High (and low) Dimensional Statistics

p.5

1] Regression is the study of the response variable Y as the predictor variable $\underset{\sim}{x} = (x_1, \ldots, x_p)^T$ vary.

2] p.1 A <u>case</u> is the set of measurements on a person or thing.

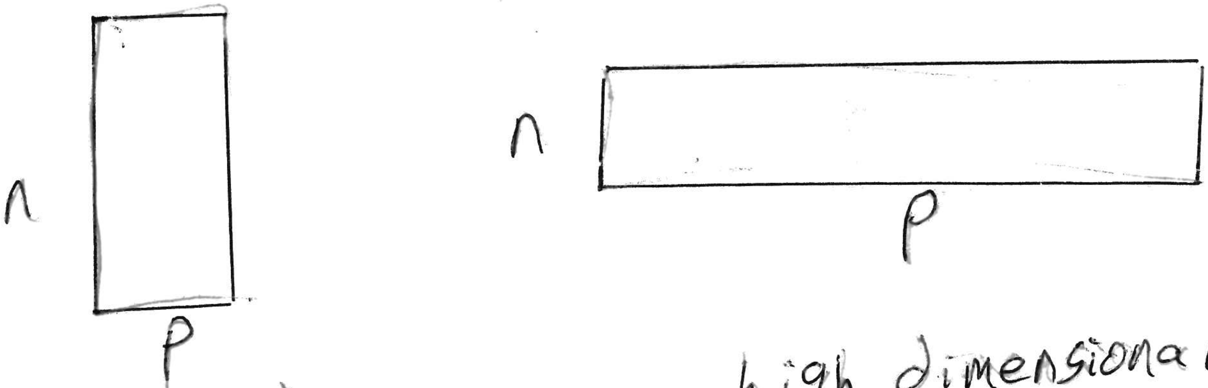
ex] $Y =$ height, $x_1 =$ span, $x_2 =$ height at shoulder

ith case $(Y_i, x_{1i}, x_{2i})^T$ for $i = 1, \ldots, n =$ sample size.

3] A regression dataset is <u>low dimensional</u> if $n \geq Jp$ with $J \geq 5$.

A regression dataset is <u>high dimensional</u> if $n < 5p$. Note that $n << p$ is possible, eg $n = 100$, $p = 1$ million.

ultrahigh dimensional statistics



classical statistics

high dimensional statistics

perhaps the hottest research topic since 2000.

4) A multivariate data set has

p measurements $\underline{x}_1, \dots, \underline{x}_n$

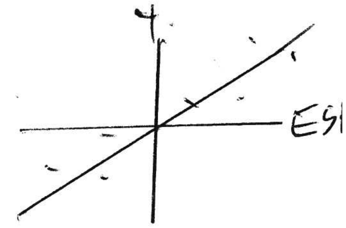
$\underline{x}_i = (x_{i1}, \dots, x_{ip})^T.$

low dimensional $n \geq Jp$, $\quad J \geq 5$

high dimensional $n < 5p$.

Things like $\text{cov}(\underline{x})$, PCA are of interest.

Math 575 586



5) multiple linear regression

$$Y_i = \alpha + \underset{\sim}{\beta}^T \underset{\sim}{x}_i + e_i = \alpha + \underset{\sim}{x}_i^T \underset{\sim}{\beta} + e_i.$$

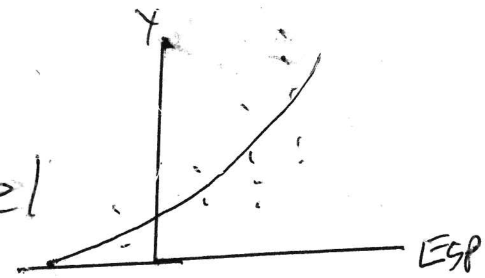
Math 484 584 586

6] The <u>sufficient predictor</u>

$$SP = \alpha + \underline{x}^T \underline{\beta} = \alpha + \underline{\beta}^T \underline{x}.$$

