MATH 2780 Chapter 3 Worksheet

Summary

A First Order (Straight-Line) Model

$y = \beta_0 + \beta_1 x + \varepsilon$

where

y = **Dependent** variable (variable to be modeled-sometimes called the **response** variable)

x = **Independent** variable (variable used as **predictor** of y)

$E(y|x) = \beta_0 + \beta_1 x$

ε = (epsilon) = Random **error** component

β_0 = (beta zero) = ***y*-intercept** of the line

β_1 = (beta one) = **Slope** of the line.

Coefficient Estimation

$$SS_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n(\overline{x})^2$$

$$SS_{xy} = \sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x}) = \sum_{i=1}^{n} x_i y_i - n\overline{xy}$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \quad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

Linear Regression Assumptions

Assumption 1 The mean of the probability distribution of ε is 0. $E(\varepsilon) = 0$

Assumption 2 The variance of the probability distribution of is constant. $\text{var}(\varepsilon) = \sigma^2$

Assumption 3 The probability distribution of ε is normal. $\varepsilon \sim N(0, \sigma^2)$

Assumption 4 The errors associated with any two observations are independent.

$$f(\varepsilon_i, \varepsilon_j) = f(\varepsilon_i) f(\varepsilon_j)$$

Regression Line Standard Error

$$s^2 = \frac{SSE}{Degrees\ \ of\ Freedom} = \frac{SSE}{n-2} \quad , \quad s = \sqrt{s^2}$$

$$SSE = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = SS_{yy} - \hat{\beta}_1 SS_{xy}$$

$$SS_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n(\overline{y})^2$$

Slope Hypothesis Test and Confidence Interval

$$H_0: \beta_1 = 0 \qquad H_a: \beta_1 \neq 0 \qquad t = \frac{\hat{\beta}_1 - 0}{s / \sqrt{SS_{xx}}} \qquad \hat{\beta}_1 \pm t_{\alpha/2} \frac{s}{\sqrt{SS_{xx}}}, \ df = n-2$$

Correlation

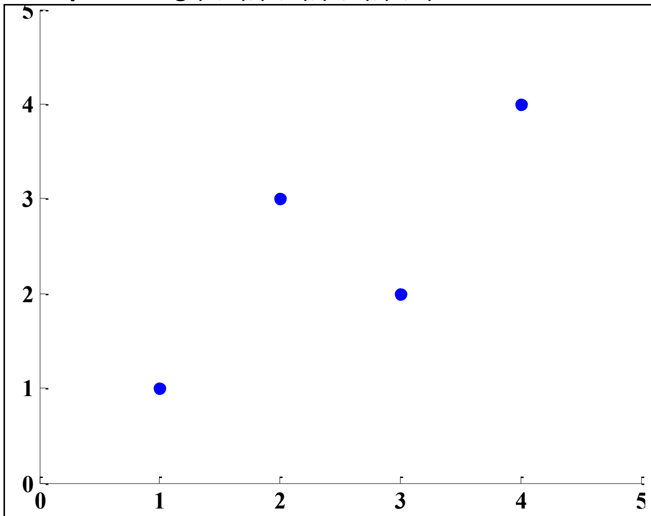
$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}} \quad SS_{xx} = \sum_{i=1}^{n} x_i^2 - n(\overline{x})^2 \quad SS_{yy} = \sum_{i=1}^{n} y_i^2 - n(\overline{y})^2 \quad SS_{xy} = \sum_{i=1}^{n} x_i y_i - n\overline{xy}$$

$$H_0: \rho = 0 \qquad H_a: \rho \neq 0 \qquad t = r\frac{\sqrt{n-2}}{\sqrt{1-r^2}}, \ df = n-2$$

$$r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = \frac{Explained\ sample\ \text{variability}}{Total\ sample\ \text{variability}}$$

MATH 2780 Chapter 3 Worksheet

Example: Using (1,1),(3,2),(2,3),(4,4).



$$\sum_{i=1}^{n} x_i =$$

$$\sum_{i=1}^{n} y_i =$$

$$\sum_{i=1}^{n} x_i^2 =$$

$$\sum_{i=1}^{n} y_i^2 =$$

$$\sum_{i=1}^{n} x_i y_i =$$

a) Compute and draw the point ($\bar{x}, \bar{y}$) on the scatterplot.

b) Draw your best guess for a least squares regression line.

c) Compute SS_{xx}, SS_{yy}, and SS_{xy}.

$$SS_{xx} =$$

$$SS_{yy} =$$

$$SS_{xy} =$$

d) Use SS_{xx}, SS_{yy}, and SS_{xy} to compute estimates of β_0 and β_1.

$$\hat{\beta}_1 =$$

$$\hat{\beta}_0 =$$

e) Write your regression equation and draw your regression Line on the Graph.

$$\hat{y} = \qquad + \qquad x$$

f) Compute SSE and s^2.

$$SSE =$$

$$s^2 =$$

$$s =$$

g) Test the hypothesis that H_0: $\beta_1 = 0$ *vs*. H_a: $\beta_1 \neq 0$. $df = n - 2$.

$t =$ $\qquad\qquad\qquad$ $t_{\alpha/2, n-2} =$

h) Compute a 95% confidence interval for β_1.

$$r =$$

i) Test the hypothesis that H_0: $\rho = 0$ *vs*. H_a: $\rho \neq 0$. $df = n - 2$.

$t =$ $\qquad\qquad\qquad$ $t_{\alpha/2, n-2} =$

j) Compute r^2.