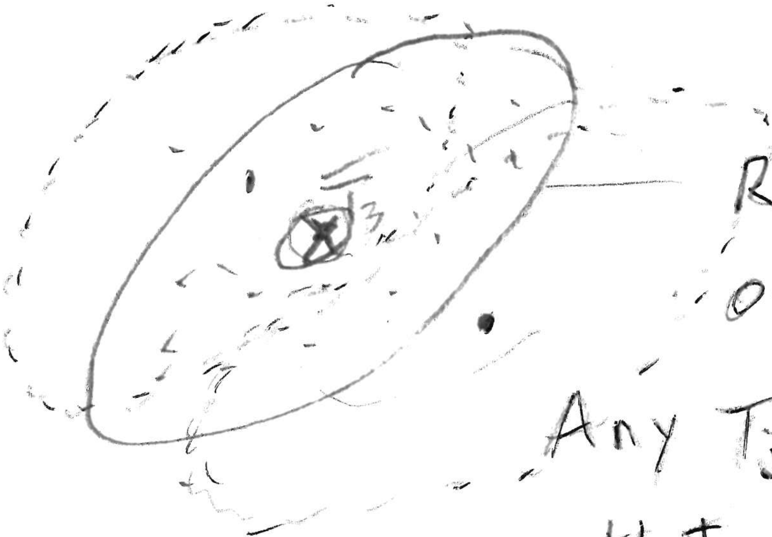
LS 80



R_B contains $\simeq 95\%$ of the T_i

Any T_j in R_B results in a new region that contains $\overline{T}_B$. Any $T_j \notin R_B$ results in a new region that does not contain $\overline{T}_B$.

$$R_B = \{ \underline{w} : (\underline{w} - \overline{T}_B)^T S_T^{-1} (\underline{w} - \overline{T}_B) \le D^2_{(U_B)} \}$$

new region $\{ \underline{w} : (\underline{w} - T_j)^T S_T^{-1} (\underline{w} - T_j) \le D^2_{(U_B)} \}$

$T_j \in R_B$ iff $\overline{T}_B \in$ new region

since $D^2_{T_j}(\overline{T}_B, S_T) = D^2_{\overline{T}_B}(T_j, S_T)$.

If B is large a hyperellipsoid centered at $\overline{T}_B$ contains $\underline{\theta}$ with high prob where the volume of the hyperellipsoid $\to 0$ as $B \to \infty$.

Hence the new region is a 95% confidence region for $\underline{\theta}$.

The prediction region method takes the entire "iid data cloud" and shifts it to be centered at $\overline{T}_B^* \approx T_n$ in that $\sqrt{n}(T_n - \underline{\theta}) \xrightarrow{D} \underline{U}$ and $\sqrt{n}(T_i^* - T_n) \xrightarrow{D} \underline{U}$.

So the bootstrap sample $T_1^*, \ldots, T_B^*$ is like $T_1, \ldots, T_B$ shifted to $\overline{T}_B^* \approx T_n$.
center

---

ex} CI: Let $X_1, \ldots, X_n$ be iid $U(0, \theta)$, $\theta > 0$. want a CI for θ.
Let $Y = X_{(n)} = \max(X_1, \ldots, X_n)$.

a) Let $CI = [aY, bY]$, $1 \leq a < b$.
Find the confidence coef of this CI.

b) Let $CI = [Y + c, Y + d]$ $\quad 0 \leq c < d$.
Find the confidence coef of this CI.

LS 81

Soln $F_X(x) = \frac{x}{\theta}$, $0 < x < \theta$

$f_X(x) = \frac{1}{\theta}$, $0 < x < \theta$.

$Y = X_{(n)}$ has pdf $n[F_X(x)]^{n-1} f_X(x)$

$= n \left(\frac{x}{\theta}\right)^{n-1} \frac{1}{\theta} = \frac{n x^{n-1}}{\theta^n} = \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}$, $0<x<\theta$

a) $P(aY \le \theta \le bY) = P\left(a \le \frac{\theta}{Y} \le b\right)$

$= P\left(\frac{1}{b} \le \frac{Y}{\theta} \le \frac{1}{a}\right)$.

