for lack of fit, we would compute the test statistic F_0 and conclude that the regression function is not linear if $F_0 > F_{\alpha,m-p,n-m}$.

This test procedure may be easily incorporated into the analysis of variance. If we conclude that the regression function is not linear, then the tentative model must be abandoned and attempts made to find a more appropriate equation. Alternatively, if F_0 does not exceed $F_{\alpha,m-p,n-m}$, there is no strong evidence of lack of fit and MS_{PE} and MS_{LOF} are often combined to estimate σ^2. Example 6.6 is a very complete illustration of this procedure, where the replicate runs are center points in a 2^2 factorial design.

10.9 Problems

10.1. The tensile strength of a paper product is related to the amount of hardwood in the pulp. Ten samples are produced in the pilot plant, and the data obtained are shown in the following table.

Strength	Percent Hardwood	Strength	Percent Hardwood
160	10	181	20
171	15	188	25
175	15	193	25
182	20	195	28
184	20	200	30

(a) Fit a linear regression model relating strength to percent hardwood.

(b) Test the model in part (a) for significance of regression.

(c) Find a 95 percent confidence interval on the parameter β_1.

10.2. A plant distills liquid air to produce oxygen, nitrogen, and argon. The percentage of impurity in the oxygen is thought to be linearly related to the amount of impurities in the air as measured by the "pollution count" in parts per million (ppm). A sample of plant operating data is shown below:

Purity (%)		**93.3**	**92.0**	**92.4**	**91.7**	**94.0**	**94.6**	**93.6**
Pollution count (ppm)		1.10	1.45	1.36	1.59	1.08	0.75	1.20
Purity (%)	**93.1**	**93.2**	**92.9**	**92.2**	**91.3**	**90.1**	**91.6**	**91.9**
Pollution count (ppm)	0.99	0.83	1.22	1.47	1.81	2.03	1.75	1.68

(a) Fit a linear regression model to the data.

(b) Test for significance of regression.

(c) Find a 95 percent confidence interval on β_1.

10.3. Plot the residuals from Problem 10.1 and comment on model adequacy.

10.4. Plot the residuals from Problem 10.2 and comment on model adequacy.

10.5. Using the results of Problem 10.1, test the regression model for lack of fit.

10.6. A study was performed on wear of a bearing y and its relationship to x_1 = oil viscosity and x_2 = load. The following data were obtained:

y	x_1	x_2
193	1.6	851
230	15.5	816
172	22.0	1058
91	43.0	1201
113	33.0	1357
125	40.0	1115

(a) Fit a multiple linear regression model to the data.

(b) Test for significance of regression.

(c) Compute t statistics for each model parameter. What conclusions can you draw?

10.7. The brake horsepower developed by an automobile engine on a dynamometer is thought to be a function of the engine speed in revolutions per minute (rpm), the road octane number of the fuel, and the engine compression. An experiment is run in the laboratory and the data that follow are collected:

Brake Horsepower	rpm	Road Octane Number	Compression
225	2000	90	100
212	1800	94	95
229	2400	88	110
222	1900	91	96
219	1600	86	100
278	2500	96	110

246	3000	94	98
237	3200	90	100
233	2800	88	105
224	3400	86	97
223	1800	90	100
230	2500	89	104

(a) Fit a multiple regression model to these data.

(b) Test for significance of regression. What conclusions can you draw?

(c) Based on t-tests, do you need all three regressor variables in the model?

10.8. Analyze the residuals from the regression model in Problem 10.7. Comment on model adequacy.

10.9. The yield of a chemical process is related to the concentration of the reactant and the operating temperature. An experiment has been conducted with the following results.

Yield	Concentration	Temperature
81	1.00	150
89	1.00	180
83	2.00	150
91	2.00	180
79	1.00	150
87	1.00	180
84	2.00	150
90	2.00	180

(a) Suppose we wish to fit a main effects model to this data. Set up the $\mathbf{X}'\mathbf{X}$ matrix using the data exactly as it appears in the table.

(b) Is the matrix you obtained in part (a) diagonal? Discuss your response.

(c) Suppose we write our model in terms of the "usual" coded variables

$$x_1 = \frac{\text{Conc} - 1.5}{0.5} \quad x_2 = \frac{\text{Temp} - 165}{15}$$

Set up the $\mathbf{X}'\mathbf{X}$ matrix for the model in terms of these coded variables. Is this matrix diagonal? Discuss your response.

(d) Define a new set of coded variables

$$x_1 = \frac{\text{Conc} - 1.0}{1.0} \quad x_2 = \frac{\text{Temp} - 150}{30}$$

Set up the $\mathbf{X}'\mathbf{X}$ matrix for the model in terms of this set of coded variables. Is this matrix diagonal? Discuss your response.

(e) Summarize what you have learned from this problem about coding the variables.

10.10. Consider the 2^4 factorial experiment in Example 6.2. Suppose that the last observation is missing. Reanalyze the data and draw conclusions. How do these conclusions compare with those from the original example?

10.11. Consider the 2^4 factorial experiment in Example 6.2. Suppose that the last two observations are missing. Reanalyze the data and draw conclusions. How do these conclusions compare with those from the original example?

10.12. Given the following data, fit the second-order polynomial regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \epsilon$$

y	x_1	x_2
26	1.0	1.0
24	1.0	1.0
175	1.5	4.0
160	1.5	4.0
163	1.5	4.0
55	0.5	2.0
62	1.5	2.0
100	0.5	3.0
26	1.0	1.5
30	0.5	1.5
70	1.0	2.5
71	0.5	2.5

After you have fit the model, test for significance of regression.

10.13.

(a) Consider the quadratic regression model from Problem 10.12. Compute t statistics for each model parameter and comment on the conclusions that follow from these quantities.

(b) Use the extra sum of squares method to evaluate the value of the quadratic terms x_1^2, x_2^2, and x_1x_2 to the model.

10.14. ***Relationship between analysis of variance and regression.*** Any analysis of variance model can be expressed in terms of the general linear model $\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where the $\mathbf{X}$ matrix consists of 0s and 1s. Show that the single-factor model $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, $i = 1, 2, 3$, $j = 1, 2, 3, 4$ can be written in general linear model form. Then,

(a) Write the normal equations $(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ and compare them with the normal equations found for this model in Chapter 3.

(b) Find the rank of $\mathbf{X}'\mathbf{X}$. Can $(\mathbf{X}'\mathbf{X})^{-1}$ be obtained?

(c) Suppose the first normal equation is deleted and the restriction $\Sigma_{i=1}^{3} n\hat{\tau}_i = 0$ is added. Can the resulting system of equations be solved? If so, find the solution. Find the regression sum of squares $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$, and compare it to the treatment sum of squares in the single-factor model.