Course 6 : Multiple linear regression

[Source : • Linear models in statistics[1], Alvin C. Rencher and G. Bruce Schaalje.

I Random vector and matrices (chap 3, p. 69)

1) Mean, Variance, covariance, correlation

ex : $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad , i = 1, \dots, n$

x variable considered as constant

as a vector $\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ and $\underline{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$ are considered as random

Only $\underline{y}$ and x are observed, $\underline{\varepsilon}$ is unknown

$$\sigma^2 = var(y) = E[y^2] - E[y]^2$$

$$\sigma_{ij} = cov(y_i, y_j) = E[(y_i - \mu_i)(y_j - \mu_j)] \text{ where } \mu_i = E(y_i), \mu_j = E(y_j)$$

$$= E[y_i y_j] - \mu_i \mu_j$$

If y_i and y_j are independant :

1. $E(y_i y_j) = E(y_i) E(y_j)$

2. $\sigma_{ij} = cov(y_i, y_j) = 0$

Coefficient of linear correlation

$$\rho_{ij} = corr(y_i, y_j) = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \in [-1, 1]$$

2) Mean vectors and covariance matrices for random vectors

a) Mean vectors

$$E(\underline{y}) = \begin{pmatrix} E(y_1) \\ \vdots \\ E(y_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \underline{\mu}$$

Rq if $\underline{x}$ and y are random vectors

$$E(\underline{x} + y) = E(\underline{x}) + E(y)$$

b) Covariance Matrix

$$\Sigma = \mathrm{cov}(y) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2p} \\ \vdots & \vdots & \ddots & \\ \sigma_{p1} & \sigma_{p2} & & \sigma_{pp} \end{pmatrix}$$

With $\sigma_{ii} = \mathrm{Var}(y_i)$ and $\sigma_{ij} = \sigma_{ji} = \mathrm{cov}(y_i, y_j)$

If $\underline{Z}$ is a random matrix

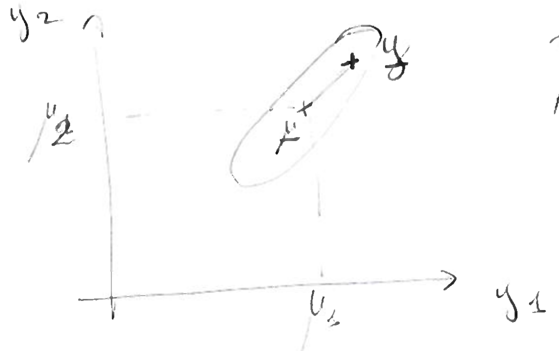
$$E(\underline{Z}) = \begin{pmatrix} E(z_{11}) & E(z_{12}) & \cdots & E(z_{1p}) \\ \vdots & \vdots & & \vdots \\ E(z_{m1}) & E(z_{m2}) & & E(z_{mp}) \end{pmatrix}$$

$$\Sigma = E\left[(y-\mu)(y-\mu)'\right]$$ where $'$ stands for the transposition

(cf exemple p. 76)

Mahalanobis distance

$$(y-\mu)'\Sigma^{-1}(y-\mu) = \|y-\mu\|^2_{\Sigma^{-1}}$$



Distance relation to the shape of the data distribution (cf links with normal distribution)

3) Linear functions of random vectors

Le $y \in \mathbb{R}^p$ a random vector, and $\underline{a} = (a_1, \ldots, a_p)'$ a vector of constants and define z:

$$z = a_1 y_1 + a_2 y_2 + \cdots + a_p y_p = \underline{a}' y$$

z is a linear combination of y. z is random!

a) Means

Th 3.6a $E[z] = E[\underline{a}' y] = a' E[y] = \underline{a}' \mu$

More generaly let $\underset{\in \mathbb{R}^k}{\underline{z}} = \underset{\mathcal{M}_{k,p}(\mathbb{R})}{A}\, y \in \mathbb{R}^p$ with $\underline{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_k \end{pmatrix}$ are

several linear combinations of y

Th3.6b a, b vectre of constants, A, B matrices of constants, y a random vector, X a random matrix. Assuming that sizes of the matrix are conformal:

