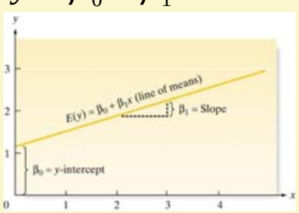
Chapter 11

Simple Linear Regression

Probabilistic Models

First Order (Straight-Line) Probabilistic Model

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



2

Probabilistic Models

5 steps of Simple Linear Regression

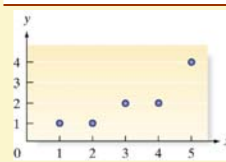
1. Hypothesize the deterministic component
2. Use sample data to estimate unknown model parameters
3. Specify probability distribution of ε, estimate standard deviation of the distribution
4. Statistically evaluate model usefulness
5. Use for prediction, estimatation, once model is useful

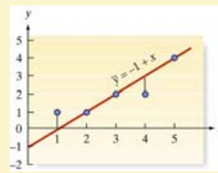
3

Fitting the Model: The Least Squares Approach

Reaction Time versus Drug Percentage

Subject	Amount of Drug x (%)	Reaction Time y (seconds)
1	1	1
2	2	1
3	3	2
4	4	2
5	5	4





4

Fitting the Model: The Least Squares Approach

Least Squares Line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ has:

- Sum of errors (SE) = 0
- Sum of Squared errors (SSE) is smallest of all straight line models

Formulas:

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

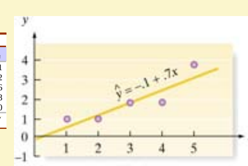
$$SS_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$SS_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

Fitting the Model: The Least Squares Approach

Preliminary Computations

	x_i	y_i	x_i^2	$x_i y_i$
	1	1	1	1
	2	1	4	2
	3	2	9	6
	4	2	16	8
	5	4	25	20
Totals	$\sum x_i = 15$	$\sum y_i = 10$	$\sum x_i^2 = 55$	$\sum x_i y_i = 37$



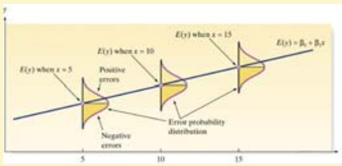
Comparing Observed and Predicted Values for the Least Squares Prediction Equation

x	y	$\hat{y} = -.1 + .7x$	$(y - \hat{y})$	$(y - \hat{y})^2$
1	1	.6	.4	.16
2	1	1.3	-.3	.09
3	2	2.0	0.0	.00
4	2	2.7	-.7	.49
5	4	3.4	.6	.36
			Sum of Errors = 0	SSE = 1.10

6

Model Assumptions

1. Mean of the probability distribution of ε is 0
2. Variance of the probability distribution of ε is constant for all values of *x*
3. Probability distribution of ε is normal
4. Values of ε are independent of each other



7

An Estimator of σ^2

Estimator of σ^2 for a straight-line model

$$s^2 = \frac{SSE}{Degrees\ of\ freedom\ for\ error} = \frac{SSE}{n-2}$$

$$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$$

$$SS_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(y_i)^2}{n}$$

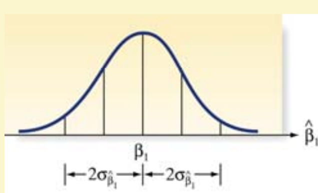
$$s = \sqrt{s^2} = Estimated\ Standard\ Error\ of\ the\ Regression\ Model$$

8

Assessing the Utility of the Model: Making Inferences about the Slope β_1

Sampling Distribution of $\hat{\beta}_1$

$$\sigma_{\hat{\beta}_1} = \frac{\sigma}{\sqrt{SS_{xx}}}$$



9

Assessing the Utility of the Model: Making Inferences about the Slope β_1

A Test of Model Utility: Simple Linear Regression

One-Tailed Test	Two-Tailed Test
H_0: $\beta_1=0$	H_0: $\beta_1=0$
H_a: $\beta_1<0$ (or H_a: $\beta_1>0$)	H_a: $\beta_1\neq0$

$$Test\ statistic: t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = \frac{\hat{\beta}_1}{s/\sqrt{SS_{xx}}}$$

Rejection region: $t< -t_\alpha$ (or $t< -t_\alpha$ when H_a: $\beta_1>0$)	Rejection region: $\|t\|> t_{\alpha/2}$

Where t_α and $t_{\alpha/2}$ are based on (n-2) degrees of freedom

Assessing the Utility of the Model: Making Inferences about the Slope β_1

A 100(1-α)% Confidence Interval for β_1

$$\hat{\beta}_1 \pm t_{\alpha/2} s_{\hat{\beta}_1}$$ where

$$s_{\hat{\beta}_1} = \frac{s}{\sqrt{SS_{xx}}}$$

11