6) p189
For the cell means model,

$\underset{\sim}{X}$ is full rank, $\underset{\sim}{1}$ is not a column of $\underset{\sim}{X}$, but $\underset{\sim}{1} \in C(\underset{\sim}{X})$

since if $\underset{\sim}{X} = (\underset{\sim}{v_1}, \ldots, \underset{\sim}{v_p})$, then $\underset{\sim}{1} = \sum_{i=1}^{p} \underset{\sim}{v_i}$.

$$\underset{p\times n}{X'}\underset{n\times p}{X} = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & & 0 & 1 & & 1 & & 0 & & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & & 0 & & 1 & & 1 \end{pmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & & \\ 0 & 1 & & \\ \vdots & \vdots & & \\ 0 & 1 & & \\ 0 & \vdots & & 0 \\ \vdots & 0 & & 1 \\ 0 & \vdots & & \vdots \\ 0 & 0 & & 1 \end{bmatrix}$$

$$= \begin{pmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & 0 \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & 0 & n_p \end{pmatrix},$$

$$(X'X)^{-1} = \operatorname{diag}\left(\frac{1}{n_1}, \ldots, \frac{1}{n_p}\right).$$

$$X'\underset{\sim}{Y} = \left(\sum_{j=1}^{n_1} Y_{1j} \quad \sum_{j=1}^{n_2} Y_{2j} \quad \cdots \quad \sum_{j=1}^{n_p} Y_{pj}\right)' = (Y_{1 0} \; Y_{2 0} \cdots Y_{p 0})^T$$

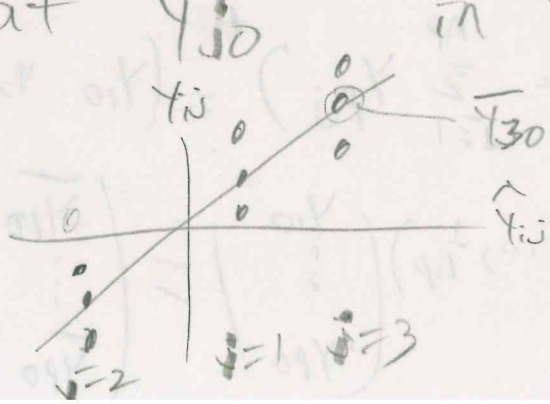
$$\text{so } \hat{\underset{\sim}{\mu}} = \hat{\beta} = (X'X)^{-1}X'\underset{\sim}{Y} = \operatorname{diag}\left(\frac{1}{n_1}, \ldots, \frac{1}{n_p}\right)\begin{pmatrix} Y_{10} \\ \vdots \\ Y_{p0} \end{pmatrix} = \begin{pmatrix} \overline{Y}_{10} \\ \vdots \\ \overline{Y}_{p0} \end{pmatrix}$$

$$\hat{\underline{Y}} = \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y} = \underline{X}\,\hat{\underline{\mu}} =$$

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & 0 & 1 \\ \vdots & & & \\ 0 & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} \overline{Y}_{10} \\ \overline{Y}_{20} \\ \vdots \\ \overline{Y}_{p0} \end{pmatrix} = \begin{pmatrix} \overline{Y}_{10} \\ \vdots \\ \overline{Y}_{10} \\ \overline{Y}_{20} \\ \vdots \\ \overline{Y}_{20} \\ \vdots \\ \overline{Y}_{p0} \\ \vdots \\ \overline{Y}_{p0} \end{pmatrix}$$

So $\hat{Y}_{ij} = \overline{Y}_{i0}$, $\quad i=1,\dots,p, \; j=1,\dots,n_i$.

7) Hence the dot plot for the jth treatment crosses the identity line at $\overline{Y}_{j0}$ in the response plot.



8] One way Anova F test

H_0 $\mu_1 = \cdots = \mu_p$ $\quad H_A$ not H_0

(not all of the p means are equal)

9] For $p=2$ this is the pooled ~~t~~ test

H_0 $\mu_1 = \mu_2$ $\quad H_A$ $\mu_1 \neq \mu_2$.

p 189

10] If H_0 is true, let $\mu_1 = \cdots = \mu_p \equiv \mu$

Then the Y_{ij} are iid,

$Y_{ij} = \mu + \varepsilon_{ij}$ and $\hat{\mu} = \overline{Y}_{00} = \dfrac{\sum\sum_{ij} Y_{ij}}{n}$.

