

Unit 2 Review

Is there a relationship between the raises administrators at the University of South Florida receive and their performance on the job? A faculty group at USF obtained the following data for 15 USF administrators:

Administrator	Raise	Job Rating (5 pt scale)
1	\$18,000	2.76
2	16,700	1.52
3	15,787	4.40
4	10,608	3.10
5	10,268	3.83
6	9,795	2.84
7	9,513	2.10
8	8,459	2.38
9	6,099	3.59
10	4,557	4.11
11	3,751	3.14
12	3,718	3.64
13	3,652	3.36
14	3,227	2.92
15	2,808	3.00

Graph and label the scatterplot. To the right, analyze the scatter. Show all calculations.

mean correlation value

$$r = \frac{1}{n-1} \sum \left(\frac{x-\bar{x}}{s_x} \right) \left(\frac{y-\bar{y}}{s_y} \right)$$

• $r = -0.263$
 • negative
 • weak
 • linear

The scatterplot displaying the relationship between job ratings and raise (dollars) is negatively formed with a weak association. This because of the calculated correlation value is -0.263 . Based on the residual plot a linear form is best fit.

Job Ratings

Find the LSRL. Show calculations. Write the equation below. Draw the line into the scatterplot above.

$$a = \bar{y} - b\bar{x}$$

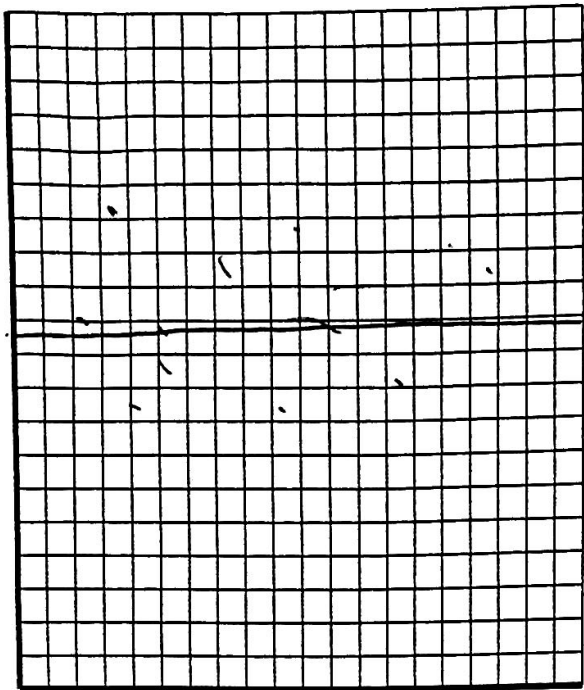
$$b = r \left(\frac{s_y}{s_x} \right)$$

$$\hat{y} = 14012 + -1700x$$

\hat{y} = predicted raise

x = job ratings

To the side of the graph below, calculate the residuals for each observation. Make a residual plot.



$$y - \hat{y} = \text{Residual}$$

7 # 2

2nd 3rd

Comment on the LSRL based on the residual plot.

Comment on the value of the coefficient of determination (r^2) $r^2 = .06$

The CoD of about 6% shows that only 6% of the predicted raises are accounted for by this LSRL model relating y to x .

Verify all calculations and graphs using your calculator.

Unit 2 Multiple Choice Practice

1. Which of the following would *not* be a correct interpretation of a correlation coefficient of $r = -.30$?

- C
- (a) the variables are inversely related ✓
 - (b) the coefficient of determination is .09 ✓
 - (c) 30% of the variation between the variables is linear.
 - (d) there exists a weak relationship between the variables
 - (e) all are correct

2. If the correlation coefficient of a bivariate set of data (x, y) is r , then which of the following is true?

- B
- (a) the variables x and y are linearly related.
 - (b) the correlation coefficient of the set (y, x) is also r
 - (c) the correlation coefficient of the set (x, ay) is ar
 - (d) the correlation coefficient of the set (ax, ay) is ar
 - (e) none of these.

3. Which of the following is *not* true of a correlation coefficient of a set of bivariate data?

- A
- (a) the higher the correlation coefficient, the steeper the line of best fit
 - (b) the absolute value of the correlation is independent of the slope of the line
 - (c) a low correlation does not necessarily indicate a weak relationship between variables
 - (d) two sets of bivariate data can have approximately equal correlation but very different scatterplots
 - (e) all of these are true

4. A coefficient of determination is found to be .81. Which of the following is true?

- C
- (a) 81% of the variation between the variables is accounted for in the linear relationship
 - (b) 81% of the data points lie on a line
 - (c) the correlation coefficient is approximately ± 0.9
 - (d) 19% of the variation between the variables is accounted for in the linear relationship
 - (e) all of these are true

5. A sample of 50 families has produced the following statistics: number of children in a family: $\bar{x} = 2.1$, $s_x = 1.4$; Annual Gross Income: $\bar{y} = 34,250$, $s_y = 10,540$; $r = .75$.

- A
- (a) $\text{Income} = 5,646 (\text{number of children}) + 22,392$
 - (b) $\text{Income} = 34,250 + .001 (\text{number of children})$
 - (c) $\text{Income} = .0001 (\text{number of children}) - 1.312$
 - (d) $\text{Number of children} = 5,646 (\text{income}) + 22,392$
 - (e) Equation cannot be determined from given information

6. A study of a sample of 40 college students studied their hours of part-time work per week and grade point average and found that the correlation between the variables was -0.43 . If the resulting linear regression equation is $\text{GPA} = 3.75 - 0.05(\text{number of hours})$, which of the following is not a correct statement?

$$\frac{-0.43}{-0.05} = \frac{-0.60}{x}$$

- (a) the average GPA of students who don't work is approximately 3.75
(b) if the correlation were -0.60 , the slope of the regression equations would be approximately -0.07
(c) students who work 40 hours per week have a mean GPA of approximately 1.75
(d) the value of the correlation coefficient and steepness of the regression line are not related
(e) all of these are correct statements

7. A bivariate set of data relates the amount of annual salary raise and previous performance rating. The LSRL is $\hat{y} = 1,400 + 2,000x$ where \hat{y} is the estimated raise and x is the performance rating. Which of the following statements is not correct?

- (a) for each increase of one point in performance rating, the raise will increase on average by \$2,000
(b) this equation produces predicted raises with an average error of 0
(c) a rating of 0 will yield a predicted raise of \$1,400
(d) the correlation coefficient of the data is positive
(e) all of the above are true

8. Linear regression usually employs the method of least squares. Which of the following is the quantity that is minimized by the least squares process?

- (a) y_i
(b) $x_i - \bar{x}_i$
(c) $\sum (y_i - \hat{y}_i)^2$
(d) (\bar{x}_i, \bar{y}_i)
(e) $\sum (x - \bar{x}_i)^2$

$$y - \hat{y} = \text{residual}$$