because $\Sigma \mu_i = 75$.

Therefore, the mean average = (1/5)75 =15 and,

$T1 = \mu 1 - \mu = 11 - 15 = -4$

$T2 = \mu 2 - \mu = 12 - 15 = -3$

$T3 = \mu 3 - \mu = 15 - 15 = 0$

$T4 = \mu 4 - \mu = 18 - 15 = 3$

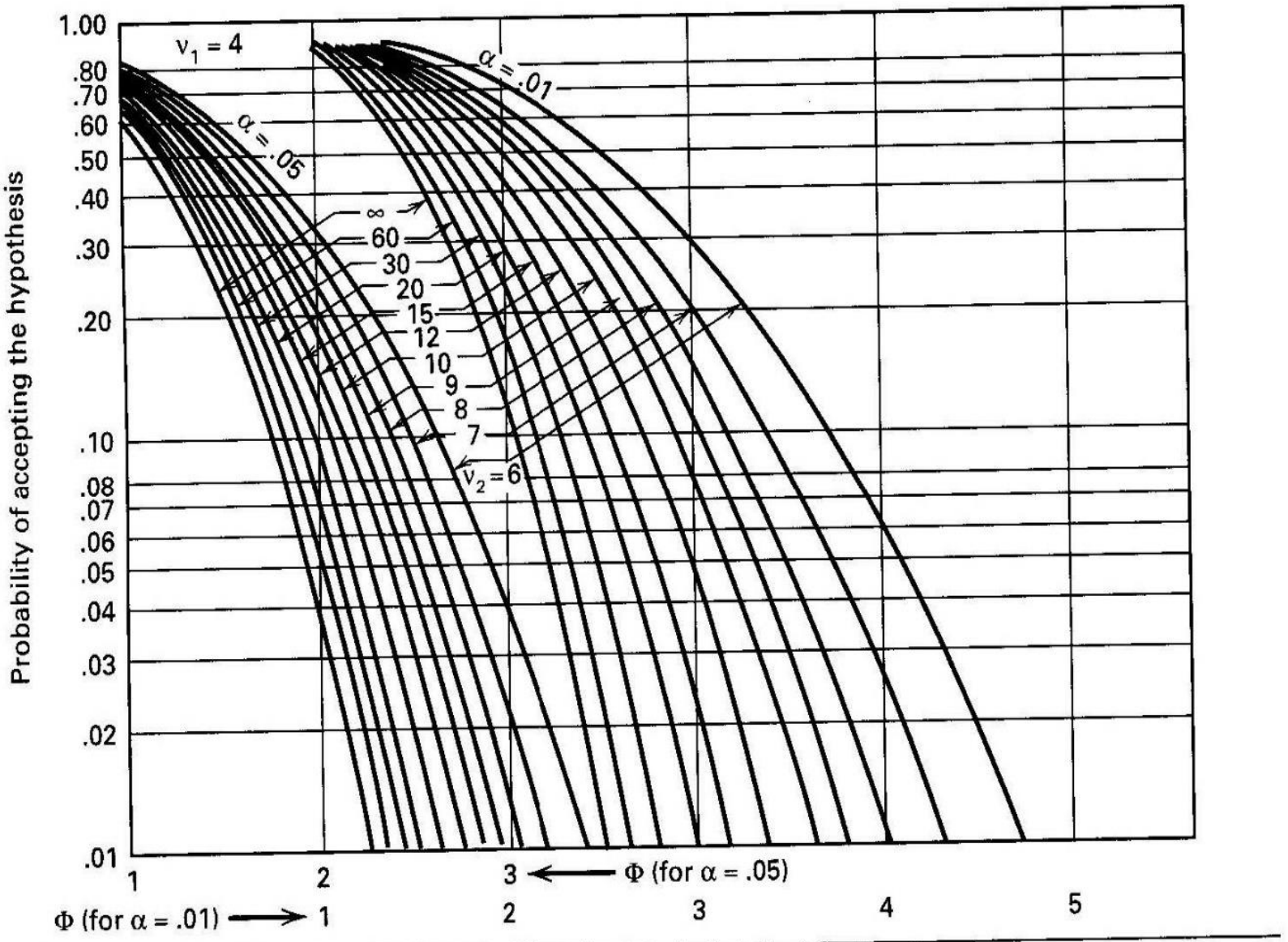
$T5 = \mu 5 - \mu = 19 - 15 = 4$

Thus, $\Sigma_i T_i^2 = 50$. Suppose the experimenter feels that the standard deviation of tensile strength at any particular level of cotton weight percentage will be no larger than $\sigma = 3$ psi. Then, by using the Equation, we have:

