

## 38 Lecture 38: April 28

### Last time

- Theoretical background of linear model

### Today

- Course evaluation (7/17)
- Typo in HW3\_keys
- Theoretical background of linear models cont.
  - Projections
  - Geometry of least squares solution
  - Multivariate normal distribution
  - Independence and Cochran's theorem

### Additional reference

[Course notes](#) by Dr. Hua Zhou

“A Primer on Linear Models” by Dr. John F. Monahan

### Projection

- A matrix  $\mathbf{P} \in \mathbb{R}^{m \times n}$  is a projection onto a vector space  $\mathcal{V}$  if and only if
  1.  $\mathbf{P}$  is idempotent
  2.  $\mathbf{P}\mathbf{x} \in \mathcal{V}$  for any  $\mathbf{x} \in \mathbb{R}^n$
  3.  $\mathbf{P}\mathbf{z} = \mathbf{z}$  for any  $\mathbf{z} \in \mathcal{V}$ .
- Any idempotent matrix  $\mathbf{P}$  is a projection onto its own column space  $\mathcal{C}(\mathbf{P})$ .

*Proof:*

- $\mathbf{A}\mathbf{A}^+$  is a projection onto the column space  $\mathcal{C}(\mathbf{A})$ .

*Proof:*

- The projection matrix

$$\mathbf{P}_{\mathbf{X}} = \underset{n \times n}{\mathbf{X}} \underset{n \times p}{(\mathbf{X}^T \mathbf{X})^{-1}} \underset{p \times p}{\mathbf{X}^T} \underset{p \times n}{\mathbf{X}}$$

is unique.

*Proof:*

- Proposition: Let  $\mathbf{X}, \mathbf{A}, \mathbf{B}$  be matrices, then  $\mathbf{X}^T \mathbf{X} \mathbf{A} = \mathbf{X}^T \mathbf{X} \mathbf{B}$  if and only if  $\mathbf{X} \mathbf{A} = \mathbf{X} \mathbf{B}$ .

*Proof:*

- $\mathbf{P}_{\mathbf{X}} \mathbf{X} = \mathbf{X}$   

$$\begin{matrix} n \times n & n \times p & n \times p \end{matrix}$$
*Proof:*

- Predicted values  $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{b}}_{ls}$  are invariant to choice of solution to the normal equation, where

$$\hat{\mathbf{b}}_{ls} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

is not necessarily unique.

*Proof:*

- Start with  $\mathbf{P}_{\mathbf{X}} \mathbf{X} = \mathbf{X}$ , we have  $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} = \mathbf{X}$ . Therefore,  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$  is a generalized inverse of  $\mathbf{X}$  which is sometimes called the least-squares inverse. And  $\mathbf{P}_{\mathbf{X}}$  is a projection onto  $\mathcal{C}(\mathbf{X})$ .

## Geometry of least squares

- $\mathbf{P}_{\mathbf{X}}^2 = \mathbf{P}_{\mathbf{X}}$  and  $\hat{\mathbf{Y}} = \mathbf{P}_{\mathbf{X}} \mathbf{Y}$  is unique.
- Recall the column space of  $\mathbf{X}$  is  $\mathcal{C}(\mathbf{X}) = \left\{ \mathbf{y}_{n \times 1} : \mathbf{y} = \mathbf{X} \mathbf{b}_{p \times 1} \text{ for some } \mathbf{b} \right\}$
- The vector in  $\mathcal{C}(\mathbf{X})$  that is closest in terms of squared norm ( $L_2$  norm:  $\|\mathbf{a} - \mathbf{b}\|_2 = \sqrt{(\mathbf{a} - \mathbf{b})^T (\mathbf{a} - \mathbf{b})}$ ) to  $\mathbf{Y}$  is given by  $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{b}}_{ls} = \mathbf{P}_{\mathbf{X}} \mathbf{Y}$ .  
*Proof:*
- $\hat{\mathbf{Y}} \in \mathcal{C}(\mathbf{X})$
- $\hat{\mathbf{e}}_{n \times 1} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{P}_{\mathbf{X}}) \mathbf{Y} \in \mathcal{N}(\mathbf{X}^T)$  where  $\mathcal{N}(\mathbf{X}^T) = \left\{ \mathbf{v}_{n \times 1} : \mathbf{X}^T \mathbf{v} = \mathbf{0} \right\}$  is the null space of  $\mathbf{X}^T$ .  
*Proof:*

