USE YOUR KNOWLEDGE

11.1 Describing a multiple regression. Traditionally, demographic and high school academic variables have been used to predict college academic success. One study investigated the influence of emotional health on GPA.[2] Data from 242 students who had completed their first two semesters of college were obtained. The researchers were interested in describing how students' second-semester grade point averages are explained by gender, a standardized test score, perfectionism, self-esteem, fatigue, optimism, and depressive symptomatology.

(a) What is the response variable?

(b) What is n, the number of cases?

(c) What is p, the number of explanatory variables?

(d) What are the explanatory variables?

USE YOUR KNOWLEDGE

11.2 Understanding the fitted regression line. The fitted regression equation for a multiple regression is

$$\hat{y} = -1.8 + 6.1x_1 - 1.1x_2$$

(a) If $x_1 = 3$ and $x_2 = 1$, what is the predicted value of y?

(b) For the answer to part (a) to be valid, is it necessary that the values $x_1 = 3$ and $x_2 = 1$ correspond to a case in the data set? Explain why or why not.

(c) If you hold x_2 at a fixed value, what is the effect of an increase of two units in x_1 on the predicted value of y?

(d) Test the hypothesis that the coefficient of the variable $(PA - 8.614)^2$ is equal to 0. Report the t statistic, degrees of freedom, and P-value. Does the quadratic term contribute significantly to the fit? Explain your answer.

11.22 Architectural firm billings. A summary of firms engaged in commercial architecture in the Indianapolis, Indiana, area provides firm characteristics such as total annual billing in billions of dollars and the number of architects, engineers, and staff employed by the firm.[6] Consider developing a model to predict total billing. BILLING

(a) Using numerical and graphical summaries, describe the distributions of total billing and the number of architects, engineers, and staff.

(b) For each of the 6 pairs of variables, use graphical and numerical summaries to describe the relationship.

(c) Carry out a multiple regression. Report the fitted regression equation and the value of the regression standard error s.

(d) Analyze the residuals from the multiple regression. Are there any concerns?

(e) The firm HCO did not report its total billing but employs 3 architects, 1 engineer, and 17 staff members. What is the predicted total billing for this firm?

The following six exercises use the MOVIES data file. This data set contains an SRS of 35 movies released in the same year. This sample was collected from the Internet Movie Database (IMDb) to see if information available soon after a movie's theatrical release can successfully predict total revenue.[7] All dollar amounts are measured in millions of U.S. dollars. MOVIES

11.23 Predicting movie revenue—preliminary analysis. The response variable is a movie's total U.S. revenue (USRevenue). Let's consider as explanatory variables the movie's budget (Budget); opening-weekend revenue (Opening); the number of theaters (Theaters) the movie was in for the opening weekend; and the movie's IMDb rating (Opinion), which is on a 1 to 10 scale (10 being best). While this rating is updated continuously, we'll assume that the current rating is the rating at the end of the first week.

(a) Using numerical and graphical summaries, describe the distribution of each explanatory variable. Are there any unusual observations that should be monitored?

(b) Using numerical and graphical summaries, describe the relationship between each pair of explanatory variables.

11.24 Predicting movie revenue—simple linear regressions. Now let's look at the response variable and its relationship with each explanatory variable.

(a) Using numerical and graphical summaries, describe the distribution of the response variable, USRevenue.

(b) This variable is not Normally distributed. Does this violate one of the key model assumptions? Explain.

(c) Generate scatterplots of each explanatory variable and USRevenue. Do all these relationships look linear? Explain what you see.

11.25 Predicting movie revenue—multiple linear regression. Now consider fitting a model using all the explanatory variables.

(a) Write out the statistical model for this analysis, making sure to specify all assumptions.

(b) Run the multiple regression model and specify the fitted regression equation.

(c) Obtain the residuals from part (b) and check assumptions. Comment on any unusual residuals or patterns in the residuals.

(d) What percent of the variability in USRevenue is explained by this model?

11.26 A simpler model. In the multiple regression analysis using all four explanatory variables, Theaters and Budget appear to be the least helpful (given that the other two explanatory variables are in the model).

(a) Perform a new analysis using only the movie's opening-weekend revenue and IMDb rating. Give the estimated regression equation for this analysis.

(b) What percent of the variability in USRevenue is explained by this model?

(c) In this chapter we discussed the F test for a collection of regression coefficients. In most cases, this capability is provided by the software. When it is not, the test can be performed using the R^2-values from the full and reduced models. The test statistic is

$$F = \left(\frac{n - p - 1}{q}\right)\left(\frac{R_1^2 - R_2^2}{1 - R_1^2}\right)$$

with q and $n - p - 1$ degrees of freedom. R_1^2 is the value for the full model and R_2^2 is the value for the reduced model. Here $n = 35$ movies, $p = 4$ variables in the full model, and $q = 2$ variables were removed to form the reduced model. Plug in the values of R^2 from part (b) of this exercise and part (d) of the previous exercise, and compute the test statistic and P-value. Do Theaters and Budget combined add any significant predictive information beyond what is already contained in Opening and Opinion?

11.27 Predicting U.S. movie revenue. Refer to the previous two exercises. *Get Smart* was released in the same year, had a budget of $80.0 million dollars, was shown in