$\Phi^2 = n\Sigma\tau_i^2 / a\sigma^2 = 1.11n$

We use the operating characteristic curve for $a - 1 = 5 - 1 = 4$ with, $N - a = a(n - 1) = 5(n - 1)$ error degrees of freedom and $\alpha = 0.01$ (see OC curves at Appendix). As a first guess at the required sample size, try $n = 4$ replicates. This yields $\Phi^2 = 1.11(4) = 4.44$, $\Phi = 2.11$, and $5(3) = 15$ error degrees of freedom. Consequently,

- from Chart V, we find that $ß = 0.30$. Therefore, the power of the test is approximately
- $1-ß = 1- 0.30 = 0.70$, which is less than the required 0.90, and so we conclude that
- $n = 4$ replicates are not sufficient. Proceeding in a similar manner, we can construct the
- following display:


Probability of accepting the hypothesis
1.00
.80
.70
.60
.50
.40
.30
.20
.10
.08
.07
.06
.05
.04
.03
.02
.01
$\nu_1 = 4$
$\alpha = .01$
$\alpha = .05$
∞
60
30
20
15
12
10
9
8
7
$\nu_2 = 6$
1
2
3
4
5
Φ (for $\alpha = .05$)
Φ (for $\alpha = .01$)

n	Φ^2	Φ	a(n -1)	ß	Power (1- ß)
4	4.44	2.11	15	0.30	0.70
5	5.55	2.36	20	0.15	0.85
6	6.66	2.58	25	0.04	0.96

Thus, at least $n = 6$ replicates must be run to obtain a test with the required power.

The only problem with this approach to using the operating characteristic curves is that it is usually difficult to select a set of treatment means on which the sample size decision should be based. 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 hy pothesis should be rejected.If the difference between any two treatment means is as large as *D*, it can be shown that the minimum value of Φ^2 is: $\Phi^2 = nD^2/ 2a\sigma 2$

To illustrate this approach, suppose that in the tensile strength experiment Example, the experimenter wished to reject the null hypothesis with probability at least 0.90 if any two treatment means differed by as much as 10 psi. Then, assuming that $\sigma = 3$ psi, we find the minimum value of Φ^2 to be:

$$\Phi^2 = n(10)^2 / 2(5)(3^2) = 1.11\, n$$

from the analysis in Example 3-11, we conclude that $n = 6$ replicates are required to give the desired sensitivity when $\alpha = 0.01$.

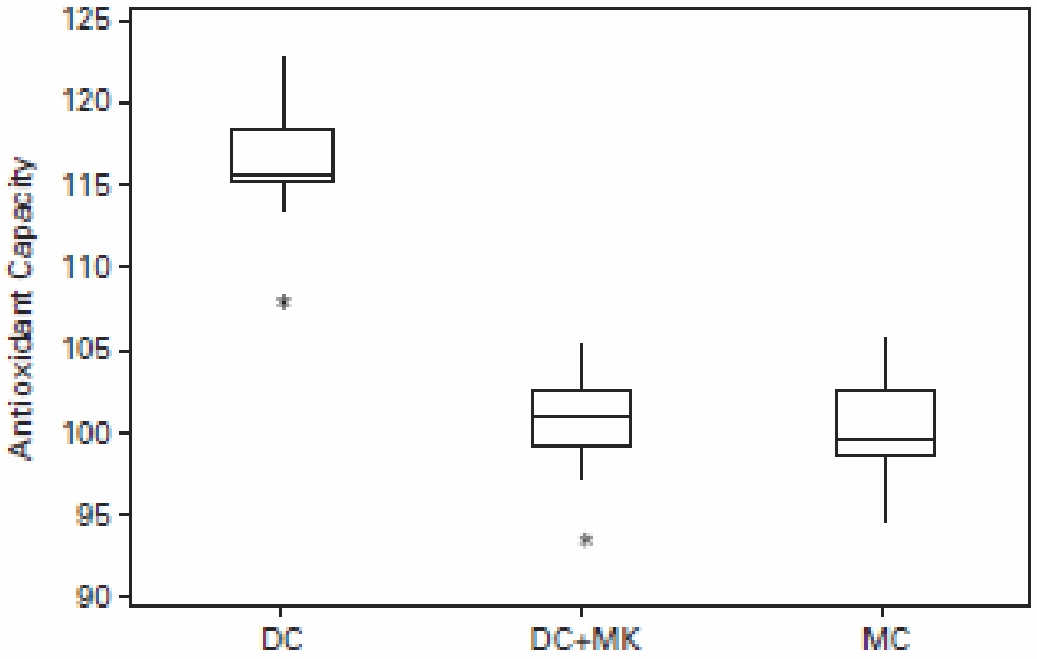
3.8 Other Examples of Single-Factor Experiments

