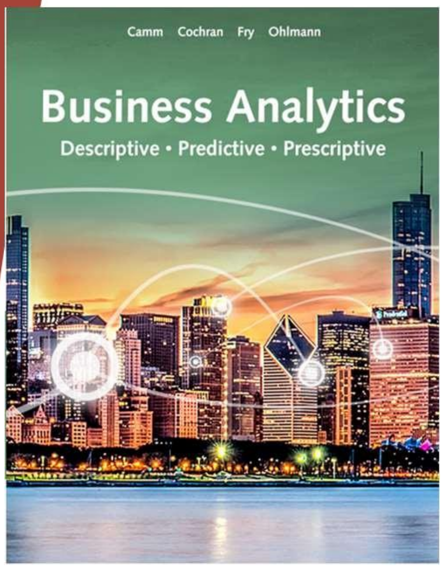


Business Analytics

Chapter 7 Linear Regression

HAMLINE UNIVERSITY SCHOOL OF BUSINESS

4

Introduction

- ☐ Managerial decisions based on:
  - ☐ Relationship between two or more variables:
- ☐ Example:
  - ☐ **Paying for ads** vs. **getting sales**
- ☐ Intuition can be useful but…
- ☐ If you have data
  - ☐ **USE regression analysis** to show how the variables are related.

5

Simple Linear Regression Model

Regression Model

Estimated Regression Equation

6

Simple Linear Regression Model

BEST FIT LINEAR REGRESSION

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

SLOPE INTERCEPT FORM

$$y = mx + b$$

Best Fit Regression Model:

- The equation that describes how y is related to x with an intercept.
  - Slope = β_1
  - Intercept = β_0
  - Y = dependent variable
  - X = independent variable

Equation of a line:

- The equation that describes how y is related to x with an intercept
  - Slope = m
  - Intercept = b
  - Y = dependent variable
  - X = independent variable

7

Simple Linear Regression Model

- The equation that describes the "True Relationship" between y and x and an error term.

SIMPLE LINEAR REGRESSION MODEL

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

Simple Linear Regression Model:

- Parameters: The characteristics of the population, β_0 and β_1.
  - Slope = β_1
  - Intercept = β_0
  - ε = error term
    - Variability in y that cannot be explained by the relationship between x and y
    - Assume ε is normally distributed with mean 0 and constant variance

Example:

How is travel time (y) of a delivery truck related to number miles traveled (x)

8

Simple Linear Regression Model

Estimated Regression Equation:

- The "true" parameter values are usually not known
  - must be estimated using sample data.
- Sample Statistics - b_0 and b_1 are calculated as estimates of β_0 and β_1
  - We plug in b_0 and b_1 and drop the error term
    - Expected value of ε is = 0

ESTIMATED SIMPLE LINEAR REGRESSION EQUATION

$$\hat{y} = b_0 + b_1 x \quad \textbf{(7.2)}$$

9

Simple Linear Regression Model

In the estimated simple linear regression equation:

$$\hat{y} = b_0 + b_1 x$$

$\hat{y}$ = Estimate for the mean value of y corresponding to a given value of x.

b_0 = Estimated y-intercept.

b_1 = Estimated slope.

- The graph of the estimated simple linear regression equation is called the estimated regression line.
- "In general, $\hat{y}$ is the point estimator of $E(y|x)$ the mean value of y for a given value of x

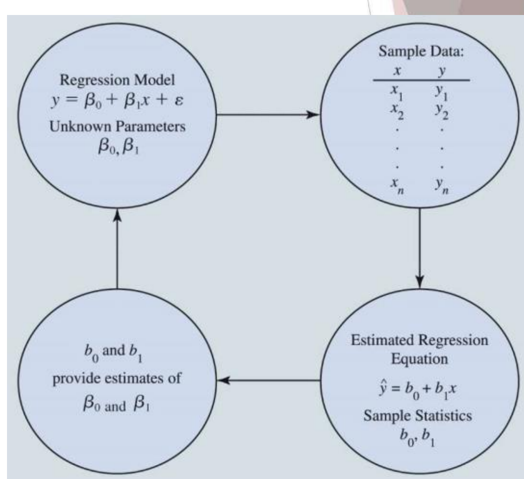
10

Simple Linear Regression Model

Example

- Butler Trucking
- How is travel time (y) of a delivery truck related to number miles traveled (x)
- We need data….

