1. (a) The fitted model is

\[ Y = 48.8645 - 0.0273x_1 + 0.0150x_2 + 8.1483x_3. \]

The model explains only 17.77% of the variability in \( Y \).

(b) A 95% C.I. for \( \beta_2 \) is

\[ 0.01504 \pm t_{38}(0.975) \times 0.02485 = 0.015 \pm 0.050. \]

(c) The residual vs fitted value plot is used to check the common variance assumption. The plot should not show any strong pattern if the model is correct. There should be an even spread of points as the fitted values vary. The residual vs fitted value plot seems fine here. The normal Q-Q plot is approximately linear (except for the lower extreme point) indicating that the model assumptions (NID(0, \( \sigma^2 \)) error terms) are reasonable.

(d) A point is a high leverage point in this case if

\[ h_{ii} > 2 \times 4/42 = 0.1905. \]

There is one high leverage point, [10]. High leverage points have extreme \( x \) vectors. In the case of [10], it has the largest \( x_1 \) and \( x_2 \) values.

(e) From the Cook’s distance plot we see that [10] has the largest value around 0.25, but this is not close to \( F_{4,38}(0.5) \) so there are no outliers in this data set.

(f) Test \( H_0 : \beta_1 = \beta_2 = 0 \). The statistic is

\[ f = \frac{(S_1 - S_0)/2}{S_0/38} = \frac{(3878.414 - 3793.872)/2}{3793.872/38} = 0.423. \]

\[ p\text{-value} = P(F_{2,38} \geq 0.423) > 0.5. \] Thus the data are consistent with \( H_0 \), i.e. we can drop both \( x_1 \) and \( x_2 \) from the model.
(g) A 95% C.I. for $\sigma$, using lm3 is

$$
\left( \sqrt{\frac{S_1}{b}}, \sqrt{\frac{S_1}{a}} \right),
$$

where $b = 59.342$ and $a = 24.433$ are the quantiles of the $\chi^2_{40}$ distribution. $S_1 = 3878.414$ so the CI is $(8.084, 12.599)$.

(h) The predicted difference in average HDL levels between people with and without SPB is $\hat{\beta}_3 = 8.377$.

2. (a) (i) This is bookwork.

(ii) The residuals are

$$R_i = Y_i - (\hat{\beta}_0 - \hat{\beta}_1 x_i) = (Y_i - \bar{Y}) - \hat{\beta}_1(x_i - \bar{x}).$$

Since $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ and $\sum_{i=1}^{n}(Y_i - \bar{Y}) = 0$ it follows that the residuals sum to 0.

(b) (i) Test $H_0 : \beta_1 = 1$ vs $H_1 : \beta_1 \neq 1$. Base the test on

$$t = \frac{|\hat{\beta}_1 - 1|}{\text{Est.Std.Err.}} = \frac{1 - 0.8153}{0.1141} = 1.6187.$$

Thus the $p$-value is $P(|t_6| \geq 1.6187) > 0.10$, from the tables. Thus the data are consistent with the claim that the slope is 1.

(ii) A 90% confidence region for the height of a person with arm span 64 inches is

$$64.402 \pm t_6(0.95) \times \sqrt{0.42177^2 + 1.090378^2} = 64.402 \pm 3.15089$$

Thus the region is (61.25, 67.55).