3.8.1 Chocolate and Cardiovascular Health

An article in *Nature* describes an experiment to investigate the effect of consuming chocolate on cardiovascular health ("Plasma Antioxidants from Chocolate," *Nature*, Vol. 424, 2003, pp. 1013). The experiment consisted of using three different types of chocolates: 100 g of dark chocolate, 100 g of dark chocolate with 200 mL of full-fat milk, and 200 g of milk chocolate. Twelve subjects were used, 7 women and 5 men, with an average age range of 32.2 $\pm$ 1 years, an average weight of 65.8 $\pm$ 3.1 kg, and body-mass index of 21.9 $\pm$ 0.4 kg m^{-2}. On different days a subject consumed one of the chocolate-factor levels and one hour later the total antioxidant capacity of their blood plasma was measured in an assay. Data similar to that summarized in the article are shown in Table 3.12.

■ **TABLE 3.12**
Blood Plasma Levels One Hour Following Chocolate Consumption

	Subjects (Observations)											
Factor	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
DC	118.8	122.6	115.6	113.6	119.5	115.9	115.8	115.1	116.9	115.4	115.6	107.9
DC+MK	105.4	101.1	102.7	97.1	101.9	98.9	100.0	99.8	102.6	100.9	104.5	93.5
MC	102.1	105.8	99.6	102.7	98.8	100.9	102.8	98.7	94.7	97.8	99.7	98.6



■ **FIGURE 3.15** **Box plots of the blood antioxidant capacity data from the chocolate consumption experiment**

■ **TABLE 3.13**

Minitab ANOVA Output, Chocolate Consumption Experiment

One-way ANOVA: DC, DC+MK, MC

```
Source  DF       SS      MS       F       P
Factor   2   1952.6   976.3   93.58   0.000
Error   33    344.3    10.4
Total   35   2296.9

S = 3.230    R-Sq = 85.01%    R-Sq(adj) = 84.10%

                            Individual 95% CIs For Mean Based on
                            Pooled StDev
Level    N     Mean   StDev   ---+---------+---------+---------+------
DC      12   116.06    3.53                                  (---*---)
DC+MK   12   100.70    3.24    (--*---)
MC      12   100.18    2.89   (--*---)
                              ---+---------+---------+---------+------
                             100.0     105.0     110.0     115.0

Pooled StDev = 3.23

Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons
Simultaneous confidence level = 88.02
DC subtracted from:

         Lower   Center    Upper      -+---------+---------+---------+---
DC+MK  -18.041  -15.358  -12.675       (---*----)
MC     -18.558  -15.875  -13.192      (----*---)
                                      -+---------+---------+---------+---
                                   -18.0     -12.0     -6.0       0.0

DC+MK subtracted from:

     Lower   Center  Upper      -+---------+---------+---------+--------
MC  -3.200  -0.517  2.166                                (---*----)
                                -+---------+---------+---------+--------
                              -18.0     -12.0     -6.0       0.0
```

Conclusions?

Designed experiments have had tremendous impact on manufacturing industries, including the design of new products and the improvement of existing ones, development of new manufacturing processes, and process improvement. In the last 15 years, designed experiments have begun to be widely used outside of this traditional environment. These applications are in financial services, telecommunications, health care, e-commerce, legal services, marketing, logistics and transporation, and many of the nonmanufacturing components of manufacturing businesses. These types of businesses are sometimes referred to as the real economy. It has been estimated that manufacturing accounts for only about 20 percent of the total US economy, so applications of experimental design in the real economy are of growing importance. In this section, we present an example of a designed experiment in marketing.