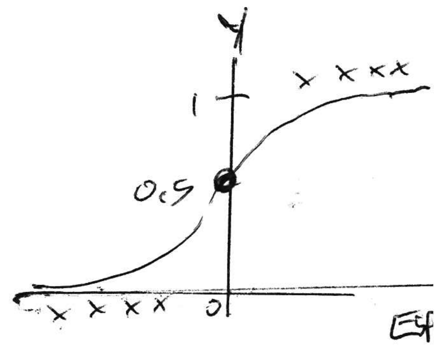
7] The <u>estimated sufficient predictor</u>

$$ESP = \hat{\alpha} + \underline{x}^T \underline{\hat{\beta}}.$$



8] A <u>Poisson regression</u> model

is $Y \sim \text{Poisson}(e^{SP})$.



9] A <u>binary regression</u> model

is $Y \sim \text{bin}(1, \rho(SP))$.

$$\rho(SP) = \underline{P}(Y=1) = \frac{e^{SP}}{1+e^{SP}}$$ corresponds

to binary logistic regression.

M485, 586

10] A <u>binomial regression</u> model

is $Y_i \sim \text{bin}(m_i, \rho(SP_i))$.

$m_i \equiv 1 \rightarrow$ binary regression

11] With $\alpha = 0$, a Weibull proportional hazards regression model is $Y | SP \sim W(\gamma = \frac{1}{\sigma}, \lambda_0 e^{SP})$

where Y has a Weibull (γ, λ) dist if the pdf of Y is

$$f(y) = \lambda \gamma y^{\gamma - 1} \exp[-\lambda y^{\gamma}], \underbrace{y > 0}_{\text{support}}$$

$(f(y) = 0, \text{else})$

M473

12) §2.1 A model for regression variable selection is

$$\underset{\sim}{x}^T \underset{\sim}{\beta} = \underset{\sim}{x}_S^T \underset{\sim}{\beta}_S + \underset{\sim}{x}_E^T \underset{\sim}{\beta}_E = \underset{\sim}{x}_S^T \underset{\sim}{\beta}_S$$

the active predictors

where $\underset{\sim}{x} = (\underset{\sim}{x}_S^T, \underset{\sim}{x}_E^T)^T$, $\underset{\sim}{x}_S$ is $a_S \times 1$

the model you want to select
often sparse in the literature

$\underset{\sim}{x}_E$ is $(p - a_S) \times 1$. Given

$\underset{\sim}{x}_S$ is in the model, $\underset{\sim}{\beta}_E = \underset{\sim}{0}$

and E denotes the terms

that can be eliminated given S is

in the model.

§3.1 section

13] For MLR, we may use

the model MLR1) $Y_i = \beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + e_i$

$$= \underset{\sim}{x}_i^T \underset{\sim}{\beta} + e_i \quad \text{for } i = 1, \ldots, n$$

$$\underset{\sim}{Y} = \mathbb{X} \underset{\sim}{\beta} + \underset{\sim}{e} \quad \text{in matrix form.}$$

or model MLR2) $Y_i = \alpha + x_{i1}\beta_1 + \cdots + x_{ip}\beta_p + e_i$
for $i = 1, \ldots, n$

$$Y_i = \alpha + \underset{\sim}{x}_i^T \underset{\sim}{\beta} + e_i$$

Let $\underset{\sim}{\phi} = \begin{pmatrix} \alpha \\ \underset{\sim}{\beta} \end{pmatrix}$

$\mathbb{X} = (\underset{\sim}{1} \;\; \mathbb{X}_1)$,

$$\underset{\sim}{Y} = \alpha \underset{\sim}{1} + \mathbb{X}_1 \underset{\sim}{\beta} + \underset{\sim}{e} = \mathbb{X} \underset{\sim}{\phi} + \underset{\sim}{e}.$$

14] For MLR1), $\underset{\sim}{Y} = \mathbb{X}\underset{\sim}{\beta} + \underset{\sim}{e}$,

$\hat{\underset{\sim}{\beta}}_{OLS} = \hat{\underset{\sim}{\beta}} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\underset{\sim}{Y}$ assuming

$n > p$ and $\mathbb{X}$ has full rank p.

