Chapter 7
LINEAR REGRESSION

REGRESSION EQUATION

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

where *x* is the independent variable or predictor variable

$\hat{y}$ is the predicted value

2

REGRESSION

✖ In our model, we have a slope (b_1):

+ The slope is built from the correlation and the standard deviations:

$$b_1 = r \frac{s_y}{s_x}$$

+ Our slope is always in units of *y* per unit of *x*.

3

REGRESSION

- In our model, we also have an intercept (b_0).
  - The intercept is built from the means and the slope:

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

  - Our intercept is always in units of y.

4

RESIDUALS

- The difference between the observed value and its associated predicted value is called the residual.
- To find the residuals, we always subtract the predicted value from the observed one:

  $$residual = observed - predicted = y - \hat{y}$$

5

RESIDUALS

- A negative residual means the predicted value's too big (an overestimate).
- A positive residual means the predicted value's too small (an underestimate).
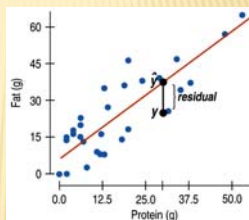
From *Stats Modeling the World* by Bock, Velleman, & De Veaux, 2010, p. 173.

6

LEAST SQUARES

× The regression equation represents the line that fits the points best by minimizing the sum of the squares of the residuals.