#1

a) Linearly dependence

b) Yes, when weight increase 1 unit, body fat increases 0.2434 / unit

c) RSS should be still

$\boxed{(5.142)^2 \cdot 247}$

d) Height + 3 · Height did not do anything in previous model, so when you took this out, nothing would've changed...

e) H_0: Height coef = ~~wrist coef~~ = 0.
↳ dropped two variables... so need to use F-stat

$$\frac{RSS(m) - RSS(M) / 2}{RSS(M) / 247}$$

$(5.142)^2 \cdot 247$

$(5,696)^2 \cdot 249$

(?) But how to get a p-value...?
↳ use table

(2)

Malbw's C_p for m

$$\Rightarrow \frac{RSS(m)}{\hat{\sigma}^2} - (n-2-2p(m))$$

$$= \frac{(5.696)^2 \cdot 249}{(5.142)^2} - (252-2-4)$$

RSS(m) ≈ $\boxed{59.546}$

Malbw's C_p for M ⇒ choose full model

$$\Rightarrow \frac{RSS(M)}{\frac{RSS(M)}{n-p-1}} - (n-2p-2)$$

$= n-p-1 - n+2p+2$

$= p+1 = 5$

M is better
↳ also adj R^2 is higher

#2

a) ① $\frac{0.23708}{0.02746} \approx \boxed{8.633}$

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

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

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

$= 1 - \frac{(5.155)^2 \cdot 245}{15077.897} \cdot \frac{251}{248}$

$= 1 - 0.6054 \approx \boxed{0.554}$

$\frac{RSS}{n-p-1} = 6$, RSS =

☆ $R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{(5.155)^2 \cdot 248}{TSS} = 0.8682$

↳ $1 - 0.8682 = \frac{(5.155)^2 \cdot 245}{TSS}$

↳ $TSS = \boxed{15077.897}$

③ D.F ⇒ $\boxed{6}$ variables

④ $n-p-1 = \boxed{245}$

⑤ $\frac{R^2 / 6}{1-R^2 / 245}$

Regss ⇒ $TSS \cdot R^2$

$RSS = TSS(1-R^2)$

$= \boxed{53.732}$

b) standardized residual

$\frac{\hat{e}_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$

It seems like it will be both an outlier & influential (High leverage & outlier)

weight ↑

c) Height is too short compared to weight

↳ (3) How to know if I don't see a data figure

d) adj R^2 ↑

F-stat also goes up meaning that ~~our model~~ I cannot take out all variables.

p-value for ~~???~~ height goes down

↳ calculate p-value for where....?

$\hat{e}_{(i)} = y_i - x_i^T \hat{\beta}_{(i)}$

↳ $t_i = \frac{\hat{e}_{(i)}\sqrt{1-h_{ii}}}{\sqrt{RSS_{(i)}/n-2}}$

we know it

picture

#3

a) No, It's biased

$E[\hat{e}^T\hat{e}]$

$\Rightarrow$ use $E[z^TAz] = tr[A\Sigma] + \mu^TA\mu$

$\Rightarrow tr[I \cdot cov(\hat{e})] + \mu^TA\mu$

So,
$cov(\hat{e}) = cov((I-H)y) = \sigma^2(I-H)$

$E(\hat{e}) = E((I-H)y) = (I-H)E(y)$

$= (I-H)(X\beta + rz)$

$= (I - X(X^TX)^{-1}X^T)X\beta$

$= (X\beta - X\beta) + (I-H)rz$

$= (I-H)rz$

$\Rightarrow tr(\sigma^2(I-H)) + r^2z^T(I-H)(I-H)z$

$= tr(\sigma^2(I-H)) + r^2z^T(I-H)z$

$= \sigma^2(n-p-1) + r^2z^T(I-H)z$

bias will be $\dfrac{r^2z^T(I-H)z}{n-p-1}$

b) $\sum \hat{e}_i = 0$?