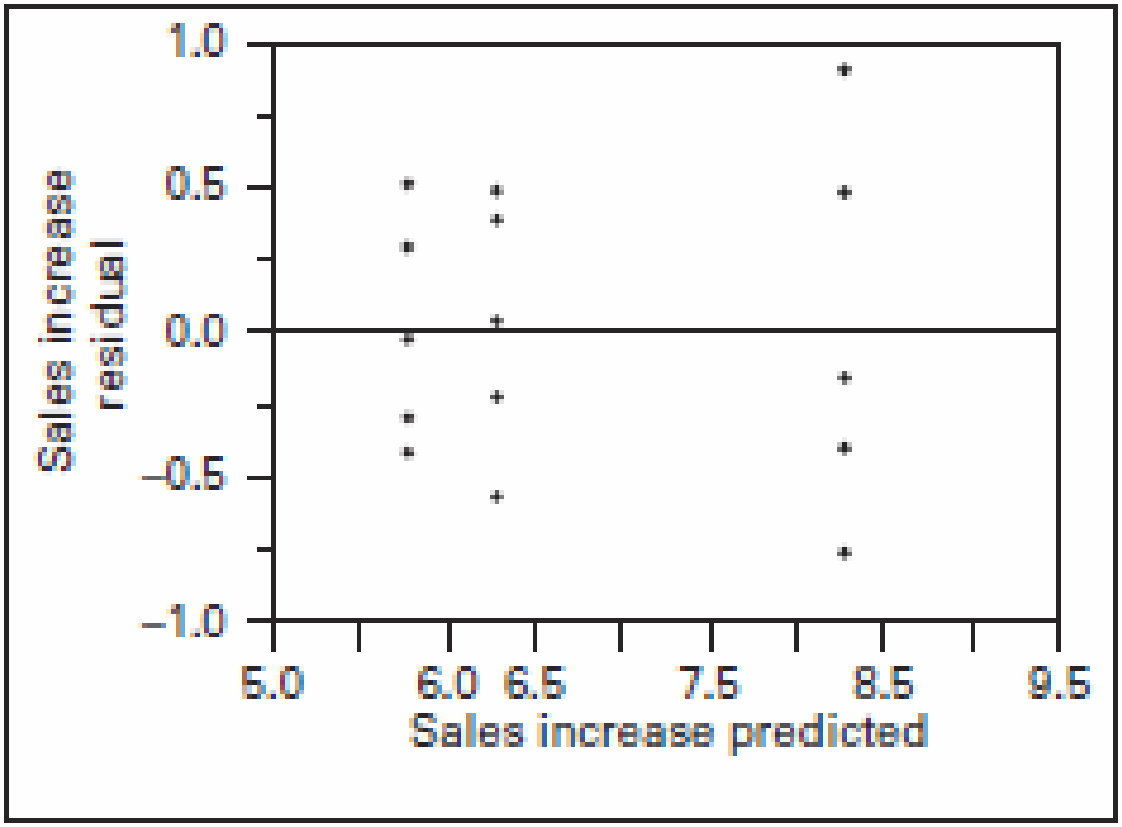
Example:

A soft drink distributor knows that end-aisle displays are an effective way to increase sales of the product. However, there are several ways to design these displays: by varying the text displayed, the colors used, and the visual images. The marketing group has designed three new end-aisle displays and wants to test their effectiveness. They have identified 15 stores of similar size and type to participate in the study. Each store will test one of the displays for a period of one month. The displays are assigned at random to the stores, and each display is tested in five stores. The response variable is the percentage increase in sales activity over the typical sales for that store when the end-aisle display is not in use. The data from this experiment are shown in Table 3.13.

■ **TABLE 3.13**
The End-Aisle Display Experimental Design

Display Design	Sample Observations, Percent Increase in Sales				
1	5.43	5.71	6.22	6.01	5.29
2	6.24	6.71	5.98	5.66	6.60
3	8.79	9.20	7.90	8.15	7.55

Residual by Predicted Plot



3.9 The Random Effects Model

- There are a large number of possible levels for the factor (theoretically an infinite number)
- The experimenter chooses '*a*' of these levels at random
- Inference will be to the entire population of levels

