Math 342 - Chapter 2: Solution

2.1

$$\begin{aligned}
\sum_{i=1}^{n} \frac{\hat{y}_1}{n} &= (\hat{\alpha} + \hat{\beta} x_i)/n \\
&= \sum_{i=1}^{n} (\bar{y} - \hat{\beta}\bar{x} + \hat{\beta} x_i)/n \\
&= \bar{y} - (\hat{\beta}/n) \sum_{i=1}^{n} (x_i - \bar{x}) \\
&= \bar{y}
\end{aligned}$$

2.2(a)

$$\begin{aligned}
\sum_{i=1}^{n} (y_i - \hat{y}_i) &= \sum_{i=1}^{n} (y_i - (\hat{\alpha} + \bar{\beta} x_i)) \\
&= \sum_{i=1}^{n} (y_i - (\bar{y} - \hat{\beta}\bar{x}) + \bar{\beta} x_i) \\
&= \sum_{i=1}^{n} [(y_i - \bar{y}) - \hat{\beta}(x_i - \bar{x})] \\
&= \sum_{i=1}^{n} (y_i - \bar{y}) - \hat{\beta} \sum_{i=1}^{n} (x_i - \bar{x}) \\
&= 0
\end{aligned}$$

2.2(b)

$$\begin{aligned}
\sum_{i=1}^{n} (y_i - \hat{y}_i) x_i &= \sum_{i=1}^{n} [(y_i - \bar{y}) - \hat{\beta}(x_i - \bar{x})] x_i \\
&= \sum_{i=1}^{n} (y_i - \bar{y}) x_i - \hat{\beta} \sum_{i=1}^{n} (x_i - \bar{x}) x_i \\
&= S_{xy} - \hat{\beta} S_{xx} \\
&= S_{xy} - (S_{xy}/S_{xx}) S_{xx} \\
&= 0
\end{aligned}$$

2.3 Show

$$\begin{aligned}
E(RSS) &= (n-2)\sigma^2 \qquad \text{(Model, } y = \alpha + \beta x + e) \\
\Rightarrow E(S_{yy} - \hat{\beta}^2 S_{xx}) &= (n-2)\sigma^2
\end{aligned}$$

$$
\begin{aligned}
(1) \qquad E(S_{yy}) &= E\left(\sum (y_i - \bar{y})^2\right) \\
&= E\left(\sum y_i^2 - n\bar{y}^2\right) \\
&= \sum E(y_i^2) - nE(\bar{y}^2) \\
&= \sum \left(Var(y_i) + (E(y_i))^2\right) - n\left(Var(\bar{y}) + (E\bar{y})^2\right) \\
&= n\sigma^2 + \sum (Ey_i)^2 - n(\frac{\sigma^2}{n} + (E\bar{y})^2) \\
&= (n-1)\sigma^2 + \sum (\alpha + \beta x_i)^2 - n(\alpha + \beta \bar{x})^2 \\
&= (n-1)\sigma^2 + \beta^2 \left(\sum x_i^2 - n\bar{x}\right) \\
&= (n-1)\sigma^2 + \beta^2 S_{xx}
\end{aligned}
$$

$$
\begin{aligned}
(2) \qquad E(\hat{\beta}^2 \ S_{xx}) &= S_{xx} \cdot E\hat{\beta}^2 \\
&= S_{xx}(Var\hat{\beta} + (E\hat{\beta})^2) \\
&= S_{xx}\left(\frac{\sigma^2}{S_{xx}} + \beta^2\right) \\
&= \sigma^2 + \beta^2 S_{xx}
\end{aligned}
$$

Combine (1) and (2)

$$
\begin{aligned}
E(S_{yy} - \hat{\beta}^2 S_{xx}) &= E(S_{yy}) - E(\hat{\beta}^2 \ S_{xx}) \\
&= (n-1)\sigma^2 + \beta^2 S_{xx} - \sigma^2 - \beta^2 S_{xx} \\
&= (n-2)\sigma^2
\end{aligned}
$$

2.5

$$
\begin{aligned}
E\left(\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2\right) &= E\left(\sum_{i=1}^{n} \left(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y}\right)^2\right) \\
&= E\left(\sum_{i=1}^{n} \left[\bar{y} + \hat{\beta}_1(x_i - \bar{x}) - \bar{y}\right]^2\right) \\
&= E\left(\sum_{i=1}^{n} \left[\hat{\beta}_1(x_i - \bar{x})\right]^2\right) \\
&= E\left(\hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2\right) \\
&= E(\hat{\beta}_1^2 \ S_{xx}) \\
&= \sigma^2 + \beta_1^2 \ S_{xx}
\end{aligned}
$$