$\frac{Y}{\theta}$ is a pivot: pdf is free of θ

$\frac{X}{\theta} \sim U(0,1)$ so $\frac{Y}{\theta}$ has pdf $\underbrace{n(u^{n-1})1 = n u^{n-1}}$, $0<u<1$

$\int_{\frac{1}{b}}^{\frac{1}{a}} n u^{n-1} du = n \frac{u^n}{n} \Big|_{\frac{1}{b}}^{\frac{1}{a}} = \left(\frac{1}{a}\right)^n - \left(\frac{1}{b}\right)^n$.

Take $a=1$ and $1 - \left(\frac{1}{b}\right)^n = 1-\delta$.

so $\left(\frac{1}{b}\right)^n = \delta$ or $b = \frac{1}{\delta^{\frac{1}{n}}}$.

So $\left[Y, \frac{Y}{\delta^{\frac{1}{n}}}\right]$ is a $100(1-\delta)\%$ CI

$0<\delta<1$.

8.5

b) $P(Y+c \le \theta \le Y+d) = P[\theta-d \le Y \le \theta-c] = (*)$

$F_Y(t) = [F_X(t)]^n = \frac{t^n}{\theta^n}$.

$(*) = F_Y(\theta-c) - F_Y(\theta-d) = \frac{(\theta-c)^n - (\theta-d)^n}{\theta^n}$

$= \left(1-\frac{c}{\theta}\right)^n - \left(1-\frac{d}{\theta}\right)^n$. Take $c=0$

to get $1 - \left(1-\frac{d}{\theta}\right)^n = 1-\delta$

or $\delta = \left(1-\frac{d}{\theta}\right)^n$ which is hard

to use since θ is unknown.

could say $\delta \approx \left(1-\frac{d}{Y_{(n)}}\right)^n$.

Regression Olive Ch6 M484, 584, 586
489,

1) multiple linear regression MLR
with $E(e_i)=0$, $V(e_i)=\sigma^2$
constant variance

$$Y_i = Y_i | \underline{x}_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$$
$$= \underline{\beta}^T \underline{x}_i + e_i = \underline{x}_i^T \underline{\beta} + e_i,$$

$i = 1, \ldots, n.$

In matrix form

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1p} \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

or $\underset{n \times 1}{\underline{Y}} = \underset{n \times p}{X} \underset{p \times 1}{\underline{\beta}} + \underset{n \times 1}{\underline{e}}$, $\underset{n \times p}{X} = \begin{pmatrix} \underline{x}_1^T \\ \vdots \\ \underline{x}_n^T \end{pmatrix}$

The ith case is $(\underline{x}_i^T, Y_i)^T$ or $(x_{i2}, \ldots, x_{ip}, Y_i)^T$
omit the $1 \equiv x_{i1}$.

2) The ordinary least squares OLS
estimator $\hat{\underline{\beta}} = \hat{\underline{\beta}}_n = \hat{\underline{\beta}}_{OLS} = (X^T X)^{-1} X^T \underline{Y}$.

3) Inference is conditional on $\underline{x}_i$ or on $\mathbb{X}$; act as if $\mathbb{X}$ and $\underline{x}_i$ are a constant matrix and constant vectors, if not, condition on $\mathbb{X}$ and $\underline{x}_i$.

4) Let $\underset{n \times n}{H} = \underset{n \times p}{\mathbb{X}} (\mathbb{X}^T \mathbb{X})^{-1} \underset{p \times n}{\mathbb{X}^T} = (h_{ij})$.

Let $h_i = h_{ii} = H_{ii}$.

5) Know
<u>OLS CLT</u>. Consider the MLR model $Y_i = \underline{x}_i^T \underline{\beta} + e_i$ where the e_i are iid, $E(e_i) = 0$, $V(e_i) = \sigma^2$. Assume $\max(h_1, \ldots, h_n) \xrightarrow{P} 0$ as $n \to \infty$. Assume $\frac{\mathbb{X}^T \mathbb{X}}{n} \to W^{-1}$ as $n \to \infty$.

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

6) $MSE = \frac{1}{n-p} \sum_{i=1}^{n} r_i^2 = \hat{\sigma}^2$ is a consistent
defined later

estimator of σ^2, $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$, under mild conditions.

$$\hat{\beta}_{OLS} \sim AN_p\left(\beta, MSE \ (X^T X)^{-1}\right)$$

$$\widehat{\sigma^2 W} = MSE \ n (X^T X)^{-1}.$$

7) Let $X = [\underset{\substack{\uparrow \\ v_1}}{1}, v_2, \ldots, v_p]$.

