$\beta_0 + \beta_{\varphi} = \Lambda^T \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{\varphi} \end{bmatrix}$

$\Lambda^T = [1\ 0\ 0\ 0\ 1]$

$\Lambda = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ $X^T = [\ //\ ]$

Linearly dependence $\Rightarrow \beta_\varphi \Rightarrow NA$ in R

Estimable

#1

a) $\Lambda^T \beta = p^T X \beta \Rightarrow \Lambda^T = p^T X$

$\Updownarrow$

$\Lambda = X^T p$

$\Updownarrow$

$\Lambda \in span(X^T)$

$\beta_1 = \Lambda^T \beta$

$\Lambda^T = [0\ 1\ 0\ 0\ 0]$

$\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

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

Not estimable

b) $\hat{\sigma} = RSE = \sqrt{\frac{RSS}{n-p-1}}$

$= \sqrt{\frac{\sum\sum (y_{ij} - \bar{y}_i)^2}{n-t}}$

$sd = \frac{1}{n_i - 1} \sum (y_{ij} - \bar{y}_i)^2$

$s_0 = \sqrt{\frac{\sum_{i=1}^{4} (n_i - 1) sd_i}{80-4}}$

$= \sqrt{\frac{19(5.026) + 19(4.438) + 19(5.265) + 19(4.193)}{76}}$

$\beta_2 + \beta_3 - 2\beta_1 = \Lambda^T \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_4 \end{bmatrix}$

$\Lambda^T = [0\ -2\ 1\ 1\ 0]$

$\Lambda = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \\ 0 \end{bmatrix}$ $X^T = [\ //\ ]$

Estimable

Scanned by CamScanner

c)

$$y_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$$

same

$$\frac{RSS(m) - RSS(M) / t-1-2}{RSS(M) / n-t}$$

$$\Rightarrow \underset{\overset{\|}{TSS}}{RSS(m)} = \underbrace{\sum\sum (y_{ij} - \bar{y}_i)^2}_{=RSS} + \underbrace{\sum n_i (\bar{y}_i - \bar{y})^2}_{RegSS}$$

$H_0 \Rightarrow$

- $\beta_1 = \beta_2 \rightarrow \beta_1 - \beta_2 = 0$
- $\beta_2 = \beta_3 \rightarrow \beta_2 - \beta_3 = 0$
- $\beta_3 = \beta_4 \rightarrow \beta_3 - \beta_4 = 0$

$$L: \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_4 \end{bmatrix} \quad c = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

use

$$\frac{(L\hat{\beta}-c)^T [L(X^TX)^{-1}X^T]^{-1}(L\hat{\beta}-c)}{\lambda\hat{\sigma}^2}$$

$\lambda \rightarrow 3$, $\hat{\sigma}^2 \rightarrow$ the one from b)

$\sim F_{3,\ 80-4}$

d) Last year $\Downarrow$ This year $\rightarrow$ prediction

$\hat{\tau}_2 = \bar{y}_2$

$\hat{y}_{i0} = \hat{\mu}_i$

$\bar{y}_i = \hat{\mu} + \hat{\tau}_i$

Mike $\sim x_0^T\hat{\beta} \pm \hat{\sigma}\sqrt{1+x_0(X^TX)^{-1}x_0}\ t_{n-p-1}(97.5\%)$

$x_0^T\hat{\beta} \rightarrow \hat{\mu}_2$

$\Rightarrow$

$\hat{\tau}_1 = \bar{y}_1 - \hat{\mu}$

$$x_0^T = [1\ 0\ 1\ 0] \in R^{1\times 4}$$

$$\hat{\beta} = \begin{bmatrix} 66.2 \\ 58.84 - 66.23 \\ 64.73 - 66.23 \\ 64.06 - 66.23 \end{bmatrix} \in R^{4\times 1}$$

Then $x_0(X^TX)^{-1}x_0$ will be the same

and x_0 will be [0 0 1 0]

or change to X with no intercept

$\boxed{\#2.}$

a)

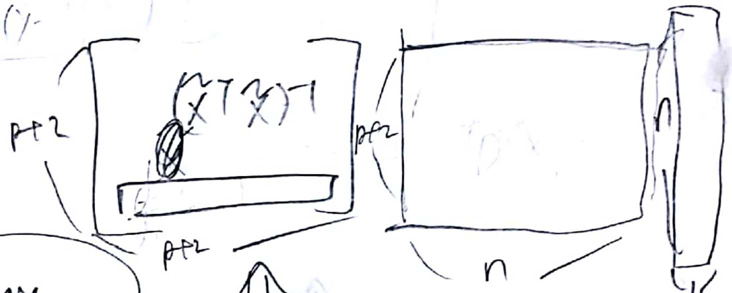
$y_1 = x_1^T\beta + e_1$

$\vdots$

$y_k = x_k^T\beta + r + e_k$

$\vdots$

$y_p = x_p^T\beta + e_p$



$p_i = 0$ anyway...