2.7 $\quad S_{yy} = 1815.07,\ S_{xx} = 1262.06,\ S_{xy} = 770.27$

$$
\begin{aligned}
\hat{\beta}_1 &= S_{xy}/S_{xx} = 770.27/1262.06 = 0.61033 \\
\hat{\sigma}^2 &= \frac{S_{yy} - \hat{\beta}_1^2 \ S_{xx}}{n-2} = \frac{1815.07 - 0.61033^2(1262.06)}{52} = 25.864 \\
Var(\hat{\beta}_1) &= \hat{\sigma}^2/S_{xx} = 25.864/1262.06 = 0.02049 \\
\Rightarrow t &= \frac{\hat{\beta}_1 - 0}{\sqrt{Var(\hat{\beta}_1)}} = 0.61033/\sqrt{0.02049} = 4.26
\end{aligned}
$$

2.8 Model: $y_i = \beta_1 x_i + e_i$, where $e_i \sim N(0, \sigma^2)$ iid

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

$$
\begin{aligned}
\frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2 &= 2 \sum_{i=1}^{n} (y_i - \beta_1 x_i)(-x_i) \\
\frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2 &= 0 \\
\Rightarrow \quad 2 \sum_{i=1}^{n} (y_i - \hat{\beta}_1 x_i)(-x_i) &= 0 \\
\Rightarrow \quad \sum_{i=1}^{n} (x_i y_i - \hat{\beta}_1 x_i^2) &= 0 \\
\Rightarrow \quad \hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \\
E(\hat{\beta}_1) &= E\left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \right) \\
&= \left(\frac{1}{\sum_{i=1}^{n} x_i^2} \right) \sum_{i=1}^{n} x_i E(y_i) \\
&= \left(\frac{1}{\sum_{i=1}^{n} x_i^2} \right) \sum_{i=1}^{n} x_i E(\beta x_i + e_i) \\
&= \beta_1
\end{aligned}
$$

$$
\begin{aligned}
Var(\hat{\beta}_1) &= Var\left(\sum_{i=1}^{n} x_i y_i / \sum_{i=1}^{n} x_i^2 \right) \\
&= \left(\frac{1}{\sum_{i=1}^{n} x_i^2} \right)^2 Var\left(\sum_{i=1}^{n} x_i y_i \right) \\
&= \left(\frac{1}{\sum_{i=1}^{n} x_i^2} \right)^2 \sum_{i=1}^{n} x_i^2 Var(y_i) \\
&= \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}
\end{aligned}
$$

$$\begin{aligned} Var[\widehat{E(y|x_0)}] &= Var(\hat{\beta}_1 x_0) \qquad \text{where } \widehat{E(y|x_0)} = \hat{\beta}_1 x_0 \\ &= x_0^2 Var(\hat{\beta}_1) \\ &= \frac{\sigma^2 x_0^2}{\sum_{i=1}^n x_i^2} \end{aligned}$$

$$\text{It is noted that} \sum_{i=1}^n (y_i - \hat{\beta} x_i)^2/\sigma^2 \sim \chi^2_{(n-1)}$$

$100(1-\alpha)\%$C.I. on $E(y|x_0)$ is given by

$$\begin{aligned} \widehat{E(y|x_0)} &\pm t_{\alpha/2,n-1}\sqrt{Var[\widehat{E(y|x_0)}]} \\ \Rightarrow \qquad \hat{\beta}_1 x_0 &\pm t_{\alpha/2,n-1}\ \hat{\sigma}\sqrt{x_0^2/\sum_{i=1}^n x_i^2} \qquad \text{(note that } \hat{\sigma}^2 = s^2\text{)} \end{aligned}$$

2.9 $\sum_{i=1}^n x_i = 0.8734$, $\sum_{i=1}^n y_i = 0.0975$, $\sum_{i=1}^n x_i^2 = 0.0426$, $\sum_{i=1}^n x_i y_i =$ -0.0474, $\sum_{i=1}^n y_i^2 = 0.0975$, n = 20

