Applied Regression Analysis and Other Multivariable Methods

Chapter 5.7 - 6

Yi Zhou

2015-11-11

Outline

5-7 **Inferences** concerning the slope and intercept

- The origin of *inference*:
  - To assess the fitted line
  - To consider the uncertainties from the sample
- CIs and test statistical hypotheses for the unknown parameters
- **Assumption**: the random variable Y has a normal distribution at each fixed value of X.
- Given $\mu_{Y|X} = \beta_0 + \beta_1 X$, $\hat{\beta}_0$ and $\hat{\beta}_1$ are normally distributed with respective means β_0 and β_1.
- *Theorem*: linear functions of independent normal distributed observations are themselves normally distributed.
- $H_0 : \beta_1 = \beta_1^{(0)}, H_0 : \beta_0 = \beta_0^{(0)}$, using t distribution with $(n-2)$ d.f. (see eq.(5.12), eq.(5.13)).

5-8 **Interpretations** of test for slope and intercept

- Assumption: normality, independence, variance homogeneity hold.
- Zero slope: $H_0 : \beta_1 = 0$
  - Accept: X provides little or no help;
    The relationship is not linear.
  - Reject: X provides significant information;
    There is a definite linear component and might be a curvilinear term.
- Zero intercept: $H_1 : \beta_0 = 0$

5-9 Inferences concerning the regression line

- $\mu_{Y|X_0} = \mu^{(0)}_{Y|X_0}$
- Confidence intervals for $\mu_{Y|X_0}$, confidence band.
- A better estimation of $\mu_{(Y|X)}$ will be done near $\bar{X}$
- Prediction interval(PI) for $\hat{Y}_{X_0}$, prediction band.

Multivariate situation

$$\mu_{Y|X} = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

- Ordinary Least Square methods: $\dfrac{\partial Q}{\partial \beta_i}$
- $H_0 : \beta_i = \beta_i^{(0)}$, using t distribution with $(n - p - 1)$ d.f.
- $H_0 : \beta_i = 0$
- $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$, F test, $F_{p,n-p-1}$.
- $H_0 : \beta_1 = \beta_2 = \cdots = \beta_r = 0$, part of X_i, F test, $F_{r,n-p-1}$

6-1.2 Definition of r

- The sample correlation coefficient: r
- The population correlation coefficient: ρ
- Mathematical properties of r is on pp.81
- Interpretation of r
- positive / negative quadrants

6-3 Bivariate Normal Distribution

- A bell-shaped surface
- The conditional distribution of Y at X, that is the distribution of Y for fixed X, is univariate normal.
- Joint p.d.f

$$P(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\{\frac{-z}{2(1-\rho^2)}\}$$

with $z = \frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}$

and $\rho = cor(x_1, x_2) = \frac{V_{12}}{\sigma_1\sigma_2}$

- Marginal probability:

$P(x_1) = \int_{-\infty}^{\infty} P(x_1, x_2)dx_2 = \frac{1}{\sigma_1\sqrt{2\pi}} exp\{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\}$

$P(x_2) = \int_{-\infty}^{\infty} P(x_1, x_2)dx_1 = \frac{1}{\sigma_2\sqrt{2\pi}} exp\{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\}$

6-4 r and the strength of the straight-line relationship

- SSY and SSE

$$r^2 = \frac{SSY - SSE}{SSY}$$

- If $SSE = 0$, then $r^2 = 1$
- If $r^2 = 0$, then it implies the absence of linear relationship.
- **Misconceptions** of r^2
  - Not a measure of magnitude of the slope
  - Not a measure of the appropriateness of the straight-line model

6-6.7 Test of hypotheses and CIs for the correlation coefficient

- $H_0 : \rho = 0$
- $H_0 : \rho = \rho_0, \ \rho_0 \neq 0$
- CI
- $H_0 : \rho_1 = \rho_2$, independent random samples
- $H_0 : \rho_{12} = \rho_{13}$, not independent