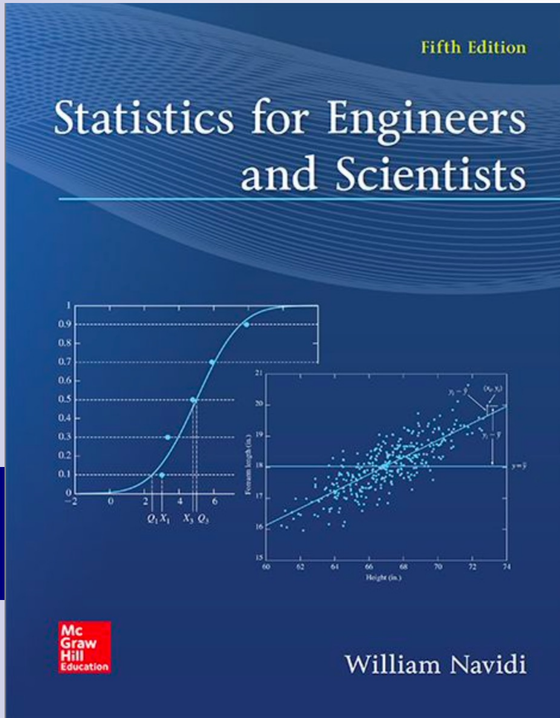


Chapter 9

Factorial Experiments

Chapter 9 Overview

9-1 One-Factor Experiments

NOTE: Sections 9-2, 9-3, 9-4 and 9-5 are not required in this course.

Introduction

- Experiments are essential to the development and improvement of engineering and scientific methods.
- Only through experimentation can different variants of a method be compared to see which are most effective.
- An experiment must be designed properly, and the data it produces must be analyzed correctly.
- Chapter 9 discusses the design of and the analysis of data from a class of experiments known as **factorial experiments**.

9-1 One-Factor Experiment

- Let us analyze the results of hardness measurements, on the Brinell scale, of five welds using each of four fluxes, presented in Table 9.1 (p.670).

TABLE 9.1 Brinell hardness of welds using four different fluxes

Flux	Sample Values					Sample Mean	Sample Standard Deviation
A	250	264	256	260	239	253.8	9.7570
B	263	254	267	265	267	263.2	5.4037
C	257	279	269	273	277	271.0	8.7178
D	253	258	262	264	273	262.0	7.4498

"An Investigation of the CaCO3-CaF2-K2SiO3-SiO2-Fe Flux System Using the Submerged Arc Welding Process on HSLA-100 and AISI-1081 Steels" (G. Fredrickson, M.S. Thesis, Colorado School of Mines, 1992)

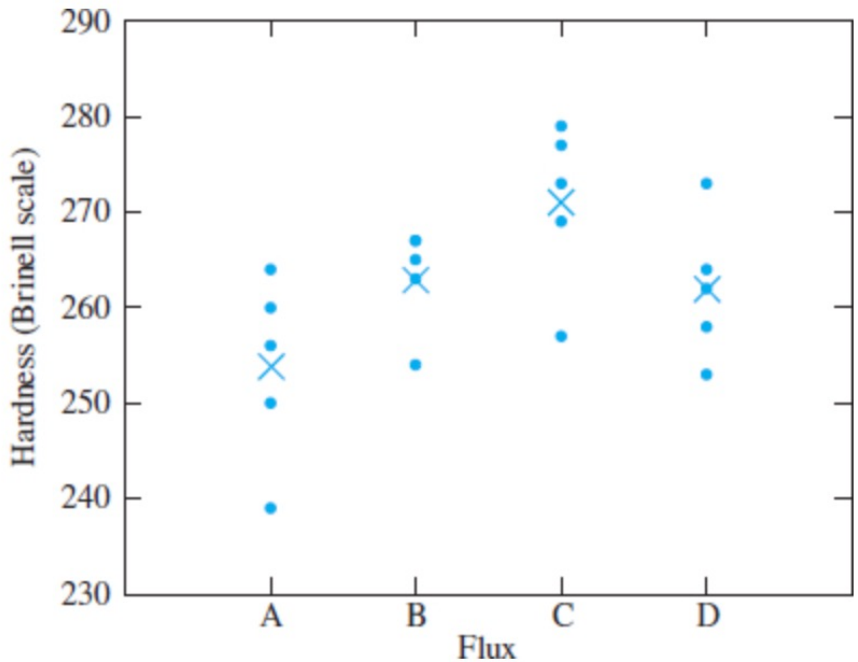


FIGURE 9.1 Dotplots for each sample in Table 9.1. Each sample mean is marked with an "X." The sample means differ somewhat, but the sample values overlap considerably.