ε_{ij} iid $N(0, \sigma^2)$

From p100, $F = \dfrac{(RSS(H) - RSS)/q}{RSS/(n-p)} \sim F_{q,\, n-p}$ $\quad q = p-1$

To show $q = p-1$, need full rank

Need $\underset{q \times p}{A} \ni A\,\underline{\mu} = \underline{0}$ is equivalent to H_0.

$$\underset{(p-1)\times p}{A} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & \vdots & \vdots \\ \vdots & & & \ddots & 0 \\ 1 & 0 & \cdots & 0 & -1 \end{pmatrix}$$ works since $A\begin{pmatrix}\mu_1 \\ \vdots \\ \mu_p\end{pmatrix} = \begin{pmatrix}\mu_1 - \mu_2 \\ \vdots \\ \mu_1 - \mu_p\end{pmatrix}$.

$RSS = \sum_i \sum_j (Y_{ij} - \overline{Y}_{i0})^2$, $RSS(H) = \sum_i \sum_j (Y_{ij} - \overline{Y}_{00})^2$
$= SSTO$

11) p192 total sum of squares $SSTO = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{00})^2$

treatment sum of squares $SSTR = \sum_{i=1}^{p} n_i (\overline{Y}_{i0} - \overline{Y}_{00})^2$

residual or error sum of squares $SSE = \sum_i \sum_j (Y_{ij} - \overline{Y}_{i0})^2$
$SSTO = SSTR + SSE$, $RSS(H) - RSS = SSTO - SSE = SSTR$.

one way ANOVA table $MS = SS/df$

source	df	SS	MS	F
between or treatment	$p-1$	SSTR	MSTR	$\frac{MSTR}{MSE}$
error	$n-p$	SSE	MSE	

H_0 $\mu_1 = \cdots = \mu_p$ H_A not H_0

$pval = P(F_{p-1, n-p} > F)$.

reject H_0 if $pval < \delta$

fail to reject H_0 if $pval \geq \delta$

12) Rule of thumb: Let $R_1, \ldots, R_p$ be the ranges of the p dot plots. If

$\max(R_1, \ldots, R_p) \leq 2 \min(R_1, \ldots, R_p)$ then the one way Anova F test pval $\approx$ correct.

if the response and residual plots suggest that the remaining one way Anova model assumptions are reasonable.

So the test has some robustness to the assumption $V(\varepsilon_{ij}) \equiv \sigma^2$.

Note: p195 CIs are less robust to this assumption.

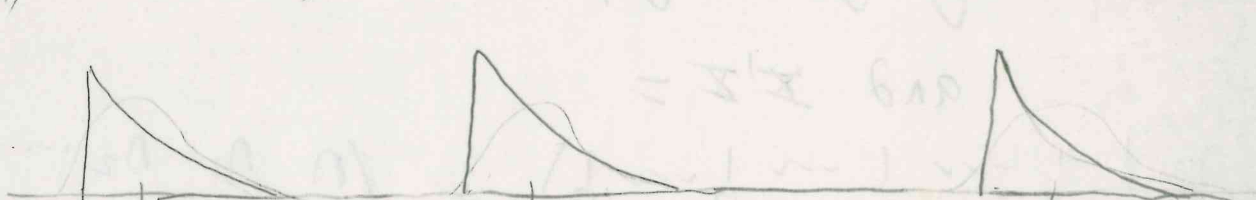
13) $Y_{ij} = \mu_i + \varepsilon_{ij}$, ε_{ij} iid, $E(\varepsilon_{ij}) = 0$, $V(\varepsilon_{ij}) = \sigma^2$

implies that the ε_{ij} have a pdf

$f_\varepsilon(z)$. So the Y_{ij} $(j = 1, \ldots, n_i)$ have pdf

$\underbrace{f_\varepsilon(z - \mu_i)}$ for $i = 1, \ldots, p$ with $\underbrace{\sigma_i^2 \equiv \sigma^2}_{\text{strong assumption}}$

location family, same shape, but different means $\mu_i = E(Y_{ij})$.



The $V(\varepsilon_{ij}) \equiv \sigma^2$ assumption is much stronger for the Oneway Anova model than for the MLR model. The F test is a large sample test if $V(\varepsilon_{ij}) \equiv \sigma^2$.

