

9.3

Regression

Regression

Definition

❖ Regression Equation

Given a collection of paired data, the regression equation

$$\hat{y} = b_0 + b_1x$$

algebraically describes the relationship between the two variables

❖ Regression Line

(line of best fit or least-squares line)

the graph of the regression equation

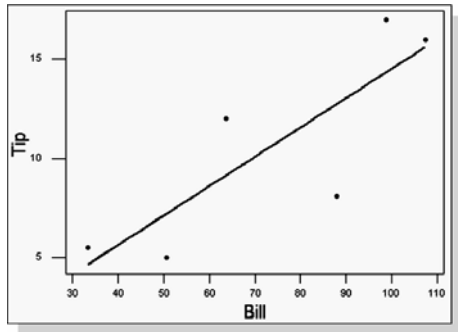
The Regression Equation

x is the independent variable
(predictor variable)

\hat{y} is the dependent variable
(response variable)

$\hat{y} = b_0 + b_1x$	$b_0 = y$ - intercept
$y = mx + b$	$b_1 =$ slope

Regression Line Plotted on Scatter Plot



Chapter 9. Section 9-1 and 9-2. Triola, Essentials of Statistics, Second Edition. Copyright 2004. Pearson/Addison Wesley

4

Assumptions

1. We are investigating only linear relationships.
2. For each x value, y is a random variable having a normal (bell-shaped) distribution. All of these y distributions have the same variance. Also, for a given value of x , the distribution of y -values has a mean that lies on the regression line. (Results are not seriously affected if departures from normal distributions and equal variances are not too extreme.)

Chapter 9. Section 9-1 and 9-2. Triola, Essentials of Statistics, Second Edition. Copyright 2004. Pearson/Addison Wesley

5

Notation for Regression Equation

	<u>Population Parameter</u>	<u>Sample Statistic</u>
y-intercept of regression equation	β_0	b_0
Slope of regression equation	β_1	b_1
Equation of the regression line	$y = \beta_0 + \beta_1 x$	$\hat{y} = b_0 + b_1 x$

Chapter 9. Section 9-1 and 9-2. Triola, Essentials of Statistics, Second Edition. Copyright 2004. Pearson/Addison Wesley

6

Formula for b_1 and b_0

Formula 9-2 $b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$ (slope)

Formula 9-3 $b_0 = \bar{y} - b_1 \bar{x}$ (y-intercept)

calculators or computers can compute these values

Chapter 9. Section 9-1 and 9-2. Triola, Essentials of Statistics, Second Edition. Copyright 2004. Pearson/Addison Wesley 7

Rounding the y-intercept b_0 and the slope b_1

- ❖ Round to three significant digits
- ❖ If you use the formulas 9-2 and 9-3, try not to round intermediate values or carry to at least six significant digits.

Chapter 9. Section 9-1 and 9-2. Triola, Essentials of Statistics, Second Edition. Copyright 2004. Pearson/Addison Wesley 8

Example: Lengths and Weights of Male Bears

x Length (in.) 53.0 67.5 72.0 72.0 73.5 68.5 73.0 37.0

y Weight (lb) 80 344 416 348 262 360 332 34

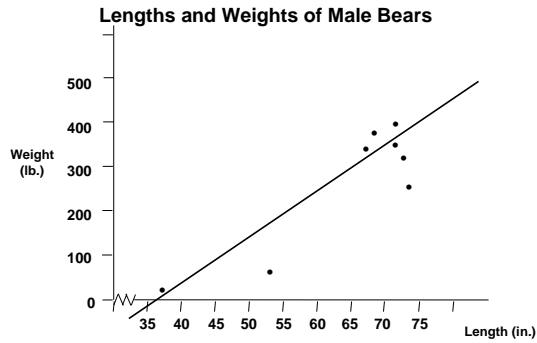
$b_0 = -352$ (rounded)

$b_1 = 9.66$ (rounded)

$\hat{y} = -352 + 9.66x$

Chapter 9. Section 9-1 and 9-2. Triola, Essentials of Statistics, Second Edition. Copyright 2004. Pearson/Addison Wesley 9

Scatter Diagram of Paired Data



Chapter 9. Section 9-1 and 9-2. Triola, Essentials of Statistics, Second Edition. Copyright 2004. Pearson/Addison Wesley

10

**The regression line
fits the sample
points best.**

Chapter 9. Section 9-1 and 9-2. Triola, Essentials of Statistics, Second Edition. Copyright 2004. Pearson/Addison Wesley