Can we conclude that there are differences in the population means among the four flux types?

9-1 One-Factor Experiment…

- This is an example of a factorial experiment.
- In general, a factorial experiment involves several variables.
- One variable is the response variable, which is sometimes called the outcome variable or the dependent variable.
- The other variables are called factors.

- The question addressed by a factorial experiment is whether varying the levels of the factors produces a difference in the mean of the response variable.

9-1 One-Factor Experiment…

- In the experiment described in Table 9.1, the hardness is the response, and there is one factor: flux type.
- Since there is only one factor, this is a one-factor experiment.
- There are four different values for the flux-type factor in this experiment.
- These different values are called the levels of the factor and can also be called **treatments**.

9-1 One-Factor Experiment…

- The objects upon which measurements are made are called **experimental units**.
- The units assigned to a given treatment are called **replicates**.
- In the preceding experiment, the welds are the experimental units, and there are five replicates for each treatment.

TABLE 9.1 Brinell hardness of welds using four different fluxes

Flux	Sample Values					Sample Mean	Sample Standard Deviation
A	250	264	256	260	239	253.8	9.7570
B	263	254	267	265	267	263.2	5.4037
C	257	279	269	273	277	271.0	8.7178
D	253	258	262	264	273	262.0	7.4498

9-1 One-Factor Experiment…

- In this welding experiment, the four particular flux compositions were chosen deliberately by the experimenter, rather than at random from a larger population of fluxes.
- Such an experiment is said to follow a fixed effects model.
- In some experiments, treatments are chosen at random from a population of possible treatments.
- In this case the experiment is said to follow a random effects model.

Completely Randomized Experiments

Definition

A factorial experiment in which experimental units are assigned to treatments at random, with all possible assignments being equally likely, is called a **completely randomized experiment**.

TABLE 9.1 Brinell hardness

Flux	Sample Values				
A	250	264	256	260	239
B	263	254	267	265	267
C	257	279	269	273	277
D	253	258	262	264	273

Completely Randomized Experiments…

- In a completely randomized experiment, it is appropriate to think of each treatment as representing a population, and the responses observed for the units assigned to that treatment as a simple random sample from that population.

- The data from the experiment consist of several random samples, each from a different population.

- The population means are called treatment means.

One-Way Analysis of Variance

- The questions of interest concern the treatment means is whether they are all equal, and if not, which ones are different, how big the differences are.

- To make a formal determination as to whether the treatment means differ, a **hypothesis test** is needed.

One-Way Analysis of Variance…

- We have I samples, each from a different treatment. The treatment means are denoted $\mu_1, \ldots, \mu_I$ …
- The sample sizes are $J_1, \ldots, J_I$
- The total number in all the samples combined is $N = J_1 + J_2 + \ldots + J_I$
- The hypotheses are
  $H_0: \mu_1 = \cdots = \mu_I$
  H_1: two or more of the μ_i are different

One-Way Analysis of Variance…

- If there were only two samples, we could use the two-sample *t test* (see Section 6.7) to test the null hypothesis.

- Since there are more than two samples, we use a method known as **one-way analysis of variance (ANOVA)**.

- To define the test statistic for one-way ANOVA, we need to use a different notation for the sample observations.

One-Way Analysis of Variance…

- Since there are several samples, we use a double subscript to denote the observations.

- Specifically, we let X_{ij} denote the *j*th observation in the *i*th sample.

One-Way Analysis of Variance…

- The sample mean of the *i*th sample is

$$\overline{X}_{i.} = \frac{\sum_{j=1}^{J_i} X_{ij}}{J_i} \qquad (9.1)$$

- The sample grand mean, is the average of all the sampled items taken together

$$\overline{X}_{..} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J_i} X_{ij}}{N} \qquad (9.2)$$

- So, the sample grand mean is also a weighted average of the sample means:

$$\overline{X}_{..} = \frac{\sum_{i=1}^{I} J_i \overline{X}_{i.}}{N} \qquad (9.3)$$

Example 9.1

For the data in Table 9.1, find $I, J_1, \ldots, J_I, N, X_{23}, \overline{X}_{3.}, \overline{X}_{..}$.