14) Another X matrix for the one way Anova model adds a constant and deletes the last column of the X_c for the cell means model.

$$\underline{Y} = X\,\underline{B} + \underline{\epsilon}, \quad \underline{B} = \begin{pmatrix} B_0 \\ B_1 \\ \vdots \\ B_{p-1} \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & & 0 \\ 1 & 0 & 1 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & & 0 \\ 1 & 0 & 0 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & & 1 \\ 1 & 0 & 0 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & & 0 \end{pmatrix}. \quad \text{so } X'\underline{Y} = \begin{pmatrix} Y_{00} \\ Y_{10} \\ \vdots \\ Y_{p-1,0} \end{pmatrix}$$

and $X'X =$

$$\begin{pmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 & \cdots & 1 & \cdots & 1 \\ 1 & \cdots & 1 & 0 & \cdots & 0 & & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 1 \cdots 1 & 0 & \cdots & 0 \end{pmatrix} X = \begin{pmatrix} n & n_1 & n_2 & \cdots & n_{p-1} \\ n_1 & n_1 & 0 & \cdots & 0 \\ n_2 & 0 & n_2 & 0 \cdots & 0 \\ \vdots & & & & \\ n_{p-1} & 0 & \cdots & 0 & n_{p-1} \end{pmatrix}.$$

$$\underset{\sim}{X}'\underset{\sim}{X} = \left(\begin{array}{c|c} n & (n_1 \; n_2 \; \cdots \; n_{p-1}) \\ \hline \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_{p-1} \end{pmatrix} & \text{diag}(n_1, \dots, n_{p-1}) \end{array}\right)$$

$$(\underset{\sim}{X}'\underset{\sim}{X})^{-1} = \frac{1}{n_p}\begin{bmatrix} 1 & -1 & \cdots & -1 \\ -1 & 1+\frac{n_p}{n_1} & \cdots & 1 \\ \vdots & 1 & 1+\frac{n_p}{n_2} & \vdots \\ & \vdots & \ddots & 1 \\ & & & 1+\frac{n_p}{n_{p-1}} \\ -1 & 1 & \cdots \; 1 & \end{bmatrix}$$

$$= \frac{1}{n_p}\begin{pmatrix} 1 & -\underset{\sim}{1}' \\ -\underset{\sim}{1} & \underset{\sim}{1}\,\underset{\sim}{1}' + \underbrace{\text{diag}\left(\frac{n_p}{n_1}, \dots, \frac{n_p}{n_{p-1}}\right)}_{\text{typo in 134)}} \end{pmatrix}.$$

$\frac{n_p}{n_i}$ not $1+\frac{n_p}{n_i}$

$$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y} = \begin{pmatrix} \bar{Y}_{p0} \\ \bar{Y}_{1\cdot} - \bar{Y}_{p0} \\ \vdots \\ \bar{Y}_{p-1,\cdot} - \bar{Y}_{p0} \end{pmatrix}, \quad \underset{\sim}{\beta} = \begin{vmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{vmatrix} = \begin{pmatrix} \mu_p \\ \mu_1 - \mu_p \\ \vdots \\ \mu_{p-1} - \mu_p \end{pmatrix}$$

15) This model is interesting since the oneway Anova F test

$H_0\ \mu_1 = \cdots = \mu_p$ vs H_A not H_0

corresponds to the MLR Anova F test

$H_0\ \beta_1 = \cdots = \beta_{p-1} = 0$ vs H_A not H_0.

16) p192 A contrast $\theta = \sum_{i=1}^{p} c_i \mu_i = \underset{\sim}{c}' \underset{\sim}{\mu}$.
where $\sum_{i=1}^{p} c_i = 0$.

A $100(1-\delta)\%$ CI for θ is

$$\sum_{i=1}^{p} c_i \overline{Y}_{i0} \pm t_{n-1, 1-\frac{\delta}{2}} \sqrt{MSE} \sqrt{\sum_{i=1}^{p} \frac{c_i^2}{n_i}}.$$

Assuming groups are independent,

$$V\left(\sum_{i=1}^{p} c_i \overline{Y}_{i0}\right) = \sum_{i=1}^{n} c_i^2 V(\overline{Y}_{i0}) = \sum_{i=1}^{n} c_i^2 \frac{\sigma^2}{n_i},$$

so $SE\left(\sum_{i=1}^{p} c_i \overline{Y}_{i0}\right) = \sqrt{MSE} \sqrt{\sum_{i=1}^{p} \frac{c_i^2}{n_i}}$.

p193
17) Can use Bonferonni CIs and Scheffe CIs for k contrasts where want P(all k CIs contain θ_j, $j = 1, \ldots, k$) $\geq 1-\delta$.

