1.

a) design matrix $X = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}$ since $X_1, \cdots, X_n$ are not all constant

Column vectors of X are linearly independent

Then $\text{rank}(X) = 2$, full rank, so X^TX is invertible, so β_0 and β_1 are estimable

b) $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \Rightarrow \hat{y} = \bar{y} - \hat{\beta}_1 (x - \bar{x})$

plug $x = \bar{x}$, we get $\hat{y} = \bar{y}$.

c) $\text{cov}(\hat{\beta}) = \sigma^2 (X^TX)^{-1}$

$$X^TX = \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}$$

Since $\bar{X} = 0$, $\frac{1}{n} \sum X_i = 0$

$$= \begin{pmatrix} n & 0 \\ 0 & \sum X_i^2 \end{pmatrix}$$

$$(X^TX)^{-1} = \frac{1}{n \sum X_i^2} \begin{pmatrix} \sum X_i^2 & 0 \\ 0 & n \end{pmatrix} = \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum X_i^2} \end{pmatrix}$$

$$= \begin{pmatrix} Var(\hat{\beta}_0) & cov(\hat{\beta}_0, \hat{\beta}_1) \\ cov(\hat{\beta}_1, \hat{\beta}_0) & Var(\hat{\beta}_1) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{\sum X_i^2} \end{pmatrix}$$

Therefore $\text{cov}(\hat{\beta}_1, \hat{\beta}_0) = 0$, hence uncorrelated

d) since $\hat{\beta}_0$ and $\hat{\beta}_1$ are uncorrelated, and $e_1, \cdots, e_n$ are jointly normal, this implies y_i, is jointly normal, and this implies $\hat{\beta}$ is jointly normal.

uncorrelated and jointly normal can imply independence

2.

a) design matrix $X = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & & X_{2p} \\ \vdots & & & \\ 1 & X_{n1} & & X_{np} \end{pmatrix}$ let X_i denote ith row vector of X

$$\beta_0 + \beta_1 \bar{X}_1 + \cdots + \beta_p \bar{X}_p = (1, \bar{X}_1, \cdots, \bar{X}_p) \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} \quad \lambda = (1, \bar{X}_1, \cdots, \bar{X}_p)^T$$

We can see $\lambda = \frac{1}{n} \sum_{i=1}^{n} X_i$ so $\lambda \in$ Row space of X, thus estimable

b) $\hat{y}_1 = \hat{\beta}_0 + \hat{\beta}_1 X_{11} + \cdots + \hat{\beta}_p X_{1p}$

$\vdots$

$\hat{y}_n = \hat{\beta}_0 + \hat{\beta}_1 X_{n1} + \cdots + \hat{\beta}_p X_{np}$

sum up and divide n $\Rightarrow$ $\bar{\hat{y}} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_1 + \cdots + \hat{\beta}_p \bar{X}_p$

We have $y = \hat{y} + e$

$\mathbb{E}(y) = \mathbb{E}(\hat{y}) + \mathbb{E}(e) \Rightarrow \bar{y} = \bar{\hat{y}} + 0$

$\bar{\hat{y}} = \bar{y}$ is lease square estimate of $\beta_0 + \beta_1 \bar{X}_1 + \cdots + \beta_p \bar{X}_p$

c) $Var(\bar{\hat{y}}) = Var(\bar{y}) = Var(\bar{X\beta} + \bar{e}) = Var(\bar{e}) = \frac{\sigma^2}{n}$

$$\hat{\sigma^2} = \frac{1}{n-p-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

3.

a) Because Weight + 3·Height is the linear combination of Weight and Height. β_4's corresponding explanatory variable is dependent with β_2's and β_1's.

b) The model we assume is

$$\text{Bodyfat} = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{weight} + \beta_3 \text{Height} + \beta_4 (\text{Weight} + 3\,\text{Height}) + \beta_5 \text{wrist} + e$$

$$= \beta_0 + \beta_1 \text{Age} + (\beta_2 + \beta_4)\text{weight} + (\beta_3 + 3\beta_4)\text{Height} + \beta_5 \text{wrist} + e$$

So 0.24341 is estimate for $\beta_2 + \beta_4$, not β_2

c) design matrix X in both model span the same column space. So RSS are the same.

$$\hat{\sigma} = \sqrt{\frac{RSS}{n-p-1}} \qquad RSS = 5.142^2 \cdot 247$$

d) $df(M) = 247$ $\qquad RSS(m) = 5.696^2 \cdot 249$

$df(m) = 249$

$$f\text{-statistic}_{(2,247)} = \frac{RSS(m) - RSS(M) / df(m) - df(M)}{RSS(M) / df(M)}$$

P-val is $P\left(F_{2,247} > f\text{-statistic}_{(2,247)}\right)$

4 a) $\text{cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$ $\qquad \dfrac{\hat{\beta}_i}{\text{s.e}(\hat{\beta}_i)} \sim t_{n-p-1}$

$\text{s.e}(\hat{\beta}_i) = \sqrt{\text{Var}(\hat{\beta}_i)}$. $\text{Var}(\hat{\beta}_i) = \text{cov}(\hat{\beta})_{ii} = \sigma^2 (X^T X)^{-1}_{ii}$

We use $\hat{\sigma}^2$ as a estimate of σ^2 in order to get standard error for $\hat{\beta}_i$

$\text{s.e}(\hat{\beta}_0) = 10.30553 = \sqrt{\hat{\sigma}^2 \, 3.74021202i} \Rightarrow \hat{\sigma}^2 = 28.395 \quad \hat{\sigma} = 5.3287$