The linear statistical model is

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases} \tag{3.47}$$

where both the treatment effects τ_i and ϵ_{ij} are random variables. We will assume that the treatment effects τ_i are NID $(0, \sigma_\tau^2)$ random variables[1] and that the errors are NID $(0, \sigma^2)$, random variables, and that the τ_i and ϵ_{ij} are independent. Because τ_i is independent of ϵ_{ij}, the variance of any observation is

$$V(y_{ij}) = \sigma_\tau^2 + \sigma^2$$

Variance components

The basic ANOVA sum of squares identity

$$SS_T = SS_{\text{Treatments}} + SS_E \tag{3.48}$$

is still valid. That is, we partition the total variability in the observations into a component that measures the variation between treatments ($SS_{\text{Treatments}}$) and a component that measures the variation within treatments (SS_E). Testing hypotheses about individual treatment effects is not very meaningful because they were selected randomly, we are more interested in the **population** of treatments, so we test hypotheses about the variance component σ_τ^2.

$$\begin{aligned} H_0&: \sigma_\tau^2 = 0 \\ H_1&: \sigma_\tau^2 > 0 \end{aligned} \tag{3.49}$$

If $\sigma_\tau^2 = 0$, all treatments are identical; but if $\sigma_\tau^2 = 0$, variability exists between treatments. As before, SS_E/σ^2 is distributed as chi-square with $N - a$ degrees of freedom and, under the null hypothesis, $SS_{\text{Treatments}}/\sigma^2$ is distributed as chi-square with $a - 1$ degrees of freedom. Both random variables are independent. Thus, under the null hypothesis $\sigma_\tau^2 = 0$, the ratio

$$F_0 = \frac{\dfrac{SS_{\text{Treatments}}}{a-1}}{\dfrac{SS_E}{N-a}} = \frac{MS_{\text{Treatments}}}{MS_E} \tag{3.50}$$

is distributed as F with $a - 1$ and $N - a$ degrees of freedom. However, we need to examine the expected mean squares to fully describe the test procedure.

$$E(MS_{\text{Treatments}}) = \frac{1}{a-1}E(SS_{\text{Treatments}}) = \frac{1}{a-1}E\left[\sum_{i=1}^{a}\frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}\right]$$

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

$$= \frac{1}{a-1}[N\mu^2 + N\sigma_\tau^2 + a\sigma^2 - N\mu^2 - n\sigma_\tau^2 - \sigma^2]$$

$$E(MS_{\text{Treatments}}) = \sigma^2 + n\sigma_\tau^2$$

$$E(MS_E) = \sigma^2$$

ANOVA *F*-test is identical to the fixed-effects case

Estimating the variance components using the ANOVA method:

$$MS_{\text{Treatments}} = \sigma^2 + n\sigma_\tau^2$$

$$MS_E = \sigma^2$$

$$\hat{\sigma}^2 = MS_E$$

$$\hat{\sigma}_\tau^2 = \frac{MS_{\text{Treatments}} - MS_E}{n}$$

- The ANOVA variance component estimators are **moment estimators**
- Normality not required
- They are unbiased estimators
- Finding confidence intervals on the variance components is “clumsy”
- Negative estimates can occur – this is “embarrassing”, as variances are always non-negative

EXAMPLE 3.11

A textile company weaves a fabric on a large number of looms. It would like the looms to be homogeneous so that it obtains a fabric of uniform strength. The process engineer suspects that, in addition to the usual variation in strength within samples of fabric from the same loom, there may also be significant variations in strength between looms. To investigate this, she selects four looms at random and makes four strength determinations on the fabric manufactured on each loom. This experiment is run in random order, and the data obtained are shown in Table 3.17. The ANOVA is conducted and is shown in Table 3.18. From the ANOVA, we conclude that the looms in the plant differ significantly.

■ **TABLE 3.17**
Strength Data for Example 3.11

	Observations				
Looms	**1**	**2**	**3**	**4**	$y_{i\cdot}$
1	98	97	99	96	390
2	91	90	93	92	366
3	96	95	97	95	383
4	95	96	99	98	388

$1527 = y_{\cdot\cdot}$

The variance components are estimated by $\hat{\sigma}^2 = 1.90$ and

$$\hat{\sigma}_\tau^2 = \frac{29.73 - 1.90}{4} = 6.96$$

Therefore, the variance of any observation on strength is estimated by

$$\hat{\sigma}_y = \hat{\sigma}^2 + \hat{\sigma}_\tau^2 = 1.90 + 6.96 = 8.86.$$

Most of this variability is attributable to differences *between* looms.

