

PROOFS

EDWARD O'CALLAGHAN

CONTENTS

1. Coefficient of determination (Proof) 2

1. COEFFICIENT OF DETERMINATION (PROOF)

Given that:

Definition 1.1 (total sum of squares). $SS_{total} = \sum_{i=1}^n (y_i - \bar{y})^2$

Definition 1.2 (regression sum of squares). $SS_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$

Definition 1.3 (residual sum of squares). $SS_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$

Problem 1.4 (Show that:). $SS_{total} = SS_{reg} + SS_{res}$

A direct proof is given:

Proof.

$$\begin{aligned}
 \sum_i (y_i - \bar{y})^2 &= \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i (y_i - \hat{y}_i)^2 \\
 &= \sum_i \underbrace{(\hat{y}_i - \bar{y})^2}_a + \underbrace{(y_i - \hat{y}_i)^2}_b \quad (\text{and recall that } a^2 + b^2 = (a + b)^2 - 2ab) \\
 &= \sum_i (y_i - \bar{y})^2 - 2 \underbrace{\sum_i (\hat{y}_i - \bar{y})(y_i - \hat{y}_i)}_{\text{show this } = 0}
 \end{aligned}$$

and so,

$$\begin{aligned}
 \implies \sum_i (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) &= \sum_i (\cancel{\bar{y}} + b_1(x_i - \bar{x}) - \cancel{\bar{y}})(y_i - \bar{y} - b_1(x_i - \bar{x})) \\
 &\text{by, } \hat{y}_i = b_0 + b_1 x_i = (\bar{y} - b_1 \bar{x}) + b_1 x_i = \bar{y} + b_1(x_i - \bar{x}) \\
 &= b_1 \sum_i (x_i - \bar{x})[y_i - \bar{y} - b_1(x_i - \bar{x})] \\
 &= b_1 [S_{xy} - b_1 S_{xx}] \\
 &= b_1 \left[S_{xy} - \frac{S_{xy}}{\cancel{S_{xx}}} \cdot \cancel{S_{xx}} \right] = 0 \quad \square
 \end{aligned}$$