15] For MLR2) $Y = \alpha + \underset{\sim}{x}^T\underset{\sim}{\beta} + e$.

OLS = ordinary least squares.

Let $\text{Cov}(\underset{\sim}{x}) = E[(\underset{\sim}{x} - E(\underset{\sim}{x}))(\underset{\sim}{x} - E(\underset{\sim}{x}))^T]$

$= \Sigma_{\underset{\sim}{x}}$ and

Sigma

$\text{Cov}(\underset{\sim}{x}, Y) = E[(\underset{\sim}{x} - E(\underset{\sim}{x}))(Y - E(Y))] = \Sigma_{\underset{\sim}{x}Y} = \Sigma_{\underset{\sim}{x}}$

If the cases $(\underset{\sim}{x}_i, Y_i)$ are iid from some population where $\Sigma_{\underset{\sim}{x}Y}$ exists and Σ_x is nonsingular, then

$\alpha = \alpha_{OLS} = E(Y) - \underset{\sim}{\beta}^T E(\underset{\sim}{x})$ and

$\underset{\sim}{\beta} = \underset{\sim}{\beta}_{OLS} = \Sigma_{\underset{\sim}{x}}^{-1}\Sigma_{\underset{\sim}{x}Y}$.

16] §3.1 Let the sample covariance matrices

be $\hat{\Sigma}_{\underset{\sim}{x}} = \frac{1}{n-1}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \overline{\underset{\sim}{x}})(\underset{\sim}{x}_i - \overline{\underset{\sim}{x}})^T$

and $\hat{\Sigma}_{\underset{\sim}{x}Y} = \frac{1}{n-1}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \overline{\underset{\sim}{x}})(Y_i - \overline{Y})$

Let the method of moments estimators $\tilde{\Sigma}_{\underset{\sim}{x}} = \frac{1}{n}\sum_{i=1}^{n}(\underline{x}_i - \underline{\bar{x}})(\underline{x}_i - \underline{\bar{x}})^T$

and $\tilde{\Sigma}_{xY} = \frac{1}{n}\sum_{i=1}^{n}(\underline{x}_i - \underline{\bar{x}})(Y_i - \bar{Y}) = \frac{1}{n}\sum_{i=1}^{n}\underset{\sim}{x}_i Y - \underline{\bar{x}}\bar{Y}.$

§3.1

17] Under MLR2 $Y = \alpha + \underline{\beta}^T \underset{\sim}{x} + e$,

$\hat{\underline{\phi}} = \hat{\underset{\sim}{\phi}}_{OLS} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underset{\sim}{Y}$.

Second way to compute $\hat{\underline{\phi}} = \hat{\underset{\sim}{\phi}}_{OLS}$:

If $\hat{\Sigma}_x^{-1}$ exists, then

$\hat{\alpha} = \bar{Y} - \hat{\underset{\sim}{\beta}}^T \underset{\sim}{\bar{x}}$ and

$\hat{\underline{\beta}} = \hat{\Sigma}_{\underline{x}}^{-1} \hat{\Sigma}_{\underline{x}Y}$.

18] If $(\underline{x}_i, Y_i)$ are iid, and σ_Y^2, $\Sigma_{\underline{x}}^{-1}$ and $\Sigma_{\underline{x}Y}$ exist, then

$\hat{\alpha} \xrightarrow{P} \alpha$ and

$\hat{\underset{\sim}{\beta}} \xrightarrow{P} \underset{\sim}{\beta}$ as $n \to \infty$ with

α and $\underline{\beta}$ as in 15].

$\hat{\underset{\sim}{\theta}} \xrightarrow{P} \underset{\sim}{\theta}$ means $\hat{\underset{\sim}{\theta}}$ is a consistent estimator of $\underset{\sim}{\theta}$.

