Monday, October 8, 2018

- Warm-up
  - Once re-expressed, the model between mortgage amounts (in millions) and the years since 1970 at the Republic National Bank is:
    $\log(mortgages) = 0.146 + 0.074(years)$
    - What is the expected mortgage amount in 1990?
    - In what year do we expect mortgages reach 30 million?
- Finish Barbie Bungee Lab
- Check work from last week
- MC Practice

Objectives

Content: I will use linear regression to make accurate predictions.
Social: I will involve my entire group in problem solving.
Language: I will use correct vocabulary while analyzing the linear regression of bungee jumping.

Warm-up

- Once re-expressed, the model between mortgage amounts (in millions) and the years since 1970 at the Republic National Bank is:
  $\log(mortgages) = 0.146 + 0.074(years)$
  - What is the expected mortgage amount in 1990?
  - In what year do we expect mortgages reach 30 million?

$\log(m) = 0.146 + 0.074(20)$

$\log(m) = 1.626$

$10^{\log(m)} = 10^{1.626}$

$m = 42.26$ millions

$\log(30) = 0.146 + 0.074y$

$-0.146 \quad -0.146$

$\frac{\log(30) - 0.146}{0.074} = \frac{0.074y}{0.074}$

$17.9881 = y$

1988

Barbie Bungee

- And the final distance is…

16.1 feet

Objectives

Content: I will use linear regression to make accurate predictions.

Social: I will involve my entire group in problem solving.

Language: I will use correct vocabulary while analyzing the linear regression of bungee jumping.

"FRAPPY"

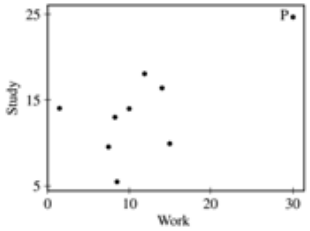
{Free Response AP Problem...Yay!}

2003 Problem #1

The following problem is taken from an actual Advanced Placement Statistics Examination. Your task is to generate a complete, concise statistical response in 15 minutes. You will be graded based on the AP rubric and will earn a score of 0-4. After grading, keep this problem in your binder for your AP Exam preparation.

A simple random sample of 9 students was selected from a large university. Each of these students reported the number of hours he or she had allocated to studying and the number of hours allocated to work each week. A least squares regression was performed and part of the resulting computer output is shown below.

Predictor	Coef	StDev	T	P
Constant	8.107	2.731	2.97	0.021
Work	0.4919	0.1950	2.52	0.040
S = 4.349	R-Sq = 47.6%	R-Sq (adj) = 40.1%		



Scoring:

The scatterplot displays the data that were collected from the 9 students.

E P I

(a) After point P, labeled on the graph, was removed from the data, a second linear regression was performed and the computer output is shown below.

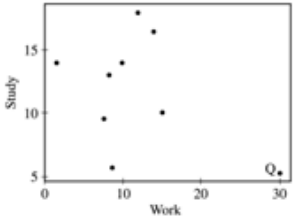
Predictor	Coef	StDev	T	P
Constant	11.123	3.986	2.79	0.032
Work	0.1500	0.3834	0.39	0.709
S = 4.327	R-Sq = 2.5%	R-Sq (adj) = 0.0%		

Does point P exercise a large influence on the regression line? Explain.

E P I

(b) The researcher who conducted the study discovered that the number of hours spent studying reported by the student represented by P was recorded incorrectly. The corrected data point for this student is represented by letter Q in the scatterplot.

Explain how the least squares regression line for the corrected data (in this part) would differ from the least squares regression line for the original data.



Part (a):

The point P does have a large influence on the regression line. When P is removed from the data set, the slope of the line changes from 0.4919 to 0.1500, the intercept changes from 8.107 to 11.123, and the value of R^2 drops from 47.6% to 2.5%. Also, the slope is significantly different from 0 when the point P is included in the data set and is not significantly different from 0 when the point P is excluded from the data set.

Part (a) is

Essentially correct (E) if the student

1. identifies the point P as influential
   AND
2. explains that there have been changes in at least 2 of the following:
   Slope
   Statistical significance of the slope
   Intercept
   Regression equation
   Value of R^2 (or R^2_{adj})

OR

mentions the change in one of the values above and discusses clearly how point P is extreme in the *x* direction (if just "extreme," needs to also explain why that implies the potential for influence).

NOTE: *r* (0.69 to 0.16) can be mentioned as well, but is not counted separately from R^2 unless the student provides clearly distinguishable interpretations of each.

Partially correct (P) if the student does one of the following

- Identifies Point P as influential but with weak justification (something changes).
- Identifies Point P as influential but only one change is noted.
- Confuses S as the slope on the computer output and thus states that the point P is not influential since the slope doesn't change much.

Incorrect (I) if the student only answers "yes" or refers to all of the numbers changing but provides no indication of understanding of what the numbers represent.

Part (b):

The regression line for the corrected data will have a negative slope rather than a positive slope, and the intercept would be much larger for the corrected data.

Part (b) is

Essentially correct (E) if the student indicates that the sign of the slope would change from positive to negative. The student should explicitly compare the 2 graphs. The student does not need to comment on the change in the intercept.

Partially correct (P) if the student does one of the following

- Indicates that the slope will change (including "slope is lower"), but fails to explicitly state that the sign changes from positive to negative.
- Comments only that the value of the correlation changes from positive to negative.
- Comments that the line "flattens."

Incorrect (I) if

- Student only comments on the intercept.
- Response is very poorly communicated (e.g., "line is negative," "data are positive," "data are weak").

4	**Complete Response (EE)** Both parts essentially correct.
3	**Substantial Response (EP or PE)** One part essentially correct and the other part partially correct
2	**Developing Response (EI or IE or PP)** One part essentially correct and the other part incorrect OR Both parts partially correct
1	**Minimal Response (PI or IP)** One part partially correct

Multiple Choice Practice

Homework

p 250 (29-31)