Study Session Week of 2/20

Objectives:

- I will apply previous knowledge to solve problems involving linear regression.

Agenda:

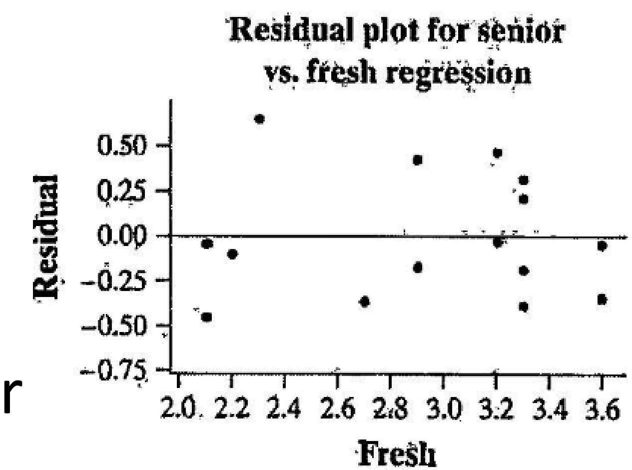
- MC practice problems
- FR practice problems

A high school guidance counselor wonders if it is possible to predict a student's GPA in their senior year from their GPA in the first marking period of their freshman year. She selects a random sample of 15 seniors from the graduating class of 468 and records both full-year GPA in their senior year ("Senior") and first=marking-period GPA in their freshman year ("Fresh"). A computer regression analysis and a residual plot of these data are given below

Predictor	Coef	SE Coef	T	P
Constant	1.6310	0.5328	3.06	0.009
Fresh	0.5304	0.1789	2.96	0.011

S = 0.3558 R-Sq = 40.3% R-Sq(adj) = 35.7%

1. Is this a linear model a good fit for this data?



A. Yes, the linear regression analysis was completed.
B. Yes, the residual plot has no clear pattern.
C. Yes, the residual plot has a clear patter
D. No, the residual plot has a clear pattern.
E. No, the residual plot has no clear pattern.

A high school guidance counselor wonders if it is possible to predict a student's GPA in their senior year from their GPA in the first marking period of their freshman year. She selects a random sample of 15 seniors from the graduating class of 468 and records both full-year GPA in their senior year ("Senior") and first=marking-period GPA in their freshman year ("Fresh"). A computer regression analysis and a residual plot of these data are given below

Predictor	Coef	SE Coef	T	P
Constant	1.6310	0.5328	3.06	0.009
Fresh	0.5304	0.1789	2.96	0.011

S = 0.3558   R-Sq = 40.3%   R-Sq(adj) = 35.7%

2. The equation for the least squares regression line from this sample is:

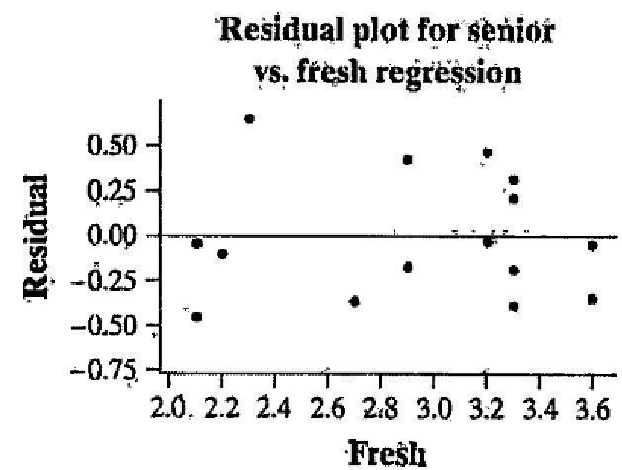
A. $\widehat{Fresh} = 1.6310Senior + 0.5304$

B. $\widehat{Fresh} = 1.6310 + 0.5304Senior$

C. $\widehat{Senior} = 1.6310Fresh + 0.5304$

D. $\widehat{Senior} = 1.6310 + 0.5304Fresh$

E. $\widehat{Senior} = 0.5304Fresh + 1.6310$



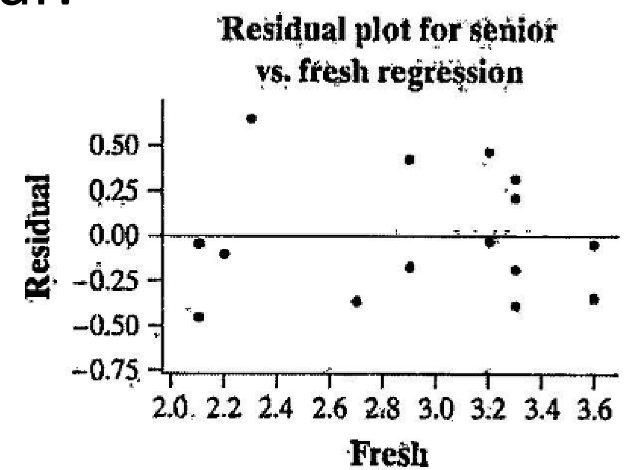
A high school guidance counselor wonders if it is possible to predict a student's GPA in their senior year from their GPA in the first marking period of their freshman year. She selects a random sample of 15 seniors from the graduating class of 468 and records both full-year GPA in their senior year ("Senior") and first=marking-period GPA in their freshman year ("Fresh"). A computer regression analysis and a residual plot of these data are given below