The Estimation Process in Simple Linear Regression

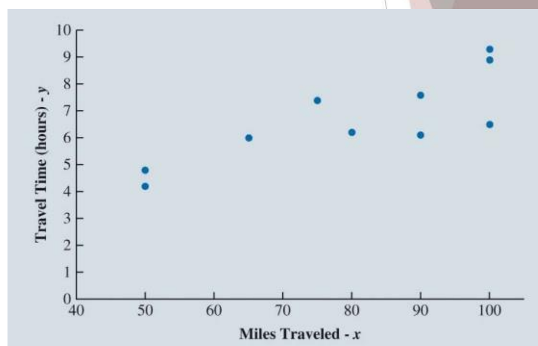


11

Least Squares Method

Miles Traveled and Travel Time for 10 Butler Trucking Company Driving Assignments

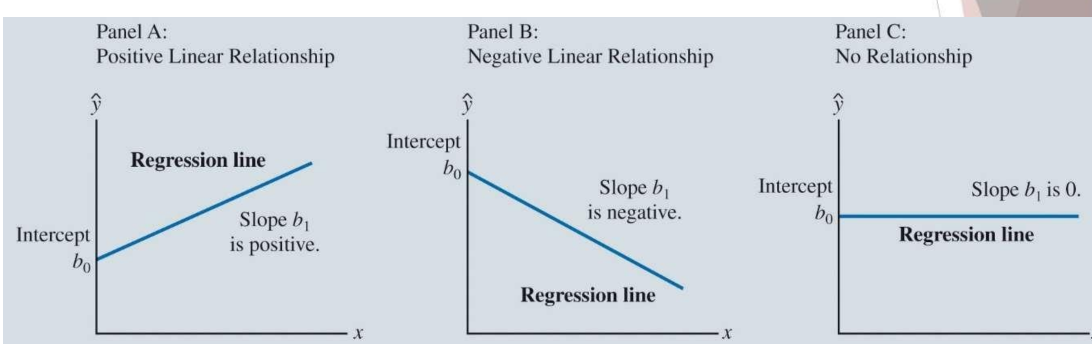
Driving Assignment *i*	*x* = Miles Traveled	*y* = Travel Time (hours)
1	100	9.3
2	50	4.8
3	50	8.9
4	100	6.5
5	50	4.2
6	80	6.2
7	75	7.4
8	65	6.0
9	90	7.6
10	90	6.1



12

Simple Linear Regression Model

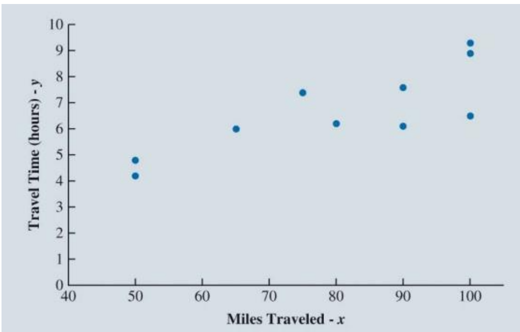
Figure 7.2: Possible Regression Lines in Simple Linear Regression

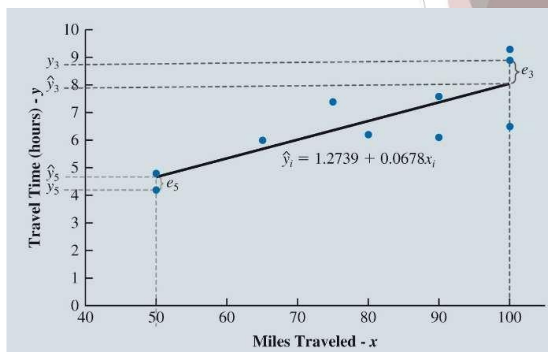


13

Least Squares Method

Scatter Chart of Miles Traveled and Travel Time for Butler Trucking Company Driving Assignments with Regression Line Superimposed





14

Least Squares Method

Least Squares Estimates of the Regression Parameters:

- For the Butler Trucking Company data
  - Estimated slope of b_1 = 0.0678
  - Estimated y-intercept of b_0= 1.2739

The estimated simple linear regression model:

$$\hat{y} = 1.2739 + 0.0678x_1$$

What do these numbers mean?

- b_1 = 0.0678 – As the trip increases 1 more mile,
  - the average travel time increases 0.0678 hours (4 Minutes)
- b_0= 1.2739 – When the trip is 0 miles,
  - the estimated travel time is 1.2 hours (76 minutes)

15

Examples

What is the expected travel time (y) of a delivery truck that travels 75 miles on deliveries?

= 1.2739 + 0.0678 (75) = 6.35 hours

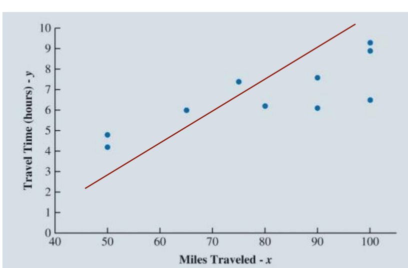
What is the expected travel time (y) of a delivery truck that travels 100 miles on deliveries?

= 1.2739 + 0.0678 (100) = 8.05 hours

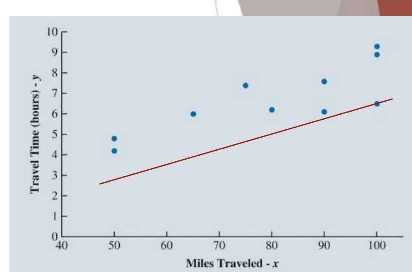
16

How do we know what line is best?

