A Lack-of-Fit Test for Simple Linear Regression

Suppose 8 field plots were used in an experiment to examine the relationship between amount of fertilizer and wheat yield. Four fertilizer amounts (5, 10, 15, and 20) were randomly assigned to the 8 plots so that each fertilizer amount was applied to two plots. The yield on each plot was recorded at harvest. The data are provided below.

	Plot								Summary Statistics		
	1	2	3	4	5	6	7	8	Mean	SD	Correlation
Amount of Fertilizer	5	5	10	10	15	15	20	20	12.5	5.97614	0.978
Yield	22	26	29	34	40	43	52	56	37.75	12.22118	

1. Name two potentially appropriate methods for analyzing this data.

2. Under what circumstances would one-way analysis of variance be more appropriate than simple linear regression for data like that considered here? Why?

3. Under what circumstances would simple linear regression be more appropriate than one-way analysis of variance for data like that considered here? Why?

4. The total sum of squares for this data set is SSTO=1045.5. If a one-way ANOVA model is fit to the data, the residual sum of squares is 33.0. If a simple linear regression model is fit to the data, the residual sum of squares is 45.5. We can conduct several tests of hypotheses using this information and the general formula

$$F = \frac{[\text{RSS(red.)} - \text{RSS(full)}]/[\text{df}_{\text{RSS(red.)}} - \text{df}_{\text{RSS(full)}}]}{\text{RSS(full)}/\text{df}_{\text{RSS(full)}}}$$

   (a) Is the mean yield the same for all levels of fertilizer, or does a model that allows each fertilizer to have its own mean yield appear to fit the data better?

   (b) Is the mean yield the same for all levels of fertilizer, or does a model that allows the mean yield to vary linearly with the amount of fertilizer appear to fit the data better?

(c) Do the mean yields vary linearly with the amount of fertilizer, or does a model that allows the mean yields to follow some other pattern appear to fit the data better?

We can organize all our calculations in a single ANOVA table. The difference between the residual sum of squares for the simple linear regression model and the residual sum of squares for the one-way ANOVA model is called the **sum of squares for lack of fit**. The words "lack of fit" refer to the fact that the simple linear regression model may not adequately fit the data. If the sum of squares for lack of fit is large, there is evidence that the simple linear regression model is not appropriate because the treatment means do not appear to have a straight-line relationship with the amount of the treatment factor. Fill in the missing entries in the ANOVA table.

Source	D.F.	SS	MS	F	P-Value
Between	-----------	-----------	-----------	-----------	-----------
Regression	-----------	-----------	-----------	-----------	-----------
Lack of Fit	-----------	-----------	-----------	-----------	-----------
Within	-----------	-----------	-----------		
Total	-----------	-----------			

5. How does the F-test for regression in this ANOVA table compare with the F-test computed on the front page? How does it compare with a contrast test for a linear trend?