Predictor	Coef	SE Coef	T	P
Constant	1.6310	0.5328	3.06	0.009
Fresh	0.5304	0.1789	2.96	0.011

S = 0.3558 R-Sq = 40.3% R-Sq(adj) = 35.7%

3. One student who had a freshman GPA of 2.5 had a senior GPA of 3.2. What is their residual?

A. 2.957
B. -2.937
C. 0.243
D. -0.243
E. 0.403



$$\widehat{Senior} = 0.5304Fresh + 1.6310$$

A high school guidance counselor wonders if it is possible to predict a student's GPA in their senior year from their GPA in the first marking period of their freshman year. She selects a random sample of 15 seniors from the graduating class of 468 and records both full-year GPA in their senior year ("Senior") and first=marking-period GPA in their freshman year ("Fresh"). A computer regression analysis and a residual plot of these data are given below

Predictor	Coef	SE Coef	T	P
Constant	1.6310	0.5328	3.06	0.009
Fresh	0.5304	0.1789	2.96	0.011

S = 0.3558   R-Sq = 40.3%   R-Sq(adj) = 35.7%

4. Another student had a residual of -0.5 with a freshman year GPA of 2.5. What was their senior year GPA?

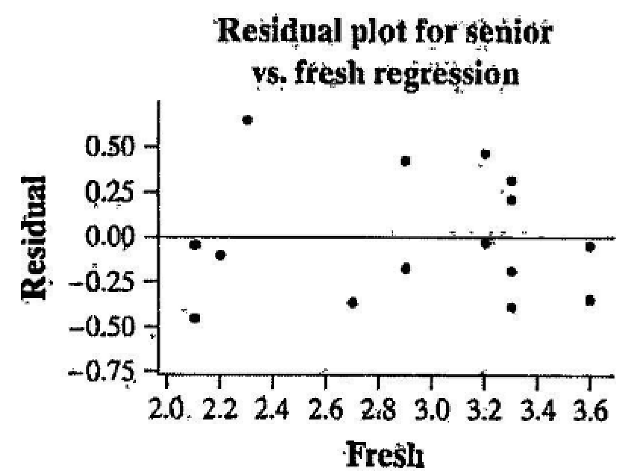
A. 2.957

B. 2.457

C. 3.457

D. 2.00

E. 3.00



$$\widehat{Senior} = 0.5304Fresh + 1.6310$$

A high school guidance counselor wonders if it is possible to predict a student's GPA in their senior year from their GPA in the first marking period of their freshman year. She selects a random sample of 15 seniors from the graduating class of 468 and records both full-year GPA in their senior year ("Senior") and first=marking-period GPA in their freshman year ("Fresh"). A computer regression analysis and a residual plot of these data are given below

Predictor	Coef	SE Coef	T	P
Constant	1.6310	0.5328	3.06	0.009
Fresh	0.5304	0.1789	2.96	0.011

S = 0.3558   R-Sq = 40.3%   R-Sq(adj) = 35.7%

5. What does the quantity R-Sq = 40.3% represent?

A. The correlation between freshman GPA and senior GPA – a measure of the strength of the linear relationship between the two variables.

B. The average deviation of observed senior GPA from the predicted senior GPA ratings, expressed as a percentage of the predicted history rating.

C. The average deviation of observed freshman GPA from the predicted senior GPA ratings, expressed as a percentage of the predicted history rating.

D. The percentage of variation in freshman gpa that can be explained by the regression equation.

E. The percentage of variation in senior gpa that can be explained by the regression equation.

2002 #4

4. Commercial airlines need to know the operating cost per hour of flight for each plane in their fleet. In a study of the relationship between operating cost per hour and number of passenger seats, investigators computed the regression of operating cost per hour on the number of passenger seats. The 12 sample aircraft used in the study included planes with as few as 216 passenger seats and planes with as many as 410 passenger seats. Operating cost per hour ranged between \$3,600 and \$7,800. Some computer output from a regression analysis of these data is shown below.

SCATTERPLOT WITH
LEAST SQUARES REGRESS

Scoring

The student response should include the following elements:

1. the correct equation of the least squares regression line with variables correctly defined;
2. the correct value for the correlation coefficient;
3. a correct interpretation of the given correlation coefficient; and
4. a complete explanation of why the given least-squares line would not be appropriate for describing the relationship over the restricted range.

Each element is scored as essentially correct (E), partially correct (P), or incorrect (I).