TABLE 9.1 Brinell hardness of welds using four different fluxes

Flux	Sample Values					Sample Mean	Sample Standard Deviation
A	250	264	256	260	239	253.8	9.7570
B	263	254	267	265	267	263.2	5.4037
C	257	279	269	273	277	271.0	8.7178
D	253	258	262	264	273	262.0	7.4498

SOLUTION:

- There are four samples, so I = 4.
- Each sample contains five observations, so $J_1 = J_2 = J_3 = J_4 = 5$.
- The total number of observations is N = 20.
- The quantity X_{23} is the third observation in the second sample, which is 267.

Example 9.1

For the data in Table 9.1, find $I, J_1, \ldots, J_I, N, X_{23}, \overline{X}_{3.}, \overline{X}_{..}$.

TABLE 9.1 Brinell hardness of welds using four different fluxes

Flux	Sample Values					Sample Mean	Sample Standard Deviation
A	250	264	256	260	239	253.8	9.7570
B	263	254	267	265	267	263.2	5.4037
C	257	279	269	273	277	271.0	8.7178
D	253	258	262	264	273	262.0	7.4498

SOLUTION (cont.):

- The $\bar{X}_{3.}$ is the mean of the third sample, so $\bar{X}_{3.} = 271.0$
- The sample grand mean $\bar{X}_{..}$ can be computed from Eq. (9.3)

$$\overline{X}_{..} = \frac{\sum_{i=1}^{I} J_i \overline{X}_{i.}}{N} = \frac{(5)(253.8) + (5)(263.2) + (5)(271.0) + (5)(262.0)}{20} = 262.5$$

One-Way Analysis of Variance…

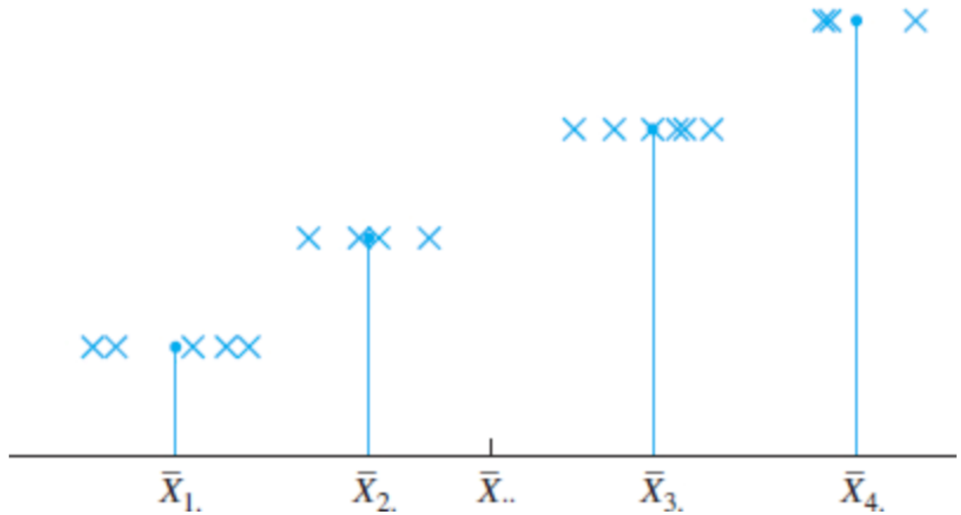


FIGURE 9.2 The variation of the sample means around the sample grand mean can be due both to random uncertainty and to differences among the treatment means. The variation within a given sample around its own sample mean is due only to random uncertainty.

- One-way ANOVA provides a way to measure this spread.
- If the sample means are highly spread out, then it is likely that the treatment means are different, and we will reject H_0.

