

Variable Types

- Explanatory variable* The variable that drives the relationship; the “independent variable.”
Graphed as the x-axis.
- Response variable* The variable changed by the explanatory variable.
Graphed as the y-axis.

Strength of Relationship

Coefficient of Correlation, r (range: -1...1)

Indicates the direction and strength of the correlation between the explanatory & response variables.

$$r = \frac{1}{n-1} \sum Z_x Z_y$$

n = number of points; Z_x, Z_y = Z scores of x- and y-values

- $r = -1$ Strong negative correlation
- $r = 0$ No correlation
- $r = +1$ Strong positive correlation

Many texts refer to r as simply the “Correlation”

Least-Squares Regression

- Derives an equation expressing the response variable as a linear function of the explanatory variable.
 - ▷ $\hat{y} = a + bx$ \hat{y} = predicted value; a = predicted value when $x = 0$; b = slope of line
 - ▷ Note this is just slope-intercept form from Algebra 1.

Residuals

- A *residual* is the difference between a particular observed value and the value predicted by the regression line.
 - ▷ $\text{residual} = y - \hat{y}$ y = measured value; \hat{y} = predicted value
- A *residual plot* is a scatter plot of the residual values
 - ▷ Can be used to determine whether a linear interpretation is appropriate for the data.
 - ▷ If the residual plot shows a distinct pattern, a linear interpretation is not appropriate.

Regression Line from r

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

$$a = \bar{y} - b\bar{x}$$

Yes, this really is just the coefficient of correlation squared.

Goodness-of-fit

Coefficient of determination, r^2 (range: 0...1)

$$r^2 = 1 - \frac{\sum(\text{residuals}^2)}{\sum(y_i - \bar{y})^2}$$

s = standard deviation; n = number of points

- ▶ A measure of how well the regression line matches the data.
0 = Horribly; 1 = Perfectly
- ▶ Also the proportion of the variations in y that can be explained by x .
Usually expressed as a percentage.

Standard deviation of the residuals, s

- ▶ The typical distance between predicted and measured values.

$$s = \sqrt{\frac{\sum(\text{residuals}^2)}{n - 2}}$$

s = standard deviation; n = number of points