§1.3

19] ~~Let~~ $\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \in \mathbb{R}^p$.

$$E(\underline{x}) = \underset{\sim}{\mu} = \begin{pmatrix} E(x_1) \\ \vdots \\ E(x_p) \end{pmatrix}$$

$$\operatorname{cov}(\underline{x}) = \Sigma = E\left[(\underset{\sim}{x} - E(\underline{x}))(\underline{x} - E(\underline{x}))^T\right] = (\sigma_{ij})$$

where $\sigma_{ij} = \operatorname{cov}(x_i, x_j)$ and

$\operatorname{cov}(x_i, x_i) = \sigma_i^2 = \sigma_{ii} = V(x_i)$.

20] ~~Know~~ If $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are $p \times 1$ random vectors ($\underset{\sim}{R}V_s$), $\underset{\sim}{a}$ a conformable constant vector and A and B conformable constant matrices, then

$$E(\underset{\sim}{a} + \underset{\sim}{x}) = \underset{\sim}{a} + E(\underset{\sim}{x})$$

$$E(\underset{\sim}{x} + \underset{\sim}{y}) = E(\underset{\sim}{x}) + E(\underset{\sim}{y})$$

$E(A\underline{x}) = A E(\underline{x})$

$E(A\underline{x}B) = \underset{q\times p}{A} \; \underset{p\times 1}{E(\underline{x})} \; \underset{1\times k}{B}$

$\mathrm{cov}(\underline{a} + A\underline{x}) = \mathrm{cov}(A\underline{x}) = A\,\mathrm{cov}(\underline{x})\,A^T.$

21) Know Similar formulas hold for $\overline{\underline{x}}$, $\hat{\Sigma}_{\underline{x}}$ and $\tilde{\Sigma}_{\underline{x}}$ (and $\hat{\Sigma}_{\underline{x}\underline{y}}$, $\tilde{\Sigma}_{\underline{x}\underline{y}}$).

Note that $\overline{\underline{x}} = E(\underline{x})$ and $\hat{\Sigma}_{\underline{x}} = \mathrm{cov}(\underline{x})$

if $P(\underline{x} = \underline{x}_i) = \frac{1}{n}$, $i = 1, \ldots, n$ (empirical dist).

So if $\underline{w}_i = A\underline{x}_i$, then

$\overline{\underline{w}} = A\overline{\underline{x}}$, $\hat{\Sigma}_{A\underline{x}} = \hat{\Sigma}_{\underline{w}} = A\hat{\Sigma}_{\underline{x}}A^T$

$\tilde{\Sigma}_{A\underline{x}} = \tilde{\Sigma}_{\underline{w}} = A\tilde{\Sigma}_{\underline{x}}A^T$,

22)* $\mathrm{cov}\left(\sum_{i=1}^{n} \underset{\sim}{x}_i, \sum_{j=1}^{m} \underset{\sim}{z}_j\right) =$ see HW1 B).

$\sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{cov}(\underset{\sim}{x}_i, \underset{\sim}{z}_j)$

cov of 2 sums = double sum of cov's

§1.3

23) A $p \times 1$ random vector $\underline{x}$ has a p-dimensional multivariate normal (MVN) distribution $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ iff (if and only if)

$\underset{\sim}{t}^T \underset{\sim}{x}$ has a univariate normal dist for any $p \times 1$ vector $\underline{t}$.

(univariate normal $p=1$, $w \sim N(\mu, \sigma^2)$)

$E(\underline{x}) = \underline{\mu}$, $cov(\underline{x}) = \Sigma$. (write N_1 as N)

24) If $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then all subsets of $\underline{x}$ are MVN: $(x_{i_1}, \ldots, x_{i_q})^T \sim N_q(\underset{\sim}{\tilde{\mu}}, \tilde{\Sigma})$

where $\tilde{\mu}_j = E(x_{i_j})$ and $\tilde{\Sigma}_{jk} = cov(x_{i_j}, x_{i_k})$.

ex) $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \sim N_3\left[\begin{pmatrix} 1 \\ 17 \\ 8 \end{pmatrix}, \begin{pmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}\right]$.

