Statistics 328 | Final Examination | Name ____________________
Fall 1999 | | **You must show all of your work**

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. Complete concise explanations are preferred.

1. Simply stated, the interpretation of a confidence interval (e.g., for a regression coefficient in a simple model with no interaction) can be expressed as "We are 95% confident that the interval contains the true rate of change." What is the real, technical, meaning of this statement?

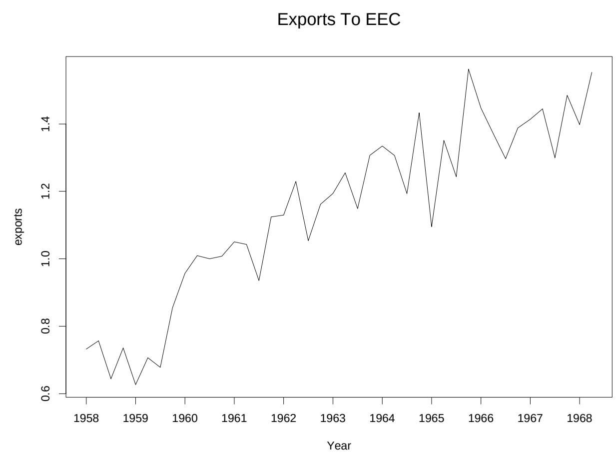
2. Briefly describe the role and importance of residual plots in regression modeling.

3. A study was conducted to investigate the relationship between sales on a given day for a department store and the amount spent on radio advertising on the previous evening. The advertising values were varied according to a pre-specified randomized pattern over a period of 50 days. A simple regression model was fit to these data. A careful examination of the residuals revealed that in most cases (about 66%) a positive residual was followed by another positive residual and that negative residuals were, in most cases (again about 66%) followed by another negative residual. Which of the standard regression assumptions was probably violated in this example? Briefly explain.

4. You have been given a computer file with relevant data and have been asked to find a model relating the selling price of houses (Y) to the assessed value of the land (X_1) and the assessed value of the improvements on the land (X_2). The correlation between X_1 and X_2 is .70. Compute the VIF for a regression model containing only X_1 and X_2. Is the correlation between X_1 and X_2 causing a serious amount of multicollinearity?

5. Briefly explain why it is that looking at the correlations between pairs of variables (as obtained from a "correlation matrix" or as depicted in a scatter-plot matrix) is not sufficient to identify all forms of multicollinearity.

6. The following figure shows quarterly data on the amount of exports from the US to the EEC from the first quarter of 1958 to the second quarter of 1968 (42 observations).



The following model was fit to the data

$$Y_t = \beta_0 + \beta_1 \text{Time} + \beta_2 \text{Q2} + \beta_3 \text{Q3} + \beta_4 \text{Q4} + \epsilon_t, \quad \epsilon_t \sim \text{nid}(0, \sigma^2)$$

where Y_t is exports in billions of dollars, Time is year $(1958.00, 1958.25, \ldots, 1968.50)$ and Q2 is 1 in quarter 2 and 0 otherwise, Q3 is 1 in quarter 3 and 0 otherwise, and Q4 is 1 in quarter 4 and 0 otherwise. The results of the regression fit using ordinary least squares showed $\widehat{\beta}_0 = -145, \widehat{\beta}_1 = .0749, S_{\widehat{\beta}_1} = .015, \widehat{\beta}_2 = -8.40, \widehat{\beta}_3 = -.44, \widehat{\beta}_4 = -23.24, S = .09$.

(a) Give a simple, clear, non-mathematical description of the fitted model.

(b) What is the *practical* interpretation of the parameter β_1 in this model?

(c) What is the *practical* interpretation of the parameter β_3 in this model?

(d) What is the *practical* interpretation of the parameter σ in this model?

(e) Based on the above model, give an expression for a simple approximate prediction interval for the amount of exports in the third quarter of 1968 (set it up, actual calculations not needed).

(f) Compute a 95% confidence interval for β_1.

(g) Write down a model, similar to the one above, that includes terms describing interaction between the time trend and the seasonal pattern. Explain clearly, but briefly, from a practical point of view, how this model differs from the one without the interaction terms. Draw a simple picture to aid your explanation.

7. Consider a statistical problem where there are two explanatory variables: X_1 and X_2. Is it possible for there to be substantial interaction between X_1 and X_2 if there is no correlation between X_1 and X_2? *Explain clearly*, using an appropriate formula or formulas and and/or an appropriate plot.

8. A study was conducted to build a model that can be used to predict Y, the number of paid visitors to a pool-water park in a resort district of New Hampshire. Let

$$X_1 = \begin{cases} 1 & \text{for a weekday} \\ 0 & \text{otherwise} \end{cases} \qquad X_2 = \begin{cases} 1 & \text{for a sunny day} \\ 0 & \text{otherwise} \end{cases} \qquad X_3 = \text{Degrees F}$$

The following three models were fit to the data from 30 consecutive days in the summer of 1992:

$$\begin{aligned} \text{Model 1} \quad Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \\ \text{Model 2} \quad Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_3 + \epsilon \\ \text{Model 3} \quad Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_5 X_1 X_2 + \epsilon \end{aligned}$$

giving the following results

Model	SS_{YY}	SSE
1	1762646	77212
2	1762646	76727
3	1762646	76292

(a) For Model 2, write down for model for the following special cases:

- Sunny weekday:
- Cloudy weekday:
- Sunny weekend day:
- Cloudy weekend day:

(b) In Model 2, what is the expected increase in paid visitors for an additional degree F in temperature?

(c) Test the null hypothesis that $\beta_4 = 0$ in Model 2, using $\alpha = .05$. In *practical* terms, what is your conclusion?

(d) Briefly explain practical difference between Model 1 and Model 2.

(e) For Model 2, Briefly explain whether you would expect β_1 and β_4 to be positive or negative and, for each, give the reason why.

(f) The Durbin-Watson statistic for Model 2 was 1.23. What does this suggest?

9. In your opinion, what is the most important need for successful implementation of a six-sigma program in a manufacturing company?