$M_2 = \begin{bmatrix} 1 & x_{11} & \cdots & r \\ \vdots & & & \vdots \\ 1 & x_{n1} & \cdots & \end{bmatrix}$ → only kth r matters

No.

~~Yes~~ => so basically when we think r makes the observation being an outlier it is a reasonable thought.

But problem is how to get $(\tilde{X}^T\tilde{X})^{-1}$....

and, $\hat{r}$ will be the last element of $\tilde{\beta} \in R^{p+2,1}$

b) min $\|y - X\beta - Dr\|^2$ → # → vector

But since r is just # we can just cbind and say $\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \\ r \end{bmatrix}$

and get $\tilde{\beta} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^Ty$ where X is $n \times (p+2)$ matrix

$\hookrightarrow (p+2, 1)$

#C)

$\hat{e}_{[k]} = \frac{\hat{e}_i}{1-h_{ii}} \rightarrow \frac{\hat{e}_{[k]}\sqrt{1-h_{ii}}}{\hat{\sigma}}$

$\hat{\beta}_{[k]} = Y_k - x_k^T \hat{\beta}_{[k]}$

$\frac{\hat{e}_{[k]}\sqrt{1-h_{kk}}}{\hat{\sigma}} \sim t_k$

$\|$

$Y_k - x_k^T \hat{\beta}_{[k]}$

$\frac{\hat{e}_{[k]}\sqrt{1-h_{kk}}}{\sqrt{RSS_{[k]}/n-p-2}} \rightarrow M_1$

$M_2 \Rightarrow \frac{\hat{r}}{SE(\hat{r})}$ $\frac{\hat{e}_k}{\hat{\sigma}\sqrt{1-h_{kk}}}$

$\boxed{\text{should use } \hat{r} \text{ we got in b) } \ldots\ldots}$

$\frac{\hat{e}_k}{\hat{\sigma}\sqrt{1-h_{kk}}} = \frac{\hat{\sigma}(I-H)}{\hat{r}\hat{\sigma}}$

#3.

a)

① $\frac{-0.05529}{0.28110} \approx \boxed{-0.1967}$

② $\hat{\sigma} = \sqrt{\frac{RSS}{246}}$

$= \sqrt{\frac{7571.938}{246}} \approx \boxed{5.669}$

③

$\text{adj } R^2 = 1 - \frac{RSS/n-p-1}{TSS/n-1}$

$= 1 - \frac{7571.938}{TSS} \cdot \frac{n-1}{n-p-1}$

$= 1 - \frac{7571.938}{TSS} \cdot \frac{251}{245}$

$= \boxed{0.5687}$

so, $0.43096 = \frac{7571.938}{TSS}$

④ $TSS = 17578.4979$

$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{7571.938}{17578.4979}$

$\approx \boxed{0.569}$

$\frac{RSS(m) - RSS(M)/6}{RSS(M)/n-p-1}$

⑤ $\frac{R^2}{1-R^2} \cdot \frac{281}{6}$

$\boxed{53.965}$ $\boxed{56.228}$

6) c)

1. False → only one variable included ...
2. True → intercept always included ...
3. True
4. True
5. False

6. $\frac{RSS(m)}{\hat{\sigma}^2} - (n-2-2p(m))$ ↳ $\sqrt{\frac{RSS(M)}{n-p-1}}$ $248 = n-p-1$

$= \frac{10985.974}{(5.559)^2} - (252-2-2\cdot 1)$

$\approx \boxed{107.505}$

7. $adj\,R^2 = 1 - \frac{RSS/n-p-1}{TSS/n-1}$ 252

$= 1 - \frac{RSS(m)/(252-3-1)}{TSS/251} = 1 - \frac{8237.182/248}{(19578.9060/251)}$

d)

4-th model

⇒ Age, weight, height, wrist

lowest mallow C_p

highest Adj R^2

e) $F = \frac{RSS(m)-RSS(M)/3}{RSS(M)/248} \sim F_{3,248}$

$H_0: \beta_3 = \beta_4 = \beta_6 = 0$

$\frac{8237.182 - 7571.938/3}{7571.938/248}$

$\approx \boxed{7.1749}$ → Reject the null

8. ~~$1 - \frac{RSS}{TSS} = 0.46$~~ $\boxed{RSS = 9411.313799}$

$1 - \frac{RSS/252-2-1}{19578.90607/251} = 0.463249$

$\frac{RSS \cdot 281}{19578.90607 \cdot 249}$

$\boxed{0.525748}$

Scanned by CamScanner

#4.

