CHAPTER 10

Reserved Problems

10.1R 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.

10.2R Plot the residuals from Problem 10.1R and comment on model adequacy.

10.3R 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?

10.4R Consider the 2_{IV}^{4-1} design discussed in Example 10.5.

(a) Suppose that you elect to augment the design with the single run selected in that example. Find the variances and covariances of the regression coefficients in the model (ignoring blocks):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 + \beta_{34} x_3 x_4 + \epsilon$$

(b) Are there any other runs in the alternate fraction that would de-alias *AB* from *CD*?

(c) Suppose you augment the design with the four runs suggested in Example 10.5. Find the variances and covariances of the regression coefficients (ignoring blocks) for the model in part (a).

(d) Considering parts (a) and (c), which augmentation strategy would you prefer and why?

Minitab Output A:
Regression Analysis: y versus x1, x2, x3, x4, x5, x6, x7, x8, x9

```
The regression equation is
y = 14.9 + 1.92 x1 + 7.00 x2 + 0.149 x3 + 2.72 x4 + 2.01 x5 - 0.41 x6 - 1.40 x7
                    - 0.0371 x8 + 1.56  x9
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	14.928	5.913	2.52	0.024	
x1	1.925	1.030	1.87	0.083	7.0
x2	7.001	4.300	1.63	0.126	2.8
x3	0.1492	0.4904	0.30	0.765	2.5
x4	2.723	4.360	0.62	0.542	3.8
x5	2.007	1.374	1.46	0.166	1.8
x6	-0.410	2.379	-0.17	0.866	11.7
x7	-1.403	3.396	-0.41	0.686	9.7
x8	-0.03715	0.06672	-0.56	0.586	2.3
x9	1.559	1.937	0.80	0.434	1.9

```
S = 2.949                  R-Sq = 85.3%               R-Sq(adj) = 75.9%
PRESS = 393.492            R-Sq(pred) = 52.54%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	707.298	78.589	9.04	0.000
Residual Error	14	121.748	8.696		
Total	23	829.046			

Minitab Output B:
Regression Analysis: y versus x1, x2, x6

```
The regression equation is
y = 12.5 + 2.90 x1 + 6.54 x2 - 0.629 x6
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	12.489	4.543	2.75	0.012	
x1	2.9017	0.5567	5.21	0.000	2.2
x2	6.538	3.266	2.00	0.059	1.8
x6	-0.6293	0.8784	-0.72	0.482	1.7

```
S = 2.825               R-Sq = 80.7%               R-Sq(adj) = 77.9%
PRESS = 237.407         R-Sq(pred) = 71.36%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	669.43	223.14	27.96	0.000
Residual Error	20	159.62	7.98		
Total	23	829.05			

10.5R Consider Minitab output A. Provide an explanation of the t-test statistic for the regressor variable x1, using the computer output provided. Specifically, state the hypothesis that is tested, and give a general formula that shows how the statistic was calculated.

10.6R Consider Minitab output A. State the hypothesis that is tested by the F-statistic in the analysis of variance portion of the output. What are your conclusions relative to this hypothesis?

10.7R Consider Minitab output A. Note that not all regressors contribute to the model. Suppose that you wish to remove one regressor from this model. Which regressor would you remove? Why?

10.8R Consider Minitab output A. Show how the adjusted R^2 was computed.

10.9R Consider Minitab output B. Notice that this is the same data as in output A, but some regressors have been removed. Based on the summary statistics, are you happy with this model? Why?

10.10R Which model would you prefer, the one in Minitab output A or Minitab output B? Why?

10.11R Suppose that you wish to test the hypothesis that none of the regressors x3, x4, x5, x7, x8, x9 contribute significantly to the model. Show how this can be done with an F-test. Provide a numerical value for the F-statistic and draw conclusions.

10.12R Consider Minitab output B. Show how the R-Sq(pred) on the output was calculated.

10.13R Consider Minitab output B. Suppose that you wanted to use a partial F statistic to determine the contribution of x2 to the model. What numerical value of the F-statistic would you obtain?

10.14R Explain why the prediction interval on a future observation at a point of interest is always wider than the confidence interval on the mean response at the same point.

10.15R The variance of an interval estimate about the mean response is dependent on the sample size and distance of the points from the centroid.

(a) True

(b) False

10.16R In a hypothesis test for significance of regression, rejection of H_0: $\beta_1 = \beta_2 = \ldots\ \beta_{k=0}$ implies that a non-linear model must be used.

(a) True

(b) False

10.17R A regression data set has four candidate regressors and $n = 25$ observations. The total sum of squares is $SS_T = S_{yy} = 100$. When the full model with all 4 regressors was fit to the data, the residual mean square was 1.25. You have currently fit a subset regression model containing two regressors, and for this model the model sum of squares is $SS_{Model} = 65$. What is the value of C_p for the subset model?

10.18R Consider the situation described in problem 10.17R. Calculate the value of R^2 and the adjusted R^2 for the subset regression model.

10.19R Consider the situation described in problem 10.17R. An additional regressor is added to the model, resulting in a new model sum of squares of $SS_{Model} = 68$. What is the value of the residual mean square for this new regression model? Do you think that the addition of this new variable has improved the model?