11

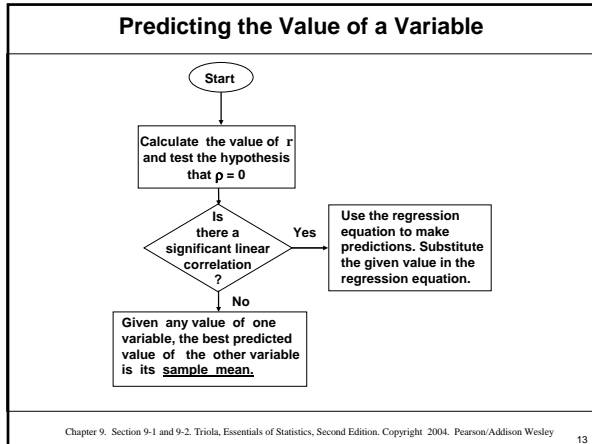
Predictions

In predicting a value of y based on some given value of x ...

1. If there is not a significant linear correlation, the best predicted y -value is \bar{y} .
2. If there is a significant linear correlation, the best predicted y -value is found by substituting the x -value into the regression equation.

Chapter 9. Section 9-1 and 9-2. Triola, Essentials of Statistics, Second Edition. Copyright 2004. Pearson/Addison Wesley

12



- ### Guidelines for Using The Regression Equation
1. If there is no significant linear correlation, don't use the regression equation to make predictions.
 2. When using the regression equation for predictions, stay within the scope of the available sample data.
 3. A regression equation based on old data is not necessarily valid now.
 4. Don't make predictions about a population that is different from the population from which the sample data was drawn.
- Chapter 9. Section 9-1 and 9-2. Triola, Essentials of Statistics, Second Edition. Copyright 2004. Pearson/Addison Wesley 14

Example: Lengths and Weights of Male Bears

x Length (in.)	53.0	67.5	72.0	72.0	73.5	68.5	73.0	37.0
y Weight (lb.)	80	344	416	348	262	360	332	34

$$\hat{y} = -352 + 9.66x$$

What is the weight of a bear that is 60 inches long?

Since the data does have a significant positive linear correlation, we can use the regression equation for prediction.

Chapter 9. Section 9-1 and 9-2. Triola, Essentials of Statistics, Second Edition. Copyright 2004. Pearson/Addison Wesley 15

Example: Lengths and Weights of Male Bears

x Length (in.) 53.0 67.5 72.0 72.0 73.5 68.5 73.0 37.0

y Weight (lb.) 80 344 416 348 262 360 332 34

$$\hat{y} = -352 + 9.66(60)$$

$$\hat{y} = 227.6 \text{ pounds}$$

Example: Lengths and Weights of Male Bears

x Length (in.) 53.0 67.5 72.0 72.0 73.5 68.5 73.0 37.0

y Weight (lb.) 80 344 416 348 262 360 332 34

A bear that is 60 inches long will weigh approximately 227.6 pounds.

Example: Lengths and Weights of Male Bears

x Length (in.) 53.0 67.5 72.0 72.0 73.5 68.5 73.0 37.0

y Weight (lb.) 80 344 416 348 262 360 332 34

If there were no significant linear correlation, to predict a weight for any length:

Example: Lengths and Weights of Male Bears

x Length (in.) 53.0 67.5 72.0 72.0 73.5 68.5 73.0 37.0
 y Weight (lb.) 80 344 416 348 262 360 332 34

If there were no significant linear correlation, to predict a weight for any length: use the average of the weights (y-values)
 $\bar{y} = 272$ lbs

Residuals and the Least-Squares Property Definitions

❖ Residual

for a sample of paired (x,y) data, the difference $(y - \hat{y})$ between an observed sample y-value and the value of \hat{y} , which is the value of y that is predicted by using the regression equation.

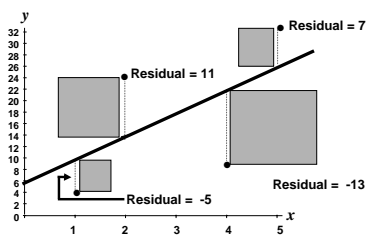
❖ Least-Squares Property

A straight line satisfies this property if the sum of the squares of the residuals is the smallest sum possible.

Residuals and the Least-Squares Property

x	1	2	4	5
y	4	24	8	32

$$\hat{y} = 5 + 4x$$



What is the best predicted size of a household that discards 0.50 lb of plastic?

Data from the Garbage Project

x Plastic (lb)	0.27	1.41	2.19	2.83	2.19	1.81	0.85	3.05
y Household	2	3	3	6	4	2	1	5

Using a calculator:

$$b_0 = 0.549$$

$$b_1 = 1.48$$

$$y = 0.549 + 1.48(0.50)$$

$$y = 1.3$$

A household that discards 0.50 lb of plastic has approximately one person.
