Statistics 328 Examination 1 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. Keep your explanations brief.

1. The marketing manager for a department store has obtained data on a sample of customers who have a credit account with the store. The data set contains the amount of purchases for each customer during the past year (in dollars) as well as the customer's yearly income (in thousands of dollars). The marketing manager wants to determine whether income can be used to predict the amount of money that a customer will spend in the department store over given period of time. The data are:

Customer	1	2	3	4	5	6	7	8	9	10	11	12
Purchases	370	680	640	820	550	740	340	500	720	760	780	1040
Income	17.0	35.7	23.0	36.5	21.3	55.3	15.8	18.5	48.8	42.5	36.0	89.0

   (a) Is the marketing manager concerned with a population or a process? Explain.

   (b) Use the attached graph paper to plot the data in an appropriate manner to begin to address the marketing manager's question.

   (c) Draw a line of best fit. Use this line to obtain a graphical estimate of the slope. What do you conclude?

   (d) Explain the interpretation of the slope for *this problem*.

2. Confidence intervals are an important tool for quantifying uncertainty in an estimate of a population or a process parameter. Why might one prefer a 95% confidence interval over a 99.99% confidence interval?

3. Statistical studies can, in some cases, produce imprecise results. Explain why. Provide a general description of how one can evaluate the precision of a proposed statistical procedure.

4. Under the usual assumptions of regression analysis, a $100(1-\alpha)\%$ prediction interval for a future value of Y for a particular value of x can be computed as

$$\widehat{y} \quad \pm \quad t_{(\alpha/2,n-2)} S_{(y-\widehat{y})}$$

but, under certain circumstances, an approximate prediction interval can be computed from

$$\widehat{y} \quad \pm \quad 2 \times S$$

Explain why and when this approximation works and the indicate the approximate level of confidence associated with the interval. Note that $S_{(y-\widehat{y})}$ depends on the particular value of x but that S does not.

5. In the quadratic regression model, the mean of Y is

$$\mathrm{E}(Y) = \beta_0 + \beta_1 x + \beta_2 x^2.$$

This is the equation of a parabola. If $\beta_2 = 0$, the equation is a straight line with a rate of change β_1. Otherwise, the rate of change of $\mathrm{E}(Y)$ with respect to x depends on level of x. The rate of change (also known as the derivative or tangent slope) of this equation is $\beta_1 + 2\beta_2 x$. What is the interpretation of the *sign* of the coefficient β_2?

6. A regression model was fit to quantify the relationship between Y, the monthly rental cost (in of dollars) and x, the number of square feet of floor space from a sample of apartments that were being offered for rent. The square footage of the apartments in the sample ranged from 700 to 1000. The data, when plotted seemed to suggest a linear relationship. Below is an edited summary of some output from MINITAB.

```
The regression equation is
Rent = 184 + 0.314 Footage

Predictor         Coef        StDev           T         P
Constant        183.70        51.12        ????     0.002
Footage        0.31364      0.05823        5.39     0.000


Analysis of Variance

Source              DF           SS          MS          F         P
Regression           1       zzzzzz      xxxxxx      29.01     0.000
Residual Error      18       5575.1       309.7
Total               19      14561.7

Predicted Values

MyFootage      Fit     StDev Fit          95.0% CI                  95.0% PI
   600       371.89       16.51     ( 337.20,  406.57)  (  321.19,  422.58) XX
   900       465.98        4.19     ( 457.18,  474.78)  (  427.97,  503.99)
  1200       560.07       19.31     ( 519.51,  600.64)  (  wwwwww,  614.96) XX
X  denotes a row with X values away from the center
XX denotes a row with very extreme X values
```

(a) Compute the quantity marked ???? in the above output.

(b) Compute the quantity marked zzzzzz in the above output.

(c) Compute the quantity marked wwwwww in the above output.

(d) How many apartments were in the sample? Show how you computed this number.

(e) Does the estimate $\widehat{\beta}_0$ have a practical interpretation in the problem? Explain why or why not.

(f) Explain the practical interpretation of the estimated slope coefficient $\widehat{\beta}_1$ for this example. Be careful to use the units on the variables as you carefully word your answer.

(g) A claim has been made that apartment price is not related to floor size. As foolish as this might sound, you have been asked to determine if the data provides enough evidence to refute this claim. Show how to do this.

(h) One of the assumptions of the simple linear regression model is that σ, the standard deviation of the residuals does not depend on the explanatory variable. For this example, use the information in the Analysis of Variance table to compute an estimate of this standard deviation.

7. When estimating β_1, the slope of a line in a simple regression model, under the standard assumptions, the standard deviation of the sampling distribution of $\widehat{\beta}_1$ is

$$\sigma_{\widehat{\beta}_1} = \sigma\sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

and, based on sample data, an estimate of this quantity is

$$S_{\widehat{\beta}_1} = S\sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

where

$$S = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n-2}}$$

(a) Explain the concept of the "sampling distribution of $\widehat{\beta}_1$."

(b) Explain the interpretation of the quantity $\sum_{i=1}^{n}(x_i - \bar{x})^2$ (also known as SS_{xx} in the text).

(c) What does the formula for $\sigma_{\widehat{\beta}_1}$ (or $S_{\widehat{\beta}_1}$) suggest for a general guideline for planning an experiment in which simple regression will be used as the method of analysis?

8. Explain the difference in interpretation between S and $S_{\widehat{y}}$ in a simple regression problem.