Then $X^T X = (U_{ij})$ where $U_{ij} = v_i^T v_j$

is the matrix of sums and cross products.

$$X^T X = \begin{bmatrix} 1^T 1 & 1^T v_2 & \cdots & 1^T v_p \\ v_2^T 1 & v_2^T v_2 & \cdots & v_2^T v_p \\ \vdots & & & \\ v_p^T 1 & v_p^T v_2 & \cdots & v_p^T v_p \end{bmatrix} =$$

$$\begin{bmatrix} n & \sum_i x_{i2} & \cdots & \sum_i x_{ip} \\ \sum_i x_{i2} & \sum_i x_{i2}^2 & \sum_i x_{i2} x_{i3} \cdots & \sum_i x_{i2} x_{ip} \\ \vdots & & & \\ \sum_i x_{ip} & \sum_i x_{ip} x_{i2} \cdots & \sum_i x_{ip} x_{i,p-1} & \sum_i x_{ip}^2 \end{bmatrix}.$$

83.5

8) Suppose $x_i \perp\!\!\!\perp x_j$ $i \neq j$ ($1 \perp\!\!\!\perp x_j$ always)

$E(x_i) = 0$ $i = 2, \ldots, p$ $E(x_i x_j) = 0, i \neq j$ $i > 1$

$E(x_i^2) = V(x_i) = \tau_i^2, \; i = 2, \ldots, p.$

Then $\dfrac{X^T X}{n} \xrightarrow{P} W^{-1} = \begin{pmatrix} 1 & 0 & \cdots & & 0 \\ 0 & \tau_2^2 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & \cdots & & 0 & \tau_p^2 \end{pmatrix}$

If the nonconstant $x_2, \ldots, x_p$ are iid,

then $\dfrac{X^T X}{n} \xrightarrow{P} W^{-1} = \begin{pmatrix} 1 & \underline{0}^T \\ \underline{0} & \tau^2 I_{p-1} \end{pmatrix}$ $(= I_p$ if $\tau^2 = 1),$

where $E(x_k^2) = V(x_k) = \tau^2, \; k = 2, \ldots, p-1.$

If $x_2, \ldots, x_p$ are RVs, then

$$\frac{X^T X}{n} \xrightarrow{P} W^{-1} = (w_{ij}) = \begin{pmatrix} 1 & E(x_2) & \cdots & E(x_p) \\ E(x_2) & E(x_2^2) & \cdots & E(x_2 x_p) \\ \vdots & & & \\ E(x_p) & E(x_p x_2) & \cdots & E(x_p^2) \end{pmatrix}.$$

$w_{ij} = E(x_i x_j)$

where $i = 1$ or $j = 1$ means $x_i \equiv 1$ or $x_j \equiv 1.$

9) Change of notation:

Let $Y_i = \alpha + \underline{\beta}^T \underset{\substack{\uparrow \\ p\times 1}}{\underline{x}} + e$.

So $\underline{Y} = \mathbb{X} \begin{pmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \underline{e} = \mathbb{X}\, \underset{\sim}{\phi} + \underline{e}$

$\underset{\sim}{\phi} = \begin{pmatrix} \alpha \\ \underset{\sim}{\beta} \end{pmatrix}$. Assume $\begin{pmatrix} Y_i \\ \underline{x}_i \end{pmatrix}$ are iid,

$\underbrace{\underline{x}_i = (x_1, \ldots, x_p)^T}$.

nontrivial predictors were $(x_2, \ldots, x_p)$ for OLS.

This assumption is often reasonable if the $\underline{x}_i$ are random vectors drawn from some population.

84.5

10) $\hat{\underline{\phi}}_{OLS} = \begin{pmatrix} \hat{\alpha}_{OLS} \\ \hat{\underline{\beta}}_{OLS} \end{pmatrix}_{(p+1)\times 1} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\underline{Y}$ } first way

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

11) Second way to compute $\hat{\underline{\phi}}_{OLS}$.

