Statistical Methods: Homework 3

Cameron McIntyre

November 12, 2018

1 11.2.18

A graph of the luxury suite data in Question 8.2.5 suggests that the xy-relationship is linear. Moreover, it makes sense to constrain the fitted line to go through the origin, since $x = 0$ suites will necessarily produce $y = 0$ revenue.

a. Find the equation of the least squares line, $y = bx$. (Hint: Recall Question 11.2.14.)

b. How much revenue would 120 suites be expected to generate?

Answer:

a.

$$\sum x = 545$$

$$\sum y = 33.9$$

$$\sum x^2 = 54437$$

$$\sum xy = 3329.4$$

So, using our formulas:

$$\frac{\sum xy}{\sum x^2} = \frac{3329.4}{54437} = .061$$

The constraint is the that $\beta_0 = 0$ So our equation is:

$$y = .061x$$

b. Using x = 120.

$$y = .061 * 120 = 7.32$$

So revenue is projected at 7.32 million.

2 11.2.26

Among mammals, the relationship between the age at which an animal develops locomotion and the age at which it first begins to play has been widely studied. The table below lists "onset" times for locomotion and for play in eleven different species (46). Fit the data to the $y = ax^b$ model.

Species	**Locomotion Begins, x (days)**	**Play Begins, y (days)**
Homo sapiens	360	90
Gorilla gorilla	165	105
Felis catus	21	21
Canis familiaris	23	26
Rattus norvegicus	11	14
Turdus merula	18	28
Macaca mulatta	18	21
Pan troglodytes	150	105
Saimiri sciurens	45	68
Cercocebus alb.	45	75
Tamaisciureus hud.	18	46

Answer

We take logarithms.

$$y = ax^b \leftrightarrow \ln(y) = \ln(a) + b\ln(x) \leftrightarrow \hat{y} = \hat{a} + b\hat{x}$$

$$b = \frac{n\sum x_i * y_i - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2} = \frac{11 * 161.37 - 40.86 * 41.49}{11 * 164.71 - 40.86^2} = 0.5606$$

$$\hat{a} = \frac{\sum y_i}{-} b \sum x_i n = \frac{41.491 - 0.5606 * 40.86}{11} = 1.6892$$

$$b = 0.5606$$

$$a = e^{1.6892} = 5.415$$

So our equation is:

$$\hat{y} = 5.415 * x^{0.5606}$$

3 11.3.2

The best straight line through the Massachusetts funding/graduation rate data described in Question 11.2.7 has the equation $y = 81.088 + 0.412x$,where s= 11.78848.

a Construct a 95% confidence interval for β_1.

b What does your answer to part (a) imply about the outcome of testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ at the $\alpha = 0.05$ level of significance?

c Graph the data and superimpose the regression line. How would you summarize these data, and their implications, to a meeting of the state School Board?

Answer:

a Construct a 95% confidence interval for β_1.

$$\sum x_i = 360$$

$$\sum x_i^2 = 5365.08$$

$$\sum x_i y_i = 31402$$

$$\sum y_i^2 = 188255.3$$

$$s = 11.78848$$

$DF = 26 - 2 = 24$ and $T_{.025,24} = 2.0639$

$$C.I. = \big[\hat{\beta}_1 - t_{.025,24}\frac{s}{\sqrt{\sum(x-\bar{x})^2}}, \hat{\beta}_1 + t_{.025,24}\frac{s}{\sqrt{\sum(x-\bar{x})^2}}\big]$$

$$C.I. = \big[.412 - 2.0639\frac{11.78}{\sqrt{380.46}}, .412 + 2.0639\frac{11.78}{\sqrt{380.46}}\big]$$

$$C.I. = \big[-.835, 1.659\big]$$

b What does your answer to part (a) imply about the outcome of testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ at the $\alpha = 0.05$ level of significance?

The confidence interval contains 0 so we can accept H_0 and say that β_1 is not significantly different than 0.

c Graph the data and superimpose the regression line. How would you summarize these data, and their implications, to a meeting of the state School Board?
I would advise that state board that spending more money on students does not show a major boost in increasing the graduation rate.

4 11.4.2

Suppose that X and Y have the joint pdf $f_{X,Y}(x, y) = x+y, 0 < x < 1, 0 < y < 1$ Find $\rho(X, Y)$.

