

Quantitative Reasoning – LAB 413

Regression & Residuals

Week 4 Recap ■ Lab 3

Instructor: **Subin Na**

September 17, 2025

Today's Plan

1. Quick Recap

- Regression & Residuals
- Hands-on Practice — fitting regression lines in R
- Exploring Residuals — plots & interpretation
- Cautions in Regression — outliers, extrapolation, omitted variables

2. RStudio Practice

01 Regression Basics

Key Concepts

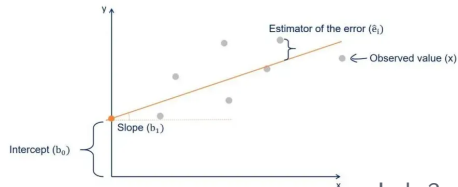
Draw the best-fitting line through the data to predict Y from X .

$$Y_i = \underbrace{\alpha}_{\text{intercept}} + \underbrace{\beta}_{\text{slope}} \cdot X_i + \underbrace{\epsilon_i}_{\text{error term}}$$

- **Intercept (α):** value of Y when $X = 0$ (starting point).
- **Slope (β):** change in Y per one-unit change in X .
- **Residuals:** prediction errors.

Linear regression model. Geometrical representation

$$\hat{y}_i = b_0 + b_1 x_i$$



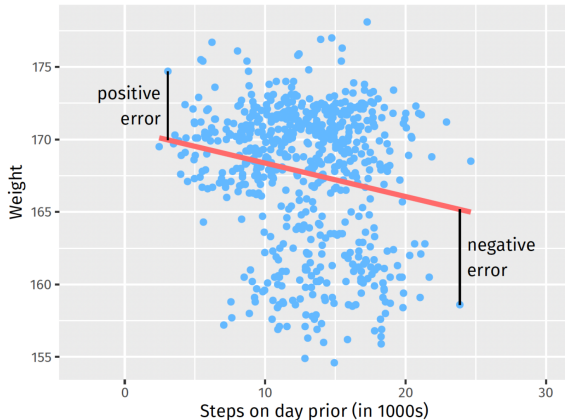
02 Least Squares Regression

Key Concepts

The best-fit line **minimizes** the vertical distances between the points and the line.

- We minimize the **Sum of Squared Residuals (SSR)**.
- The regression line is the one with the **smallest total error**.

Weight and Steps



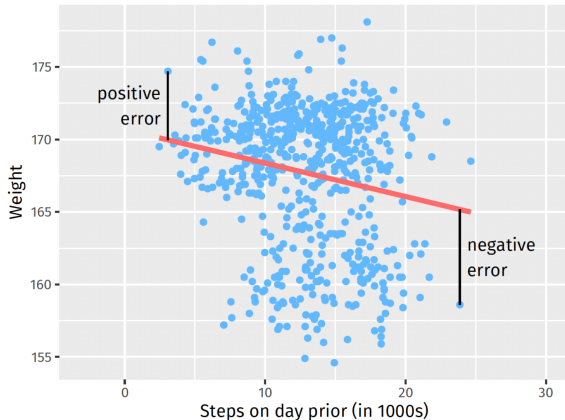
03 Residuals

Key Concepts

Residual = actual Y – predicted Y .

- Show how far the data deviate from the regression line.
- Average of residuals = 0 (a special property of the least-squares line).

Weight and Steps



Key Concepts

R^2 measures the proportion of variance in Y that is **explained** by X .

- Ranges from 0 to 1: closer to 1 means the line explains more of the variation in Y .
- Closer to 0 means the predictor X explains little of the variation.

05 Cautions (1)

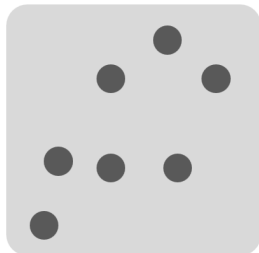
Key Concepts

- **Outliers** can strongly affect a regression.
- Avoid **extrapolation** — predicting far outside the data range.

Interpolation



Extrapolation



05 Cautions (2)

Key Concepts

- Beware of **omitted variables**.
- **Simpson's paradox**: when the trend in the whole dataset is the *opposite* of the trend within the smaller groups that make it up.

Remember: correlation $\neq$ causation

RStudio Practice

Correlation, Regression & Residuals in R

See Lab3.Rmd / lab3_master file.R