One-Way Analysis of Variance…

- The variation of the sample means around the sample grand mean is measured by a quantity called the **treatment sum of squares (SSTr)**

$$\text{SSTr} = \sum_{i=1}^{I} J_i(\overline{X}_{i.} - \overline{X}_{..})^2 \qquad (9.4)$$

- An equivalent formula for SSTr, which is a bit easier to compute by hand, is

$$\text{SSTr} = \sum_{i=1}^{I} J_i \overline{X}_{i.}^{\,2} - N\overline{X}_{..}^{\,2} \qquad (9.5)$$

One-Way Analysis of Variance…

- In order to determine whether SSTr is large enough to reject H_0, we compare it to another sum of squares, called **the error sum of squares (SSE)**

$$\mathrm{SSE} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (X_{ij} - \overline{X}_{i.})^2 \tag{9.6}$$

- In other words, the SSE is the sum of the squared residuals.

One-Way Analysis of Variance…

An equivalent formula for SSE, which is a bit easier to compute by hand, is

$$\mathrm{SSE} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} X_{ij}^2 - \sum_{i=1}^{I} J_i \overline{X}_{i.}^{\,2} \tag{9.7}$$

Another equivalent formula for SSE is based on the sample variances. Let s_i^2 denote the sample variance of the ith sample. Then

$$s_i^2 = \frac{\sum_{j=1}^{J_i} (X_{ij} - \overline{X}_{i.})^2}{J_i - 1} \tag{9.8}$$

It follows from Equation (9.8) that $\sum_{j=1}^{J_i} (X_{ij} - \overline{X}_{i.})^2 = (J_i - 1)s_i^2$. Substituting into Equation (9.6) yields

$$\mathrm{SSE} = \sum_{i=1}^{I} (J_i - 1)s_i^2 \tag{9.9}$$

Example 9.2

For the data in Table 9.1, compute SSTr and SSE.

TABLE 9.1 Brinell hardness of welds using four different fluxes

Flux	Sample Values					Sample Mean	Sample Standard Deviation
A	250	264	256	260	239	253.8	9.7570
B	263	254	267	265	267	263.2	5.4037
C	257	279	269	273	277	271.0	8.7178
D	253	258	262	264	273	262.0	7.4498

Solution

The sample means are presented in Table 9.1. They are

$$\overline{X}_{1.} = 253.8 \qquad \overline{X}_{2.} = 263.2 \qquad \overline{X}_{3.} = 271.0 \qquad \overline{X}_{4.} = 262.0$$

The sample grand mean was computed in Example 9.1 to be $\overline{X}_{..} = 262.5$.

Example 9.2

For the data in Table 9.1, compute SSTr and SSE.

Solution

The sample means are presented in Table 9.1. They are

$$\overline{X}_{1.} = 253.8 \qquad \overline{X}_{2.} = 263.2 \qquad \overline{X}_{3.} = 271.0 \qquad \overline{X}_{4.} = 262.0$$

The sample grand mean was computed in Example 9.1 to be $\overline{X}_{..} = 262.5$

$$\text{SSTr} = \sum_{i=1}^{I} J_i(\overline{X}_{i.} - \overline{X}_{..})^2 \tag{9.4}$$

$$\begin{aligned}\text{SSTr} &= 5(253.8-262.5)^2+5(263.2-262.5)^2+5(271.0-262.5)^2+5(262.0-262.5)^2 \\ &= 743.4\end{aligned}$$

Example 9.2

For the data in Table 9.1, compute SSTr and SSE.

TABLE 9.1 Brinell hardness of welds using four different fluxes

Flux	Sample Values					Sample Mean	Sample Standard Deviation
A	250	264	256	260	239	253.8	9.7570
B	263	254	267	265	267	263.2	5.4037
C	257	279	269	273	277	271.0	8.7178
D	253	258	262	264	273	262.0	7.4498

$$\mathrm{SSE} = \sum_{i=1}^{I} (J_i - 1)s_i^2 \qquad (9.9)$$

$$\begin{aligned}\mathrm{SSE} &= (5-1)(9.7570)^2 + (5-1)(5.4037)^2 + (5-1)(8.7178)^2 + (5-1)(7.4498)^2 \\ &= 1023.6\end{aligned}$$

Assumptions for One-Way ANOVA

- We can use SSTr and SSE to construct a test statistic, provided the following two assumptions are met.

