Simple Linear Regression

Ravleen Bajaj

Why Regression?

- Regression relates an **explanatory variable (X)** to a **response variable (Y)**.
- Example: Are changes in BMI associated with changes in systolic blood pressure?

Simple Linear Regression Model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

- Y_i: Systolic blood pressure (response).
- x_i: BMI (predictor).
- β_0: Intercept.
- β_1: Slope – mean change in BP for a 1-unit increase in BMI.
- ε_i: Random error term.

Interpretation of Coefficients

$$\mu_i = 95 + 1.5x_i$$

- **Intercept (95):** Mean BP if BMI = 0 (not realistic).
- **Slope (1.5):** Each 1-unit increase in BMI is associated with a 1.5 mmHg increase in mean systolic BP.

Regression Line

True line:

$$\mu_i = 95 + 1.5x_i$$

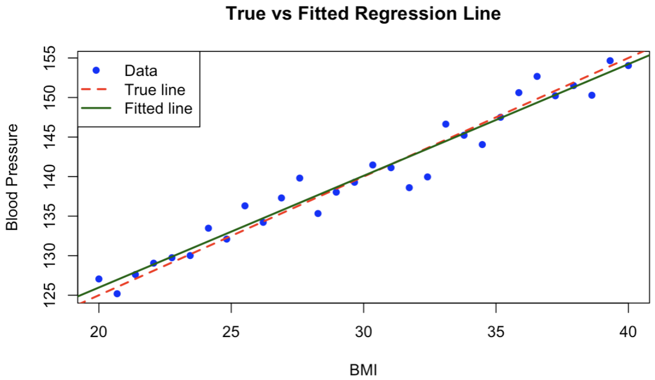
Estimated/Fitted line (from simulated sample):

$$\hat{\mu}_i = 97.72 + 1.41x_i$$

Simulated Dataset (First 10 Patients)

BMI (kg/m^2)	BP (mmHg)
20.0	129.97
20.7	124.65
21.4	133.55
22.1	143.33
22.8	126.80
23.4	127.83
24.1	146.99
24.8	139.92
25.5	128.58
26.2	139.74

True vs Fitted Regression Line



- Red dashed line: true regression line.
- Green solid line: fitted regression line.
- Blue points: simulated data.

β_0 and β_1

- OLS estimates minimize the sum of squared errors.

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

β_0 and β_1

- OLS estimates minimize the sum of squared errors.

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

- $\hat{\beta}_1$: estimated slope
- $\hat{\beta}_0$: estimated intercept

Properties of OLS Estimator

- **Unbiasedness:** $E[\tilde{\beta}_0] = \beta_0, \; E[\tilde{\beta}_1] = \beta_1.$
- **Efficiency:** OLS gives the *Best Linear Unbiased Estimator (BLUE)* where "best" implies minimum variance.

Summary & Questions

- Simple linear regression links a predictor to an outcome.
- Interpretation: intercept, slope (mean change).
- OLS formulas for $\hat{\beta}_0$ and $\hat{\beta}_1$.
- Properties $\tilde{\beta}_0, \tilde{\beta}_1$: unbiased, efficient.

Questions?