(i) $E[A\underline{y}] = A\,E(y)$

(ii) $E[a' X b] = a' E[X]\, b$

(iii) $E[AXB] = A\,E[X]\,B$

Corrolary 1 $E[Ay + b] = A\,E(y) + b$

b) Variances and covariances

Th 3.6c $z = \underline{a}' y$

$\sigma_z^2 = \mathrm{var}(\underline{a}' y) = \underline{a}' \Sigma a$

Th 3.6d $\underline{z} = A y$

(i) $\mathrm{cov}(\underline{z}) = \mathrm{cov}(A\underline{y}) = A \Sigma A'$

Do exercices p. 83

3.10: Show that $E[(y-\underline{\mu})(y-\underline{\mu})'] = E[yy'] - \mu\mu'$

3.3: Show that $\mathrm{cov}(y_i, y_j) = E(y_i y_j) - E(y_i)E(y_j)$

3.20:

3.21

II Multivariate normal distribution (p. 87)

1) Univariate normal density : $y \sim \mathcal{N}(\mu, \sigma^2)$

$$f(y) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y-\mu)^2/2\sigma^2}$$

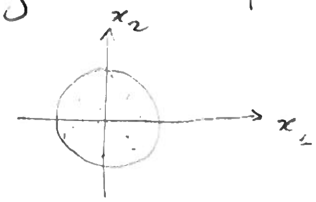
2) Multivariate normal density function

$z_1, z_2, \dots, z_p$ iid $\mathcal{N}(0,1)$

density

$$g(z_1, z_2, \dots, z_p) = g(\underline{z}) = g(z_1)g(z_2)\dots g(z_m) \quad \text{independance}$$

$$= \frac{1}{\sqrt{2\pi}^{\,p}} e^{-\underline{z}'\underline{z}/2}$$



Let define $\underline{y} = \Sigma^{1/2}\underline{z} + \underline{\mu}$ then

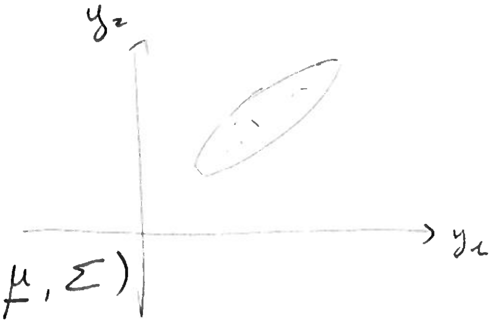
$$E(\underline{y}) = E(\Sigma^{1/2}\underline{z} + \underline{\mu}) = \Sigma^{1/2}\underbrace{E(\underline{z})}_{0} + \underline{\mu} = \underline{\mu}$$

$$\operatorname{cov}(\underline{y}) = \operatorname{cov}(\Sigma^{1/2}\underline{z} + \underline{\mu}) = \Sigma^{1/2}\operatorname{cov}(\underline{z})\Sigma^{1/2} = \Sigma^{1/2} I \Sigma^{1/2} = \Sigma$$

Moreover the variable change formula give the density of $\underline{y}$ base on the known density of $\underline{z}$

$$f(\underline{y}) = \frac{1}{(\sqrt{2\pi})^p |\Sigma|^{1/2}} e^{-(\underline{y}-\underline{\mu})'\Sigma^{-1}(\underline{y}-\underline{\mu})/2}$$

We note $\underline{y} \sim \mathcal{N}_p(\underline{\mu}, \Sigma)$



If $\underline{z} = A\underline{y} + b$ and $\underline{y} \sim \mathcal{N}_p(\underline{\mu}, \Sigma)$

Then $\underline{z} \sim \mathcal{N}(A\underline{\mu} + b, A\Sigma A')$

If $\underline{y} \sim \mathcal{N}_p(\underline{\mu}, \Sigma)$, then $y_i \sim \mathcal{N}(\mu_i, \sigma_{ii})$

Do exercises

4.2 Show that $|\Sigma^{-1/2}| = |\Sigma|^{-1/2}$

4.9 Assuming $\underset{\sim}{y} \sim N_p(\underset{\sim}{\mu}, \sigma^2 I)$ and C is an orthonormal matrix, show that $C\underset{\sim}{y} \sim N_p(C\mu, \sigma^2 I)$.