## Normal distribution in scalar case

- A random variable  $Z$  has a standard normal distribution, denoted  $Z \sim \mathcal{N}(0, 1)$ , if

$$F_Z(t) = \Pr(Z \leq t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz,$$

or equivalently  $Z$  has density

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

or equivalently,  $Z$  has moment generating function (mgf)

$$m_Z(t) = \mathbb{E}(e^{tZ}) = e^{t^2/2}, \quad -\infty < z < \infty$$

- Non-standard normal random variable

- Definition 1: A random variable  $X$  has normal distribution with mean  $\mu$  and variance  $\sigma^2$ , denoted  $X \sim \mathcal{N}(\mu, \sigma^2)$ , if

$$X = \mu + \sigma Z$$

where  $Z \sim \mathcal{N}(0, 1)$

- Definition 2:  $X \sim \mathcal{N}(\mu, \sigma^2)$  if

$$m_X(t) = \mathbb{E}(e^{tX}) = e^{t\mu + \sigma^2 t^2/2}, \quad -\infty < t < \infty$$

- In both definitions,  $\sigma^2 = 0$  is allowed. If  $\sigma^2 > 0$ , it has a density

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

### Multivariate normal distribution

- The standard multivariate normal is a vector of independent standard normals, denoted  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p)$ . The joint density is

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^{p/2}} e^{-\sum_{i=1}^p z_i^2/2}.$$

The mgf is

$$m_{\mathbf{Z}}(\mathbf{t}) = \prod_{i=1}^p m_{Z_i}(t_i) = \prod_{i=1}^p e^{t_i^2/2} = e^{\mathbf{t}^T \mathbf{t}/2}.$$

- Consider the affine transformation  $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$  where  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p)$ .  $\mathbf{X}$  has mean and variance

$$\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}, \quad \text{Var}(\mathbf{X}) = \mathbf{A}\mathbf{A}^T$$

and the moment generating function is

$$m_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}(e^{\mathbf{t}^T(\boldsymbol{\mu} + \mathbf{A}\mathbf{Z})}) = e^{\mathbf{t}^T \boldsymbol{\mu}} \mathbb{E}(e^{\mathbf{t}^T \mathbf{A}\mathbf{Z}}) = e^{\mathbf{t}^T \boldsymbol{\mu} + \mathbf{t}^T \mathbf{A}\mathbf{A}^T \mathbf{t}/2}.$$

- $\mathbf{X} \in \mathbb{R}^p$  has a multivariate normal distribution with mean  $\boldsymbol{\mu} \in \mathbb{R}^p$  and covariance  $\mathbf{V} \in \mathbb{R}^{p \times p}$ ,  $\mathbf{V} \succeq_{p.s.d.} \mathbf{0}$ , denoted  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ , if its mgf takes the form

$$m_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}^T \boldsymbol{\mu} + \mathbf{t}^T \mathbf{V} \mathbf{t}/2}, \quad \mathbf{t} \in \mathbb{R}^p$$

- if  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$  and  $\mathbf{V}$  is non-singular, then
  - \*  $\mathbf{V} = \mathbf{A}\mathbf{A}^T$  for some non-singular  $\mathbf{A}$
  - \*  $\mathbf{A}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p)$
  - \* The density of  $\mathbf{X}$  is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\mathbf{V}|^{1/2}} e^{-(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu})/2}.$$

- (Any affine transform of normal is normal) If  $\mathbf{X} \in \mathbb{R}^p$ ,  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$  and  $\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{X}$ , where  $\mathbf{a} \in \mathbb{R}^q$  and  $\mathbf{B} \in \mathbb{R}^{q \times p}$ , then  $\mathbf{Y} \sim \mathcal{N}(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\mathbf{V}\mathbf{B}^T)$ .
- (Marginal of normal is normal) If  $\mathbf{X} \in \mathbb{R}^p$ ,  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ , then any subvector of  $\mathbf{X}$  is normal too.
- A convenient fact about normal random variables/vectors is that zero correlation/covariance implies independence.  
If  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$  and is partitioned as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_m \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \cdots & \mathbf{V}_{1m} \\ \vdots & & \vdots \\ \mathbf{V}_{m1} & \cdots & \mathbf{V}_{mm} \end{bmatrix}$$

then  $\mathbf{X}_1, \dots, \mathbf{X}_m$  are jointly independent if and only if  $\mathbf{V}_{ij} = \mathbf{0}$  for all  $i \neq j$ .

*Proof:*

#### Independence and Cochran's theorem

- (Independence between two linear forms of a multivariate normal) Let  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ ,  $\mathbf{Y}_1 = \mathbf{a}_1 + \mathbf{B}_1\mathbf{X}$  and  $\mathbf{Y}_2 = \mathbf{a}_2 + \mathbf{B}_2\mathbf{X}$ . Then  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  are independent if and only if  $\mathbf{B}_1\mathbf{V}\mathbf{B}_2^T = \mathbf{0}$ .