Find the dist of $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\begin{pmatrix} x_1 \\ x_3 \end{pmatrix}$, $\begin{pmatrix} x_2 \\ x_3 \end{pmatrix}$

Soln $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N_2\left[\begin{pmatrix} 1 \\ 17 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix}\right]$

$$\begin{pmatrix} x_1 \\ x_3 \end{pmatrix} \sim N_2\left[\begin{pmatrix} 1 \\ 8 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}\right]$$

$$\begin{pmatrix} x_2 \\ x_3 \end{pmatrix} \sim N_2\left[\begin{pmatrix} 17 \\ 8 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}\right]$$

25] Let $\underset{\sim}{x} = \begin{pmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \end{pmatrix} \begin{matrix} q \times 1 \\ (p-q) \times 1 \end{matrix}$, $\underset{\sim}{\mu} = \begin{pmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{pmatrix}$,

$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. If $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, \Sigma)$,

then $\underset{\sim}{x}_1 \sim N_q(\underset{\sim}{\mu}_1, \Sigma_{11})$ and $\underset{\sim}{x}_2 \sim N_{p-q}(\underset{\sim}{\mu}_2, \Sigma_{22})$.

26] If $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then $\underset{\sim}{x}_1$ is independent of $\underset{\sim}{x}_2$, written $\underset{\sim}{x}_1 \perp\!\!\!\perp \underset{\sim}{x}_2$, iff $\Sigma_{12} = 0$ matrix of 0's.

27] pop correlation between 2 RVs x and y is $\rho(x,y) = \dfrac{\mathrm{cov}(x,y)}{\sqrt{V(x)\,V(y)}} = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$

$= \rho(y,x)$.

28] KNOW If $\begin{pmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \end{pmatrix} = \underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, \Sigma)$,

then the conditional distribution

$$\underset{\sim}{x}_1 \mid \underset{\sim}{x}_2 = \underset{\sim}{x}_2 \sim N_q\left(\underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2), \underbrace{\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}_{\text{free of } \underset{\sim}{x}_2 = \underset{\sim}{x}_2}\right).$$

Notation $\underset{\sim}{X}_1 \mid \underset{\sim}{X}_2 \sim N_q\left(\underbrace{\underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{X}_2 - \underset{\sim}{\mu}_2)}_{\text{family of means depending on the values of } \underset{\sim}{X}_2}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$

ex] $\begin{pmatrix} Y \\ X \end{pmatrix} \sim N_2\left[\begin{pmatrix} 49 \\ 100 \end{pmatrix}, \begin{pmatrix} 16 & \sigma_{12} \\ \sigma_{12} & 25 \end{pmatrix}\right]$

a) If $\sigma_{12} = 0$, find $Y \mid X$.
b) If $\sigma_{12} = 10$, find $E(Y \mid X)$.
c) If $\sigma_{12} = 10$, find $V(Y \mid X)$.

soln $\Sigma_{ij} = \sigma_{ij}$

a) $Y \mid X \sim N(49, 16)$ since $Y \perp\!\!\!\perp X$.

b) $E(Y \mid X) = \mu_Y + \Sigma_{12}\Sigma_{22}^{-1}(X - \mu_X)$

$= 49 + \frac{10}{25}(X - 10) = 49 + \frac{2}{5}(X - 100) = \boxed{9 + 0.4X}$

c) $V(Y|X) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} =$

$16 - 10\frac{1}{25}10 = 16-4 = \boxed{12}$

29] $\underset{\sim}{X}_n$ converges in distribution to $\underset{\sim}{X}$,

$\underset{\sim}{X}_n \xrightarrow{D} \underset{\sim}{X}$, if the cumulative distribution

functions (cdfs) $F_{\underset{\sim}{X}_n}(\underset{\sim}{x}) \to F_{\underset{\sim}{X}}(\underset{\sim}{x})$

for all continuity points $\underset{\sim}{x}$ of $F_{\underset{\sim}{X}}(\underset{\sim}{x})$

($\forall \underset{\sim}{x} \in \mathbb{R}^p$ if $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, Σ^{-1} exists).