Inference After Variable Selection and Lasso,
see ch4 of online notes

1] One simple method for inference after variable selection is data splitting with 2 sets: let the training set have $n_T \le \frac{n}{2}$ cases and the validation set have $n_V = n - n_T \ge \frac{n}{2}$ cases. Select the n_T cases without replacement from the n cases.

Assume the cases are independent and follow a statistical model, eg MLR,

I) Build the model I with the training set, possibly using variable selection and using the response to select predictors and predictor transformation.
Let model I have k_1 predictors.

II) Act as if I is the full model for the validation set. want $n \ge 5k$ and preferably $n \ge 10k_1$.
Need model I to be a good model for the data.

*2] Variant: use, for example, $\frac{n}{10}$ cases for the training set. If you can not get a good model, select $\frac{n}{10}$ cases from the

validation set for the new training set. Cases that remain are the new validation set.
Repeat until I is a good model

or $n_T \le \frac{n}{2}$ with $n_T \approx \frac{n}{2}$.

$\frac{n}{10}$ $\quad \frac{2n}{10}$ $\quad \frac{3n}{10}$ $\quad \frac{4n}{10}$ $\quad \frac{5n}{10}$ training

$\frac{9n}{10}$ $\quad \frac{8n}{10}$ $\quad \frac{7n}{10}$ $\quad \frac{6n}{10}$ $\quad \frac{5n}{10}$ validation.

3) Data splitting can work for $n \ge 10p$ and $n \ll 10p$. Inefficient inference is much better than invalid inference. Efficiency $\approx \frac{n_V}{n} = 1 - \frac{n_T}{n}$.

4) The bootstrap is useful if $n \ge 10p$.
The bootstrap is used for tests, CIs and confidence regions.

5) Suppose $z_1, \ldots, z_n$ are iid and there is a statistic $\underset{\substack{p \times m \\ p \times 1 \\ 1 \times 1}}{T} = T(z_1, \ldots, z_n)$,

Suppose we could gather B iid samples

$$T^{(1)} = T(\underset{\sim}{z}_1^{(1)}, \ldots, \underset{\sim}{z}_n^{(1)})$$

$\vdots$

$$T^{(B)} = T(\underset{\sim}{z}_1^{(B)}, \ldots, \underset{\sim}{z}_n^{(B)}),$$ If B is large

could examine $T^{(1)}, \ldots, T^{(B)}$ for inference.

6) The <u>empirical distribution</u> gives probability $\frac{1}{n}$

to the iid data $\underset{\sim}{z}_1, \ldots, \underset{\sim}{z}_n$
$\qquad\qquad\qquad\qquad \frac{1}{n} \qquad \frac{1}{n}$

Let $\underset{\sim}{w}$ be from the emp. dist.

$E(\underset{\sim}{w} - E\underset{\sim}{w})(\underset{\sim}{w} - E\underset{\sim}{w})^T = E\underset{\sim}{w}\underset{\sim}{w}^T - E(\underset{\sim}{w})(E\underset{\sim}{w})^T$

(so $E\underset{\sim}{w} = \sum_{i=1}^n \underset{\sim}{z}_i \frac{1}{n} = \bar{\underset{\sim}{z}}$ and $E\underset{\sim}{w}\underset{\sim}{w}^T = \sum_{i=1}^n \underset{\sim}{z}_i\underset{\sim}{z}_i^T \frac{1}{n}$), $\text{Cov}\,\underset{\sim}{w} = \frac{1}{n}\sum(\underset{\sim}{z}_i - \bar{\underset{\sim}{z}})(\underset{\sim}{z}_i - \bar{\underset{\sim}{z}})^T$

So draw $\underset{\sim}{w}$ from the empirical distribution

and $P(\underset{\sim}{w} = \underset{\sim}{z}_i) = \frac{1}{n}$ $\quad i = 1, \ldots, n$.

Sample with replacement to get

$\underset{\sim}{w}_1, \ldots, \underset{\sim}{w}_n$ which are n iid observations

from the empirical distribution.

7) Suppose iid sample is $Y_1, \ldots, Y_n$.

Then the empirical cdf is

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{I[Y_i \le x]}_{\substack{1 \text{ if } Y_i \le x \\ 0 \text{ if } Y_i > x}}$$

So $I[Y_i \le x]$ are iid binomial$(1, p)$

where $p = P(Y_i \le x) = F_Y(x)$.