Hint $CC' = I_p$

4.16 questions (a), (b), (c), (d)

III Distribution of quadratic form of y

A) Sum of squares

Ex
$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = y\left(I - \frac{1}{n}J\right)y'$$

where $J = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & 1 & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$ of dimension $n \times n$.

B) Mean of quadratic forms

Th 5.2.a If $E[y] = \mu$ and $\mathrm{cov}(y) = \Sigma$ and A a symetric matrix $E[y'Ay] = \mathrm{tr}(A\Sigma) + \mu' A \mu$

Ex (p 108) $s^2 = \dfrac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$, possible to show that $E[s^2] = \sigma^2$!!

c) Chi-square distribution

$z_1, \ldots, z_n$ iid $\mathcal{N}(0,1)$, $\underline{z} = (z_1, \ldots, z_n)'$

$$\sum_{i=1}^{n} z_i^2 = \underline{z}'\underline{z} \sim \chi^2(n)$$

d) F distribution

If $u \sim \chi^2(p)$ and $v \sim \chi^2(q)$ then $w = \dfrac{u/p}{v/q} \sim F(p,q)$: Fisher distribution.

e) t-distribution

If $z \sim \mathcal{N}(0,1)$, and $u \sim \chi^2(p)$ and z and u are independent

$$t = \frac{z}{\sqrt{u/p}} \sim t(p) \text{ also denoted by } \mathcal{T}_p$$

<u>Corollary 1</u> (p.120) If $y \sim N_p(\mu, \sigma^2 I)$ then $y'Ay$ and $y'By$ are independent if and only if $AB = 0$

<u>Exercices</u> (p122)

5.2. show that $(1/n) J$ is idempotent,

5.27 (a)

5.32

IV Simple linear regression (done) p.127

A/ Assumptions

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \;, \quad i = 1, \dots, n$$

$$\begin{cases} 1. \; E(\varepsilon_i) = 0 \\ 2. \; var(\varepsilon_i) = \sigma^2 \\ 3. \; cov(\varepsilon_i, \varepsilon_j) = 0 \end{cases}$$

B/ OLS

The least squares give $\hat{\beta}_1 = \dfrac{cov(\underline{x}, \underline{y})}{var(\underline{x})}$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

⚠ $\hat{\beta}_1$ and $\hat{\beta}_0$ are random since they depend on $\underline{y}$ which is random

C/ Hypothesis test and confidence interval

Under the additional assumption that ε_i are iid and $\varepsilon_i \sim N(0, \sigma^2)$

it is possible to get ~~the~~ closed form ~~form~~ formulaes

1. $\hat{\beta}_1 \sim N\left(\beta_1, \dfrac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right)$

2. $(n-2) s^2/\sigma^2 \sim \chi^2(n-2)$

3. $\hat{\beta}_1$ and s^2 are independant

$$\Rightarrow \quad t = \frac{\hat{\beta}_1 - \beta_1}{s/\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \sim t(n-2)$$

Thus possible to test $H_0 : \beta_1 = 0$, or to make confidence intervals on β_1.

$$r^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$SST = SSR + SSE$ with $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, thus $r^2 = 1 - \frac{SSE}{SST}$

Moreover for ~~time~~ simple linear regression

$$r^2 = \text{corr}(\underline{x}, \underline{y})^2 = \rho_{xy}^2$$

<u>exercise</u>

6.14 (p.136)

IV Multiple regression : Estimation

1/ The model

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i \ , \ i = 1, 2, \dots, n$$

Assumption 1. $E(\varepsilon_i) = 0 \Leftrightarrow E(y_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$

2. $\text{var}(\varepsilon_i) = \sigma^2$

3. $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$

Rq : Additional assumption are added to perform tests and obtain confidence intervals.

