Answer on Question #46755 – Math – Statistics and Probability

Associated with a job are two random variables: CPU time required (Y) and number of disk I/O Operations (X).

i Time(sec) y_i Number x_i

1 40 398

2 38 390

3 42 410

4 50 502

5 60 590

6 30 305

7 20 210

8 25 252

9 40 398

10 39 392

(i) Find the regression equation.

(ii) What is the estimate increase in CPU time for one additional disk operation.

(iii) Comment on the goodness of the model.

Solution

(i) Find the regression equation.

$$y = ax + b.$$

$$b = \frac{(\sum_i y_i)(\sum_i x_i^2) - (\sum_i x_i)(\sum_i x_i y_i)}{n(\sum_i x_i^2) - (\sum_i x_i)^2}; a = \frac{n(\sum_i x_i y_i) - (\sum_i x_i)(\sum_i y_i)}{n(\sum_i x_i^2) - (\sum_i x_i)^2}.$$

$$n = 10; \left(\sum_i x_i\right) = 3847; \left(\sum_i y_i\right) = 384; \left(\sum_i x_i y_i\right) = 159318; \left(\sum_i x_i^2\right) = 1591405.$$

$$a = \frac{(384)(1591405) - (3847)(159318)}{10(1591405) - (3847)^2} = \frac{-1796826}{1114641} = -1.61.$$

$$b = \frac{10(159318) - (3847)(384)}{10(1591405) - (3847)^2} = \frac{115932}{1114641} = 0.10.$$

So

$$y = 0.10 \cdot x - 1.61.$$

(ii) The estimate increase in CPU time for one additional disk operation is

$$\Delta y = y(x+1) - y(x) = (0.10 \cdot (x+1) - 1.61) - (0.10 \cdot x - 1.61) = 0.10\ s.$$

(iii) Comment on the goodness of the model.

The value r^2 is a fraction between 0.0 and 1.0, and has no units. An r^2 value of 0.0 means that knowing X does not help you predict Y. There is no linear relationship between X and Y, and the best-fit line is a horizontal line going through the mean of all Y values. When r^2 equals 1.0, all points lie exactly on a straight line with no scatter. Knowing X lets you predict Y perfectly.

$$r^2 = \frac{\left(\sum_i \left(f_i - \frac{(\sum_i y_i)}{n}\right)^2\right)}{\left(\sum_i y_i^2\right) - \frac{(\sum_i y_i)^2}{n}},$$

where $f_i = y(x_i)$.

$$\left(\sum_i y_i^2\right) = 15954;\ \frac{(\sum_i y_i)}{n} = \frac{384}{10} = 38.4;\ \left(\sum_i \left(f_i - \frac{(\sum_i y_i)}{n}\right)^2\right) = 1138.357.$$

$$r^2 = \frac{1138.357}{15954 - \frac{(384)^2}{10}} = 0.94.$$

Our value r^2 is near 1.0. Therefore the model is in good agreement with the experimental data.