BMEG 802 – Advanced Biomedical Experimental Design and Analysis

Regression

Joshua G. A. Cashaback, PhD

Recap

- Effect Size
- Power
    - Parametric
    - Numerical

Learning Objectives

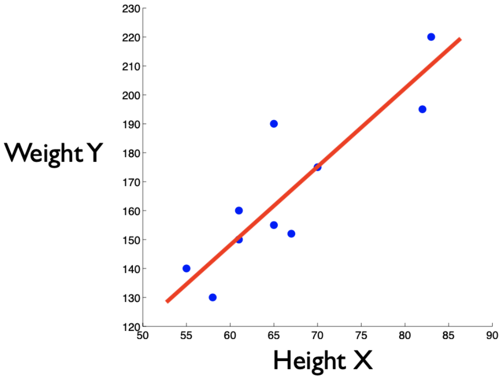
- Regression
    - Bivariate
        - Linear (Derivation)
        - Nonlinear
    - Multiple Regression
- Correlation
    - Pearson's r
    - Spearman's ρ

Regression - Bivariate

$$\hat{y}_i = B_0 + B_1 \cdot x_i$$

- want to predict $\hat{y}_i$ (e.g., height) based on x_i (e.g., weight).
    - equation of a line
        - B_1 (slope), B_1 (intercept)
    - X,Y continuous
    - relationship between X and Y

Regression - Bivariate



Height (X)	Weight (Y)
55	140
61	150
67	152
83	220
65	190
82	195
70	175
58	130
65	155
61	160

Line of best fit: $B_0 = -7.2$, $B_1 = 2.64$

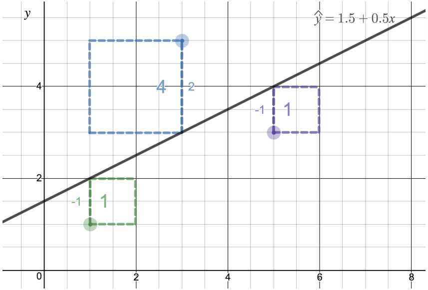
- least squares
- what do squares have to do with this???

Bivariate Regression - Least Squares

Optimization problem:

$$min(y_i - \hat{y}_i)^2$$

where $\hat{y}_i$ and y_i are respectively the predicted and actual y values.



We find B_1 (slope), B_0 (intercept) that minimize the squared differences!

Least Squares Derivation

remember, $\hat{y}_i = B_0 + B_1 \cdot x_i$ and we want to $min(\hat{y}_i - y_i)^2$

$$(y_i - \hat{y}_i)^2 = (y_i - (B_0 + B_1 \cdot x_i))^2$$
$$(\hat{y}_i - y_i)^2 = (y_i - B_0 - B_1 \cdot x_i)^2$$

In particular, we minimum the sum of squares

$$\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} (y_i - B_0 - B_1 \cdot x_i)^2$$

$$S = \sum_{i=1}^{n} (y_i - B_0 - B_1 \cdot x_i)^2$$
$$\frac{\partial S}{\partial B_0} = 0$$
$$\frac{\partial S}{\partial B_1} = 0$$

Results in two equations with two unknowns.

Least Squares Derivation - Intercept

$$\frac{\partial}{\partial B_0} \sum_{i=1}^{n} (y_i - B_0 - B_1 \cdot x_i)^2 = 0$$

We can move the sum outside:

$$\sum_{i=1}^{n} \frac{\partial}{\partial B_0} (y_i - B_0 - B_1 \cdot x_i)^2 = 0$$

Taking the derivative (note the chain rule):

$$\sum_{i=1}^{n} 2(y_i - B_0 - B_1 \cdot x_i)(-1) = 0$$

Rearranging, Taking the derivative (note the chain rule):

