Inference for Multivariate Regression

GR 5205 / GU 4205
Section 3

Columbia University
Xiaofei Shi

Recall our road map...

x - predictor (random) variables **Y - response random variable**

- Build your model:

  1) relationship: $Y = x\beta + \epsilon, \qquad \mathbb{E}[\epsilon] = 0,\ \mathrm{Var}[\epsilon] = \sigma^2 I_n.$

  2) preference: choose $\widehat{\beta}$ to minimize mean squared error

- Estimate your model parameters: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

  1) using observed data to express your preference: $Q = \sum_{i=1}^{n} (y_i - x_i^\top \beta)^2$

  2) get parameters estimation for your model:

- Understand your model:

  1) properties of estimations:

  2) predictions: $\widehat{Y}_0 = x_0^\top \widehat{\beta}, \qquad \widehat{y}_0 = x_0^\top b.$

Back to optimize $Q(\beta) = \|y - x\beta\|^2$

Taking partial derivatives wrt a vector:

$$\widehat{\beta} = (\boldsymbol{x}^\top \boldsymbol{x})^{-1} \boldsymbol{x}^\top \boldsymbol{Y}$$

$$\widehat{\boldsymbol{\sigma}}^2_{LS} = \frac{1}{n-p} ||\boldsymbol{Y} - \widehat{\boldsymbol{Y}}||^2 = \frac{1}{n-p} ||\boldsymbol{Y} - \boldsymbol{x}\widehat{\beta}||^2$$

Our model estimation

Plugging $Y = x\beta + \epsilon$ into the estimator $\widehat{\beta}$,

$$\begin{aligned}\widehat{\beta} &= (x^\top x)^{-1}x^\top Y = (x^\top x)^{-1}x^\top x\beta + (x^\top x)^{-1}x^\top \epsilon \\ &= \beta + (x^\top x)^{-1}x^\top \epsilon.\end{aligned}$$

- When only the expectation and variance of ϵ is available, we have

$$\mathbb{E}[\widehat{\beta}] = \beta, \qquad \mathrm{Var} = \sigma^2(x^\top x)^{-1}.$$

- When information about the distribution of ϵ, for example ϵ $N(0, \sigma^2 I_n)$, is known, we have

$$\widehat{\beta} \sim N(\beta, \sigma^2(x^\top x)^{-1})$$

Very useful tricks for variance and covariance of random vectors

- $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.
- For vectors a, b have the same size as random vectors X, Y respectively,

$$\mathrm{Cov}(a^\top X, b^\top Y) = a^\top \mathrm{Cov}(X, Y)(b^\top)^\top = a^\top \mathrm{Cov}(X, Y) b.$$

- Same relationship holds if substitute a, b with matrices.

Now let’s see how to do the calculation

Estimation for σ^2

- Degree of freedom: n-p
- $\widehat{\sigma}^2_{LS} = \frac{1}{n-p}\|Y - \widehat{Y}\|^2 = \frac{1}{n-p}\|Y - x\widehat{\beta}\|^2$

In-sample estimation given by the regression line: $\widehat{Y} = x\widehat{\beta} = x(x^\top x)^{-1}x^\top Y$

- Notice that this is for the whole samples, i.e. $\widehat{Y}$ is an n-dim random vector.
- $x(x^\top x)^{-1}x^\top$ is a projection matrix.
- $\mathbb{E}\left[\widehat{Y}\right] = x\beta, \qquad \mathrm{Var}\left[\widehat{Y}\right] = \sigma^2 x(x^\top x)^{-1}x^\top.$
- When focusing on a specific point at x_0, the in-sample point estimation is

$$\widehat{Y}_0 = x_0^\top \widehat{\beta}, \qquad \text{with} \qquad \mathbb{E}[\widehat{Y}_0] = x_0^\top \beta, \quad \mathrm{Var}\left[\widehat{Y}_0\right] = \sigma^2 x_0^\top (x^\top x)^{-1} x_0.$$

Out-of-sample estimation given by the regression line: $\widehat{Y}_0 = x_0^\top \widehat{\beta}$

- When focusing on a specific point at x_0, the out-of-sample point prediction is

$$\widehat{Y}_0 = x_0^\top \widehat{\beta}, \qquad \text{with} \qquad \mathbb{E}[\widehat{Y}_0] = x_0^\top \beta, \quad \mathrm{Var}\left[\widehat{Y}_0\right] = \sigma^2 + \sigma^2 x_0^\top (x^\top x)^{-1} x_0.$$

With Gaussian Assumption:

$$\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$$

- We have the distribution for the estimators

$$\widehat{\beta} = (x^\top x)^{-1} x^\top Y \sim N(\beta, \sigma^2 x (x^\top x)^{-1} x^\top)$$

$$\widehat{\sigma}^2_{LS} = \frac{1}{n-p} \left\| Y - \widehat{Y} \right\|^2 \sim \frac{\sigma^2}{n-p} \chi^2(n-p)$$

$$\widehat{\beta} \perp \widehat{\sigma}^2$$

- This is also the starting point of creating confidence interval and hypothesis testing!