$\Rightarrow \qquad S_{xy} = 0.0045,\ S_{yy} = 0.0181,\ S_{xy} = 0.0076$

(a)

$$\begin{aligned} \hat{\beta}_1 &= S_{xy}/S_{xx} = 0.0076/0.0045 = 1.69886 \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} = 0.0975/20 - 1.689(0.8734/20) = -0.13719 \end{aligned}$$

The fitted model:

$$\hat{k}_i = -0.13719 + 1.6988(\hat{N}_p)_i$$

(b)

$$\begin{aligned} \text{Total } SS &= S_{yy} = 0.0181 \\ \text{Reg } SS &= \hat{\beta}_1\ S_{xx} = (1.69886)^2(0.0045) = 0.01299 \\ \text{Res } SS &= S_{yy} - \hat{\beta}_1\ S_{xx} = 0.0181 - 0.01299 = 0.00514 \\ \Rightarrow \qquad R^2 &= \text{Reg } SS/\text{Total } SS = 0.01299/0.0181 = 0.7167 \\ \hat{\sigma}^2 &= \text{Res } SS/(n-2) = 0.00518/18 = 0.00028533 \qquad \text{(note } \hat{\sigma}^2 = s^2\text{)} \end{aligned}$$

(c) Confidence limits on mean response and residuals

observation	95%C.I. on mean response	residual
1	(-0.1342, -0.0977)	-0.009671
2	(-0.1100, -0.0836)	-0.0238
3	(-0.1061, -0.0812)	-0.002598
4	(-0.097, -0.0755)	0.005599
5	(-0.0906, -0.0712)	0.00154
6	(-0.0809, -0.064)	-0.000704
7	(-0.0809, -0.064)	-0.000084
8	(-0.0809, -0.064)	0.000546
9	(-0.0751, -0.0591)	-0.0266
10	(-0.0719, -0.056)	0.0108
11	(-0.0687, -0.0528)	0.003236
12	(-0.0657, -0.0494)	0.000051
13	(-0.0647, -0.0483)	0.0365
14	(-0.0609, -0.0436)	0.008492
15	(-0.0590, -0.0412)	0.0301
16	(-0.052, -0.0313)	-0.009626
17	(-0.0437, -0.0183)	-0.000244
18	(-0.0358, -0.004929)	-0.0284
19	(-0.0251, 0.0141)	-0.0133
20	(-0.0704, -0.0545)	0.0162

C.I. $= \hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2,n-2}\ \hat{\sigma}\sqrt{\frac{1}{n} + (x_0 - \bar{x})^2/S_{xx}}$
Residual $= y_i - \hat{y}_i$

2.10 (a) The fitted line:

$$\hat{y} = 90.72172 - 0.05103x$$

(b)

$$H_0 : \beta_1 = 0 \text{ vs } H_1 : \beta_1 \neq 0$$

$$t = \frac{-0.05103 - 0}{0.00787} = -6.48$$

$$t_{0.025,(24-2)} = 2.074$$

$\therefore$ we reject H_0 and conclude that the time it takes to run a distance of two miles have significant influence on maximum oxygen uptake.

(c) 95% confidence interval on mean max volume of O_2
at $x = 750$: $52.4503 \pm 2.07387(0.9345) = (50.5123, 54.3884)$
at $x = 775$: $51.1746 \pm 2.07387(0.8215) = (49.4709, 52.8784)$
at $x = 800$: $49.8989 \pm 2.07387(0.7443) = (48.3552, 51.4428)$
at $x = 825$: $48.6232 \pm 2.07387(0.7148) = (47.1412, 50.1052)$
at $x = 850$: $47.3475 \pm 2.07387(0.7381) = (45.8168, 48.8782)$

2.11 Let right leg be x_{i1}, left leg be x_{i2}.

$$\begin{aligned}
\sum_{i=1}^{n} x_{i1} &= 1920, \qquad \sum_{i=1}^{n} x_{i2} = 1870, \qquad \sum_{i=1}^{n} y_i = 1926.92 \\
\sum_{i=1}^{n} x_{i1}^2 &= 289800, \qquad \sum_{i=1}^{n} x_{i2}^2 = 275300, \qquad \sum_{i=1}^{n} x_{i1}y_i = 290215.6 \\
\sum_{i=1}^{n} x_{i2}y_i &= 282499.7, \qquad \sum_{i=1}^{n} y_i^2 = 293719.57, \qquad n = 13 \\
\Rightarrow S_{x_1x_1} &= 6230.77, \qquad S_{x_1y} = 5624.34 \\
S_{x_2x_2} &= 6307.69, \qquad S_{x_2y} = 5319.67
\end{aligned}$$

(a) Model with pd as the response and right leg strength as the independent variable.

