REVIEW AP Statistics Name:

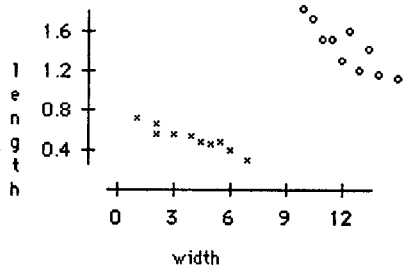
Directions: *Do all of your work on these sheets.*

Part 1: Multiple Choice. *Circle the letter corresponding to the best answer.*

1. There is a positive association between the number of drownings and ice cream sales. This is an example of an association likely caused by:
   (a) Coincidence
   (b) Cause and effect relationship
   (b) Confounding factor
   (d) Common response
   (e) None of the above

2. If the correlation between body weight and annual income were high and positive, we could conclude that:
   (a) High incomes cause people to eat more food.
   (b) Low incomes cause people to eat less food.
   (c) High-income people tend to spend a greater proportion of their income on food than low-income people, on average.
   (d) High-income people tend to be heavier than low income people, on average.
   (e) High incomes cause people to gain weight.

3. A study examined the relationship between the sepal length and sepal width for two varieties of an exotic tropical plant. Varieties A and B are represented by x's and o's, respectively, in the following plot:



Which of the following statements is FALSE?
   (a) Considering variety A alone, there is a negative correlation between sepal length and sepal width.
   (b) Considering variety B alone, the least squares regression line for predicting sepal length from sepal width has a negative slope.
   (c) Considering both varieties together, there is a positive correlation between sepal length and sepal width.
   (d) Considering each variety separately, there is a positive correlation between sepal length and sepal width.
   (e) Considering both varieties together, the least squares regression line for predicting sepal length from sepal width has a positive slope.

4. From tax records, it is relative easy to determine the amount of liquor consumed per capita and the number of cigarettes consumed per capita for each of the 10 provinces of Canada. These are plotted on a scatterplot and a high positive correlation is found. Which of the following is correct?
   (a) This implies that heavy smoking causes people to drink more.
   (b) This implies that heavy drinking causes people to smoke more.
   (b) We cannot conclude cause and effect, but this also implies that there is a high positive correlation between cigarette smoking and alcohol consumption for individuals.
   (d) This could be an example of a correlation caused by a common cause because both activities are highly correlated with average family income and average income varies widely among the provinces.
   (e) We cannot conclude cause and effect, but this also implies that the same individuals both smoke and consume liquor.

Part 2: Free Response

Answer completely, but be concise. Write sequentially and show all steps.

5. Suppose that two-variable data has been plotted and that the points show a clearly curved pattern. In this situation, several methods can be used to transform the data. In each of the following, data have been transformed to obtain a good model. In each case, what would be the equation that best fits the untransformed data?

   (a) $\sqrt{d} = 22.19\,t + 0.12$

   (b) $\ln c = 0.105\,d + 0.01$

   (c) $\log y = -7.43 + 2.49 \log x$

6. A business school conducted a survey of companies in its state. They mailed a questionnaire to 200 small companies, 200 medium-sized companies, and 200 large companies. The rate of nonresponse is important in deciding how reliable survey results are. Here are the data on response to this survey:

	Small	Medium	Large
Response	125	81	40
No Response	75	119	160
Total	200	200	200

   (a) What was the overall response rate?

   (b) Describe how nonresponse is related to the size of the business. (Use percents to make your statements precise.)

   (c) Draw a bar graph to compare the nonresponse percents for the three size companies.

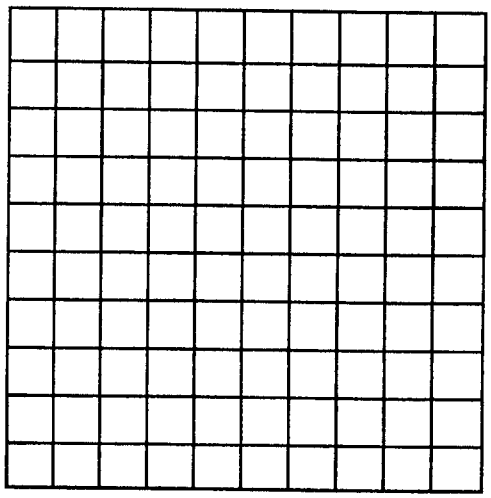
7. According to data from the U.S. Health Care Financing Administration, the national expenditures for drugs and other medical nondurables (in billions of dollars) for selected years from 1970 to 1997 are as follows: (Note that Year is coded: 1970 is recorded simply as 70.)

Year	70	80	85	87	89	90	91	92	93	94	95	97
Spent	8.8	21.6	37.1	43.2	50.6	59.9	65.6	71.2	75	77.7	83.4	108.9

(a) Apply a test to show that the national expenditures for drugs and other medical nondurables are increasing exponentially.

(b) Calculate the logarithms of the y-values and extend the table above to show the transformed data.

(c) Plot the transformed data on the grid provided. Label the axes completely.



(d) You want to construct a model to predict the national drug expenditures in the near future. Perform linear regression on the transformed data and write your least squares equation.

(e) Now transform your linear equation back to obtain a model for the national drug expenditures data. It should be in the form $y = (\text{constant}) \cdot (10^{bx})$ Write the equation for this model.

