Tuesday, October 2, 2018

- **Warm-up**
  - **Enter the following data in your calculator and perform linear regression on the data comparing the distance of a threat in feet to the barks per 30 seconds**
    - **Write the linear equation**
    - **R**
    - **R^2**
- **Check/collect Homework**
- **Computer output**

Bark Freq of Black-Tailed Prairie Dogs

Distance From burrow	Bark Frequency
10.00	81
20.00	79
30.00	78
40.00	73
50.00	72
60.00	71
70.00	59
80.00	71
90.00	67
100.00	64
110.00	57
120.00	55
130.00	41

Homework: p p239 (5, 6)

5. Models.

a)

$$\ln \hat{y} = 1.2 + 0.8x$$
$$\ln \hat{y} = 1.2 + 0.8(2)$$
$$\ln \hat{y} = 2.8$$
$$\hat{y} = e^{2.8} = 16.44$$

b)

$$\sqrt{\hat{y}} = 1.2 + 0.8x$$
$$\sqrt{\hat{y}} = 1.2 + 0.8(2)$$
$$\sqrt{\hat{y}} = 2.8$$
$$\hat{y} = 2.8^2 = 7.84$$

c)

$$\frac{1}{\hat{y}} = 1.2 + 0.8x$$
$$\frac{1}{\hat{y}} = 1.2 + 0.8(2)$$
$$\frac{1}{\hat{y}} = 2.8$$
$$\hat{y} = \frac{1}{2.8} = 0.36$$

d)

$$\hat{y} = 1.2 + 0.8 \ln x$$
$$\hat{y} = 1.2 + 0.8 \ln(2)$$
$$\hat{y} = 1.75$$

e)

$$\log \hat{y} = 1.2 + 0.8 \log x$$
$$\log \hat{y} = 1.2 + 0.8 \log(2)$$
$$\log \hat{y} = 1.440823997...$$
$$\hat{y} = 10^{1.4408...}$$
$$\hat{y} = 27.59$$

Homework: p p239 (5, 6)

6. More models.

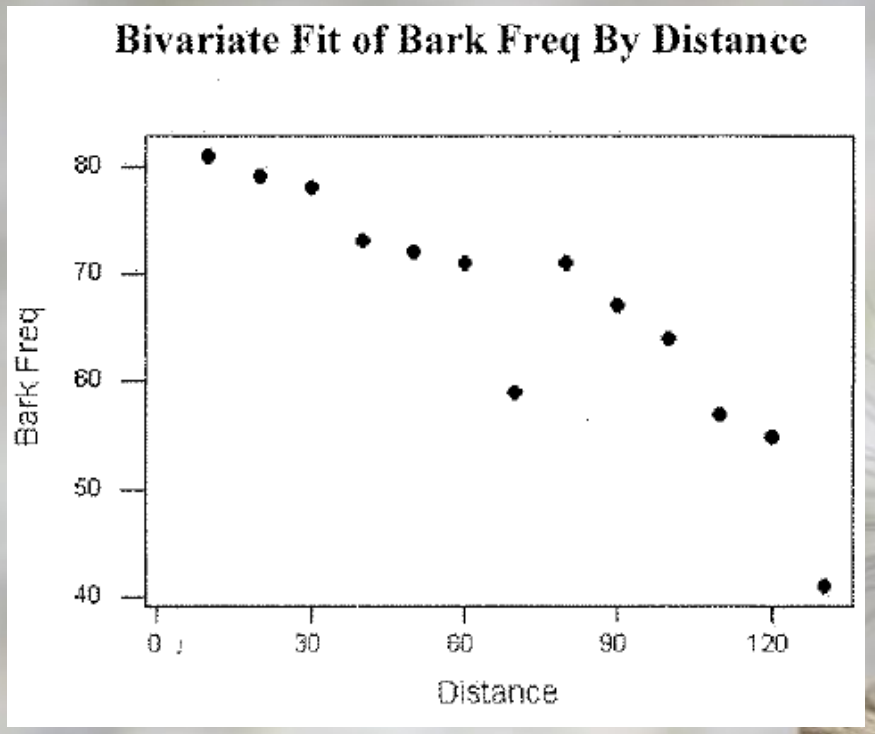
a)
$$\hat{y} = 1.2 + 0.8\log x$$
$$\hat{y} = 1.2 + 0.8\log(2)$$
$$\hat{y} = 1.44$$

