Study Session Week of 10/30

Objectives:

- I will apply previous knowledge to solve miscellaneous problems involving concepts connected with linear regression
- I will review information about linear regression including R^2, correlation, residuals, slope and y-intercept.

Agenda:

- MC practice problems

You need a calculator, clicker, AND formula sheets

1

For the years 1950–1980, the number of heart disease deaths per 100,000 people in the United States were recorded. The regression line below was computed using a statistical software package. Which statement is the correct interpretation of the slope?

```
The regression equation is
Number of deaths=7387−3.63 year
```

(A) The number of heart disease deaths per 100,000 people has been dropping by an estimated 3.63 deaths per year on the average.
(B) There is an increase of approximately 7387 deaths per year.
(C) For every 3.63 years there is a decrease on the average of 1 death due to heart disease per 100,000 people.
(D) The regression line estimates that for every 3.63 years there is an average increase of 1 death due to heart disease per 100,000 people.
(E) Heart disease will be cured in the year 2036.

2

Which of the following statements about a linear regression model is true?

I. The sum of the residuals is always zero.
II. If $r^2 = 0$, the regression line is a horizontal line.
III. No pattern in the residual plot is an indication that a nonlinear model will show a better fit to the data than a straight-line regression model.

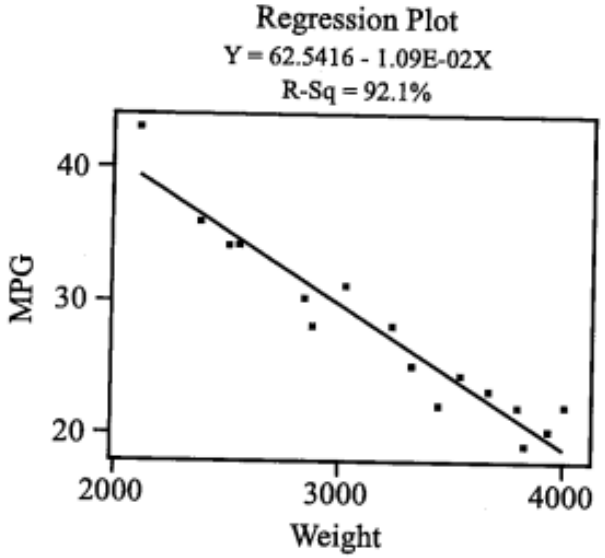
(A) I only
(B) I and II only
(C) I and III only
(D) II and III only
(E) I, II, and III

3

In a statistics course, a linear regression equation was computed to predict the final exam score based on the score on the first test of the term. The equation was: $\hat{y} = 25 + 0.7x$ where y is the final exam score and x is the score on the first test. George scored 80 on the first test. On the final exam George scored 85. What is the value of his residual?

(A) -4
(B) 4
(C) 4.5
(D) 5
(E) 81

4 A student was interested in the relationship between weight of a car and gas consumption measured in mpg. He selected sixteen different automobiles and recorded their weights along with their advertised mpg. The regression equation and regression plot are shown below.



What affect would the addition of the point (4,300 lbs., 15.63mpg) have on the value of r^2?

A) It will have no effect on r^2 because it lies on the line.
B) It will have no effect on r^2 because its value is the same as $(\overline{x}, \overline{y})$.
C) It will decrease r^2 because it is an outlier.
D) It will increase r^2 because it is an influential point which lies on the least squares line.
E) It will increase r^2 because every additional point will increase the percent variation in y due to the relationship with the least squares line.

5

The equation of the least squares regression line for a set of points in a scatterplot is given by $\hat{y} = 2.2 + 0.81x$. The point (5, 7) is one point on this scatterplot. Which of the following is the residual for the point (5, 7)?

(A) 0.71
(B) 0.75
(C) 4.05
(D) 6.25
(E) 7.87

6

There is a linear relationship between the duration x (in seconds) of an eruption of a geyser and the interval of time y (in minutes) until the next eruption. A least-squares regression line of data collected by a geologist is represented by, $\hat{y} = 41.9 + 0.18x$, $100 < x < 300$. What is the estimated increase in the interval of time until the next eruption that corresponds to an increase of 60 seconds in the duration?

(A) 0.18 minutes
(B) 3.6 minutes
(C) 10.8 minutes
(D) 36.0 minutes
(E) 41.9 minutes

7

Which of the following statements related to residuals are true?

