

Foundations of data science - Solutions to comprehension questions on regression

QUESTION 2

Researchers capture 94 squirrels, and determine that the sample mean of the squirrels is 352.0g with a standard deviation of 23.0g. They then standardise this variable. What is the standardised weight of a squirrel weighing 386.5g?

Sample mean $\bar{x} = 352.0$

Sample s.d. $s = 23 \text{ g}$

Weight of squirrel $x = 386.5 \text{ g}$

Standardised weight of squirrel

$$z = \frac{x - \bar{x}}{s} = \frac{386.5 - 352.0}{23}$$

$$= \frac{34.5}{23}$$

$$\underline{\underline{z = 1.5}}$$

QUESTION 3

Researchers capture 94 squirrels, and determine that the sample mean of the squirrels is 352.3g with a standard deviation of 23.3g. They then standardise this variable. What is the standard deviation of this standardised variable?

The standard deviation of a standardised variable is 1, by definition.

QUESTION 7

Suppose the correlation coefficient of the length of time spent writing an exam and the final mark is -0.2. A lecturer decides to try to predict marks from the time taken. The first steps are to plot standardise the variables and then to fit a regression line to the standardised variables. What is the gradient of the regression line?

From the lecture notes we know that in the linear regression

$$y = \hat{\beta}_0 + \hat{\beta}_1 x$$

the slope coefficient is equal to

$$\hat{\beta}_1 = r \frac{s_y}{s_x}$$

If we standardise both x and y , then $s_y = s_x = 1$, so $\hat{\beta}_1 = r$.

\Rightarrow In this case $\hat{\beta}_1 = r = -0.2$

QUESTION 8

In the above example, what is the coefficient of determination?

In simple linear regression we have the following relationship between the coefficient of determination and the correlation coefficient:

$$R^2 = r^2$$

\Rightarrow In this case $R^2 = (-0.2)^2 = 0.04$