*Proof:*

- Consider the normal linear model  $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\mathbf{b}, \sigma^2\mathbf{I}_n)$

- Using  $\mathbf{A} = (1/\sigma^2)(\mathbf{I} - \mathbf{P}_\mathbf{X})$ , we have

$$SSE/\sigma^2 = \|\hat{\boldsymbol{\epsilon}}\|_2^2/\sigma^2 = \mathbf{y}^T \mathbf{A} \mathbf{y} \sim \chi_{n-r}^2,$$

where  $r = \text{rank}\mathbf{X}$ . Note the noncentrality parameter is

$$\phi = \frac{1}{2}(\mathbf{X}\mathbf{b})^T (1/\sigma^2)(\mathbf{I} - \mathbf{P}_\mathbf{X})(\mathbf{X}\mathbf{b}) = 0 \quad \text{for all } \mathbf{b}.$$

- Using  $\mathbf{A} = (1/\sigma^2)\mathbf{P}_\mathbf{X}$ , we have

$$SSR/\sigma^2 = \|\hat{\mathbf{y}}\|_2^2/\sigma^2 = \mathbf{y}^T \mathbf{A} \mathbf{y} \sim \chi_r^2(\phi),$$

with the noncentrality parameter

$$\phi = \frac{1}{2}(\mathbf{X}\mathbf{b})^T (1/\sigma^2)\mathbf{P}_\mathbf{X}(\mathbf{X}\mathbf{b}) = \frac{1}{2\sigma^2} \|\mathbf{X}\mathbf{b}\|_2^2.$$

- The joint distribution of  $\hat{\mathbf{y}}$  and  $\hat{\boldsymbol{\epsilon}}$  is

$$\begin{bmatrix} \hat{\mathbf{y}} \\ \hat{\boldsymbol{\epsilon}} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_\mathbf{X} \\ \mathbf{I}_n - \mathbf{P}_\mathbf{X} \end{bmatrix} \mathbf{y} \sim \mathcal{N} \left( \begin{bmatrix} \mathbf{X}\mathbf{b} \\ \mathbf{0}_n \end{bmatrix}, \begin{bmatrix} \sigma^2\mathbf{P}_\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \sigma^2(\mathbf{I} - \mathbf{P}_\mathbf{X}) \end{bmatrix} \right).$$

So  $\hat{\mathbf{y}}$  is independent of  $\boldsymbol{\epsilon}$ . Thus  $\|\hat{\mathbf{y}}\|_2^2/\sigma^2$  is independent of  $\|\hat{\boldsymbol{\epsilon}}\|_2^2/\sigma^2$  and

$$F = \frac{\|\hat{\mathbf{y}}\|_2^2/\sigma^2/r}{\|\hat{\boldsymbol{\epsilon}}\|_2^2/\sigma^2/(n-r)} \sim F_{r, n-r} \left( \frac{1}{2\sigma^2} \|\mathbf{X}\mathbf{b}\|_2^2 \right).$$

- (Independence between linear and quadratic forms of a multivariate normal) Let  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ . Then  $\mathbf{A}$  is symmetric with rank  $s$ . If  $\mathbf{BVA} = \mathbf{0}$ , then  $\mathbf{BX}$  and  $\mathbf{X}^T \mathbf{AX}$  are independent.

*Proof:*

- (Independence between two quadratic forms of a multivariate normal) Let  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ ,  $\mathbf{A}$  be symmetric with rank  $r$ , and  $\mathbf{B}$  be symmetric with rank  $s$ . If  $\mathbf{BVA} = \mathbf{0}$ , then  $\mathbf{X}^T \mathbf{AX}$  and  $\mathbf{X}^T \mathbf{BX}$  are independent.

*Proof:*

- (Cochran's theorem) Let  $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$  and  $\mathbf{A}_i$ ,  $i = 1, \dots, k$  be symmetric idempotent matrix with rank  $s_i$ . If  $\sum_{i=1}^k \mathbf{A}_i = \mathbf{I}_n$ , then  $(1/\sigma^2) \mathbf{y}^T \mathbf{A}_i \mathbf{y}$  are independent  $\chi_{s_i}^2(\phi_i)$ , with  $\phi_i = \frac{1}{2\sigma^2} \boldsymbol{\mu}^T \mathbf{A}_i \boldsymbol{\mu}$  and  $\sum_{i=1}^k s_i = n$ .

*Proof:*

- Application to the one-way ANOVA:  $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ . We have the classical ANOVA table

Source	df	Projection	SS	Noncentrality
Mean	1	$\mathbf{P}_1$	$SSM = n\bar{y}^2$	$\frac{1}{2\sigma^2} n(\mu + \bar{\alpha})^2$
Group	$a - 1$	$\mathbf{P}_X - \mathbf{P}_1$	$SSA = \sum_{i=1}^a n_i \bar{y}_i^2 - n\bar{y}^2$	$\frac{1}{2\sigma^2} \sum_{i=1}^a n_i (\alpha_i - \bar{\alpha})^2$
Error	$n - a$	$\mathbf{I} - \mathbf{P}_X$	$SSE = \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	0
Total	$n$	$\mathbf{I}$	$SST = \sum_i \sum_j y_{ij}^2$	$\frac{1}{\sigma^2} \sum_{i=1}^a n_i (\mu + \alpha_i)^2$