

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

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

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

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

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 ε 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

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

$$SSE = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = SS_{yy} - \hat{\beta}_1 SS_{xy}$$

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

Slope Hypothesis Test and Confidence Interval

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

Correlation

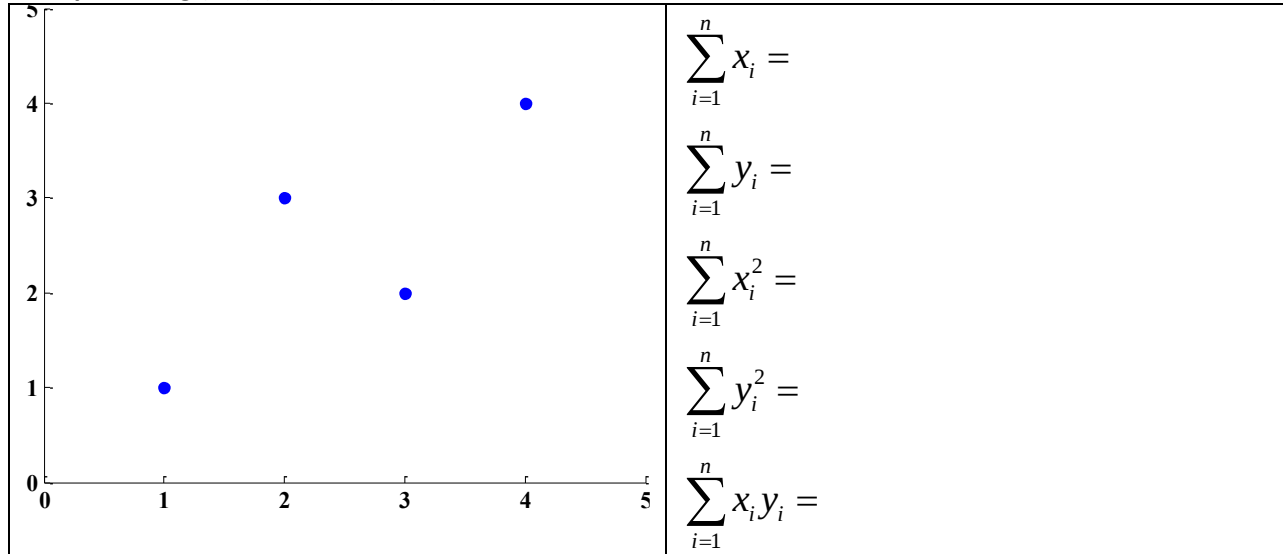
$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} \quad SS_{xx} = \sum_{i=1}^n x_i^2 - n(\bar{x})^2 \quad SS_{yy} = \sum_{i=1}^n y_i^2 - n(\bar{y})^2 \quad SS_{xy} = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}$$

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

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

MATH 2780 Chapter 3 Worksheet

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



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.

$$\hat{y} = \underline{\hspace{1cm}} + \underline{\hspace{1cm}} x$$

f) Compute SSE , s^2 , and s .

$$SSE =$$

$$s^2 =$$

$$s =$$

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

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

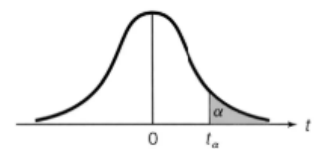
h) Compute a 95% confidence interval for β_1 .

$$L = \hspace{2cm} U =$$

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

$$r = \hspace{2cm} t = \hspace{2cm} t_{\alpha/2, n-2} =$$

j) Compute r^2 .



ν	$t_{.100}$	$t_{.050}$	$t_{.025}$
1	3.078	6.314	12.706
2	1.886	2.920	4.303
3	1.638	2.353	3.182
4	1.533	2.132	2.776
5	1.476	2.015	2.571