$\hat{\beta}_1 = S_{x_1y}/S_{x_1x_1} = 5624.34/6230.77 = 0.90267$

$\hat{\beta}_2 = \bar{y} - \hat{\beta}_1\bar{x}_1 = 1926.92/13 - 0.90267(1920/13) = 14.90696$

The fitted model:

$$\hat{y}_i = 14.90696 + 0.90267x_{i1}$$

(b) Model wit pd as the response and left leg strength as the independent variable.

$\hat{\beta}_1 = S_{x_2y}/S_{x_2x_2} = 5319.67/6307.69 = 0.84336$

$\hat{\beta}_2 = \bar{y} - \hat{\beta}_1\bar{x}_2 = 1926.92/13 - 0.84336(1870/13) = 26.91021$

The fitted model:

$$\hat{y}_i = 26.91021 + 0.84336x_i$$

2.12

$$\begin{aligned}
\sum_{i=1}^{22} x_i &= 14154, \qquad \sum_{i=1}^{22} y_i = 56156, \qquad \sum_{i=1}^{22} x_i^2 = 18361344 \\
\sum_{i=1}^{22} y_i^2 &= 292495554, \qquad \sum_{i=1}^{22} x_i y_i = 73280101 \\
\Rightarrow \quad S_{xx} &= 9255175.09, \qquad S_{xy} = 36919761.73, \qquad S_{yy} = 147311087.82
\end{aligned}$$

(a) $\hat{\beta}_1 = S_{xy}/S_{xx} = 3.98909, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 2.47119$

The fitted model:

$$\hat{y}_i = 2.47119 + 3.98909x_i$$

(b) $R^2 = \text{Reg SS/Total SS} = \hat{\beta}_1^2\ S_{xx}/S_{yy} = 0.9998$
(R^2 = % of variation in y explained by the model = coeff. of determination)

For the zero intercept model:

$$\begin{aligned}
\hat{\beta}_{1(0)} &= \sum_{i=1}^{22} x_i y_i / \sum_{i=1}^{22} x_i^2 = 3.991 \qquad (3.990998753) \\
\text{Total SS}_{(0)} &= \sum_{i=1}^{22} (y_i - \beta_0)^2 = \sum_{i=1}^{22} y_i^2 = 292495554 \\
\text{Reg SS}_{(0)} &= \sum_{i=1}^{22} (\hat{y}_i - \beta_0)^2 = \sum_{i=1}^{22} \hat{y_i}^2 = \sum_{i=1}^{22} \hat{\beta}_{1(0)} x_i^2 = 292460792 \\
\Rightarrow \quad R^2_{(0)} &= \text{Reg SS}_{(0)} / \text{Total SS}_{(0)} = 0.9999
\end{aligned}$$

$$\begin{aligned}
s^2 &= \hat{\sigma}^2 = (S_{yy} - \hat{\beta}_1^2 S_{xx})/(n-2) = 34695/20 = 1734.75 \\
s^2_{(0)} &= \hat{\sigma}^2_{(0)} = \left(\sum_{i=1}^{22} y_i^2 - \hat{\beta}^2_{1(0)} \sum_{i=1}^{22} x_i^2 \right) / (n-1) \\
& \quad ((n-1) \text{ since one unknown parameter only under given } \beta_0 = 0) \\
&= 34762/21 = 1655.33
\end{aligned}$$

(d) C.I. for $E(y|x_i)$:

Intercept model: $\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2,n-1}\ \hat{\sigma} \sqrt{\frac{1}{n} + (x_0 - \bar{x})^2 / S_{xx}}$

No Intercept model: $\hat{\beta}_1 x_0 \pm t_{\alpha/2,n-1}\ \hat{\sigma}_{(0)} \sqrt{x_0^2 / \sum_{i=1}^{n} x_i^2}$