<u>Matrix form</u> $\underline{y} = X\beta + \underline{\varepsilon}$

1. $E(\underline{\varepsilon}) = 0 \Rightarrow E(y) = X\beta$

2. $\text{cov}(\underline{\varepsilon}) = \sigma^2 I \Rightarrow \text{cov}(\underline{y}) = \sigma^2 I$

<u>Criterion</u> : We seek to $\beta_0, \beta_1, \dots, \beta_k$ to minimize

$$C(\hat{\beta}) = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$= \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik})^2 = \| \underline{y} - X\hat{\beta} \|^2$$

$$= (\underline{y} - X\hat{\beta})'(\underline{y} - X\hat{\beta})$$

Theorem 7.3a If X is of rank $k+1<n$ then

$$\boxed{\hat{\beta} = (X'X)^{-1} X' y}$$

Rq Obtained by solving the normal equation

$$X'X\hat{\beta} = X'y$$ (obtained after derivation of $C(\beta)$)

Note compare to intro ML !!!

Properties of the OLS

$$E(\hat{\beta}) = \beta \quad \text{: Unbiased}$$

$$\operatorname{cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

Gauss-Markov Theorem

If $E(y) = X\beta$ and $\operatorname{cov}(y) = \sigma^2 I$, $\hat{\beta}_j$, $j = 0, 1, \dots, k$ have minimum variance among all unbiased estimators

Estimator of σ^2

$$s^2 = \frac{1}{n-k-1} \| y - X\hat{\beta} \|^2 = \frac{SSE}{n-k-1}$$

Th $E[s^2] = \sigma^2$: Unbiased

Corollary $\widehat{\operatorname{cov}}(\hat{\beta}) = s^2 (X'X)^{-1}$

D/ Normal model, not needed previously but need for CI and tests...

$$y \sim \mathcal{N}_n(X\beta, \sigma^2 I) \iff \underline{\varepsilon} \sim \mathcal{N}_n(0, \sigma^2 I)$$

MLE for β and σ^2

$$\hat{\beta}^{MLE} = \hat{\beta}^{OLS}, \quad \hat{\sigma}^2 = \frac{1}{n} \| y - X\hat{\beta} \|^2 : \text{Biased estimator } \sigma^2.$$

Th 7.6b (p 159)

(i) $\hat{\beta} \sim \mathcal{N}_{k+1}(\beta, \sigma^2 (X'X)^{-1})$

(ii) $n\hat{\sigma}^2/\sigma^2 \sim \chi^2(n-k-1)$, or equivalently, $(n-k-1)s^2/\sigma^2 \sim \chi^2(n-k-1)$

(iii) $\hat{\beta}$ and $\hat{\sigma}^2$ (or s^2) are independant

multiple correlation coefficient

$R^2 = \frac{SSE}{SST}$, pb it increase with the number of variable thus need to consider R_a^2 the adjusted R^2

$$R_a^2 = \frac{(n-1)R^2 - k}{n-k-1}$$

Problems : ~~solve~~ do

7.1, 7.2, 7.21, 7.53 (a) (b) (d), 7.54, 7.55 (a), (c)

V Multiple regression hypothesis testing and confidence intervals

$y \sim \mathcal{N}(X\beta, \sigma^2 I)$, X is $n\times(k+1)$ of rank $k+1<n$

A/ Test of overall regression

$\underline{\beta_1} = (\beta_1, \dots, \beta_k)'$ and $\underline{\beta} = \begin{pmatrix} \beta_0 \\ \underline{\beta_1} \end{pmatrix}$

We want to test $H_0 : \underline{\beta_1} = 0$ against $H_1 : \exists j \in \{1,\dots,k\}, \beta_j \neq 0$

$\Leftrightarrow \beta_1 = \beta_2 = \dots = \beta_k = 0$

Th 8.1.d

$$F = \frac{SSR/k}{SSE/(n-k-1)}$$

If $H_0 : \underline{\beta_1}$ is true then $F \sim F(k, n-k-1)$

the Fisher distribution with k an $n-k-1$ degrees of freedom

B/ Test on a subset of β's

C/ General linear hypothesis tests for

$H_0 : C\beta = 0$ and $C\beta = t$