1. The treatment populations must be normal.

2. The treatment populations must all have the same variance, which we will denote by σ^2.

One-Way Analysis of Variance…

- The mean of SSTr satisfies the condition

$$\mu_{\text{SSTr}} = (I-1)\sigma^2 \quad \text{when } H_0 \text{ is true} \tag{9.10}$$

$$\mu_{\text{SSTr}} > (I-1)\sigma^2 \quad \text{when } H_0 \text{ is false} \tag{9.11}$$

- The likely size of SSE, and thus its mean, does not depend on whether H_0 is true.

- The mean of SSE is given by

$$\mu_{\text{SSE}} = (N-I)\sigma^2 \quad \text{whether or not } H_0 \text{ is true} \tag{9.12}$$

One-Way Analysis of Variance…

- The quantities I − 1 and N − I are the **degrees of freedom** for SSTr and SSE, respectively.

- When a sum of squares is divided by its degrees of freedom, the quantity obtained is called **a mean square**.

- The treatment mean square is denoted MSTr, and the error mean square is denoted MSE.

$$\text{MSTr} = \frac{\text{SSTr}}{I-1} \qquad \text{MSE} = \frac{\text{SSE}}{N-I} \tag{9.13}$$

One-Way Analysis of Variance…

$$\mu_{\mathrm{MSTr}} = \sigma^2 \quad \text{when } H_0 \text{ is true} \tag{9.14}$$

$$\mu_{\mathrm{MSTr}} > \sigma^2 \quad \text{when } H_0 \text{ is false} \tag{9.15}$$

$$\mu_{\mathrm{MSE}} = \sigma^2 \quad \text{whether or not } H_0 \text{ is true} \tag{9.16}$$

Equations (9.14) and (9.16) show that when H_0 is true, MSTr and MSE have the same mean. Therefore, when H_0 is true, we would expect their quotient to be near 1. This quotient is in fact the test statistic. The test statistic for testing $H_0 : \mu_1 = \cdots = \mu_I$ is

$$F = \frac{\mathrm{MSTr}}{\mathrm{MSE}} \tag{9.17}$$

Summary

The *F* test for One-Way ANOVA

To test $H_0: \mu_1 = \cdots = \mu_I$ versus H_1 : two or more of the μ_i are different:

1. Compute $\mathrm{SSTr} = \sum_{i=1}^{I} J_i(\overline{X}_{i.} - \overline{X}_{..})^2 = \sum_{i=1}^{I} J_i\overline{X}_{i.}^{\,2} - N\overline{X}_{..}^{\,2}$.
2. Compute $\mathrm{SSE} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij} - \overline{X}_{i.})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J_i} X_{ij}^2 - \sum_{i=1}^{I} J_i\overline{X}_{i.}^{\,2} = \sum_{i=1}^{I}(J_i - 1)s_i^2$.
3. Compute $\mathrm{MSTr} = \dfrac{\mathrm{SSTr}}{I-1}$ and $\mathrm{MSE} = \dfrac{\mathrm{SSE}}{N-I}$.
4. Compute the test statistic: $F = \dfrac{\mathrm{MSTr}}{\mathrm{MSE}}$.
5. Find the *P*-value by consulting the *F* table (Table A.8 in Appendix A) with $I - 1$ and $N - I$ degrees of freedom.

Example 9.3

For the data in Table 9.1, compute MSTr, MSE, and *F*. Find the *P*-value for testing the null hypothesis that all the means are equal. What do you conclude?

Solution

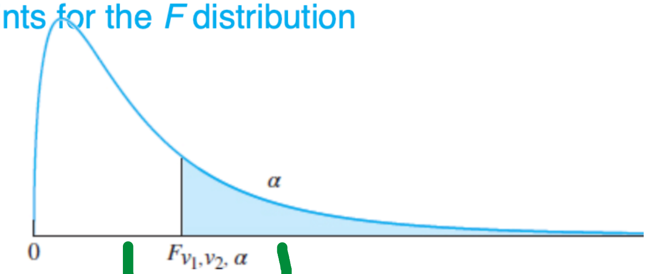
From Example 9.2, SSTr = 743.4 and SSE = 1023.6. We have $I = 4$ samples and $N = 20$ observations in all the samples taken together. Using Equation (9.13),

$$\text{MSTr} = \frac{743.4}{4-1} = 247.8 \qquad \text{MSE} = \frac{1023.6}{20-4} = 63.975$$