95% C.I. for $E(y|x_i)$

Observation	Intercept model	No intercept model
1	(36.5176, 88.0975)	(59.5688, 60.1612)
2	(76.6065, 127.7906)	(99.2813, 100.2686)
3	(204.8793, 254.8198)	(226.3614, 228.6124)
4	(244.9609, 294.5201)	(266.0739, 268.7199)
5	(765.8372, 810.8081)	(782.3368, 790.1167)
6	(641.6618, 687.6597)	(659.228, 665.7836)
7	(625.6376, 671.7712)	(643.343, 649.7406)
8	(501.4374, 548.6476)	(520.2341, 525.4075)
9	(609.6129, 655.8831)	(627.458, 633.6976)
10	(942.0451, 985.6405)	(957.072, 966.5894)
11	(1574, 1614)	(1585, 1600)
12	(2086, 2124)	(2093, 2114)
13	(2110, 2147)	(2117, 2138)
14	(2230, 2267)	(2236, 2258)
15	(2230, 2267)	(2236, 2258)
16	(3700, 3741)	(3701, 3738)
17	(3915, 3957)	(3916, 3955)
18	(4054, 4097)	(4055, 4095)
19	(6523, 6591)	(6525, 6590)
20	(7878, 7963)	(7883, 7961)
21	(6511, 6579)	(6513, 6578)
22	(8504, 8596)	(8510, 8595)

(e) Firstly, test $H_0 : \beta_0 = 0$

$$t_{obs} = \frac{\hat{\beta}_0 - 0}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} = \frac{2.47119}{\sqrt{1734.75}\sqrt{\frac{1}{22} + \frac{(14154/22)^2}{9255175.09}}} = 0.1976$$

$$t_{0.025,(22-2)} = 2.086$$

So we can't reject H_0, zero intercept model is reasonable.
Secondly, comparing with non-zero intercept model, it has larger R^2, smaller s^2, narrow C.I., and hence zero intercept model is better.

2.13 Model: $y_i = \beta_1 x_i + e_i, \qquad i = 1, \dots, n$

LSE:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \\
Var(\hat{\beta}_1) &= \sigma^2 / \sum_{i=1}^n x_i^2 \\
E(\hat{\beta}_1) &= \beta_1, \qquad \text{so } \hat{\beta}_1 \sim N(\beta_1, \sigma^2 / \sum_{i=1}^n x_i^2) \\
\hat{\sigma}^2 &= \frac{RSS}{n-1} = \sum_{i=1}^n y_i^2 - \hat{\beta}_1^2 \sum_{i=1}^n x_i^2 \\
& \text{and hence } \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/\sqrt{\sum_{i=1}^n x_i^2}} \sim t_{(n-1)}
\end{aligned}$$

Therefore, C.I. for β_1 is

$$\hat{\beta}_1 \pm t_{\alpha/2,n-1}\ \hat{\sigma}/\sqrt{\sum_{i=1}^n x_i^2}$$

2.15 (a)

$$\begin{aligned}
\sum_{i=1}^{25} x_i &= 778.7, \qquad \sum_{i=1}^{25} y_i = 2049.1, \qquad \sum_{i=1}^{25} x_i^2 = 26591.63 \\
\sum_{i=1}^{25} y_i^2 &= 172762.85, \qquad \sum_{i=1}^{25} x_i y_i = 65125.31, \qquad n = 25 \\
\Rightarrow \quad S_{xx} &= 2336.68, \qquad S_{xy} = 1299.94, \qquad S_{yy} = 4810.42 \\
\hat{\beta}_1 &= S_{xy}/S_{xx} = 1299.94/2336.68 = 0.55632 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\
&= 2049.1/25 - 0.55632(7787.7/25) = 64.63574
\end{aligned}$$

The fitted model: $\hat{y}_i = 64.63574 + 0.55632 x_i$

(b) Will be discussed in Chapter 3

(c) Will be discussed in Chapter 5

2.16 (a)

$$\begin{aligned}
\sum_{i=1}^{12} x_i &= 2700, \qquad \sum_{i=1}^{12} y_i = 1037.8, \qquad \sum_{i=1}^{12} x_i^2 = 645000 \\
\sum_{i=1}^{12} y_i^2 &= 90265.52, \qquad \sum_{i=1}^{12} x_i y_i = 237875, \qquad n = 12 \\
\Rightarrow \qquad S_{xx} &= 37500, \qquad S_{xy} = 4370, \qquad S_{yy} = 513.12 \\
\hat{\beta}_1 &= S_{xy}/S_{xx} = 4370/37500 = 0.11653 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\
&= 1037.8/12 - 0.11653(2700/12) = 60.26333 \\
\hat{\sigma}^2 &= (S_{yy} - \hat{\beta}_1^2 S_{xx})/(n-2) = 0.3866
\end{aligned}$$

