BREAKDOWN TIMES

9. The Lack-of-Fit F-Test

In the insulating fluid experiment replicate values of the time to breakdown are obtained at all the voltage levels. In this case a formal test of the adequacy of the straight-line regression model is available. This test called the *lack-of-fit* F-test compares the regression model to the more general separate-means (one-way analysis of variance) model.

More precisely, the lack-of-fit test tests the null hypothesis

H_0: $\mu\{Ln(Time\,|\,Voltage)\} = \beta_0 + \beta_1 * Voltage$ for some values β_0, β_1,

(the simple linear regression model fits) versus the alternative

H_a: H_0 is not true.

The sum of squared residuals from the more general separate-means model (ANOVA) is smaller than the sum of squared residuals from the simple linear regression model. The lack-of-fit F-statistic measures the reduction in the sum of squared residuals resulting from a generalization from the simple linear regression model to the separate-means (ANOVA) model.

Specifically, the lack-of-fit F-statistic is defined as follows

$$F = \frac{[SS\,\mathrm{Re}\,s_{LR} - SS\,\mathrm{Re}\,s_{SM}]/[d.f._{LR} - d.f._{SM}]}{\hat{\sigma}^2_{SM}},$$

where $SS\,\mathrm{Re}\,s_{LR}$, and $SS\,\mathrm{Re}\,s_{SM}$ are the sums of squares of residuals from the simple linear regression and separate-means models, respectively; and $d.f._{LR}$ and $d.f._{SM}$ are the degrees of freedom associated with these residual sums of squares. The denominator of the F-statistic is the estimate of σ^2 from the separate-means model. The p-value of the test is found as the proportion of values from an F distribution that exceeds the F-statistic. The numerator degrees of freedom are and $d.f._{LR}$ - $d.f._{SM}$, and the denominator degrees of freedom are $d.f._{SM}$.

Some statistical computer packages will automatically compute the lack-of-fit F-test statistic within the simple regression procedure or perform the test if requested. Unfortunately, neither of these options is available in SPSS, so we must compute the F-statistic manually based on the analysis of variance tables from both the regression and ANOVA models.

The two tables displayed below are the analysis of variance tables for the insulating fluid data from a simple linear regression analysis (Section 6) and from a separate-means (one-way ANOVA) analysis (Section 8).

Analysis of Variance Table From a Simple Linear Regression

	DF	Sum of Squares	Mean Square
Regression	1	190.15149	190.15149
Residual	74	180.07484	2.43344

F = 78.14090 Signif F = .0000

Analysis of Variance from a One-Way ANOVA

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	6	196.4774	32.7462	13.0043	.0000
Within Groups	69	173.7489	2.5181		
Total	75	370.2263			

From the above display, $SS\,\mathrm{Re}\,s_{LR} = 180.07484$ and $SS\,\mathrm{Re}\,s_{SM}$ =173.7489, $d.f._{LR}$ =74, and $d.f._{SM}$ = 69. So the F-statistic is

$$F = \frac{[180.07484 - 173.7489]/[74 - 69]}{2.5181} = 0.502.$$

The proportion of values from an F-distribution with 5 and 69 degrees of freedom that exceed 0.502 is 0.78. This large p-value provides no evidence of lack-of-fit to the simple linear regression model.