So $E[I(Y_i \le x)] = F_Y(x)$ and $V[I(Y_i \le x)] = F_Y(x)(1 - F_Y(x))$,

For fixed x, $\hat{F}(x)$ is a sample mean, so

by the CLT $\sqrt{n}\left(\hat{F}(x) - F_Y(x)\right) \xrightarrow{D} N\Big[0, \underbrace{F_Y(x)(1 - F_Y(x))}_{\in [0, \frac{1}{4}]}\Big]$

So $\hat{F}(x)$ is a $\sqrt{n}$ consistent estimator of $F_Y(x)$,

g) For many statistics $T(\underline{z}_1, \ldots, \underline{z}_n) - T(\underline{w}_1, \ldots, \underline{w}_n) \xrightarrow{P} 0$

where the $\underset{\sim}{w}_i$ are drawn with replacement from the empirical distribution of $\underline{z}_1, \ldots, \underline{z}_n$,

9) Suppose T is a vector (or T is $p \times m$ and look at T_{ij}). Let

$T = T(\underline{Z}_1, \ldots, \underline{Z}_n)$ where $\underline{Z}_1, \ldots, \underline{Z}_n$ are iid.

Let $\underline{Z}_{11}^*, \underline{Z}_{12}^*, \ldots, \underline{Z}_{1n}^*$, $T_1^* = T(\underline{Z}_{11}^*, \ldots, \underline{Z}_{1n}^*)$

$\underline{Z}_{21}^*, \underline{Z}_{22}^*, \ldots, \underline{Z}_{2n}^*$, $T_2^* = T(\underline{Z}_{21}^*, \ldots, \underline{Z}_{2n}^*)$

$\vdots$

$\underline{Z}_{B1}^*\ \underline{Z}_{B2}^* \cdots \underline{Z}_{Bn}^*$, $T_B^* = T(\underline{Z}_{B1}^*, \ldots, \underline{Z}_{Bn}^*)$

Bth bootstrap sample

ex) data $Z_1, \ldots, Z_7 = 1, 2, 3, 4, 5, 6, 7$

$T = \text{median}(Z_1, \ldots, Z_7) = 4$

Let $B = 2$

(2, 2, 2, 3, 3, 5, 6 ← ordered)

1st: 3, 2, 3, 2, 5, 2, 6 $\quad T_1^* = 3$

2nd: 3, 5, 3, 4, 3, 5, 7 $\quad T_2^* = 4$

(3 3 3 4 5 5 7 ← ordered)

Let $T_{1n}, \ldots, T_{Bn}$ be iid with same dist as statistic T_n.

10) $\sqrt{n}(T_1^* - T_n), \ldots, \sqrt{n}(T_B^* - T_n)$ are often pseudodata for $\sqrt{n}(T_{1n} - \underline{\theta}), \ldots, \sqrt{n}(T_{Bn} - \underline{\theta})$

$p \times 1 \rightarrow \overline{T}^* = \frac{1}{B} \sum_{i=1}^{B} T_i^*$ $\quad S_{T^*}^2 = \frac{1}{B-1} \sum (T_i^* - \overline{T}^*)(T_i^* - \overline{T}^*)^T$

11] P=1
($I+P=1$) $T^*_{(1)}, T^*_{(2)}, \ldots, T^{(*)}_{(B)}$

a) $100(1-\delta)\%$ percentile CI for $E(T)=\theta$

discards smallest and largest $100\frac{\delta}{2}\%$ of $T^*_{(i)}$

or b) compute the shorth$(c=\lceil B(1-\delta)\rceil)$ interval of the $T^*_{(i)}$

$\left[T^*_{(\lceil B\frac{\delta}{2}\rceil)}, T^*_{(\lceil B(1-\frac{\delta}{2})\rceil)}\right]$ eg $1-\delta=0.9$, $\delta=0.1$, $\frac{\delta}{2}=0.05$

$\left[T^*_{(\lceil B(0.025)\rceil)}, T^*_{(\lceil B\,0.975\rceil)}\right]$ is 95% CI for θ

reject $H_0\ \theta=\theta_0$ if θ_0 is outside the CI.

12] Sample indices $1,\ldots,n$ with replacement

get $\underset{\sim}{i}_1$, $\ldots$ $\underset{\sim}{i}_n$ eg $i_1=$

nonparametric bootstrap (empirical, naive, rowwise).

ex $n=6$ $\underset{\sim}{i}_1 = 3, 2, 3, 2, 5, 6$

so use $(Y_3, \underset{\sim}{x}_3), (Y_2, \underset{\sim}{x}_2)\ (Y_3, \underset{\sim}{x}_3), (Y_2, \underset{\sim}{x}_2)\ (Y_5, \underset{\sim}{x}_5)\ (Y_6, \underset{\sim}{x}_6)$

in the 1st bootstrap sample.