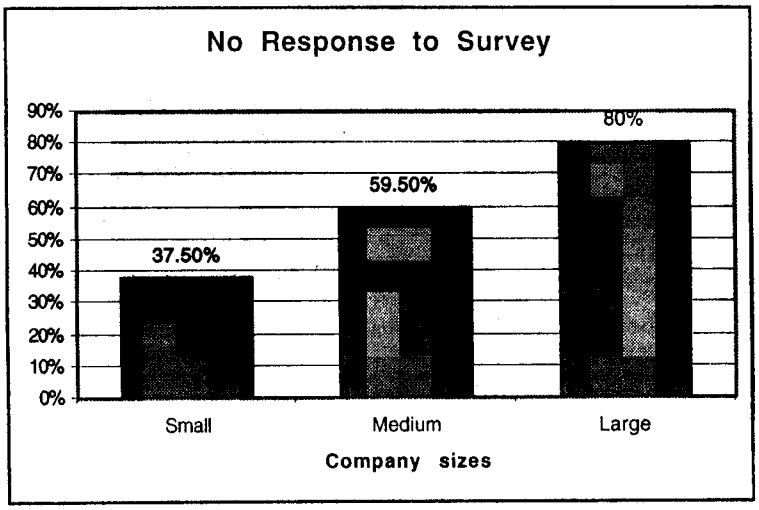
(f) Predict the national drug expenditure for this year. Do you have confidence in this result? Why or why not?

8. Foresters are interested in predicting the amount of usable lumber they can harvest from various tree species. The following data have been collected on the diameter of Ponderosa pine trees, measured at chest height, and the yield in board feet. Note that a board foot is defined as a piece of lumber 12 inches by 12 inches by 1 inch. Construct an appropriate model for these data. Then comment on the quality of your model.

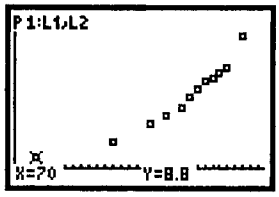
Diameter	Bd Feet
36	192
28	113
28	88
41	294
19	28
32	123
22	51
38	252
25	56
17	16
31	141
20	32
25	86
19	21
39	231
33	187
17	22
37	205
23	57
39	265

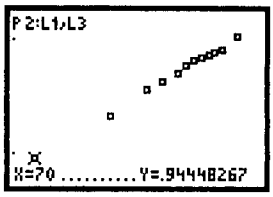
Ch 4 Review

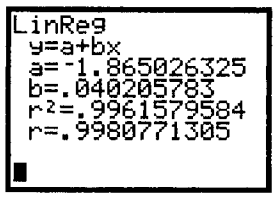
(1) d. (2) d. (3) d. (4) e. (5a) $d = 492.3961\,t^2 + 5.3256\,t + 0.0144$. (5b) $c = 1.01\,e^{0.105d}$. (5c) $y = (3.71535\times10\text{–}8)\,x^{2.49}$. (6a) (75 + 119 + 160)/600 = 59% did not respond. (6b) 75/200 = 37.5% of small businesses, 119/200 = 59.5% of medium sized businesses, and 160/200 = 80% of large businesses did not respond. Generally the large the business, the less likely they are to respond.

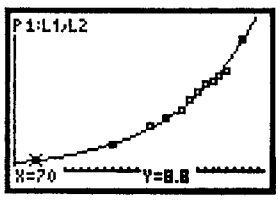


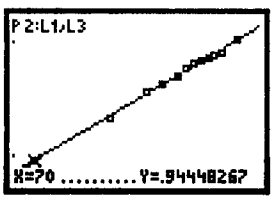
(7a) Looking only at the *odd* years, the ratios of each odd year y-value to the previous odd year y-value are, starting with 1987: 1.2, 1.2, 1.3, 1.2, 1.1, 1.3 (for 1997). (7b) The first 4 logarithms (base 10) are: 0.9445, 1.3345, 1.5694, and 1.6355. (7c) The scatterplot of National expenditures for drugs and other medical nondurables by year is screen shot #1 below. The points are straightened by plotting log(Expenditures) vs. Year (picture #2). Regression is performed on the transformed data; the correlation is 0.998, and the equation of the least squares line is *log(Expenditures)* = –1.8662 + 0.0402 (*Year*). (See pictures #3 and #4.) Finally, we back-transform to obtain the exponential curve *Expenditures* = (10^–1.8662)(10^(0.0402 *Year*)). See picture 5.











(8) The scatterplot (picture 1 below) shows a clearly curved pattern. To determine the model to use, we note that diameter is one-dimensional, and board feet is 3-dimensional. Board feet should be proportional to the cube of the diameter. We hypothesize a power function of the form $y = a\,x^b$ and plot log(BoardFeet) vs. log(Diameter). The plot of the transformed data appears linear (picture 2), so we perform least squares regression on the transformed data (picture 3). The fitted line appears in picture 4. The residual plot (picture 5) shows no pattern, so we judge the line to be an acceptable model for the transformed data. Note that the correlation is 0.988 ($r^2 = 97.6\%$) which indicates a very strong association. Back-transforming, we obtain the power function model, BoardFeet = (10^-2.5691) (x^3.13667). Note that the power of x is very close to 3, which affirms our hypothesis.