a) $\hat{\beta} \sim N(\beta, (X^TWX)^{-1})$

$$\beta^{(m+1)} = \beta^{(m)} - (\underbrace{H\ell(\beta^{(m)})}_{-X^TWX})^{-1} \underbrace{\nabla \ell(\beta^{(m)})}_{=X^T(y-p)}$$

$$= \beta^{(m)} + (X^TWX)^{-1} X^T(y-p)$$

$$\hat{\beta} - \beta \simeq (X^TWX)^{-1} X^T(y-p)$$

$$\mathrm{cov}(\hat{\beta}) \simeq (X^TWX)^{-1}$$

① $\dfrac{4.11947}{0.36342} \simeq 11.335287$

② $\dfrac{xxx}{0.028} = 12.348 \rightarrow 0.3467$

③ $-2\sum y_i \ln \hat{p}_i + (1-y_i)\ln(1-\hat{p}_i)$

$= -2\sum \cancel{y_i} \log \hat{p}_i + (1-\cancel{y_i}) \ln(1-\hat{p}_i)$

$\hat{p}_i = \dfrac{e^{\hat{\beta}_0}}{1+e^{\hat{\beta}_0}} = \bar{y}$

$= -2n[\bar{y}\ln\bar{y} + (1-\bar{y})\ln(1-\bar{y})]$

$\dfrac{1813}{4601} = \bar{y}$ $\quad n = 4601$

$\Rightarrow$ so, 6170.153 $\rightarrow$ d.f. $n-1$

④ $= 4600$

⑤ $n-p+1$

$= n-7$

$= 4601-7$

$= 4594$

⑥ AIC $\Rightarrow$

Residual dev $+ 2(p(m)+1)$

$= 3245.1 + 2(7)$

$= 3259.1$

b) $\hat{p}_i = \dfrac{e^{x_i^T\hat{\beta}}}{1+e^{x_i^T\hat{\beta}}}$ $= 22.8814$

$= e^{[1 \ \ln(167) \ \ \ln(0.869) \ \ \ln(2.895)]}$

$\ln(0.001)$
$\ln(0.001)$
$\ln(0.001)$

$$\hat{\beta} = \begin{bmatrix} 4.11947 \\ 0.30228 \\ 0.32686 \\ 0.40981 \\ 0.34863 \\ 0.18947 \\ -0.11418 \end{bmatrix}$$

$0.968 \quad 1$

Scanned by CamScanner

#5.

a)

on the plot

(1) 246

(2) 17

(3) 418

(4) 0.2297

(5) 0.7702

b) $RSS(T) = \sum\sum (y_{ij} - \bar{y}_i)^2$

$C_\alpha(T) = \underbrace{C_0(T)}_{=RSS} + \alpha\, \underbrace{C(T_{root})}_{=TSS} |T|$

$\frac{\#error}{n}$

c) still # of variables = 6

so residual deviance is analogous to AIC!!!

I prefer M1.

d) $Recall = \frac{a}{a+b}$

$precision = \frac{a}{a+c}$

pre / obs	T	F
T	a	b
F	c	d

$\frac{RSS}{TSS}$

$\frac{RSS}{TSS}$

$M_1 \Rightarrow$ High Recall !!! & High precision

$\frac{246 + 106 + 12 + 17 + 383}{\#601}$ ≈ 0.138

In printcp ⇒ $RSS \cdot \left(\frac{RSS}{TSS}\right)$

c) RSS

not really...

~~You~~ can do better with $cp = 0.010000$

d) (b) Don't they have 6 have 3 splits?

They are on boundary...

⇒ If I do 0.011884, I will get 3 splits!!!

pre / obs	T	F
T	a	b
F	c	d

$\Rightarrow$ precision $\Rightarrow \frac{a}{a+c} \Rightarrow \boxed{0.8287}$

recall $\Rightarrow \frac{a}{a+b} \Rightarrow \boxed{0.80585}$

pre / obs	Y	N	
Y	69+395+997	246+106	1813
N	48+121+133	2174+312	2788
			4601

#6.

a)

(1) $\boxed{14}$

(2) $\boxed{10}$

(3) $\boxed{15}$

(4) $\frac{5}{16}$

(5) $\frac{11}{16}$

b) $C_0(T)$ in $C_\alpha(T) = C_0(T) + \alpha\, C(T_{root})|T|$

$$\frac{1+14+2+5+2+14}{212}$$

$\approx \boxed{0.17924}$

c) root tree (?)

