Chapter 7-10 Practice Test Name________________________

Part I: Multiple Choice (Questions 1-10) - Circle the answer of your choice.

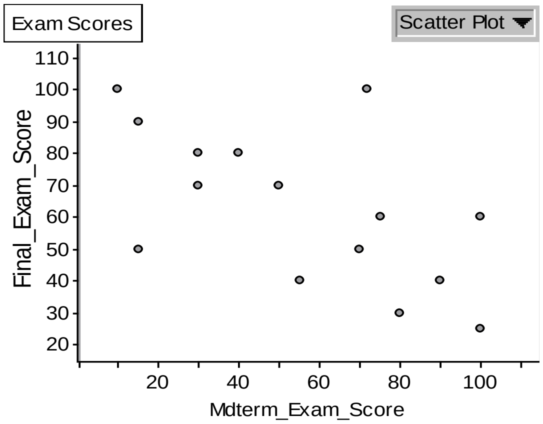
1. Foresters use regression to predict the volume of timber in a tree using easily measured quantities such as diameter. Let y be the volume of timber in cubic feet and x be the diameter in feet (measured at 3 feet above ground level). One set of data gives $\hat{y} = -30 + 60x$. The predicted volume for a tree of 18 inches is:

(A) 1050 cubic feet
(B) 600 cubic feet
(C) 105 cubic feet
(D) 90 cubic feet
(E) 60 cubic feet

E; 18 inches = 1.5 feet
$\hat{y} = -30 + 60(1.5) = 60$

2. Consider the following scatterplot of midterm and final exam scores for a class of 15 students. Which of the following are true statements?

I. The same number of students scored 100 on the midterm exam as scored 100 on the final exam.
II. Students who scored higher on the midterm exam tended to score higher on the final exam.
III. The scatterplot shows a moderate negative correlation between midterm and final exam scores.



(A) I and II
(B) I and III
(C) II and III
(D) I, II, and III
(E) None of the above gives the complete set of complete true responses.

B; I is true; 2 students scored a 100 on the midterm and 2 scored a 100 on the final. II is false; there is a negative relationship between the two exams. III is true.

3. The relationship between population (*y*) and year (*x*) was determined to be exponential. The least-squares regression equation of the appropriately transformed data was $\hat{y} = 0.05 + 0.004x$. What would be the predicted population in the year 1990?

(A) 8.01
(B) 288,403,150
(C) 3.21
(D) 102,329,299
(E) There is insufficient information to make a prediction.

D; $\hat{y} = 0.05 + 0.004(1990) = 8.01$. $10^{8.01} = 102329299.2$ *(It is possible that we should have used e; that choice does not lead to a given answer).*

4. Data are obtained for a group of college freshman examining their SAT scores (math plus verbal) from their senior year of high school and their GPAs during their first year of college. The resulting regression equation is $\hat{y} = 1.35 + 0.00161x$ with $S_x = 120$, and $S_y = 0.3057$. What percentage of the variation in GPAs can be explained by looking at SAT scores?

(A) 0.161%
(B) 16.1%
(C) 39.9%
(D) 63.2%
(E) This value cannot be computed from the information given.

C;

$$b_1 = r\left(\frac{s_y}{s_x}\right)$$

$$.00161 = r\left(\frac{.3053}{120}\right)$$

$$r = .632$$

Since the problem is asking for R^2*, square .632.*

5. Given the least-squares regression line,

$Cost\ of\ \ Monopoly\ Property = 67.3 + 6.78\left(Spaces\ \ from\ Go\right)$,

determine the residual for Reading Railroad which costs \$200 and is 5 spaces from Go.

(A) –98.8
(B) –9.88
(C) 98.8
(D) –1418.3
(E) A residual has no meaning since one of the variables is categorical.

C;

$\hat{y} = 67.3 + 6.78\left(5\right) = 101.2$

$residual = y - \hat{y}$

$residual = 200 - 101.2$

6. Suppose that the scatterplot of (log *x*, log *y*) shows a strong positive correlation close to $r = 1$. Which of the following are true?

I. The variables *x* and *y* also have a correlation close to 1.
II. A scatterplot of (*x, y*) shows a nonlinear pattern.
III. The residual plot of the variables *x* and *y* shows a random pattern.

