Class Exercise

Formulas:

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2}\sqrt{n \sum y^2 - (\sum y)^2}}$$

$$slope: \quad b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \qquad intercept: \quad a = \frac{\sum y - b \sum x}{n}$$

1. **Suitcases**

Bernard Sadow who is credited with inventing the wheeled suitcase in the 1970s struggled to find buyers. He was told that men would consider rolling a suitcase *unmanly* and women would expect their luggage to be carried by men.

The table below shows the approximate weight (in kg) of different suitcases and the corresponding force (in Newtons) required to pull them across a flat surface with wheels:

Weight (kg)	8	10	13	15	19	21	22	25	30	32
Force (N)	9	13	14	17	21	23	23	26	29	30

To help you answer the following questions, we have computed the following values for you:

$$\sum x = 195 \qquad \sum y = 205$$
$$\sum x^2 = 4393 \qquad \sum y^2 = 4651 \qquad \sum xy = 4507$$

a. Calculate the coefficient of correlation, r, for the given data. Comment on the strength and direction of the relationship.

Solution

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2}\sqrt{n \sum y^2 - (\sum y)^2}} = \frac{10(4507) - (195)(205)}{\sqrt{10(4393) - (195)^2}\sqrt{10(4651) - (205)^2}}$$
$$= 0.9900$$

Comment: There is a strong positive (near perfect) linear correlation between the weight of the suitcases and the force required to pull it. The heavier the bag, the more force required.

b. Calculate the coefficient of determination, R^2, for the given data. Comment on the amount of variation that is accounted for by this data.

Solution

$$R^2 = r^2 = (0.9900)^2 = 0.9802 \qquad ; \qquad 1 - R^2 = 1 - 0.9802 = 0.0198$$

Comment: About 98.08% of the variation in the pulling force (N) can be explained by its linear relationship with the weight of the suitcase (kg) (i.e. the regression model using suitcase weight as the predictor does an excellent job of accounting for differences in pulling force). Only about 1.92% of the variation in force is not accounted for by this linear model. This could be due to other factors such as surface friction, suitcase wheel quality, measurement error, etc.

c. Calculate the slope of the least squares line and interpret it in the context of the problem.

Solution

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{10(4507) - (195)(205)}{10(4393) - (195)^2} = 0.8628$$

Interpretation: For every 1 kg increase in suitcase weight, the predicted force required to pull it increases by about 0.8628 N

d. Calculate the y-intercept of the least squares line and interpret it in the context of the problem.

Solution

$$a = \frac{\sum y - b \sum x}{n} = \frac{205 - (0.8628)(195)}{10} = 3.6749$$

Interpretation: When the suitcase weight is 0 kg, the regression model predicts that the force required to pull it would be about 3.6749 Newtons.

Note: a suitcase with zero weight doesn't make sense in reality; but statistically, the intercept represents the model's baseline starting value. In context, it can be thought of as the small amount of force needed to overcome friction or to get the wheels moving, even for a very light suitcase.

e. Make a prediction for a suitcase that weighs 20 kg. Is this estimate trustworthy? Why?

Solution

$$y = 0.8628x + 3.6749 \qquad \Rightarrow \qquad \begin{aligned} y &= 0.8628(20) + 3.6749 \\ &= 20.9309\,N \end{aligned}$$

Comment: This estimate is an interpolation and is trustworthy since $20 \in [8,\,32]$

f. Make a prediction for a suitcase that weighs 3 kg. Is this estimate reliable? Why?

Solution

$$y = 0.8628x + 3.6749 \qquad \Rightarrow \qquad \begin{aligned} y &= 0.8628(3) + 3.6749 \\ &= 6.2633\,N \end{aligned}$$

Comment: This estimate is an extrapolation and is not trustworthy since $3 \notin [8,\,32]$

g. A suitcase that weighs 18 kg actually requires 20 N of force to pull. Calculate the residual and interpret the result.

Solution

$$y = 0.8628x + 3.6749 \qquad \Rightarrow \qquad \begin{aligned} y &= 0.8628(18) + 3.6749 \\ &= 19.2053\,N \end{aligned}$$

Comment: the model underestimates the amount of force required to pull an 18 kg suitcase by $0.7947\,N$