

When asked to explain something, or to provide an interpretation for a quantity, provide an explanation that could be understood by someone who does not have formal training in statistical methods. Keep your explanations brief.

1. The process of choosing how many and which model terms to put on the right-hand side of a regression equation is somewhat subjective. Make a brief comment (no more than two sentences for each) on the role and usefulness of each of the following statistics that are often outputs from regression analyses.
 - S (which has the same model-ordering as the adjusted R^2):

 - t -ratios

 - C_p

 - Residual plots and other diagnostics.

2. Consider the antique clock auction data discussed in class. In this example the highest bid price for the clocks (the response) was to be modeled as a function of the clock's age and the number of bidders (the X s). In a two-variable problem like this we might be concerned with both "correlation between the X s" and "interaction between the X s." They are *not* the same.
 - (a) Draw a simple graph or plot to illustrate the meaning of "correlation between the X s" in the above problem. Make sure to label your axes.

 - (b) Draw a simple graph or plot to illustrate the meaning of "interaction between the X s" in the above problem. Make sure to label your axes.

3. Data were collected over a period of 25 years to study the profitability of savings and loan companies. The response was percent profit margin (Y) and the two explanatory variables were percent revenues per dollar on deposit (X_1) and interest rate (X_2). Suppose that the observed correlation between X_1 and X_2 was .6, indicating some degree of multicollinearity. The regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

was to found to provide a good fit to the data. There was no evidence for the need for any interaction or squared terms.

- (a) Briefly explain how the multicollinearity will impact the ability to make inferences about the effect that percent revenues per dollar on deposit has on percent profit margin.
- (b) Briefly explain how the multicollinearity will impact the ability to make percent profit margin for given values of percent revenues per dollar on deposit and interest rate.
- (c) Note that the data in this experiment were time series data. What is an important model assumption that one might expect to have been violated in this example.
4. Consider the model $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$ where ϵ has a normal distribution with a mean of 0 and a standard deviation σ .
- (a) How many different levels of x are needed to fit the model (i.e., to estimate β_0 , β_1 , and β_2 ?)
- (b) How large of a sample is needed to estimate β_0 , β_1 , β_2 , and σ ?
- (c) How large does the sample size need to be to estimate only σ ? Briefly explain.

5. The manufacturer of a clothes washer wants to establish the wash-cycle time for a new model. An experiment was conducted using three different detergents and three different cycle times, with three replications in each cell. The order of the experimental trials was randomized. Clothes in each run were soiled in a kind of “mud soak” followed by drying. Then the clothes were weighed, washed, dried, and weighed again. The response in each experiment was the amount of weight change (presuming that the change was due to dirt removal) in pounds times 100. A summary of the data, giving a table of cell means, as well as a table of some regression output for the two different regression models is given below.

Detergent Brand	Washing Cycle Time (Minutes)		
	16	18	20
A	10.33	16.33	17.67
B	8.67	13.67	17.00
C	6.67	10.67	14.67

The following two regression models were fit to the data:

Model	Terms	SSE	SST
1	Brand and Cycle Time	15.704	379.407
2	Brand and Cycle Time with interaction	9.333	379.407

- (a) Plot the data in a manner that will allow you to assess whether there might be interaction between Brand and Cycle Time. What can you conclude from the plot?
- (b) What would be the *practical* interpretation if there were interaction in this example?
- (c) Write down the dummy variable regression equation for Model 1 and Model 2. Carefully define all dummy variables.

(d) Do a statistical test to see if there is strong evidence of interaction or not. Use $\alpha = .05$. State your conclusion.

6. In testing an hypothesis that the coefficient of a regression equation is equal to zero, the p -value for the test, given my Minitab output was .00012. How would you interpret this quantity?

7. Explain briefly how Cook's D statistics can be used to help better interpret the results of a regression analysis and modeling project.

8. List two important elements of Six-Sigma implementation.

-

-