The value of the test statistic *F* is therefore

$$\boxed{F = \frac{247.8}{63.975} = 3.8734}$$

TABLE A.8 Upper percentage points for the F distribution



ν_2	aa	ν_1 = 1	2	3	4	5	6	7	8	9
1	**0.100**	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	0.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
	0.010	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47
	0.001	405284	500012	540382	562501	576405	585938	592874	598144	603040
2	**0.100**	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
	0.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	0.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39
3	**0.100**	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	0.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
4	**0.100**	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	0.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47

- To find the P-value, we consult the F table (Table A.8).
- The degrees of freedom are 4 − 1 = 3 for the numerator and 20 − 4 = 16 for the denominator.

		v_1								
v_2	aa	1	2	3	4	5	6	7	8	9
15	**0.100**	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
	0.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
	0.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
	0.001	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26
16	**0.100**	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
	0.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
	0.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
	0.001	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.19	5.98
17	**0.100**	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
	0.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
	0.010	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
	0.001	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75

- Looking at the F table under 3 and 16 degrees of freedom, we find that the upper 5% point is 3.24 and the upper 1% point is 5.29.
- Therefore, the **P-value is between 0.01 and 0.05**

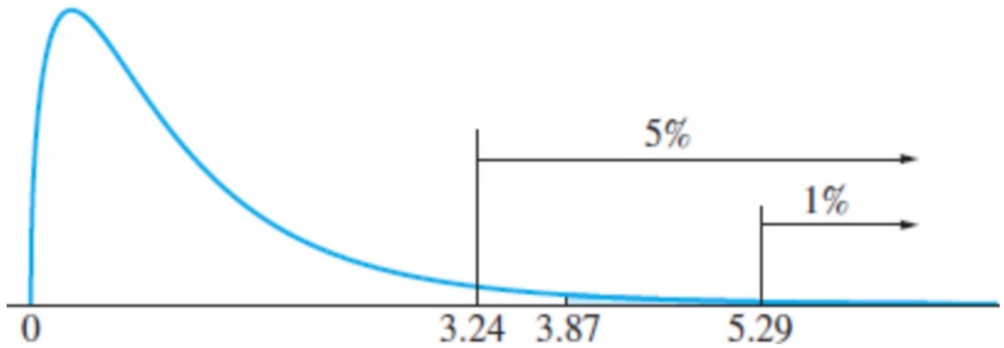


FIGURE 9.3 The observed value of the test statistic is 3.87. The upper 5% point of the $F_{3,16}$ distribution is 3.24. The upper 1% point of the $F_{3,16}$ distribution is 5.29. Therefore the *P*-value is between 0.01 and 0.05. A computer software package gives a value of 0.029.

It is reasonable to conclude that **the population means are not all equal**, and, thus, that flux composition does affect hardness.

The ANOVA Table

The following output (from MINITAB) shows the analysis of variance for the weld data presented in Table 9.1

```
One-way ANOVA: A, B, C, D

Source     DF          SS          MS       F          P
Factor      3      743.40     247.800    3.87      0.029
Error      16     1023.60      63.975
Total      19     1767.00

S = 7.998   R-Sq = 42.07%   R-Sq(adj) = 31.21%


                         Individual 95% CIs For Mean Based on
                         Pooled StDev
Level   N      Mean    StDev   ----+--------+--------+--------+----
A       5    253.80     9.76   (-------*-----)
B       5    263.20     5.40             (-----*------)
C       5    271.00     8.72                    (------*------)
D       5    262.00     7.45            (------*------)
                               ----+--------+--------+--------+----
                               250     260     270      280

Pooled StDev = 8.00
```

Example 9.5

In the article "Review of Development and Application of CRSTER and MPTER Models" (R. Wilson, *Atmospheric Environment*, 1993:41–57), several measurements of the maximum hourly concentrations (in $\mu g/m^3$) of SO_2 are presented for each of four power plants. The results are as follows (two outliers have been deleted):

Plant 1:	438	619	732	638		
Plant 2:	857	1014	1153	883	1053	
Plant 3:	925	786	1179	786		
Plant 4:	893	891	917	695	675	595

Can you conclude that the maximum hourly concentrations differ among the plants?