	Coef	StDev	T	P
Constant	1136	1226	0.93	0.376
Seats	14.673	4.027	3.64	0.005

S = 845.3 R-Sq = 57.0% R-Sq (adj) = 52.7%

(a) What is the equation of the least squares regression line that describes the relationship between operating cost per hour and number of passenger seats in the plane? Define any variables used in this equation.

(b) What is the value of the correlation coefficient for operating cost per hour and number of passenger seats in the plane? Interpret this correlation.

(c) Suppose that you want to describe the relationship between operating cost per hour and number of passenger seats in the plane for planes only in the range of 250 to 350 seats. Does the line shown in the scatterplot still provide the best description of the relationship for data in this range? Why or why not?

(a) What is the equation of the least squares regression line that describes the relationship between operating cost per hour and number of passenger seats in the plane? Define any variables used in this equation.

Part (a):

Predicted cost = 1136 + 14.673 (number of passenger seats)

OR

$\hat{y} = 1136 + 14.673x$ where y = operating cost per hour and x = number of passenger seats

Part (a) addresses the first element.

Element one is:

essentially correct if the solution has the correct equation and variables are defined correctly.

partially correct if only the equation is correct.

incorrect if the equation is not stated correctly.

(b) What is the value of the correlation coefficient for operating cost per hour and number of passenger seats in the plane? Interpret this correlation.

Part (b):

- The value of the correlation coefficient

  $r = +\sqrt{0.570} = 0.755$ (r is positive because the scatterplot shows a positive association)

- The interpretation of correlation

  There is a moderate (or strong) positive linear relationship between operating costs per hour and number of passenger seats.

  OR

  Fifty-seven percent of the variability in operating cost per hour can be explained by a linear relationship between cost and number of passenger seats AND the relationship is positive.

Part (b) addresses the second and third elements.

Element two is:

essentially correct if the student's solution states that $r = 0.755$.

partially correct if the student's solution only states that $r = \pm 0.755$.

incorrect if the student states any other value of r including $r = 0.726$ (square root of R-Sq (adj)).

(b) What is the value of the correlation coefficient for operating cost per hour and number of passenger seats in the plane? Interpret this correlation.

Part (b):

- The value of the correlation coefficient

  $r = +\sqrt{0.570} = 0.755$ (*r* is positive because the scatterplot shows a positive association)

- The interpretation of correlation

  There is a moderate (or strong) positive linear relationship between operating costs per hour and number of passenger seats.

  OR

  Fifty-seven percent of the variability in operating cost per hour can be explained by a linear relationship between cost and number of passenger seats AND the relationship is positive.

Element three is:

essentially correct if the student's solution

addresses, based on a correct understanding of the correlation coefficient, three or four of the following:

- type of relationship
- strength
- direction
- context

OR

states, based on a correct understanding of r^2:

- that 57 percent of the variability in operating cost per hour can be explained by a linear relationship between cost and number of passenger seats

  AND

- that the relationship is positive.

Note: If the student gives a correct interpretation of *r* but then incorrectly explains r^2, this is considered a parallel solution and is incorrect.

partially correct if the student's solution

addresses exactly two of the following – type of relationship (linear), strength, direction, and context (based on a correct understanding of the correlation coefficient).

OR

only states that 57 percent of the variability in operating cost per hour can be explained by a linear relationship between cost and number of passenger seats (based on a correct understanding of r^2) – BUT – does not state that the relationship is positive.

Suppose that you want to describe the relationship between operating cost per hour and number of passenger seats in the plane for planes only in the range of 250 to 350 seats. Does the line shown in the scatterplot still provide the best description of the relationship for data in this range? Why or why not?

Part (c):

No. The equation of the least-squares regression line is influenced by the three points in the upper right-hand corner and the two points in the lower left-hand corner of the scatterplot. The seven remaining points (with number of seats in the 250 to 350 range) would have a negative correlation. Hence, the slope of the recalculated least-squares regression line is negative.

Part (c) addresses the fourth element.

Element four is **essentially correct** if the student's solution

states that the existing line is not a good fit for the remaining seven points and correctly explains that the restricted data has a negative correlation or the recalculated least-squares regression line has a negative slope.

Element four is **partially correct** if the student's solution

explains why the existing line is not a good fit for the remaining seven points but does not communicate that the restricted data has a negative correlation or the recalculated least-squares regression line has a negative slope.

OR

removes fewer than the specified five points, but gives a correct interpretation of the effect on the correlation or slope of the least-squares regression line.

For elements 1 through 4, essentially correct responses count as one element and partially correct responses count as one-half of an element.

4 **Complete Response**

Four elements correct.

3 **Substantial Response**

Three elements correct.

2 **Developing Response**

Two elements correct.

1 **Minimal Response**

One element correct.

If a paper is between two scores (for example, 2 1/2 elements) use a holistic approach to determine whether to score up or down depending on the strength of the response and quality of communication.