The fitted model: $\hat{y}_i = 60.26333 + 0.11653x_i$

(b) 95% C.I. on $E(y|x)$ at 4 level of temperature:

Level x=150:

$60.26333 + 0.11653(150) \pm 2.22814\sqrt{0.3866}\sqrt{1/12 + (150-225)^2/37500} = (77.0741, 78.4125)$

x=200: (83.1319, 84.0081)

x=250: (88.9586, 89.8348)

x=300: (94.5541, 958925)

2.18 Will be discussed in Chapter 5.

2.19 (a)

$$\begin{aligned}
\sum_{i=1}^{45} x_i &= 45492, \qquad \sum_{i=1}^{45} y_i = 114.63, \qquad \sum_{i=1}^{45} x_i^2 = 61976798 \\
\sum_{i=1}^{45} x_i y_i &= 124053.15, \qquad \sum_{i=1}^{45} y_i^2 = 299.54, \qquad n = 45 \\
\Rightarrow \qquad S_{xx} &= 15987418.8, \qquad S_{xy} = 8166.83, \qquad S_{yy} = 7.53 \\
\hat{\beta}_1 &= S_{xy}/S_{xx} = 0.00051083 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} = 2.03099
\end{aligned}$$

The fitted model: $\hat{y}_i = 2.03099 + 0.00051083x_i$

(b) From the result, $R^2 = RegSS/TotalSS = \hat{\beta}_1^2 S_{xx}/S_{yy} = 0.5542$, we can say that the model does not fit well. Obviously, chamber condition will affect the size of larvae as the values of "size of larvae" are quite different from other chambers with the same "Head Diametor". That means, some of the chamber will have a more suitable environment for the larvae to grow. Thus, the estimated $\beta_1, \hat{\beta}_1$, will have a big difference between different chambers.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

ϵ_i, the effect of chamber treated as a part of error (ϵ_i becomes larger $\Rightarrow$ model fits bad)

2.20

$$
\begin{aligned}
E\left(\sum_{i=1}^{n}(y_i-\hat{y}_i)^2\right) &= E(y_i-\hat{\beta}x_i)^2 \\
&= E\left[\sum_{i=1}^{n} y_i^2+\hat{\beta}^2\sum_{i=1}^{n} x_i^2-2\hat{\beta}\sum_{i=1}^{n} x_i y_i\right] \\
&= E\left(\sum_{i=1}^{n} y_i^2\right)+E\left(\hat{\beta}^2\sum_{i=1}^{n} x_i^2\right)-E\left(2\hat{\beta}\sum_{i=1}^{n} x_i y_i\right)
\end{aligned}
$$

The first term:

$$
\begin{aligned}
E\left(\sum_{i=1}^{n} y_i^2\right) &= \sum_{i=1}^{n} E\left(y_i^2\right) \\
&= \sum_{i=1}^{n}\left[Var(y_i)+[E(y_i)]^2\right] \\
&= \sum_{i=1}^{n}[\sigma^2+\beta^2 x_i^2] \\
&= n\sigma^2+\beta^2\sum_{i=1}^{n} x_i^2
\end{aligned}
$$

The second term:

$$
\begin{aligned}
E\left(\sum_{i=1}^{n}\hat{\beta}^2 x_i^2\right) &= \sum_{i=1}^{n} x_i^2 E(\hat{\beta}^2) \\
&= \sum_{i=1}^{n} x_i^2\left[Var(\hat{\beta})+[E(\hat{\beta})]^2\right] \\
&= \sum_{i=1}^{n} x_i^2\left(\sigma^2/\sum_{i=1}^{n} x_i^2+\beta^2\right) \\
&= \sigma^2+\beta^2\sum_{i=1}^{n} x_i^2
\end{aligned}
$$

The third term:

$$
\begin{aligned}
E\left(\sum_{i=1}^{n} 2x_i y_i\ \hat{\beta}\right) &= E\left[\sum_{i=1}^{n} 2x_i y_i\left(\sum_{j=1}^{n} x_j y_j/\sum_{j=1}^{n} x_j^2\right)\right] \\
&= 2\ E\left[\left(\sum_{i=1}^{n} x_i y_i\right)^2\right]/\sum_{j=1}^{n} x_j^2 \\
&= 2\left[Var\left(\sum_{i=1}^{n} x_i y_i\right)+E\left(\sum_{i=1}^{n} x_i y_i\right)^2\right]/\sum_{j=1}^{n} x_j^2 \\
&= 2\sigma^2+2\beta^2\sum_{i=1}^{n} x_i^2 \\
\Rightarrow \qquad E\left(\sum_{i=1}^{n}(y_i-\hat{y}_i)^2\right) &= n\sigma^2+\beta^2\sum_{i=1}^{n} x_i^2+\sigma^2+\beta^2\sum_{i=1}^{n} x_i^2-2\sigma^2-2\beta^2\sum_{i=1}^{n} x_i^2 \\
&= (n-1)\sigma^2
\end{aligned}
$$

