

Problem 1

The data in Table B.5 (MPV package) present the performance of a chemical process as a function of several controllable process variables.

- Fit a multiple regression model relating CO₂ product (y) to total solvent (x_6) and hydrogen consumption (x_7).
- Test for significance of regression. Calculate R^2 and R_{Adj}^2 .
- Using t tests determine the contribution of x_6 and x_7 to the model.
- Construct 95% CIs on β_6 and β_7 .
- Refit the model using only x_6 as the regressor. Test for significance of regression and calculate R^2 and R_{Adj}^2 . Discuss your findings. Based on these statistics, are you satisfied with this model?
- Construct a 95% CI on β_6 using the model you fit in part e) Compare the length of this CI to the length of the CI in part d) Does this tell you anything important about the contribution of x_7 to the model?
- Compare the values of MS_{Res} obtained for the two models you have fit (parts a and e). How did the MS_{Res} change when you removed x_7 from the model? Does this tell you anything important about the contribution of x_7 to the model?

Problem 2

Cigarette Consumption Data: A national insurance organization wanted to study the consumption pattern of cigarettes in all 50 states and the District of Columbia. The dataset can be accessed from this link (https://stats.idre.ucla.edu/wp-content/uploads/2016/02/p081.sps_.txt)

In (a)-(b) below, specify the null and alternative hypotheses, the test used, and your conclusion using a 5% level of significance.

- Test the hypothesis that the variable Female is not needed in the regression equation relating Sales to the six predictor variables.
- Test the hypothesis that the variables Female and HS are not needed in the above regression equation.
- Compute the 95% confidence interval for the true regression coefficient of the variable Income.
- What percentage of the variation in Sales can be accounted for when Income is removed from the above regression equation? Explain.
- What percentage of the variation in Sales can be accounted for by the three variables: Price, Age, and Income? Explain.
- What percentage of the variation in Sales that can be accounted for by the variable Income, when Sales is regressed on only Income? Explain.

Problem 3

For testing significance of regression in multiple linear regression, theoretically it can be shown that the F_0 statistic value is given as,

$$F_0 = \frac{R^2(n-p)}{k(1-R^2)}$$

And we reject H_0 if the computed value of F_0 exceeds $F_{\alpha, k, n-p}$. Now, suppose that a linear regression model with two predictors has been fit to $n = 25$ observations and $R^2 = 0.90$, then

- a) Test for significance of regression at $\alpha = 0.05$. Comment.
- b) What is the smallest value of R^2 that would lead to the conclusion of a significant regression if $\alpha = 0.05$? Comment on the same.