I. The mean of the least-squares residuals is always zero.
II. If one tries to fit a linear model to bivariate data, a curved pattern in a residual plot shows that the relationship between two variables is not linear.
III. A residual plot can be a scatter plot of the regression residuals against the explanatory variable.

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

8

The equation of the least squares regression line for a set of points in a scatterplot is given by $\hat{y} = 1.3 + 0.27x$. The point (3, 2) is one point on this scatterplot. Which of the following is the residual for the point (3, 2)?

(A) –0.11
(B) 0.11
(C) 0.22
(D) 1.04
(E) 1.57

9

Data are obtained from a group of high school seniors comparing age and the number of hours spent on the telephone. The resulting regression equation is

Predicted number of hours = 0.123 (age) + 2.57 with $r = 0.866$

What percentage of the variation in the number of hours spent on the telephone can be explained by this least-squares regression model?

(A) 0.75%
(B) 0.866%
(C) 75%
(D) 86.6%
(E) This value cannot be found with the information given.

10

Imagine a least squares regression model is generated that predicts y based on x of the form $\hat{y} = b_0 + b_1 x$. Which of the following statements about the residuals from this regression model are true?

I. The slope of a least squares regression line that predicts the residual based on x will be zero.
II. The mean of the residuals is always zero.
III. A curved pattern in the residual plot is an indication that a nonlinear model may be a better fit.

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

11

A researcher wishes to examine the relationship between years of schooling completed and the number of pregnancies in young women. Her research discovers a linear relationship, and a least squares fit of her data results in the model

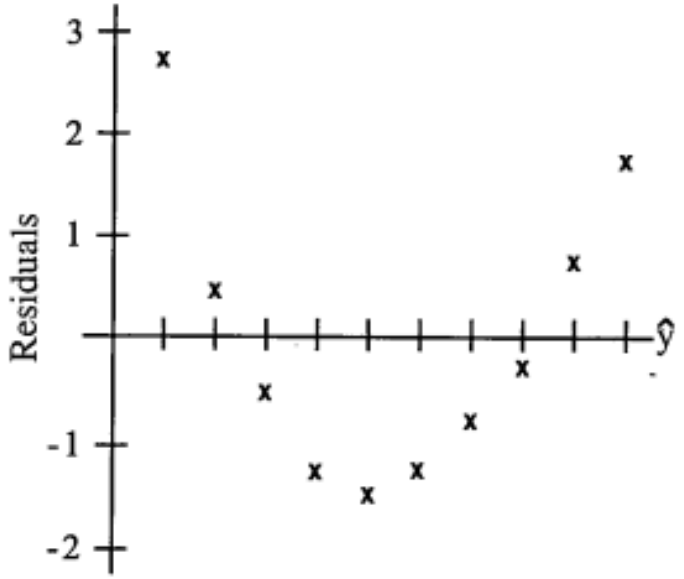
$$\hat{y} = 6.4 - 0.12x$$

where x is the number of years completed in school and y is the number of pregnancies. What is the estimated change in the number of pregnancies that corresponds to the completion of an additional 10 years of school?

(A) a decrease of 5.2
(B) an increase of 7.6
(C) a decrease of 7.6
(D) an increase of 1.2
(E) a decrease of 1.2

12

A linear model was constructed for a set of bivariate data using least squares regression techniques. Given the residual plot shown, what conclusion should be drawn?



(A) A linear model was not a good fit for the data.
(B) A linear model was a good fit for the data.
(C) The correlation between the original variables is close to 1.
(D) The data was drawn from a population that was not normally distributed.
(E) The study was poorly designed.

6) Which of the following could be the equation of the regression line shown on the scatterplot?

