Straight-line Regression Analysis

Yi Zhou

November 24, 2015

1 STRAIGHT-LINE MODEL

1. Association vs Causality
2. Statistical model vs Deterministic model
3. Assumptions
   - Existence
   - Independence
   - Linearity
   - Homoscedasticity ($\sigma^2_{Y|X_i} = \sigma^2_{Y|X_j}$)
   - Normal Distribution

1.1 MATHEMATICAL MODEL

Firstly, in the mathematical context, given that x and y are mathematical vectors, we have equation

$$y = \beta_0 + \beta_1 x \tag{1.1}$$

where β_0 is called *y-intercept*, β_1 is called *slope*.

1.2 Statistical Model

In the statistical context, considering X the only one independent variable (predictor) and Y the only one dependent variable (response),

- $Y = \beta_0 + \beta_1 X + E$
- $\mu_{Y|X} = \beta_0 + \beta_1 X$
- $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$
- $E = Y - (\beta_0 + \beta_1 X)$
- $E = Y - \mu_{Y|X}$
- Residual: $\hat{E} = Y - \hat{Y} = Y - (\hat{\beta}_0 + \hat{\beta}_1 X)$
- $SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$

Parameters to be estimated: β_0, β_1, specific values
Methods: Least-square method
Variables of interest:

- X: fixed and observable
- Y: random and observable
- E: random but unobservable, $E \sim N(0, \sigma^2)$

Then, another parameter of interest: σ^2.

1.3 Statistical Inferences

1. Estimation
2. Statistical hypotheses

1.3.1 σ^2

The estimate of σ^2: $S^2_{Y|X} =$
The definition of SSE
Another convenient computational formula concerning $S^2_X, S^2_Y, \hat{\beta}^2$ is $S^2_{Y|X} =$
where $S^2_X =$ $S^2_Y =$

1.3.2 β_0, β_1

Least-square method: $\hat{\beta}_0 = \dfrac{\partial SSE}{\partial \beta_0}$, $\hat{\beta}_1 = \dfrac{\partial SSE}{\partial \beta_1}$
Centralized transformation: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1(X - \bar{X})$

- β_1
  - $\mu_{\hat{\beta}_1}, \sigma^2_{\hat{\beta}_1}$,
  - $H_0 : \beta_1 = \beta_1^{(0)}$
  - $T_1 \sim t_{n-2}$
  - CI
  - Interpretation of the test
- β_0
  - $\mu_{\hat{\beta}_0}, \sigma^2_{\hat{\beta}_0}$,
  - $H_0 : \beta_0 = \beta_0^{(0)}$
  - $T_0 \sim t_{n-2}$
  - CI
  - Interpretation of the test

1.3.3 $\mu_{Y|X} = \beta_0 + \beta_1 X$

Given X_0,
$H_0 : \mu_{Y|X_0} = \mu^{(0)}_{Y|X_0}$
$T \sim t_{(n-2)}$
CI
Confidence band, prediction band

2 Correlation Coefficient

- Population correlation coefficient: ρ
- Sample correlation coefficient: $r = \dfrac{S_X}{S_Y}\hat{\beta}_1$
- $r^2 = \dfrac{SSY - SSE}{SSY}$

Mathematical properties
Interpretation and measurement of r

2.1 Test Hypotheses

1. $H_0 : \rho = 0 (\beta_1 = 0)$
   - $T \sim t_{n-2}$
2. $H_0 : \rho = \rho_0, \rho_0 \neq 0$
   - Fisher's Z transformation
   - $Z \sim N(0,1)$
   - CI
3. $H_0 : \rho_1 = \rho_2$
   - Fisher's Z transformation, Z_1, Z_2
   - $Z \sim N(0,1)$
   - CI
4. $H_0 : \rho_{12} = \rho_{13}$, not independent
   - $Z \sim N(0,1)$