$\underset{\sim}{X}$ is the limiting or asymptotic dist

of $\underset{\sim}{X}_n$, and does not depend on n.

ex) CLT: $X_1, \ldots, X_n$ iid, $E(X_i) = \mu$, $V(X_i) = \sigma^2 > 0$

$\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2)$

so $\bar{X} \approx N(\mu, \frac{\sigma^2}{n})$. normal approx does depend on n

Know

30] If $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, so $p\times 1$

A is a $q\times p$ constant matrix, b a constant, $\underset{\sim}{a}$ a $p\times 1$ constant vector and $\underline{d}$ a $q\times 1$ constant vector, then

i) $A\underline{x} \sim N_{\underline{q}}(A\underline{\mu}, A\Sigma A^T)$

ii) $\underline{a} + b\underset{\sim}{x} \sim N_{\underline{p}}(\underline{a} + b\underline{\mu}, b^2\Sigma)$

Note: $b\underline{x} = b I_p \underline{x}$ with $A = bI_p$.

iii) $A\underline{x} + \underline{d} \sim N_{\underline{q}}(A\underline{\mu} + \underline{d}, A\Sigma A^T)$.

Know

31] Suppose $\underset{\sim}{x}_n \xrightarrow{D} \underline{x} \sim N_p(\underline{\mu}, \Sigma)$. Then

i) $A\underline{x}_n \xrightarrow{D} A\underline{x} \sim N_q(A\underline{\mu}, A\Sigma A^T)$

ii) $\underset{\sim}{a} + b\underset{\sim}{x}_n \xrightarrow{D} N_p(\underset{\sim}{a} + b\underset{\sim}{\mu}, b^2\Sigma)$

iii) $A\underline{x}_n + \underline{d} \xrightarrow{D} N_q(A\underline{\mu} + \underline{d}, A\Sigma A^T)$.

30] and 31] are very similar

32] Multivariate Central Limit Theorem

<u>MCLT</u>: If $\underline{x}_1, \ldots, \underline{x}_n$ are iid $p \times 1$ RVs with $E(\underline{x}) = \underline{\mu}$ and $Cov(\underline{x}) = \Sigma$, then $\sqrt{n}\,(\overline{\underline{x}}_n - \underline{\mu}) \xrightarrow{D} N_p(\underline{0}, \Sigma)$.

Note: $\overline{x}_n = \frac{1}{n} \sum_{i=1}^{n} \underline{x}_i = \begin{pmatrix} \overline{x}_{1n} \\ \vdots \\ \overline{x}_{pn} \end{pmatrix}$. coordinatewise sample means

$\hat{\Sigma} = \hat{\Sigma}_{\underline{x}}$

33] If $\underline{x}_n \xrightarrow{D} \underline{x}$, then all subsets of $\underline{x}_n$ converge to subsets of $\underline{x}$:

$$\begin{pmatrix} x_{i_1,n} \\ \vdots \\ x_{i_K,n} \end{pmatrix} \xrightarrow{D} \begin{pmatrix} x_{i_1} \\ \vdots \\ x_{i_K} \end{pmatrix}.$$

34] Roughly want $n \geq Jp$ with $J \geq 10$ and often J much larger for MCLT to hold.

Suppose $n = 100$, $p = 10000$ and MCLT approx is good for $n \geq 10p$.

Then can do inference

on $\underset{d \times k}{A} \underset{\sim}{w}$ with $\underset{\sim}{w} = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ik} \end{pmatrix}$ and $k \le 10$.

$d \le k$

So can do inference on μ_i, $\mu_i - \mu_j$

etc, using $A \hat{\Sigma}_{\underset{\sim}{w}} A^T$.

MLE: inverse Fisher information matrix

35] often $\underline{x}_n \xrightarrow{D} \underset{\sim}{x} \sim N_p(\underline{\mu}, V^{-1})$

where we can't compute $\hat{V}^{-1}$ if

$n < p$. So can't find $A \hat{V}^{-1} A^T$.