b)
$$\log \hat{y} = 1.2 + 0.8x$$
$$\log \hat{y} = 1.2 + 0.8(2)$$
$$\log \hat{y} = 2.8$$
$$\hat{y} = 10^{2.8} = 630.96$$

c)
$$\ln \hat{y} = 1.2 + 0.8\ln x$$
$$\ln \hat{y} = 1.2 + 0.8\ln(2)$$
$$\ln \hat{y} = 1.7545...$$
$$\hat{y} = e^{1.7545..} = 5.78$$

d)
$$\hat{y}^2 = 1.2 + 0.8x$$
$$\hat{y}^2 = 1.2 + 0.8(2)$$
$$\hat{y}^2 = 2.8$$
$$\hat{y} = \sqrt{2.8} = 1.67$$

e)
$$\frac{1}{\sqrt{\hat{y}}} = 1.2 + 0.8x$$
$$\frac{1}{\sqrt{\hat{y}}} = 1.2 + 0.8(2)$$
$$\frac{1}{\sqrt{\hat{y}}} = 2.8$$
$$\hat{y} = \frac{1}{2.8^2} = 0.128$$


Bivariate Fit of Bark Freq By Distance
Bark Freq
80
70
60
50
40
0
30
60
90
120
Distance

Computer Analysis of the relationship

Predictor	Coef	StDev	T	P
Constant	85.269	2.942	28.98	0.000
Distance	-0.26429	0.03707	-7.13	0.000

S = 5.801 R-Sq = 82.2% R-Sq(adj) = 80.6%

y-int

variable → slope

- **Equation:** Frequency = 85.269 − 0.264 distance
- **R^2** = 82.2%
- **r** = $\sqrt{.822}$ = 0.9

sign should match slope

Below is the computer output for the appraised value and the number of rooms for houses in East Meadow, New York.

The regression equation is

value = 74.8 + 19.718 rooms

= 74.8 + 19.718(9)

Predictor	Coef	Stdev	t-ratio
Constant	74.80	19.04	3.93
rooms	19.718	2.631	7.49

R^2

$r = \sqrt{.438}$

$r = .661$

s = 29.05 R-sq = 43.8% R-sq(adj) = 43.0%

r +/- shows match slope

Analysis of Variance

SOURCE	DF	SS	MS
Regression	1	47398	47398
Error	72	60775	844
Total	73	108173	

- **Regression equation:**
- **R^2 (coefficient of determination):**
- **r:**
- **Find the appraised value of a house with 9 rooms:**

The relationship between hours of watching television in a typical day and age was examined. The data was gathered in the 1996 general Social Survey done by the National Opinion Research Center at the University of Chicago.

Predictor	Coef	SE Coef	T	P
Constant	2.1899	0.1577	13.89	0.000
age	0.017255	0.003348	5.15	0.000

S = 2.371 R-Sq = 1.4% R-Sq(adj) = 1.3%

- What is the estimated increase in average daily hours of television watching for each one-year increase in age? 0.017
- Write the regression equation for the output. $\widehat{hours} = 2.18 + 0.017 age$
- How reliable to you find this model? Why/why not?

Homework

"FRAPPY"

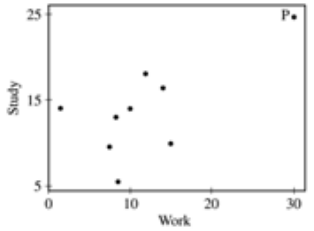
{Free Response AP Problem...Yay!}

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.