■ **TABLE 3.18**
Analysis of Variance for the Strength Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	*P*-Value
Looms	89.19	3	29.73	15.68	<0.001
Error	22.75	12	1.90		
Total	111.94	15			

- Confidence interval for the error variance:

$$\frac{(N-a)MS_E}{\chi^2_{\alpha/2,N-a}} \leq \sigma^2 \leq \frac{(N-a)MS_E}{\chi^2_{1-(\alpha/2),N-a}}$$

- Confidence interval for the interclass correlation:

$$\frac{L}{1+L} \leq \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \leq \frac{U}{1+U}$$

$$L = \frac{1}{n}\left(\frac{MS_{\text{Treatments}}}{MS_E}\frac{1}{F_{\alpha/2,a-1,N-a}} - 1\right)$$

$$U = \frac{1}{n}\left(\frac{MS_{\text{Treatments}}}{MS_E}\frac{1}{F_{1-\alpha/2,a-1,N-a}} - 1\right)$$

Estimation of the Overall Mean μ. In many random effects experiments the experimenter is interested in estimating the overall mean μ. From the basic model assumptions it is easy to see that the expected value of any observation is just the overall mean. Consequently, an unbiased estimator of the overall mean is

$$\hat{\mu} = \bar{y}_{..}$$

So for Example 3.11 the estimate of the overall mean strength is

$$\hat{\mu} = \bar{y}_{..} = \frac{y_{..}}{N} = \frac{1527}{16} = 95.44$$

It is also possible to find a $100(1-\alpha)\%$ confidence interval on the overall mean. The variance of $\bar{y}$ is

$$V(\bar{y}_{..}) = V\left(\frac{\sum_{i-1}^{a}\sum_{j-1}^{n} y_{ij}}{an}\right) = \frac{n\sigma_\tau^2 + \sigma^2}{an}$$

The numerator of this ratio is estimated by the treatment mean square, so an unbiased estimator of $V(\bar{y})$ is

$$\hat{V}(\bar{y}_{..}) = \frac{MS_{\text{Treatments}}}{an}$$

Therefore, the $100(1-\alpha)\%$ CI on the overall mean is

$$\bar{y}_{..} - t_{\alpha/2,\, a(n-1)}\sqrt{\frac{MS_{\text{Treatments}}}{an}} \le \mu \le \bar{y}_{..} + t_{\alpha/2,\, a(n-1)}\sqrt{\frac{MS_{\text{Treatments}}}{an}} \qquad (3.61)$$

3.11 Nonparametric Methods in the Analysis of Variance

3.11.1 The Kruskal–Wallis Test

In situations where the normality assumption is unjustified, the experimenter may wish to use an alternative procedure to the F test analysis of variance that does not depend on this assumption. Such a procedure has been developed by Kruskal and Wallis (1952). The Kruskal–Wallis test is used to test the null hypothesis that the a treatments are identical against the alternative hypothesis that some of the treatments generate observations that are larger than others. Because the procedure is designed to be sensitive for testing differences in means, it is sometimes convenient to think of the Kruskal–Wallis test as a test for equality of treatment means. The Kruskal–Wallis test is a **nonparametric alternative** to the usual analysis of variance.

To perform a Kruskal–Wallis test, first rank the observations y_{ij} in ascending order and replace each observation by its rank, say R_{ij}, with the smallest observation having rank 1. In

the case of ties (observations having the same value), assign the average rank to each of the tied observations. Let $R_{i.}$ be the sum of the ranks in the ith treatment. The test statistic is

$$H = \frac{1}{S^2}\left[\sum_{i=1}^{a}\frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}\right] \tag{3.67}$$

where n_i is the number of observations in the ith treatment, N is the total number of observations, and

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{a}\sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4}\right] \tag{3.68}$$

Note that S^2 is just the variance of the ranks. If there are no ties, $S^2 = N(N+1)/12$ and the test statistic simplifies to

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{a}\frac{R_{i.}^2}{n_i} - 3(N+1) \tag{3.69}$$

When the number of ties is moderate, there will be little difference between Equations 3.68 and 3.69, and the simpler form (Equation 3.69) may be used. If the n_i are reasonably large, say $n_i \geq 5$, H is distributed approximately as χ^2_{a-1} under the null hypothesis. Therefore, if

$$H > \chi^2_{\alpha,a-1}$$

the null hypothesis is rejected. The P-value approach could also be used.