$$-2 \sum_{i=1}^{n} (y_i - B_0 - B_1 \cdot x_i) = 0$$

Least Squares Derivation - Slope

$$\frac{\partial}{\partial B_1}\sum_{i=1}^{n}(y_i - B_0 - B_1 \cdot x_i)^2 = 0$$

$$\sum_{i=1}^{n}\frac{\partial}{\partial B_0}(y_i - B_0 - B_1 \cdot x_i)^2 = 0$$

$$\sum_{i=1}^{n}2(y_i - B_0 - B_1 \cdot x_i)(-x_i) = 0$$

$$-2\sum_{i=1}^{n}(y_i - B_0 - B_1 \cdot x_i)(x_i) = 0$$

Two equations and two unknowns

Least Squares Derivation - Tip Interlude

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i;$$

Thus,

$$\sum_{i=1}^{n} x_i = n\bar{x}$$

Similarly,

$$\sum_{i=1}^{n} y_i = n\bar{y}$$

When summing a constant you can multiply by n. For example,

$$\sum_{i=1}^{n} \bar{x} = n\bar{x}$$

Least Squares Derivation - Tip Interlude Cont'd

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \bar{y} - \sum_{i=1}^{n} y_i \cdot \bar{x} + \sum_{i=1}^{n} \bar{x} \cdot \bar{y} \tag{1}$$

$$= \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \bar{y} - \sum_{i=1}^{n} y_i \cdot \bar{x} + \sum_{i=1}^{n} \bar{x} \cdot \bar{y} \tag{2}$$

$$= \sum_{i=1}^{n} x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y} - n \cdot \bar{x} \cdot \bar{y} + n \cdot \bar{x} \cdot \bar{y} \tag{3}$$

$$= \sum_{i=1}^{n} x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y} \tag{4}$$

Least Squares Derivation - Tip Interlude Cont'd

$$
\begin{align}
\sum_{i=1}^{n}(x_i - \bar{x})^2 &= \sum_{i=1}^{n}(x_i^2 - 2 \cdot x_i \cdot \bar{x} + \bar{x}^2) \tag{5} \\
&= \sum_{i=1}^{n} x_i^2 - 2 \cdot \bar{x} \sum_{i=1}^{n} \cdot x_i + \sum_{i=1}^{n} \bar{x}^2 \tag{6} \\
&= \sum_{i=1}^{n} x_i^2 - 2 \cdot \bar{x} \cdot n \cdot \bar{x} + n\bar{x}^2 \tag{7} \\
&= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \tag{8}
\end{align}
$$

Least Squares Derivation - Intercept

Let's do some algebra on our previous equation:

$$-2\sum_{i=1}^{n}(y_i - B_0 - B_1 \cdot x_i) = 0$$

Divide both sides by -2,

$$\sum_{i=1}^{n}(y_i - B_0 - B_1 \cdot x_i) = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} B_0 - B_1 \cdot \sum_{i=1}^{n} x_i = 0$$

$$n \cdot \bar{y} - n \cdot B_0 - B_1 \cdot n \cdot \bar{x} = 0$$

$$\bar{y} - B_0 - B_1 \cdot \bar{x} = 0$$

$$B_0 = \bar{y} - B_1 \cdot \bar{x}$$

Least Squares Derivation - Slope

$$-2\sum_{i=1}^{n}(y_i - B_0 - B_1 \cdot x_i)(x_i) = 0$$

$$\sum_{i=1}^{n}(x_i \cdot y_i - B_0 \cdot x_i - B_1 \cdot x_i^2) = 0$$

$$\sum_{i=1}^{n} x_i \cdot y_i - B_0 \sum_{i=1}^{n} x_i - B_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i \cdot y_i - B_0 \sum_{i=1}^{n} x_i - B_1 \sum_{i=1}^{n} x_i^2 = 0$$

Substitute in B_0

$$\sum_{i=1}^{n} x_i \cdot y_i - (\bar{y} - B_1 \cdot \bar{x})n \cdot \bar{x} - B_1 \sum_{i=1}^{n} x_i^2 = 0$$

Least Squares Derivation - Slope Cont'd

$$\sum_{i=1}^{n} x_i \cdot y_i - (\bar{y} - B_1 \cdot \bar{x}) n \cdot \bar{x} - B_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y} + n \cdot B_1 \cdot \bar{x}^2 - B_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y} = B_1 \sum_{i=1}^{n} x_i^2 - n \cdot B_1 \cdot \bar{x}^2$$

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = B_1 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$B_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Derivation Summary

$$B_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$B_0 = \bar{y} - B_1 \cdot \bar{x}$$

Lets plug in the example above

do height weight example.

Next Week

Analysis of Variance (ANOVA) - between (one-way)