$\hat{\beta}_4 = 2.444 \cdot 0.14952 = 0.36542688$ $\quad \text{se}(\hat{\beta}_1) = \sqrt{28.395 \cdot 2.632523 \times 10^{-5}} = 0.0274$

$(X^T X)^{-1}_{44} \cdot 28395 = 0.14952^2$ $\qquad t\text{-val}(\hat{\beta}_2) = \dfrac{0.12373}{\text{se}(\hat{\beta}_2)} = 4.5255$

$(X^T X)^{-1}_{4,4} = 7.873298 \times 10^{-4}$

$RSS = \hat{\sigma}^2 \cdot (n-p-1) = 28.395 \cdot 247 = 7013.565$ $\quad F\text{-statistic} = \dfrac{(TSS - RSS)/4}{RSS/(n-4-1)} = 67.12994$

b) "confidence" $\quad t\text{-statistic} = qt(1 - \frac{0.05}{2}, \ df = 247)$ $\quad$ 95% CI for $X^{0T}\beta$

$X_0^T \hat{\beta} \pm t \cdot \hat{\sigma} \sqrt{X_0^T (X^T X)^{-1} X_0} \Rightarrow (14.56726, \ 16.6723)$

"prediction" $\quad$ fitted value is the same for both : 15.61978

$X_0^T \hat{\beta} \pm t \cdot \hat{\sigma} \sqrt{X_0^T (X^T X)^{-1} X_0 + 1} \Rightarrow (4.123176, \ 27.11638)$

$t \cdot \hat{\sigma} = 10.4949$

c)

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
248	7059.57				
247	7013.565	1	46.0084	1.6203	0.2042

$$F = \frac{RSS(m) - 7013.565/1}{7013.565/247} = 1.6203 \qquad RSS(m) = 7059.57$$

5 a) $RSS(M) = 0.3295^2 \times 137 = 14.8741$

$$\left(\frac{\hat{\beta}_3}{s.e(\hat{\beta}_3)}\right)^2 = \frac{RSS(m) - RSS(M)}{RSS(M)/137}$$

$$F_{1,137} = \frac{RSS(m) - RSS(M)}{RSS(M)/137} = 10.2179$$

$\Rightarrow RSS(m) = 15.9835$

The table:

Res.df	RSS	df	Sum of sq	F	P-val
138	15.9835				
137	14.8741	1	1.109355	10.2179	0.0017

(b) $X_0 = (1, 3.4, 25, 2)$ $\hat{\beta} = (1.38955\ldots, 0.41182\ldots, 0.01472, -0.08311)$

fit $\hat{Y}_0 = X_0^T \hat{\beta} = 2.9915$ diff = upr − fit = 0.07289 lwr = fit − diff

	fit	lwr	upr
"confidence"	2.9915	2.918628	3.064408
"prediction"	2.9915	2.3359	3.6471

$t\hat{\sigma} = 0.6516$

$\text{diff} = t\hat{\sigma}\sqrt{\text{unknown}}$

unknown = 0.0125

"prediction diff" $= t\hat{\sigma}\sqrt{1+0.0125}$

$= 0.65563$

6. a) False. Normal equations will have infinitely many solutions

b) True Same reason as above

c) False high leverage means outlier

d) True $\frac{RSS}{\sigma^2} \sim \chi^2_{n-p-1}$

e) False $\hat{Y}$ follows $N_n(X, \sigma^2 H)$

f) True Then this estimator must be biased, because of Gauss-Markov theorem

g) False On this model, TSS would be equal to RSS. $R^2 = 1 - \frac{1}{1} = 0$ not F distribution

h) True

$\hat{\sigma}^2 = \frac{RSS}{n-p-1}$

$Var(\hat{\sigma}^2) = Var\left(\frac{\sigma^2 RSS}{\sigma^2(n-p-1)}\right) = \frac{1}{(n-p-1)^2}\sigma^4 \cdot Var\left(\frac{RSS}{\sigma^2}\right)$ Since $\frac{RSS}{\sigma^2} \sim \chi^2_{n-p-1}$

$= \frac{1}{(n-p-1)^2} \cdot \sigma^4 \cdot 2(n-p-1) = \frac{2\sigma^4}{n-p-1}$

i) True Different permutation will have different p-value

j) True Hat-matrix

k) False $\hat{Y} = \hat{\beta}_1 X + \hat{\beta}_0$ let $\hat{\beta}_1 = 2$, then slope is greater than 1

l) True $\hat{\beta}_1 = r\frac{S_y}{S_x}$ when X, Y normalized, $S_y = S_x = 1$ and correlation between -1 an 1

m) True $\hat{\sigma} = \sqrt{\frac{RSS}{n-p-1}}$ $\hat{\sigma}$ of bigger model is always equal to or less than reduced model

n) True $X^T X$ is invertible, then β is estimable, then and linear combination of β is estimable

o) False $Var(\hat{e}_i) = 1 - h_{ii} \geq 0$ even though assume e has same variance

p) False $y = \hat{y} + \hat{e} = Hy + (I-H)y = y\, H(I-H) = 0$ always orthogonal

q) False $cov(\hat{y}, \hat{e}) = cov(Hy, (I-H)y) = H\, \sigma^2 I_n\, (I-H)^T = \sigma^2 H(I-H) = 0$ uncorrelated

r) True Under assumption of normality, we can see $\hat{\beta} \perp\!\!\!\perp \hat{e}$, if normality is Violated, then they are not independent

s) False This can only tell us that Y and X have some relations

t) False p-value is probability of the result greater or equal to the observed under the null Hypothesis is true. We can only say that very Small chance that $\beta_2 = 0$ We don't know how large

u) True $\hat{\beta}_2$ is large so we can have a small p value