2.21 (a) Find $\hat{\beta}_1$ such that $\sum_{i=1}^n (y_i - \beta_0 - \hat{\beta}_1 x_i)^2$ is minimized.

$$\begin{aligned}
\frac{\partial}{\partial \hat{\beta}_1} \sum_{i=1}^n (y_i - \beta_0 - \hat{\beta}_1 x_i)^2 &= \sum_{i=1}^n 2(y_i - \beta_0 - \hat{\beta}_1 x_i)(-x_i) = 0 \\
\Rightarrow \qquad \hat{\beta}_1 &= \frac{\sum_{i=1}^n (y_i - \beta_0) x_i}{\sum_{i=1}^n x_i^2} \\
&= \frac{\sum_{i=1}^n x_i y_i - \beta_0 \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2}
\end{aligned}$$

(b)

$$\begin{aligned}
Var(\hat{\beta}_1) &= \frac{1}{(\sum_{i=1}^n x_i^2)^2} \cdot Var\left(\sum_{i=1}^n x_i y_i - \beta_0 \sum_{i=1}^n x_i \right) \\
&= \frac{1}{(\sum_{i=1}^n x_i^2)^2} \sum_{i=1}^n x_i^2 \cdot Var(y_i) \\
&= \frac{\sigma^2}{\sum_{i=1}^n x_i^2}
\end{aligned}$$

(c) A point estimate for $E(y|x)$ is $\beta_0 + \hat{\beta}_1 x$,

$$\begin{aligned}
Var(\beta_0 + \hat{\beta}_1 x) &= x^2 Var(\hat{\beta}_1) \\
&= \frac{x^2}{\sum_{i=1}^n x_i^2} \sigma^2 \\
\hat{\sigma}^2 = \frac{\text{Res SS}}{n-1} &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \beta_0 - \hat{\beta}_1 x_i)^2
\end{aligned}$$

$100(1-\alpha)\%$ confidence interval for $E(y|x)$ is given by

$$(\beta_0 + \hat{\beta}_1 x) \pm t_{\alpha/2, n-1} \; \hat{\sigma} \; x \Big/ \sqrt{\sum_{i=1}^n x_i^2}$$

2.22

$$\frac{y_i - \hat{y}_i}{s \sqrt{1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}}} \qquad \text{where } s = \sqrt{\sum_{i=1}^n \hat{e}_i^2/(n-2)}, \quad \hat{e}_i = y_i - \hat{y}_i$$

$$Var(y_i - \hat{y}_i) = \sigma^2 \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}} \right]$$

(a)

$$\begin{aligned}
\sum_{i=1}^n Var(y_i - \hat{y}_i)/\sigma^2 &= \sum_{i=1}^n \sigma^2 \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}} \right] / \sigma^2 \\
&= n - 1 - \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{S_{xx}} \\
&= n - 1 - 1 \\
&= n - 2
\end{aligned}$$

(b) $\hat{e}_i$ & $\sum_{i=1}^{n} \hat{e_i}^2 = (n-2)s^2$ are not independent.

$\therefore$ Studentized residual does not follow t-distribution.

Remark:

$$\frac{u_i - E(u)}{\sqrt{Var(u)}} \leftarrow \text{standardized (with mean 0, variance 1)}$$
$$\frac{u_i - \bar{u}}{S_u} \leftarrow \text{studentized(student } t \text{ distribution)}$$