EXAMPLE 3.12

The data from Example 3.1 and their corresponding ranks are shown in Table 3.20. There are ties, so we use Equation 3.67 as the test statistic. From Equation 3.67

$$S^2 = \frac{1}{19}\left[2869.50 - \frac{20(21)^2}{4}\right] = 34.97$$

and the test statistic is

$$H = \frac{1}{S^2}\left[\sum_{i=1}^{a}\frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}\right]$$
$$= \frac{1}{34.97}[2796.30 - 2205]$$
$$= 16.91$$

■ **TABLE 3.20**
Data and Ranks for the Plasma Etching Experiment in Example 3.1

Power							
160		180		200		220	
y_{1j}	R_{1j}	y_{2j}	R_{2j}	y_{3j}	R_{3j}	y_{4j}	R_{4j}
575	6	565	4	600	10	725	20
542	3	593	9	651	15	700	17
530	1	590	8	610	11.5	715	19
539	2	579	7	637	14	685	16
570	5	610	11.5	629	13	710	18
$R_{i.}$	17		39.5		63.5		90

Because $H > \chi^2_{0.01,3} = 11.34$, we would reject the null hypothesis and conclude that the treatments differ. (The P-value for $H = 16.91$ is $P = 7.38 \times 10^{-4}$.) This is the same conclusion as given by the usual analysis of variance F test.

Problem. A manufacturing engineer was concerned about the density of bricks. He conducted an experiment on the brick-manufacturing process, to determine the effects of firing temperatures on the density of a certain type of brick.
Four specific firing temperatures were selected to be used in this experiment. This experiment led to the following data:

Temperature (oF)	Density				
100	15.3	15.3	15.2	15.3	15.4
130	15.7	15.4	15.5	15.5	-
160	15.9	15.8	15.8	15.6	15.5
190	15.9	15.7	15.8	15.7	-

Q1- Does the firing temperature affect the density of the brick? What the experimenter should do to decrease the Type I Error? Analyze the residuals from the experiment.

Q1-Answer: Ho: $\mu_{100} = \mu_{130} = \mu_{160} = \mu_{190}$ and H_1: Not.

ANOVA	SS	df	MS	Fo		
Treatment	0.653111111	3	0.217704	15.01405		
Error	0.203	14	0.0145			
Total	0.856111111	17			F0.05,3,14:	3.34

Therefore, reject Ho.

Q2- Assuming that the normality assumption of ANOVA is unjustified, test the null hypothesis that the 'a treatments' are identical? Did you obtain a similar decision to the one obtained from ANOVA?

Q2- Answer: Let's, first, sort the density from minimum to maximum.

Density	i	j	Rank
15.2	100	3	1
15.3	100	1	3
15.3	100	2	3
15.3	100	4	3
15.4	100	5	5.5
15.4	130	2	5.5
15.5	130	3	8
15.5	130	4	8
15.5	160	5	8
15.6	160	4	10
15.7	130	1	12
15.7	190	2	12

Density	i	j	Rank
15,7	190	4	12
15.8	160	2	15
15.8	160	3	15
15.8	190	3	15
15.9	160	1	17.5
15.9	190	1	17.5

Temperature	$Y_{i,j}$	$R_{i,j}$	$Y_{i,j}$	$R_{i,j}$	$Y_{i,j}$	$R_{i,j}$	$Y_{i,j}$	$R_{i,j}$	$Y_{i,j}$	$R_{i,j}$
100	15.3	3	15.3	3	15.2	1	15.3	3	15.4	5.5
130	15.7	12	15.4	5.5	15.5	8	15.5	8		
160	15.9	17.5	15.8	15	15.8	15	15.6	10	15.5	8
190	15.9	17.5	15.7	12	15.8	15	15.7	12		
Sum	62.8		62.2		62.3		62.1		30.9	
Ho: μ1=μ2=μ3=μ4										
H1: Not.										

There are ties.

Y1j	R1j	Y2j	R2j	Y3j	R3j	Y4j	R4j	SUMj(Rij^2)
15.3	3	15.7	12	15.9	17.5	15.9	17.5	765.5
15.3	3	15.4	5.5	15.8	15	15.7	12	408.25
15.2	1	15.5	8	15.8	15	15.8	15	515
15.3	3	15.5	8	15.6	10	15.7	12	317
15.4	5.5			15.5	8			94.25
							SUMi:	2100
			S^2=	[1/(18-1)]*[SUMi SUMj Rij^2 -(18*(19)^2)/4]=				27.97
			H=	1/S^2*(SUMi ((Ri.^2)/ni) - (18*19^2)/4)=				12.8787

$X^2_{0.05,3}=7.81$

Since H= 12.8787 > 7.81 We reject Ho.