For MLR and $T^*_i = \hat{\underset{\sim}{\beta}}^*_i$, this should work well

If $\underline{x}_i = (1, \underline{u}_i^T)^T$ and the

$\underline{z}_i = (Y_i, \underset{\sim}{u}_i^T)^T$ are iid from some distribution

with nonsingular covariance matrix $\Sigma_{\underset{\sim}{z}}$.

This regularity condition is strong eg $\underset{\sim}{z}_i \sim N_p(\underline{\mu}, \sigma^2 I)$.

13) $Y_i = E(Y_i) + \epsilon_i = \hat{Y}_i + r_i$

The residual bootstrap samples the residuals

$r_1, \ldots, r_n$ with replacement giving

$r_{11}^*, \ldots r_{1n}^*$ $\quad Y_{1i}^* = \hat{Y}_i + r_{1i}^*, i=1,\ldots,n$ $\quad \underset{\sim}{\hat{\beta}}_1^* = \hat{\beta}$ from regressing $\underset{\sim}{Y}_1^*$ on X

$\vdots$

$r_{B1}^*, \ldots, r_{Bn}^*$ $\quad Y_{Bi}^* = \hat{Y}_i + r_{Bi}^*$ $i=1,\ldots,n,$ $\underset{\sim}{\hat{\beta}}_B^* = \hat{\beta}$ from regressing $\underline{Y}_B^*$ on X,

As $B \to \infty$ $\quad \underbrace{\widehat{\text{cov}}(\underset{\sim}{\hat{\beta}})}_{\text{bootstrap est of cov}(\hat{\beta}_{OLS})} = \frac{1}{B-1} \sum_{i=1}^{B} (\underset{\sim}{\hat{\beta}}_i^* - \underset{\sim}{\bar{\beta}}^*)(\underset{\sim}{\hat{\beta}}_i^* - \underset{\sim}{\bar{\beta}}^*)^T$

$\to \frac{\sum_{i=1}^n r_i^2}{n} (X'X)^{-1}$. Usual estimator is $\underbrace{\frac{\sum_{i=1}^n r_i^2}{n-p}}_{MSE} (X'X)^{-1}$

14) Idea for variable selection

(regulization methods, lasso, ridge regression etc) where $n \geq 10p$. Let $\underline{\beta} = (\beta_1, \ldots, \beta_p)'$.

consider automated variable selection

eg model $I = I_i$ ie $I = I_{min}$ with forward selection,

often fit $\underset{\sim}{Y} = X_{I_{min}} \underset{k \times 1}{\underset{\sim}{\beta_{I_{min}}}} + \varepsilon$ with MLR

and use $\text{cov}(\hat{\beta}_{I_{min}}) = \underbrace{\left(\frac{1}{n-k} \sum_{i=1}^{n} r_{i\,I_{min}}^2 \right)}_{\widehat{MSE}(I_{min})} (X_{I_{min}}' X_{I_{min}})^{-1}$,

which is an incorrect estimator,

Instead get bootstrap data as in $\boxed{B}$

do variable selection to get model $I^*_{min,i}$, $i = 1, \ldots, B$.

Then $\underset{p \times 1}{\underset{\sim}{\hat{\beta}^*_{I_{min},0,i}}}$ has the k_i values of $\hat{\beta}_{I_{min},i}$ and $p - k_i$ 0's

eg $p = 5$ $\underset{\sim}{\hat{\beta}^*_{I_{min}}} = \begin{pmatrix} 1. \\ 3.7 \\ 4.3 \end{pmatrix}$ uses $x_1 \equiv 1$, x_3 and x_5

$\underset{p \times 1}{\hat{\beta}^*_i} = \hat{\beta}^*_{I_{min},0,i} = \begin{pmatrix} 1. \\ 0 \\ 3.7 \\ 0 \\ 4.3 \end{pmatrix}$, Compute $\hat{\beta}^*_1, \ldots, \hat{\beta}^*_B$

$\overline{\beta}^* = \frac{1}{B} \sum_{i=1}^{B} \hat{\beta}^*_i$ and $\widehat{\text{cov}}(\underset{\sim}{\hat{\beta}}) = \frac{1}{B-1} \sum_{i=1}^{B} (\hat{\beta}^*_i - \overline{\beta}^*)(\hat{\beta}^*_i - \overline{\beta}^*)^T$

Display cov