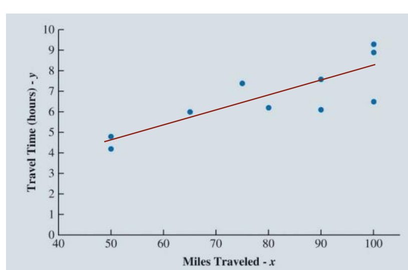
A.



B.



C.



17

Least Squares Method

Least Squares Estimates of the Regression Parameters

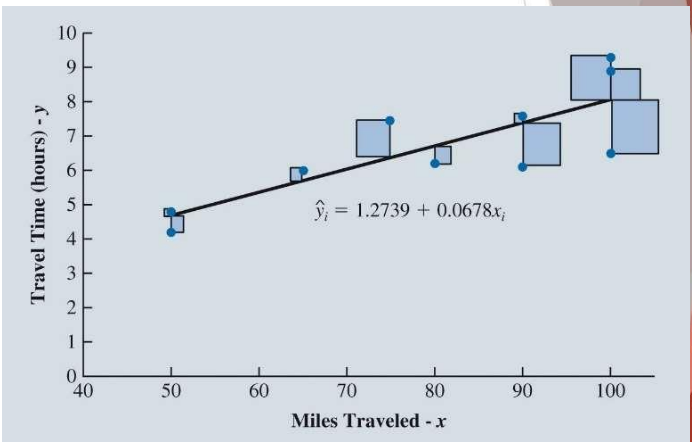
Using Excel's Chart Tools to Compute the Estimated Regression Equation

18

Least Squares Method

Least Squares Method

1. Measure the difference between each y value (y_i data point) and the estimated y value on the regression line ($y\hat{}_i$)
   Denoted as $e_i = y_i - y\hat{}_i$
   (Called Residual)
2. Square the differences
3. Add up the squared differences
4. The minimum sum of square differences is the best-fit line



19

Least Squares Method

LEAST SQUARES EQUATION

$$\min \sum_{i=1}^{n} e_i{}^2 = \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n} (y_1 - b_0 - b_1 x_1)^2 \quad \text{(7.4)}$$

where

y_i = observed value of the dependent variable for the i^{th} observation
$\hat{y}_i$ = predicted value of the dependent variable for the i^{th} observation
n = total number of observations

20

Least Squares Method

Table 7.2: Predicted Travel Time and Residuals for 10 Butler Trucking Company Driving Assignments

Driving Assignment i	x = Miles Traveled	y = Travel Time (hours)	$\hat{y}_i = b_0 + b_1 x_i$	$e_i = y_i - \hat{y}_i$	e_i^2
1	100	9.3	8.0565	1.2435	1.5463
2	50	4.8	4.6652	0.1348	0.0182
3	100	8.9	8.0565	0.8435	0.7115
4	100	6.5	8.0565	−1.5565	2.4227
5	50	4.2	4.6652	−0.4652	0.2164
6	80	6.2	6.7000	−0.5000	0.2500
7	75	7.4	6.3609	1.0391	1.0797
8	65	6.0	5.6826	0.3174	0.1007
9	90	7.6	7.3783	0.2217	0.0492
10	90	6.1	7.3783	−1.2783	1.6341
	Totals	67.0	67.0000	0.0000	8.0288

What do you notice about the actual y values (data points) and the estimated y values ($y\hat{}_i$)

What do you notice about the sum of residuals or error terms?

21

Least Squares Method – Differential Calculus

Slope Equation

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

y-Intercept Equation

$$b_0 = \bar{y} - b_1\bar{x}$$

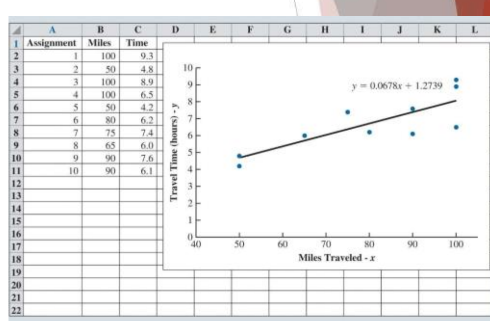
x_i = value of the independent variable for the *i*th observation.
y_i = value of the dependent variable for the *i*th observation.
$\bar{x}$ = mean value for the independent variable.
$\bar{y}$ = mean value for the dependent variable.
n = total number of observations.

22

Least Squares Method

Using Excel's Chart Tools to Compute the Estimated Regression Equation:

- After constructing a scatter chart with Excel's chart tools:
  1. Right-click on any data point and select **Add Trendline.**
  2. When the **Format Trendline** task pane appears:
     - Select **Linear** in the **Trendline Options** area.
     - Select **Display Equation on chart** in the **Trendline Options** area.



	A	B	C
1	Assignment	Miles	Time
2	1	100	9.3
3	2	50	4.8
4	3	100	8.9
5	4	100	6.5
6	5	50	4.2
7	6	80	6.2
8	7	75	7.4
9	8	65	6.0
10	9	90	7.6
11	10	90	6.1

23