

Exercise 1:

Two attributes, X and Y are found to have a correlation of 0.43

$$\bar{X} = 1.32, SD_x = 0.35, \bar{Y} = 5.30, SD_y = 1.40$$

Suppose we want to predict the value of Y for X=0

$$b = r_{xy} \left(\frac{SD_y}{SD_x} \right) \quad \hat{Y} = \bar{Y} + b(X - \bar{X})$$

First calculate b: $b = r_{xy} \left(\frac{SD_y}{SD_x} \right) = 0.43 \left(\frac{1.40}{0.35} \right) = 0.43(4) = 1.72$

Now calculate $\hat{Y}$: $\hat{Y} = \bar{Y} + b(X - \bar{X})$

$$\begin{aligned} &= 5.3 + 1.72(0 - 1.32) \\ &= 5.3 + 1.72(-1.32) \\ &= 5.3 + (-2.27) \\ &= 3.03 \end{aligned}$$

Exercise 2:

Two attributes, X and Y are found to have a correlation of 0.87

$$\bar{X} = 2.75, SD_x = 0.25, \bar{Y} = 3.14, SD_y = 0.33$$

Suppose we want to predict the value of Y for X=4.2

$$b = r_{xy} \left(\frac{SD_y}{SD_x} \right) \quad \hat{Y} = \bar{Y} + b(X - \bar{X})$$

First calculate b: $b = r_{xy} \left(\frac{SD_y}{SD_x} \right) = 0.87 \left(\frac{0.33}{0.25} \right) = 0.87(1.32) = 1.15$

Now calculate $\hat{Y}$: $\hat{Y} = \bar{Y} + b(X - \bar{X})$

$$\begin{aligned} &= 3.14 + 1.15(4.2 - 2.75) \\ &= 3.14 + 1.15(1.45) \\ &= 3.14 + 1.67 \\ &= 4.81 \end{aligned}$$

Exercise 3

Suppose the average number of rats in a set of dockside warehouses is 57 per warehouse with an SD of 12. The average number of packages of rat poison purchased per warehouse per year is 4.2 with an SD of 1.1. The correlation between rats in warehouses and packages of rat poison purchased is 0.5

- Kleensite bought 12 packages of rat poison last year. How many rats would you predict are in their warehouse?
- Remember to calculate the standard error as well as the predicted value

Hint: First work out what is Y , the thing you want to calculate. Then identify $\bar{X}$, SD_x , $\bar{Y}$, SD_y from the wording of the question.

$$b = r_{xy} \left(\frac{SD_y}{SD_x} \right) \quad \hat{Y} = \bar{Y} + b(X - \bar{X}) \quad \sigma_{est} = \left(SD_y \sqrt{1 - r_{xy}^2} \right)$$

Y is the number of rats.

Identify r_{xy} , $\bar{X}$, SD_x , $\bar{Y}$, SD_y , X

$\bar{X}$ = mean number of packets of rat poison = 4.2, $SD_x = 1.1$

$\bar{Y}$ = mean number of rats = 57, $SD_y = 12$

Correlation coefficient $r_{xy} = 0.5$ $X = 12$

First, calculate b : $b = r_{xy} \left(\frac{SD_y}{SD_x} \right) = 0.5 \left(\frac{12}{1.1} \right) = 0.5(10.9) = 5.45$

Next calculate $\hat{Y}$: $\hat{Y} = \bar{Y} + b(X - \bar{X})$
 $= 57 + 5.45(12 - 4.2)$
 $= 57 + 5.45(7.8)$
 $= 57 + 42.51$
 $= 99.51 \text{ rats}$

Finally calculate the standard error σ_{est}

$$\begin{aligned} \sigma_{est} &= SD_y \left(\sqrt{1 - r_{xy}^2} \right) \\ &= 12 \left(\sqrt{1 - 0.5^2} \right) \\ &= 12 \left(\sqrt{1 - 0.25} \right) \\ &= 12 \left(\sqrt{0.75} \right) \\ &= 12(0.87) = 10.44 \end{aligned}$$

So the expected number of rats is 99.51 ± 10.44

Exercise 4 (Homework)

Using the data in DAV Datafile 14.xls calculate the estimated number of bounces for a ping time of 42ms and the standard error related to that estimate.

$$b = r_{xy} \left(\frac{SD_y}{SD_x} \right) \quad \hat{Y} = \bar{Y} + b(X - \bar{X}) \quad \sigma_{est} = \left(SD_y \sqrt{1 - r_{xy}^2} \right)$$

Y is the number of bounces.

Use Excel to calculate r_{xy} , $\bar{X}$, SD_x , $\bar{Y}$, SD_y , X

$$\bar{X} = \text{mean ping time} = 39.5 \quad SD_x = 22.5 \quad X = 42$$

$$\bar{Y} = \text{mean bounces} = 2.5 \quad SD_y = 1.0$$

Correlation coefficient $r_{xy} = 0.9$

First, calculate b : $b = r_{xy} \left(\frac{SD_y}{SD_x} \right) = 0.9 \left(\frac{1}{22.5} \right) = 0.9(0.04) = 0.04$ ^(0.036)

Next, calculate $\hat{Y}$: $\hat{Y} = \bar{Y} + b(X - \bar{X})$

$$= 2.5 + 0.04(42 - 39.5)$$

$$= 2.5 + 0.04(2.5)$$

$$= 2.5 + 0.10$$

$$= 2.60 \text{ bounces}$$

Finally, calculate the standard error σ_{est}

$$\sigma_{est} = SD_y \sqrt{1 - r_{xy}^2}$$

$$= 1.0 \sqrt{1 - 0.9^2}$$

$$= 1.0 \sqrt{1 - 0.81}$$

$$= 1.0 \sqrt{0.19}$$

$$= 1.0(0.44)$$

$$= 0.44$$

So the expected number of bounces is 2.6 ± 0.44 .