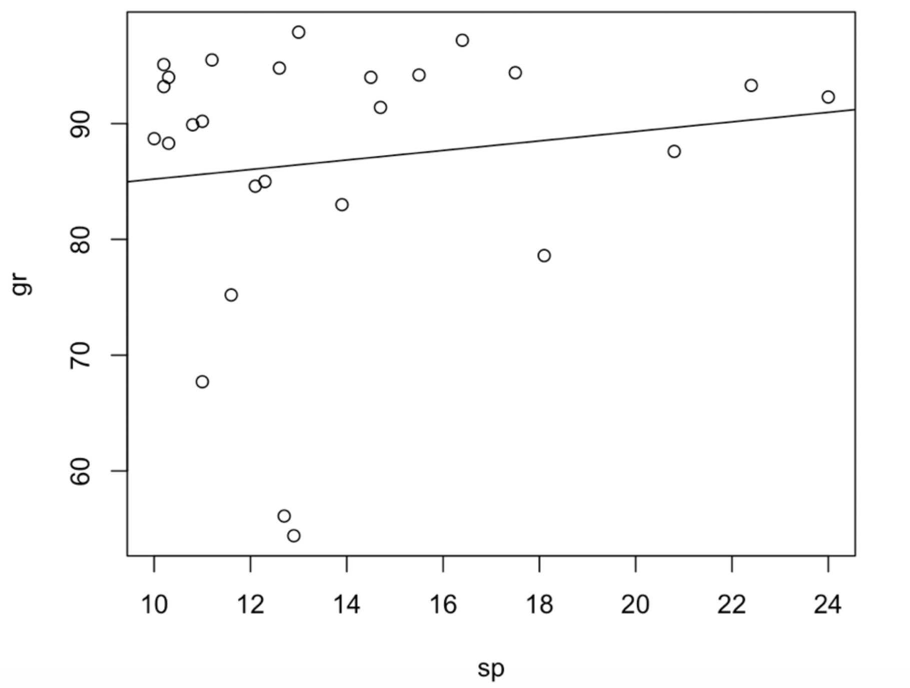

Figure 1: Data and Regression Line.

Answer:

$$\rho(x, y) = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

$$E[X] = \int_0^1 \int_0^1 x(x + y)dydx = \frac{7}{12}$$

$$E[Y] = \int_0^1 \int_0^1 y(x + y)dydx = \frac{7}{12}$$

$$E[XY] = \int_0^1 \int_0^1 xy(x + y)dydx = \frac{1}{3}$$

$$E[X^2] = \int_0^1 \int_0^1 y(x + y)dydx = \frac{5}{12}$$

$$E[Y^2] = \int_0^1 \int_0^1 y(x + y)dydx = \frac{5}{12}$$

Then we have,

$$Var[X] = E[X^2] - E[X]^2 = \frac{5}{12} - \frac{49}{144} = \frac{11}{144}$$

$$Var[Y] = E[Y^2] - E[Y]^= \frac{5}{12} - \frac{49}{144} = \frac{11}{144}$$

$$Cov(X, Y) = E[XY] - E[X]E[Y] = \frac{1}{3} - \frac{49}{144} = \frac{-1}{144}$$

Finally,

$$\rho(x, y) = \frac{Cov(X, Y)}{(Var[X])^{\frac{1}{2}}(Var[Y])^{\frac{1}{2}}} = \frac{\frac{-1}{144}}{\frac{11}{144}} = \frac{-1}{11}$$

5 11.5.4

Suppose that the random variables X and Y have a bivariate normal pdf with $\mu_X = 56, \mu_Y = 11, \sigma_X^2 = 1.2, \sigma_Y^2 = 2.6$, and $\rho = 0.6$. Compute $P(10 < Y < 10.5|x = 55)$. Suppose that n = 4 values were to be observed with x fixed at 55. Find $P(10.5 < \bar{Y} < 11|x = 55)$.

Answer:

$$E[Y|x] = \mu_y + \frac{\rho\sigma_y}{\sigma_x} = 11 + \frac{.6 * \sqrt{2.6}}{\sqrt{1.2}}(56 - 55) = 10.11$$

$$Var[Y|x] = (1 - \rho^2)\sigma^2 = (1 - .6^2) * 2.6 = 1.664$$

Therefore,

$$P(10 < y < 10.5) = P(\frac{10 - 10.1}{1.289} < \frac{y - \mu_y}{\sigma_y} < \frac{10.5 - 10.1}{1.289}) = P(-.09 < z < .3) = .1538$$

And if there are 4 measurements (N=4):

$$P(10 < \bar{y} < 10.5) = P(\frac{10 - 10.1}{\frac{1.289}{2}} < \frac{y - \mu_y}{\frac{\sigma_y}{\sqrt{n}}} < \frac{10.5 - 10.1}{\frac{1.289}{2}}) = P(.61 < z < 1.38) = .1871$$