Statistics 328 Examination 2 Name _____
Fall 1997 **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. A study was conducted to study the profitability of different kinds of construction jobs. The purpose of the study was to develop a model that could be used to predict profit. Data were gathered from the records of a sample of 26 previously completed jobs. The response variable was profit (in units of 1000's of dollars). The explanatory variables were Space (floor space in 100's of square feet) and Type (a dummy variable with 0 for residential construction and 1 for commercial construction). Three models were fit to the data.

   Model 1: $\text{Profit} = \beta_0 + \beta_1 \text{Space} + \beta_2 \text{Type}$
   Model 2: $\text{Profit} = \beta_0 + \beta_1 \text{Space} + \beta_2 \text{Type} + \beta_3 \text{Space} \times \text{Type}$
   Model 3: $\text{Profit} = \beta_0 + \beta_1 \text{Space} + \beta_2 \text{Type} + \beta_3 \text{Space} \times \text{Type} + \beta_4 \text{Space}^2 + \beta_5 \text{Type} \times \text{Space}^2$

   The following table gives the data and a summary of the results:

	Model 1	Model 2	Model 3	Space	Type	Profit
β_0	−48.5096	33.8654	−71.6500	100	0	129.
β_1	1.2706	0.8587	2.1057	100	0	92.
β_2	109.4615	−55.2885	64.9500	100	0	83.
β_3		0.8238	−0.0031	100	0	127.
β_4			−0.5972	200	0	209.
β_5			0.0036	200	0	266.
$SSE = \sum_{i=1}^{26}(y_i - \widehat{y}_i)^2$	47377.9	20235.4	17186.8	200	0	230.
				200	0	205.
$SS_{yy} = \sum_{i=1}^{26}(y_i - \bar{y}_i)^2$	383577.9	383577.9	383577.9	200	0	214.
				300	0	306.
				300	0	262.
				300	0	238.
				300	0	312.
				100	1	134.
				100	1	187.
				100	1	124.
				100	1	149.
				200	1	314.
				200	1	303.
				200	1	323.
				200	1	323.
				200	1	299.
				300	1	484.
				300	1	479.
				300	1	432.
				300	1	545.

(a) Use the attached graph paper to plot the data in a meaningful way. You should need to make only one plot.

(b) List some of the things that you have learned from your plot:

-
-
-

(c) Briefly explain why $\sum_{i=1}^{26}(y_i - \bar{y})^2$ is the same for all three models.

(d) In comparing Models 2 and 3, the terms $\beta_4 \text{Space}^2 + \beta_5 \text{Type} \times \text{Space}^2$ were added to allow for possible curvature in the relationship between Profit and Space. Do the data provide evidence for the need of these curvature terms? Do an appropriate test (use $\alpha = .05$) and briefly explain your conclusion.

(e) Compute R^2 from Model 2. Briefly explain the *practical* interpretation of this quantity.

(f) Model 2 has an interaction term and Model 1 does not. Briefly explain, in terms of this particular example, the practical difference between Model 1 and Model 2.

(g) Use an F-test to compare Model 1 and Model 2. Use $\alpha = .05$. What do you conclude?

(h) Compute s for each of the three models. What do you conclude?

(i) Briefly explain the interpretation of s for this example.

(j) In Problem 1, the values of $\widehat{\beta}_2$ changed dramatically as the model changed. Briefly explain why.

(k) Using the information available, which of the three models would you use to make predictions. Briefly explain your choice.

(l) Provide a description (use an equation or equations or numbers as necessary) of the effect that change in floor space has on the expected profit of a job.

(m) You do not have values of $s_{\widehat{y}}$ for this model and they are too difficult to compute by hand. However, it is still possible to compute an approximate 95% prediction interval for the profit of a future job. Do this for a residential construction job with Space = 400. Comment.

(n) One must use caution when predicting outside the range of one's data. Briefly explain, specifically, what this means in the context of this problem.

2. Write down the full second order model relating two explanatory variables x_1 and x_2 to a response Y. Briefly describe what kind of shape this surface could have.

3. Consider a model to describe the effect that a qualitative explanatory variable has on a response variable. Suppose that the qualitative variable has 3 levels A, B, and C.

   (a) Write the model relating the response to the qualitative variable.

   (b) Brifly explain the interpretation of the parameters in the model.

4. The output of a computer program for a regression analysis prints out p-values corresponding to all regression coefficients. In one case, $p = .0011$ was reported for the p-value corresponding to β_1. Briefly explain the interpretation of this quantity.

5. The statements "There is an interaction between x_1 and x_2" and "The explanatory variables x_1 and x_2 are correlated" mean two different things. It is possible for x_1 and x_2 to have interaction but no correlation. Also, it is possible the x_1 and x_2 to be highly correlated, but there be no interaction between x_1 and x_2.

   (a) Briefly explain the meaning of interaction between x_1 and x_2 and briefly explain how or give an example of a model and experimental design such that there is interaction but no correlation between x_1 and x_2.

   (b) Briefly explain the meaning of correlation between x_1 and x_2 and briefly explain how or give an example of a situation where x_1 and x_2 are correlated but where there is no interaction between x_1 and x_2.