$\bar{\underset{\sim}{x}}$ and $\hat{\Sigma}_{\underset{\sim}{x}\underset{\sim}{y}}$ will be exceptions.

36] Know $\operatorname{cov}(A\underline{x}, B\underline{y}) = A \operatorname{cov}(\underline{x}, \underline{y}) B^T$

$\Sigma_{A\underline{x},\underline{y}} = A \Sigma_{\underline{x}\underline{y}}$ take $B = I$

$\Sigma_{\underline{x},B\underline{y}} = \Sigma_{\underline{x}\underline{y}} B^T$ take $A = I$

37] Similar formulas hold for $\hat{\Sigma}$ and $\tilde{\Sigma}$

$\hat{\Sigma}_{A\underline{x},\underline{y}} = A \hat{\Sigma}_{\underline{x}\underline{y}}$, $\tilde{\Sigma}_{A\underline{x},\underline{y}} = A \tilde{\Sigma}_{\underline{x}\underline{y}}$

38] For iid cases, the <u>pop OLS regression</u> of Y on $\underline{w} = A\underline{x}$ is

$$\underline{\beta}(\underline{w}, Y) = \Sigma_{\underline{w}}^{-1} \Sigma_{\underline{w}Y} = (A \Sigma_{\underline{x}} A^T)^{-1} A \Sigma_{\underline{x}Y}$$

39] The OLS regression of Y on $\underline{W} = A\underline{X}$ ← don't need iid cases, 2nd way to compute $\hat{\underline{\beta}}$

is $\hat{\underline{\beta}}_{OLS}(\underline{w}, Y) = \hat{\Sigma}_{\underline{w}}^{-1} \hat{\Sigma}_{\underline{w}Y} = (A \hat{\Sigma}_{\underline{x}} A^T)^{-1} A \hat{\Sigma}_{\underline{x}Y}$

provided the inverse matrices exist.

40] §2.6 <u>OLS CLT</u> Consider MLR 1) model

$Y_i = \underline{x}_i^T \underline{\beta} + e_i$. Assume the 0 mean errors e_i are iid with $E(e_i) = 0$ and $V(e_i) = \sigma^2$. Let $H = X(X^TX)^{-1}X^T$ and assume $\max(h_{11}, \ldots, h_{nn}) \xrightarrow{P} 0$ as $n \to \infty$ and

$$\frac{X^T X}{n} \to V^{-1} \text{ as } n \to \infty. \text{ Then}$$

$$\sqrt{n}(\hat{\underline{\beta}}_{OLS} - \underline{\beta}) \xrightarrow{D} N_p(\underline{0}, \sigma^2 V)$$

as $n \to \infty$.

If $Y = \alpha + \underline{x}^T\underline{\beta} + e$
$= Y|\underline{x} = $ " "
$Y|\underline{x}^T\underline{\beta}_{OLS} = $ " ",

Cases independent, errors independent of $\underline{x}$'s: $e_i \perp\!\!\!\perp \underline{x}_i$

then under mild conditions,

$\alpha = \alpha_{OLS}$ and $\underline{\beta} = \underline{\beta}_{OLS}$.

Chapter 3

1] §3.1 OLS CLTs
For the MLR model, assume the zero mean errors e_i are iid with $E(e_i) = 0$, $V(e_i) = \sigma^2$. Assume the cases $(\underline{x}_i, Y_i)$ are independent. If the $\underline{x}_i$ are random vectors, (not constant vectors), assume $e_i \underset{\text{ind}}{\perp\!\!\!\perp} \underline{x}_i$.

Assume $\frac{X^TX}{n} \xrightarrow{P} V^{-1}$ as $n \to \infty$.

a) For model MLR 1]

$$\sqrt{n}(\hat{\underline{\beta}}_{OLS} - \underline{\beta}) \xrightarrow{D} N_p(\underline{0}, \sigma^2 V).$$

b) For model MLR 2)

$$\sqrt{n}(\hat{\underline{\phi}} - \underline{\phi}) \xrightarrow{D} N_{p+1}(\underline{0}, \sigma^2 V).$$