

# 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:  $Y = x\beta + \epsilon$ ,  $\mathbb{E}[\epsilon] = 0$ ,  $\text{Var}[\epsilon] = \sigma^2 I_n$ .

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

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

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:  $\hat{Y}_0 = x_0^\top \hat{\beta}$ ,  $\hat{y}_0 = x_0^\top b$ .



## Back to optimize

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

Taking partial derivatives wrt a vector:

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

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



# Our model estimation

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

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

- When only the expectation and variance of  $\epsilon$  is available, we have

$$\mathbb{E}[\hat{\beta}] = \beta, \quad \text{Var} = \sigma^2 (x^\top x)^{-1}.$$

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

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



## Very useful tricks for variance and covariance of random vectors

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

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

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




**Now let's see how to do the calculation**



# Estimation for $\sigma^2$

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




## In-sample estimation given by the regression line:

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

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

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





## Out-of-sample estimation given by the regression line:

$$\hat{Y}_0 = x_0^\top \hat{\beta}$$

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

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



## With Gaussian Assumption:

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

- We have the distribution for the estimators

$$\begin{aligned}\hat{\beta} &= (x^\top x)^{-1} x^\top Y \sim N(\beta, \sigma^2 (x^\top x)^{-1}) \\ \hat{\sigma}_{LS}^2 &= \frac{1}{n-p} \|Y - \hat{Y}\|^2 \sim \frac{\sigma^2}{n-p} \chi^2(n-p) \\ \hat{\beta} &\perp \hat{\sigma}^2\end{aligned}$$

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