(A) I only
(B) II only
(C) III only
(D) I and II
(E) I, II, and III

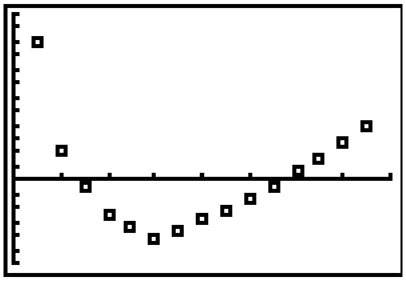
B; I is false because while the correlation between the raw data could be strong, it might not be. III is false because the reason you attempt a transformation is because the linear model is not appropriate and a random pattern would indicate otherwise.

7. Consider the three points (2,11), (3,17), and (4,29). Given any straight line, we can calculate the sum of the squares of the three vertical distances from these points to the line. What is the smallest possible value this sum can be?

(A) 6
(B) 9
(C) 29
(D) 57
(E) The sum cannot be determined from only three data points.

A; find the residual list. Square these values and sum them.

8. A study of the fuel economy for various automobiles plotted the fuel consumption (in liters of gasoline used per 100 kilometers traveled) vs. speed (in kilometers per hour). A least-squares regression line was fitted to the data and the residual plot is displayed to the right. What does the pattern of the residuals tell you about the linear model?



(A) The evidence is inconclusive.
(B) The residual plot confirms the linearity of the data.
(C) The residual plot suggests a different line would be more appropriate.
(D) The residual plot clearly contradicts the linearity of the data.
(E) None of the above.

D; the answer cannot be C because the linear is not appropriate based on the pattern in the residuals.

9. With regard to regression, which of the following statements about outliers are true?

I. Outliers have large residuals.
II. Successful prediction requires a cause and effect relationship.
III. Removal of an outlier sharply affects the regression line.

(A) I and II
(B) I and III
(C) II and III
(D) I, II, and III
(E) None of the above gives the complete set of true responses.

E; only I is true. II is false because otherwise this unit is meaningless! III is false because it should be removal of an influential point sharply affects the regression line.

10. As reported in the *Journal of the American Medical Association* (June 13, 1990), for a study of ten nonagenarians, the following tabulation shows a measure of strength versus a measure of functional mobility. What does the slope of the least-squares regression line signify?

Strength (kg)	7.5	6	11.5	10.5	9.5	18	4	12	9	3
Walk time (s)	18	46	8	25	25	7	22	12	10	48

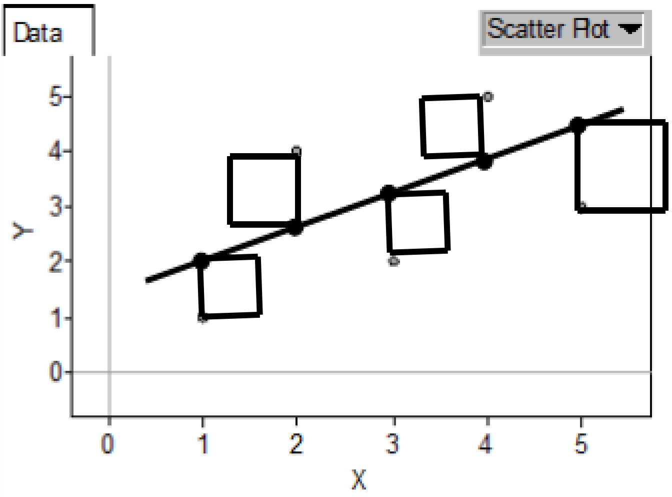
(A) The sign is positive, signifying a direct cause-and-effect relationship between strength and mobility.
(B) The sign is positive, signifying that the greater the strength, the greater the functional mobility.
(C) The sign is negative, signifying that the relationship between strength and functional mobility is weak.
(D) The sign is negative, signifying that the greater the strength, the less the functional mobility.
(E) The slope is close to zero, signifying that the relationship between strength and functional mobility is weak.

D; calculate the regression on the calculator. The slope is negative indicating that as the explanatory variable increases, the response variable decreases. Thinking about it logically, the stronger a person is, the faster their walk time.

Part II: Free Response (Questions 11-13) – Show your work and explain your results clearly.

11. Consider the scatterplot of *y* vs. *x*.

(a) Find the equation of the least-squares regression line of the data represented in the scatterplot at the right.



Using a calculator, $\hat{y} = 1.5 + .5x$.

(b) Explain (and illustrate on the scatterplot) the meaning of the term "least-squares".