One-way ANOVA: Plant 1, Plant 2, Plant 3, Plant 4

Source	DF	SS	MS	F	P
Plant	3	378610	126203	6.21	0.006
Error	15	304838	20323		
Total	18	683449			

We conclude that not all the treatment means are equal.

S = 142.6   R-Sq = 55.40%   R-Sq(adj) = 46.48%

```
                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level   N   Mean  StDev  ------+---------+---------+---------+-
1       4  606.8  122.9  (-----*------)
2       5  992.0  122.7                    (-----*----)
3       4  919.0  185.3               (------*------)
4       6  777.7  138.8          (-----*-----)
                         ------+---------+---------+---------+-
                             600       800      1000      1200
```

Pooled StDev = 142.6

Checking the Assumptions

Assumptions for One-Way ANOVA

The standard one-way ANOVA hypothesis tests are valid under the following conditions:

1. The treatment populations must be normal.
2. The treatment populations must all have the same variance, which we will denote by σ^2.

- A good way to check the normality assumption is with a normal probability plot.
- If the sample sizes are large enough, one can construct a separate probability plot for each sample.
- When the sample sizes are not large enough for individual probability plots to be informative, the residuals can all be plotted together in a single plot.

Checking the Assumptions…

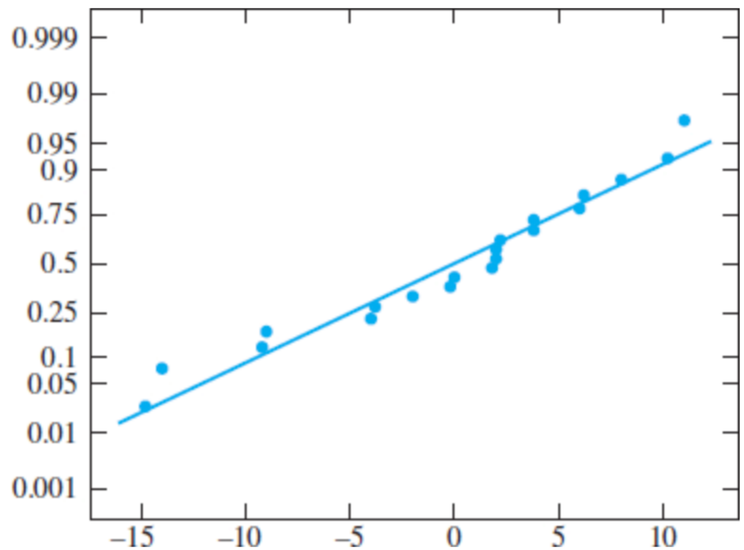


FIGURE 9.4 Probability plot for the residuals from the weld data. There is no evidence of a serious violation of the assumption of normality.

Checking the Assumptions…

- For the weld data, the sample standard deviations range from 5.4037 to 9.7570.
- It is reasonable to proceed as though the variances were equal.

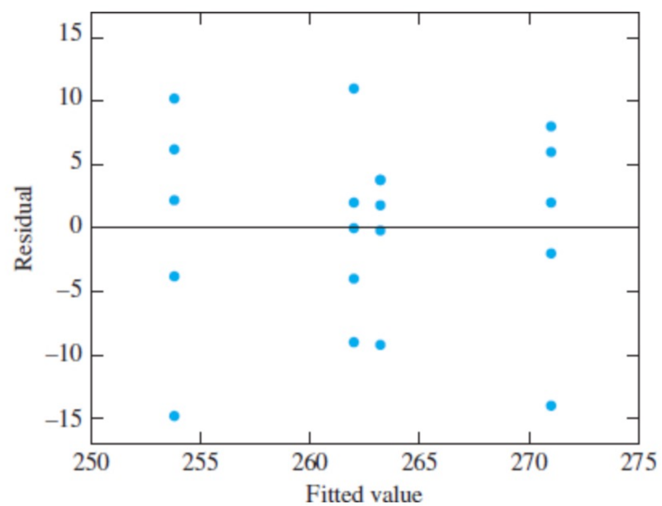


FIGURE 9.5 Residual plot of the values $x_{ij} - \overline{x}_{i.}$ versus $\overline{x}_{i.}$ for the weld data. The spreads do not differ greatly from sample to sample, and there are no serious outliers.