Statistics 328 Final Examination Name _____
Fall 1996 **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.

1. Briefly explain the difference between having two explanatory variables that interact and two explanatory variables that are correlated. Use pictures to help explain.

2. Outliers of of concern to analysts. If an outlier is caused by a data-recording error, the error should be fixed if possible. Otherwise it may be best to discard the observation. In other cases, outliers can teach us something.

   (a) Explain how one can determine whether an outlier is influential or not.

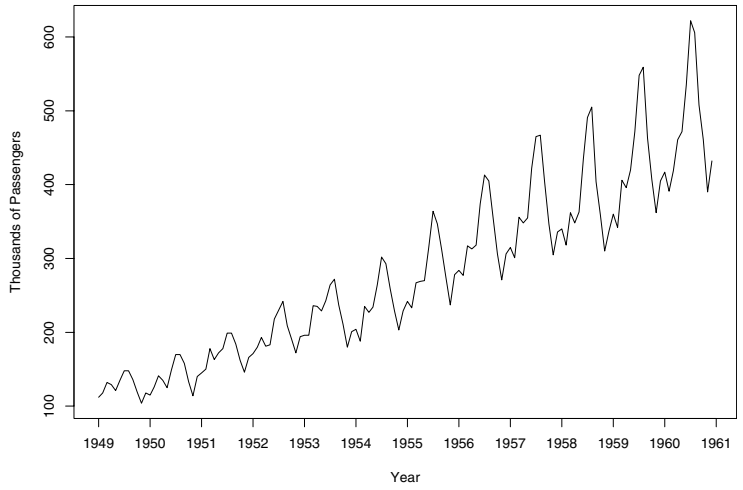
   (b) Give an example of when an outlier, known not to be an error, should still be omitted form the model fitting process.

   (c) Explain why an outlier might be extremely important to an analyst.

3. What is the most important model assumption for time series forecasting? (Hint: the answer is *not* independent observations.)

4. An analyst is working with a set of data having two explanatory variables. Fitting the model $Y = \beta_0+\beta_1 X_1+\beta_2 X_2+\epsilon$ results in an R^2 of .98 but both t-ratios less than 1. Fitting $Y = \beta_0+\beta_1 X_1+\epsilon$ or $Y = \beta_0 + \beta_1 X_1 + \epsilon$ results in t-ratios greater than 4. Explain the reason for this behavior.

5. The following figure is a plot of the number of international commercial airline passengers for the years 1949 to 1960. The data are monthly, 12 observations each year.



(a) *List* the things that you can see in this plot.

(b) Do you think that a transformation might be useful for fitting a model to these data? Explain why or why not.

(c) Write down a model that would provide a reasonable fit to these data.

6. Autocorrelation in time series data can lead to incorrect inferences or predictions if it is not recognized. Briefly explain the effect that positive autocorrelation would have on:

(a) Prediction of a future value.

(b) Testing an economic hypothesis that two variables are unrelated.

7. Assume that the response of construction job profitability is related to three explanatory variables: the engineer in charge ($x_1 = 0$ for Jones and $x_1 = 1$ for Smith), the state in which the job was completed ($x_2 = 0$ for Illinois and $x_2 = 1$ for Iowa), and job size (x_3). Each of the following three models has x_1, x_2, x_3 and x_3^2 included. The models differ according to the interactions included. Model 1 has interaction only between x_1 and x_2. Model 2 has interaction only between x_1 and x_3. Model 3 has interaction between and among x_1, x_2 and x_3. When appropriate, interaction terms include the quadratic term (i.e., if there is an interaction with x_3, then there should also be a corresponding interaction with x_3). Write down the three different models. For each of these models, draw a simple picture and use it to explain the structure of the model and how it differs from the other two models.

8. Refer to problem 7. There are 32 observations available for fitting the models.
For model 1, $\sum_{i=1}^{32}(y_i - \widehat{y}_i)^2 = 77212$.
For model 2, $\sum_{i=1}^{32}(y_i - \widehat{y}_i)^2 = 76729$.
For model 3, $\sum_{i=1}^{32}(y_i - \widehat{y}_i)^2 = 76292$.
For all of the models, $\sum_{i=1}^{32}(y_i - \bar{y})^2 = 1762646$.

   (a) Explain why $\sum_{i=1}^{32}(y_i - \bar{y})^2 = 1762646$ for all of the models.

   (b) Compute an ANOVA table for model 1.

   (c) Is there evidence in the data to support the claim that model 3 is more appropriate than model 2? Do an appropriate test using $\alpha = .05$.

9. Dummy variables can be used to fit a piecewise-linear function to data. The unit cost for a product has the following form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 (X_1 - 500) X_2 + \epsilon$$

where X_1 is lot size and $X_2 = 1$ if $X_1 > 500$ and $X_2 = 0$ if $X_1 < 500$.

(a) Write down separate models (make the form as simple as possible) for lot size greater than 500 and for lot size less than 500.

(b) Suppose that β_0=5.0, β_1=.004, and β_2=.003. Draw a picture of the overall model for lot sizes ranging between 0 and 1000 units (hint: substitute some appropriate numbers into the model equations).

(c) What is the interpretation of the parameters β_0, β_1, and β_2 in this problem?