$\hat{\alpha}_{OLS} = \bar{Y} - \hat{\underline{\beta}}^T_{OLS}\,\underset{\text{xbar}}{\bar{\underline{X}}}$

$\hat{\underline{\beta}}_{OLS} = \hat{\Sigma}_{\underline{x}}^{-1}\hat{\Sigma}_{\underline{x}Y} = \tilde{\Sigma}_{\underline{x}}^{-1}\tilde{\Sigma}_{\underline{x}Y}$ } second way

$$\hat{\alpha}_{OLS} \xrightarrow{P} \alpha_{OLS} = E(Y) - \underline{\beta}^T_{OLS}E(\underline{x})$$

$$\hat{\underline{\beta}}_{OLS} \xrightarrow{P} \underline{\beta}_{OLS} = \Sigma_{\underline{x}}^{-1}\Sigma_{\underline{x}Y}$$

Go back to $\underline{Y} = \mathbb{X}\underline{\beta} + \underline{e}$ (not $\underline{\phi} = \begin{pmatrix}\alpha \\ \underline{\beta}\end{pmatrix}$).

12) Let $H = P = \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T$.

Let $\mathbb{X}_I$ contain K columns of $\mathbb{X}$.

Let $H_I = P_I = \mathbb{X}_I(\mathbb{X}_I^T\mathbb{X}_I)^{-1}\mathbb{X}_I^T$.

$H = H^T = H^2$, $H_I = H_I^T = H_I^2$

$H H_I = H_I = H_I H$. $\quad H\mathbb{X}_I = \mathbb{X}_I$, $\mathbb{X}_I^T H = \mathbb{X}_I^T$.

$H_I\mathbb{X}_I = \mathbb{X}_I$, $\mathbb{X}_I^T H_I = \mathbb{X}_I^T$.

13) a) $\underset{\sim}{Y} = X\beta + e$, e_i iid, $E(e_i) = 0$
$V(e_i) = \sigma^2$

b) $\underset{\sim}{Y} = X\underline{\beta} + \underline{e}$, e_i iid $N(0, \sigma^2)$.

Then $\underset{\sim}{Y} = \underset{\sim}{Y} | X \sim N_n(X\underline{\beta}, \sigma^2 I_n)$.

Models a) and b) have similar large sample theory. Model b) often has exact MVN theory.

a)	b)
$\sqrt{n}(\hat{\underline{\beta}} - \underline{\beta}) \xrightarrow{D} N_p(\underline{0}, \sigma^2 W)$	" "
$\sqrt{n}(A\hat{\underline{\beta}} - A\underline{\beta}) \xrightarrow{D} N_k(\underline{0}, \sigma^2 A W A^T)$	" "

A $k \times p$ (often of full rank k)

$\hat{\underline{Y}} = X\hat{\underline{\beta}} = H\underline{Y}$ " "

$\underline{r} = \underline{Y} - \hat{\underline{Y}} = (I - H)\underline{Y}$ " "

$r_i = Y_i - \hat{Y}_i$ = ith residual in MSE b)

$\underset{\sim}{AY} \sim N_k(A\mathbb{X}\underline{\beta}, \sigma^2 A A^T)$

$\hat{\underline{Y}} = H\underline{Y} \sim N_n(\mathbb{X}\underline{\beta}, \sigma^2 H)$

Hand $I-H$ are singular for $n > p$

$H = I_n$ for $n = p$, $\underset{n \times n}{\mathbb{X}}$ full rank

$\underline{r} = (I-H)\underline{Y} \sim N_n(\underline{0}, \sigma^2 (I-H))$

$\underset{\sim}{\hat{\beta}} = \underbrace{(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T}_{A} Y \sim N_p(\underline{\beta}, \sigma^2 (\mathbb{X}^T\mathbb{X})^{-1})$

under a) $\underset{\sim}{\hat{\beta}} \sim AN_p(\underline{\beta}, \sigma^2 (\mathbb{X}^T\mathbb{X})^{-1})$.

14) <u>parametric bootstrap</u>, for MLR

$\underset{\sim}{Y^*} \sim N_n(\mathbb{X}\underline{\hat{\beta}}_n, \sigma_n^2 I_n)$

$\sim N_n(H\underset{\sim}{Y}, \sigma_n^2 I_n)$, $\sigma_n^2 = MSE_r$

uses $\underline{Y}^* = \mathbb{X}\underset{\sim}{\hat{\beta}}_n + \underline{e}^*$, $\underline{e}^* \sim N_n(0, \sigma_n^2 I)$,

e_i^* iid $N(0, \sigma_n^2)$

$\hat{\underline{Y}}^* = H\underline{Y}^* \sim N_n(\mathbb{X}\underline{\hat{\beta}}_n, \sigma_n^2 H)$