The least squares is produced by minimizing the sum of the squared residuals, i.e., squared vertical distances from the line.

(c) Calculate and interpret r .

r = .5; the relationship between y and x is moderate and positive.

(d) Calculate and interpret r^2.

25% of the variation in y can be accounted for by the model using x as the explanatory variable.

(e) Suppose the point $(10,1)$ was added to the data set. Find the equation of the least-squares regression line with this point added to the data set. Would this point be considered an outlier or an influential point? Justify your answer.

$\hat{y} = 3.213 - .131x$; *the point is influential because the slope of the LSRL changes dramatically (from positive to negative)*

12. An analysis of the relationship between the number of telephones in a household (*x*) and the annual family income (*y*) revealed the following statistics:

$\bar{x} = 3.8$	$\bar{y} = 65{,}000$	$n = 26$
$S_x = 1.2$	$S_y = 15{,}500$	$r = 0.65$

(a) Determine the least-squares regression line.

$$\hat{y} = b_0 + b_1 x$$
$$b_1 = r\left(\frac{s_y}{s_x}\right) = .65\left(\frac{15500}{1.2}\right) = 8395.833$$
$$b_0 = \bar{y} - b_1\bar{x} = 65000 - 8395.833(3.8) = 33095.834$$
$$\hat{y} = 33095.834 + 8395.833x$$

(b) Interpret the slope of the least-squares regression line.

The model predicts that for each additional telephone, the family income increases by about $8395.83.

(c) Interpret the vertical intercept of the least-squares regression line.

The model predicts a family with no telephones will have an annual family income of about $33,095.83.

(d) Determine the numerical amount that the data point $(5, 80000)$ contributes to the correlation.

$$r = \frac{1}{n-1}\sum z_x z_y$$
$$r = \frac{1}{26-1}\left(\frac{5-3.8}{1.2}\right)\left(\frac{80000-65000}{15500}\right) = .039$$

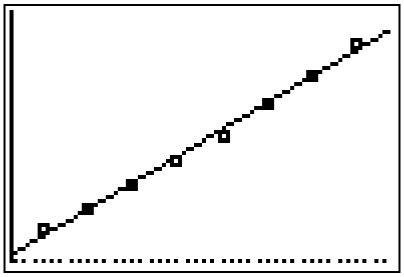
Note: all of the formulas on this page appear on the formula sheet.

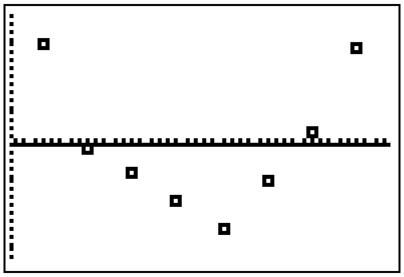
13. The following data represent the Woodward Academy Upper School enrollment over 35 years.

Year	1965	1970	1975	1980	1985	1990	1995	2000
Enrollment	650	690	740	790	840	900	960	1025

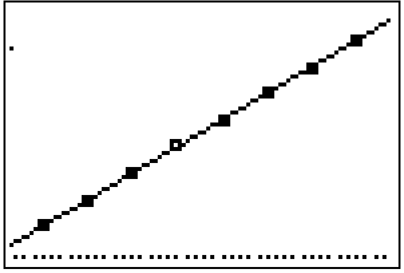
(a) Determine an appropriate model for the data. Justify your answer. Let 1965 = year 65.

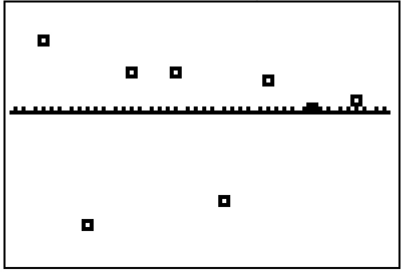
The linear model is not appropriate due to the non-linear nature of the scatterplot and the pattern in the residuals.





To attempt to "straighten" the graph up, we log the enrollment list. The resulting scatterplot and residual plot are below. Note the residuals are reasonably scattered.





The model is $Log\ Enrollment = 2.443 + .006(Year)$.

(b) Use your model to predict the enrollment in 2010. Comment on your result.

$Log\ Enrollment = 2.443 + .006(110) = 3.103$.

The enrollment is $10^{3.103} = 1267.7$.

The model predicts the enrollment will be about 1268 students. Since 2010 is outside the range of the years, this is extrapolation and this prediction should be used with caution.