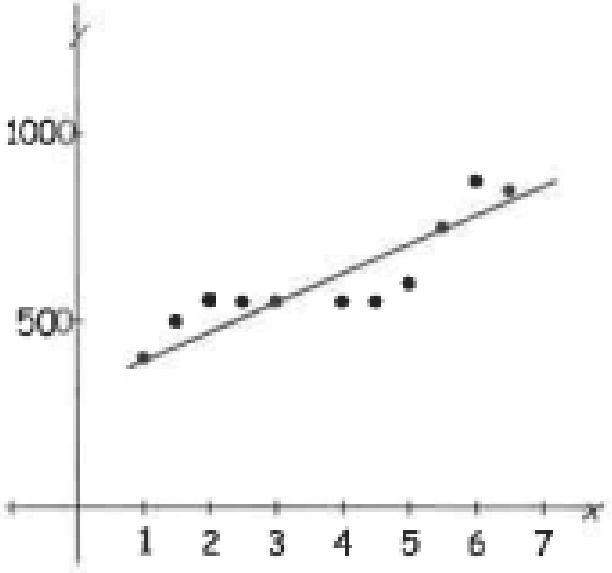
a) $y = 384.66 - 93.72x$

b) $y = 93.72 - 384.66x$

c) $y = -384.66 + 93.72x$

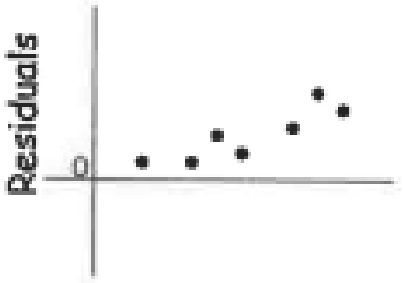
d) $y = 384.66 + 93.72x$

e) None of these are reasonable.

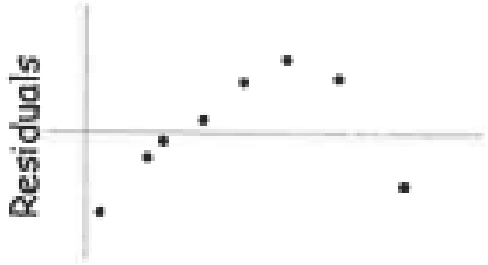


2) Which of the following residual plots indicates a reasonable fit to a given set of data?

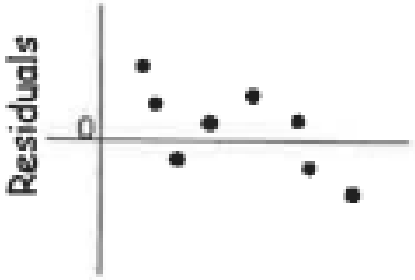
a)



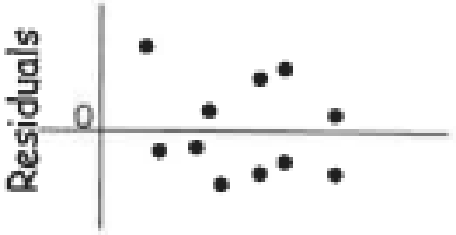
c)



b)



d)



e) None of these indicates a reasonable fit.

9) Suppose the correlation between variables is $r = .23$. What will the new correlation be if 14 is added to all values of the *x*-variable, every value of the *y*-variable is doubled, and the two variables are interchanged?

a) .23
b) .37
c) .74
d) -.23
e) -.74

7) Which of the following statements about residuals are true?

I. The mean of the residuals is always zero.

II. The regression line for a residual plot is a horizontal line.

III. A definite pattern in the residual plot is an indication that a nonlinear model will show a better fit to the data than the LSRL.

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.

2. Data was collected on two variables x and y and a least squares regression line was fitted to the data. The resulting equation is $\hat{y} = -2.29 + 1.70x$. What is the residual for point (5, 6)?

(A) −2.91
(B) −0.21
(C) 0.21
(D) 6.21
(E) 7.91

4. Given a set of ordered pairs (x, y) with $s_x = 2.5, s_y = 1.9$, $r = .63$, what is the slope of the regression line of y on x?

(A) 0.48
(B) 0.65
(C) 1.32
(D) 1.90
(E) 2.63

5. The relation between the selling price of a car (in \$1,000) and its age (in years) is estimated from a random sample of cars of a specific model. The relation is given by the following formula:

$$\text{SellingPrice} = 24.2 - (1.182)\text{Age}$$

Which of the following can be concluded from this equation?

(A) For every year the car gets older, the selling price drops by approximately \$2420.
(B) For every year the car gets older, the selling price goes down by approximately 11.82 percent.
(C) On average, a new car costs about \$11,820.
(D) On average, a new car costs about \$23,018.
(E) For every year the car gets older, the selling price drops by approximately \$1182.

1) For the model $\ln \hat{y} = 1.03 + 3.2x$, predict y when $x = 2$. Round to two decimal places.

A) 0.87
B) 1685.81
C) 2.01
D) $\hat{y}$ is undefined for $x = 2$.
E) 7.43

21

For the model $\sqrt{\hat{y}} = 4.7 + 0.9x$, predict y when x = 2. Round to two decimal places.

A) 2.55 B) 6.5 C) 43.48 D) 42.25 E) 41.44