$X^TX\beta = X^Ty$

$X^T(y - X\beta) = 0$

$\langle X, \hat{e} \rangle = 0$

$1^T\hat{e}^T$

BUT maybe zero

$E(\hat{e}) \neq 0$

it implies sum of residuals will not be zero

$\frac{1}{n}\sum \hat{e}_i \neq 0$

c) $E(\hat{e}) = (I-H)rz$

#4

$-2\log(\text{max value of like in } m)$

$Y_i \sim N(X\beta, \sigma^2)$

$\log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-X\beta}{\sigma}\right)^2}\right)$

$n\log\left(\frac{RSS(m)}{n}\right) + n\log(2\pi e)$

$\Rightarrow -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|y-X\beta\|^2$

$\Rightarrow \times -2$ then $n\log(2\pi) + n\log(\sigma^2) + \frac{1}{\sigma^2}\|y-X\beta\|^2$

$\Rightarrow -2l(\beta) = \frac{e^Te}{\sigma^2} + n\log(2\pi\sigma^2)$

$= \frac{RSS}{\sigma^2} + n\log(2\pi\sigma^2)$

And

$-2\log l(\hat\beta, \hat\sigma^2) = n + n\log(2\pi) + n\log\left(\frac{RSS}{n}\right)$

$= n\log(e) + n\log(2\pi) + n\log\left(\frac{RSS}{n}\right)$

$= n\log(2\pi e) + n\log\left(\frac{RSS}{n}\right)$

$E(y_i)$ $E(Y_i)$

$\eta(\mu_i) = \beta_0$

#5.

$\nabla l(\beta) = X^T(y-p)$

$= \sum (y_i - p_i)(1, \dots, x_{ip})$

$\langle X, Y-\hat p\rangle = 0$

inner product $= 0$ means

$X \perp Y - \hat p$

#6 -

a) 212 obs $\begin{cases} 1 \Rightarrow 79 \\ 0 \Rightarrow 133 \end{cases}$ $\begin{cases} \boxed{-0.288928} \\ \boxed{(0.3131)^2} \end{cases}$

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

$\frac{0.6864}{XXX} = 0.313$

1. $\boxed{\frac{0.6864}{0.313}} \approx 2.19297$

2. $\boxed{\frac{-0.9060}{-4.349}} \approx 0.20809$

3. residual dev $\Rightarrow \sum[y_i \log p_i + (1-y_i)\log(1-p_i)]$

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

$+2(1+p(\ldots))$

$\boxed{AIC - 2(1+p(\ldots))}$

$222.18 - 8 = \boxed{214.18}$

with d.f. $n-p-1 = 212 - 4 = \boxed{208}$

null dev $\Rightarrow -2\sum[y_i \log p_i + (1-y_i)\log(1-p_i)] = -2\sum[y_i \log p + (1-y_i)\log(1-p)]$

$p_i = \frac{e^{\beta_0}}{1+e^{\beta_0}} = p$

and $y_i \sim Ber(p)$
and
$E(y_i) = p = \bar{y}$

$-2\sum[y_i \log \bar{y} + (1-y_i)\log(1-\bar{y})]$

$=$

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

and $\bar{y} = \frac{79}{212}$

so $\boxed{279.987} \Rightarrow NM$

with d.f. $n-1 = \boxed{211}$

#7

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

$\hookrightarrow \beta^{(m+1)} = (X^TWX)^{-1}X^TWz$

where ~~$X\beta^{(m)} + X(y-p)$~~

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

b) $\hat{p}_i = \dfrac{e^{x_0^T\hat{\beta}}}{1+e^{x_0^T\hat{\beta}}}$

$\Rightarrow x_0 = \begin{bmatrix} 1 \\ \log(268) \\ \log(26) \\ 3.5 \end{bmatrix}$

$\hat{\beta} = \begin{bmatrix} 0.6864 \\ -0.9050 \\ 0.5027 \\ 1.1153 \end{bmatrix}$

Then $\boxed{\hat{p}_i \approx 0.770198}$

b) $\sum[y_i \log p_i + (1-y_i)\log(1-p_i)]$

~~$= \sum[y_i \log p + (1-y_i)\log(1-\exp(x_i^T\beta))]$~~

$= \sum[y_i(x_i^T\beta) - \log(1+e^{x_i^T\beta})]$

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

c) #6

c) Residual dev $\Rightarrow$ $\boxed{\text{decrease}}$ / $\boxed{\text{the same}}$