$\underline{r}^* = (I-H)\underline{Y}^* \sim N_n(\underline{0}, \sigma_n^2 (I-H))$

Let $\underset{k \times 1}{\hat{\underline{\beta}}_I^*} = (\mathbb{X}_I^T\mathbb{X}_I)^{-1}\mathbb{X}_I^T \underline{Y}^* \sim N_k(\underline{\hat{\beta}}_I, \sigma_n^2 (\mathbb{X}_I^T\mathbb{X}_I)^{-1})$

Since $\hat{\underline{\beta}}_I^* = A\,\underline{Y}^* \sim N(\underline{\mu}_A, \Sigma_A)$

with $\underline{\mu}_A = (X_I^T X_I)^{-1} \underbrace{X_I^T H Y}_{X_I^T Y} = \hat{\beta}_I$

and $\Sigma_A = \sigma_n^2 A I_n A^T = \sigma_n^2 A A^T =$

$\sigma_n^2 (X_I^T X_I)^{-1} X_I^T \; I \; X_I (X_I^T X_I)^{-1}$

$= \sigma_n^2 (X_I^T X_I)^{-1}$.

See Quiz 9 and Old Q9.

15) Back to $Y_i = \alpha + \beta^T \underline{x}_i + e_i$

$\underline{Y} = X \underline{\phi} + \underline{e}$

$\mathrm{cov}(A\underline{x}) = A \Sigma_x A^T = A\,\mathrm{cov}(\underline{x})\,A^T$,

$\hat{\Sigma}_{A\underline{x}} = A \hat{\Sigma}_x A^T$

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

$\mathrm{cov}(A\underline{x}, Y) = \Sigma_{A\underline{x}Y} = A \Sigma_{\underline{x}Y} = A\,\mathrm{cov}(\underline{x}, Y)$

86c5

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

16) $Y = \alpha + \underline{\beta}^T \underset{p\times 1}{\underline{x}} + e$ $(\underline{x}_i, Y_i)$ iid. e_i iid

Partial Least Squares estimators (PLS).

$Y \perp\!\!\!\perp \underline{x} \mid \underbrace{\underline{\beta}^T \underline{x}}_{w_1}$, $Y \perp\!\!\!\perp \underline{x} \mid \underline{x} = (x_1, \ldots, x_p)$

So $Y \perp\!\!\!\perp \underline{x} \mid w_1, w_2, \ldots, w_K$ $K = 1, \ldots, p$

pop PLS models

$Y = \alpha + \underbrace{\underline{\beta}_q^T}_{1\times q} \underbrace{\Lambda_q^T}_{q\times p} \underset{p\times 1}{\underline{x}} + e$ for $q = 1, \ldots, p$

where $\Lambda_p = I$, $\underline{\beta}_p = \underline{\beta}$.

$$\underline{\beta}^T = \underline{\beta}_q^T \Lambda_q^T = [\Sigma_{\underline{x}}^{-1} \Sigma_{\underline{x}Y}]^T$$

$$\Lambda_q \underline{\beta}_q = \Sigma_{\underline{x}}^{-1} \Sigma_{\underline{x}Y} = \underline{\beta} \qquad q = 1, \ldots, p.$$

Fit $\underline{\beta}_q$ with OLS. At the pop models,

let $\underline{z} = \Lambda_q^T \underline{x}$. Then $\underline{\beta}_q = \Sigma_{\underline{z}}^{-1} \Sigma_{\underline{z}Y}$

$$= (\Lambda_q^T \Sigma_{\underline{x}} \Lambda_q)^{-1} \Lambda_q^T \Sigma_{\underline{x}Y}$$

$$\underset{p\times 1}{\underline{\beta}_{qPLS}} = \Lambda_q \underline{\beta}_q = \underline{\beta}_{OLS} = \Sigma_{\underline{x}}^{-1} \Sigma_{\underline{x}Y}.$$

So $\underline{\beta}_{1PLS} = \dfrac{\underline{\Lambda}_1 \underline{\Lambda}_1^T}{\underline{\Lambda}_1^T \Sigma_{\underline{x}} \underline{\Lambda}_1} \Sigma_{\underline{x}Y} = \underline{\beta}_{OLS} = \Sigma_{\underline{x}Y} =$