Practice Problem 2

Ian Chen

February 7, 2024

Problem 1 — Find the equation for the least-squares regression line for this dataset by hand and show your work.

Answer

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$

X	Y
2	50
3	71
4	87
5	95

$\bar{X} = \frac{2+3+4+5}{4} = 3.5$

$\bar{Y} = \frac{50+71+87+95}{4} = 75.75$

$\sum (X_i - \bar{X})^2 = (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 = (-1.5)^2 + (-0.5)^2 + (0.5)^2 + (1.5)^2 = 5$

$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = (2-3.5)(50-75.75) + (3-3.5)(71-75.75) + (4-3.5)(87-75.75) + (5-3.5)(95-75.75) = (-1.5)(-25.75) + (-0.5)(-4.75) + (0.5)(11.25) + (1.5)(19.25) = 75.5$

$\hat{\beta}_1 = \frac{75.5}{5} = 15.1$

$\hat{\beta}_0 = 75.75 - 15.1(3.5) = 22.9$

$Y = 15.1X + 22.9$

Problem 2 — Using your regression model, predict the exam grade for someone who studied for 30 minutes.

Answer

$Y = \hat{\beta}_1 X + \hat{\beta}_0$

30 minutes = 0.5 hours

$Y = 15.1(0.5) + 22.9 = 30.45$

Problem 3 — Calculate the R^2 of your model by hand and show your work.

Answer

$mean = \bar{y} = \frac{50+71+87+95}{4} = 75.75$

$Var(mean) = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}$

$Var(mean) = \frac{(50-75.75)^2 + (71-75.75)^2 + (87-75.75)^2 + (95-75.75)^2}{4} = \frac{(-25.75)^2 + (-4.75)^2 + (11.25)^2 + (19.25)^2}{4} = \frac{1182.75}{4} = 295.6875$

$Var(fit) = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$

Best fit line: $f(x) = 15.1x + 22.9$

$Var(fit) = \frac{(50-f(2))^2+(71-f(3))^2+(87-f(4))^2+(95-f(5))^2}{4}$

$= \frac{(50-53.1)^2+(71-68.2)^2+(87-83.3)^2+(95-98.4)^2}{4}$

$= \frac{(-3.1)^2+(2.8)^2+(3.7)^2+(-3.4)^2}{4}$

$= \frac{9.61+7.84+13.69+11.56}{4}$

$= \frac{42.7}{4} = 10.675$

$R^2 = \frac{Var(mean)-Var(fit)}{Var(mean)} = \frac{295.6875-10.675}{295.6875} = 0.963897696$