$\hookrightarrow -2\sum[y_i\log\hat{p}_i + (1-y_i)\log(1-\hat{p}_i)]$

null dev $\Rightarrow$ $\boxed{\text{same}}$ $\Rightarrow$ only intercept

d) $X^T(y-p)=0$

$\sum y_i = \sum \hat{p}_i$

$\sum I(y_i=1)$

$$l(\beta) = \sum[y_i \ln \hat{p}_i + (1-y_i)\ln(1-\hat{p}_i)]$$

$$= \sum[y_i(x_i^T\beta) - \log(1+e^{x_i^T\beta})]$$

e) $-2\ln(\text{max value in likelihood in model})$

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

f)

~~precision~~ pre / obs	1	0
1	a	b
0	c	d

precision (= PPV) $\Rightarrow \dfrac{a}{a+c} \Rightarrow \dfrac{\sum I(\hat{y}_i=1 \cap y_i=1)}{\sum I(\hat{y}_i=1)}$

$$= \frac{\sum \hat{y}_i y_i}{\sum \hat{y}_i y_i + \sum \hat{y}_i(1-y_i)}$$

Recall (= sensitivity) $\Rightarrow \dfrac{a}{a+b} \Rightarrow \dfrac{\sum I(\hat{y}_i=1 \cap y_i=1)}{\sum I(y_i=1)}$

$$= \frac{\sum \hat{y}_i y_i}{\sum y_i \hat{p}_i + \sum (1-\hat{y}_i) y_i}$$

#8

a) $\hat{y} = X\beta$

$\hat{\beta}$ True

full model

if $y = X\beta + e$

$E(e) = 0$

$\mathrm{Var}(e) = \sigma^2 I_n$

=> Gauss-Markov

b) False => bias ↑ but variance ↓

c) False => we can get

do permutation

with no normality assumption

but p-value

will be calculated

based on property of T!!

d) True => IWLS(=IRLS) for GLM

is the same

as ~~[illegible]~~ Newton Raphson method

with cannical link??

e)

~~RSS~~

$$\frac{\hat{e}_{C(i)}\sqrt{1-h_{ii}}}{\hat{\sigma}}$$

$$= \frac{\hat{e}_{C(i)}\sqrt{1-h_{ii}}}{\sqrt{RSS_{C(i)}/n-p-2}}$$

$\hat{e}_{C(i)} = y_i - x_i^T\hat{\beta}_{C(i)} \perp RSS_{C(i)}$

$\hat{e}_{C(i)} = \frac{\hat{e}_i}{1-h_{ii}} \perp RSS_{C(i)}$

$\hat{e}_i \perp RSS_{C(i)}$

True

$\hat{p}$ binomial → logit

[illegible]

f) $\sum \hat{e}_i^2 = RSS$

$\hat{e}^T\hat{e}$

False

$\Rightarrow$ always goes down when add variables...

g) $\sum \hat{e}_{(i)}^2$ PRESS

$\hat{e}_{(i)} = y_i - x_i^T \hat{\beta}_{(i)}$

↳ Leave one out CV

$$\hookrightarrow \sum \left(\frac{\hat{e}_i}{1-h_{ii}}\right)^2$$

$$= \sum \left(\frac{\hat{e}_i}{1-\frac{p+1}{n}}\right)^2$$

$$= \left(1-\frac{p+1}{n}\right)^{-2} \sum \hat{e}_{i(m)}^2$$

$$= \left(1+2\frac{p(m)+1}{n}\right) RSS = RSS + 2(1+p(m))\frac{RSS}{n}$$

↳ MLE of $\hat{\sigma}^2$

True

h)

pre \ obs	T	F
T	a	b
F	c	d

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

Recall $\Rightarrow \frac{a}{a+b}$

a+b stays the same although cutoff changes...

FNR, TNR ↑

FPR, TPR ↓

cutoff ↑ $\Rightarrow$ b, d ↑, a, c ↓

cutoff ↓ $\Rightarrow$ b, d ↓, a, c ↑

$\frac{a}{a+c} + \frac{a}{a+b} \Rightarrow$ False

$$\frac{RSS}{\hat{\sigma}^2} - (n-2-2p(m))$$

$\approx RSS - \hat{\sigma}^2(n-2-2p(m))$

$\approx RSS + 2\hat{\sigma}^2(1+p(m))$