$\Rightarrow |T| = 1$

$C_\alpha(T_{root}) = RSS(T) + \alpha|T|TSS = \frac{79}{212} + \alpha \cdot \frac{79}{212}$

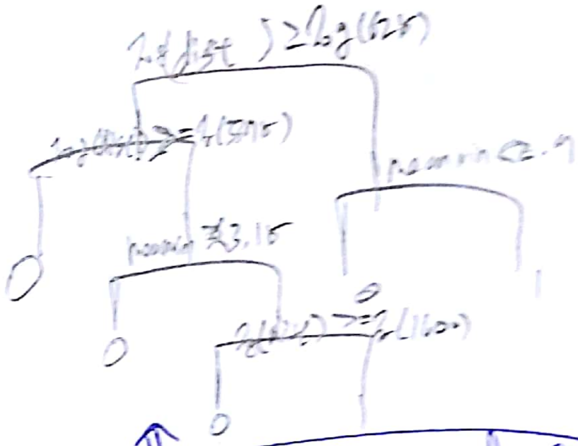
When $|T| = 0$

$C_\alpha(T_{tree}) = 0.17924 + \alpha \cdot 6 \cdot \frac{79}{212}$

So, $0.17924 + (6)\frac{79}{212}\alpha \geq \frac{79}{212}$

$\boxed{\alpha \geq \cancel{0.0865}}$ or $\boxed{0.1038}$

Scanned by CamScanner



So, it will be just by order

$\Delta I = \frac{1}{n}[n I(G) - n_R I(G_R) - n_L I(G_L)]$

For example, for the first split, distance >= 12.5!!! we got ...

So, although we log(distance) the splitting will not be that different

a b
c d =>

d)

obs \ pre	1	0	
1	11+49	1+14 +2+2	79
0	5+14	29+62 +13+10	133
			212

Precision => $\frac{a}{a+c} = \frac{11+49}{11+49+5+14}$

Recall => $\frac{a}{a+b} = \frac{11+49}{79}$

e)

X =>

Gini in R to decide which variable

$I(G_R) = \sum p_R(1-p_R)$

#7.

a)

$Y = X\beta + e$

$\mathrm{cov}(W^{-1/2}e) = W^{-1/2}\, W\, (W^{-1/2})^T$

$= I$

$W^{-1/2}Y = W^{-1/2}X\,\beta + W^{-1/2}e$

$\hat\beta = (X^TX)^{-1}X^TY$

$= (X^TW^{-1/2}W^{-1/2}X)^{-1}X^TW^{-1/2}W^{-1/2}Y$

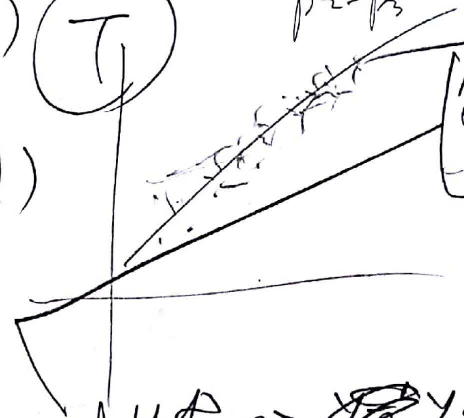
$= (X^TW^{-1}X)^{-1}X^TW^{-1}Y$

F

b) T

c) T

d)



$\beta_2 - \beta_3$

$\hat e = (I-H)Y = (I-H)e$

$\hat e_{(i)} = \frac{\hat e_i}{1-h_{ii}}$

If $H\uparrow \rightarrow Y \approx \hat Y \rightarrow \hat e \approx 0$

$\hat e_{(i)}$ might go up/down.

T

e) T

f) T

g)

$\nabla \ell(\beta) = X^T(Y-p) = 0$

$\sum y_i = \sum \hat p_i$

if there is an intercept

T

h) $X^TX\beta = X^TY$

$X^T(Y-\hat Y) = 0$

F

(?) Does fitted probability exist in linear regression?

i) F $\Rightarrow$ newton Raphson

j) $\hat{\beta}^{(m+1)} = \hat{\beta} - (H\,\ell(\beta^{(m)}))^{-1}\nabla \ell(\beta^{(m)})$

$\min(n,p)=1$

$X^T G$

F

$Z = (X^T W X)^{-1} X^T (\ell p)$

$Z = \ln\frac{p_i}{1-p_i} + \frac{y_i - p_i}{p_i(1-p_i)}$

$(X^T X)^{-1} X^T y$

lm($Z \sim X$ weight)

β

$X^T W Z$

True

$\hat{\beta} = (X^T W X)^{-1} X^T W Z$

k) True $\Rightarrow$ How we did it ...

l) True

$C = \frac{r_i^2}{p_{ii}} \cdot \frac{h_{ii}}{1-h_{ii}}$

$\Rightarrow$ 'or' should be 'and' to be true ...

m) singular = X invertible

adj $R^2 = 1 - \frac{RSS/n-p-1}{TSS/n-1}$

$RSS \Rightarrow \sum(y_i - \hat{y}_i)^2$

$TSS \Rightarrow \sum(y_i - \bar{y})^2 \rightarrow$ defind

unique

n)

precision = $\frac{T \& T}{\text{prediced } T}$

recall = $\frac{T \& T}{\text{Observed } T}$

F $\Rightarrow$ depends ...