THE ANOVA *F* TEST

Draw an independent SRS from each of I Normal populations that have a common standard deviation but may have different means. The sample from the ith population has size n_i, sample mean $\bar{x}_i$, and sample standard deviation s_i.

To test the null hypothesis that all I populations have the same mean against the alternative hypothesis that not all the means are equal, calculate the **ANOVA *F* statistic**

$$F = \frac{\text{MSG}}{\text{MSE}}$$

The numerator of F is the **mean square for groups**

$$\text{MSG} = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_I(\bar{x}_I - \bar{x})^2}{I - 1}$$

The denominator of F is the **mean square for error**

$$\text{MSE} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2}{N - I}$$

When H_0 is true, F has the ***F* distribution** with $I - 1$ and $N - I$ degrees of freedom.

sums of squares

ANOVA table

The denominators in the formulas for MSG and MSE are the two degrees of freedom $I - 1$ and $N - I$ of the F test. The numerators are called **sums of squares,** from their algebraic form. It is usual to present the results of ANOVA in an **ANOVA table.** Output from software usually includes an ANOVA table.

EXAMPLE 24.7 ANOVA calculations: software

Look again at the three outputs in Figure 24.3. The two software outputs give the ANOVA table. The calculator, with its small screen, gives the degrees of freedom, sums of squares, and mean squares separately. Each output uses slightly different language to identify the two sources of variation. The basic ANOVA table is

Source of variation	df	SS	MS	*F* statistic
Variation among samples	2	1082.87	MSG = 541.44	259.12
Variation within samples	51	106.57	MSE = 2.09	

You can check that each mean square MS is the corresponding sum of squares SS divided by its degrees of freedom df. The F statistic is MSG divided by MSE. ■

pooled standard deviation

Because MSE is an average of the individual sample variances, it is also called the *pooled sample variance,* written as s_p^2. When all I populations have the same population variance σ^2, as ANOVA assumes that they do, s_p^2 estimates the common variance σ^2. The square root of MSE is the **pooled standard deviation** s_p. It estimates the common standard deviation σ of observations in each group. The Minitab and calculator outputs in Figure 24.3 give the value $s_p = 1.446$.