2.23

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n_1}\Big(y_i - \hat{\beta_0}^{(1)*} - \hat{\beta}_1(x_i - \bar{x}_1)\Big)^2 + \sum_{i=n_1+1}^{n_1+n_2}\Big(y_i - \hat{\beta_0}^{(2)*} - \hat{\beta}_1(x_i - \bar{x}_2)\Big)^2$$

$$\begin{cases}
(1) \quad \dfrac{\partial}{\partial \hat{\beta}_0^{(1)*}} \displaystyle\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 2\sum_{i=1}^{n_1}\Big(y_i - \hat{\beta_0}^{(1)*} - \hat{\beta}_1(x_i - \bar{x}_1)\Big)(-1) \\
(2) \quad \dfrac{\partial}{\partial \hat{\beta}_0^{(2)*}} \displaystyle\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 2\sum_{i=n_1+1}^{n_1+n_2}\Big(y_i - \hat{\beta_0}^{(2)*} - \hat{\beta}_1(x_i - \bar{x}_2)\Big)(-1) \\
(3) \quad \dfrac{\partial}{\partial \hat{\beta}_1} \displaystyle\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 2\sum_{i=1}^{n_1}\Big(y_i - \hat{\beta_0}^{(1)*} - \hat{\beta}_1(x_i - \bar{x}_1)\Big)\Big(-(x_i - \bar{x}_1)\Big) \\
\qquad\qquad +2\displaystyle\sum_{i=n_1+1}^{n_1+n_2}\Big(y_i - \hat{\beta_0}^{(2)*} - \hat{\beta}_1(x_i - \bar{x}_2)\Big)\Big(-(x_i - \bar{x}_2)\Big)
\end{cases}$$

Set them equal to 0.

(1)
$$\sum_{i=1}^{n_1}\Big(y_i - \hat{\beta_0}^{(1)*} - \hat{\beta}_1(x_i - \bar{x}_1)\Big) = 0 \quad \Rightarrow$$
$$\begin{aligned}\hat{\beta_0}^{(1)*} &= \frac{1}{n_1}\sum_{i=1}^{n_1}\Big(y_i - \hat{\beta}_1(x_i - \bar{x}_1)\Big) \\ &= \bar{y_1}\end{aligned}$$

(2)
$$\sum_{i=n_1+1}^{n_1+n_2}\Big(y_i - \hat{\beta_0}^{(2)*} - \hat{\beta}_1(x_i - \bar{x}_2)\Big) = 0 \quad \Rightarrow$$
$$\begin{aligned}\hat{\beta_0}^{(2)*} &= \frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2}\Big(y_i - \hat{\beta}_1(x_i - \bar{x}_2)\Big) \\ &= \bar{y_2}\end{aligned}$$

(3)
$$\begin{aligned}
&\sum_{i=1}^{n_1}\Big(y_i - \hat{\beta_0}^{(1)*} - \hat{\beta}_1(x_i - \bar{x}_1)\Big)\Big(-(x_i - \bar{x}_1)\Big) \\
&+\sum_{i=n_1+1}^{n_1+n_2}\Big(y_i - \hat{\beta_0}^{(2)*} - \hat{\beta}_1(x_i - \bar{x}_2)\Big)\Big(-(x_i - \bar{x}_2)\Big) = 0 \quad \Rightarrow \\
&\hat{\beta}_1\left[\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2 + \sum_{i=n_1+1}^{n_1+n_2}(x_i - \bar{x}_2)^2\right] = \sum_{i=1}^{n_1}(y_i - \hat{\beta_0}^{(1)*})(x_i - \bar{x}_1) \\
&\qquad + \sum_{i=n_1+1}^{n_1+n_2}(y_i - \hat{\beta_0}^{(2)*})(x_i - \bar{x}_2) \qquad (*)
\end{aligned}$$

Sub $\hat{\beta_0}^{(1)*}$ and $\hat{\beta_0}^{(2)*}$ into $(*)$

$$
\begin{aligned}
\hat{\beta}_1 \left[\sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2 + \sum_{i=n_1+1}^{n_1+n_2} (x_i - \bar{x}_2)^2 \right] &= \sum_{i=1}^{n_1} (y_i - \bar{y_1})(x_i - \bar{x}_1) + \sum_{i=n_1+1}^{n_1+n_2} (y_i - \bar{y_2})(x_i - \bar{x}_2) \\
&= \sum_{i=1}^{n_1} (x_i - \bar{x}_1) y_i + \sum_{i=n_1+1}^{n_1+n_2} (x_i - \bar{x}_2) y_i \\
\Rightarrow \qquad \hat{\beta}_1 &= \frac{\sum_{i=1}^{n_1} (x_i - \bar{x}_1) y_i + \sum_{i=n_1+1}^{n_1+n_2} (x_i - \bar{x}_2) y_i}{\sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2 + \sum_{i=n_1+1}^{n_1+n_2} (x_i - \bar{x}_2)^2}
\end{aligned}
$$