1. (a) Let μ denote the average coffee price in Melbourne (in dollars). The null hypothesis is $\mu = 4.20$, and the alternative hypothesis is $\mu \neq 4.20$.

   (b) The test statistic is
$$t = \frac{3.8 - 4.2}{1.44/\sqrt{19}} \approx -1.21.$$
Using $T \sim t_{18}$, the p-value is
$$p = 2 \times P(T \geqslant |-1.21|), \quad T \sim t_{18}.$$
So, $p = 2 \times (1 - P(T < 1.21)) = 2 \times (1 - 0.879) = 0.242$.

   (c) Assuming a significance level of $\alpha = 0.05$, here we do not have $p < \alpha$. Hence, we do not reject the null hypothesis. The data does not suggest that mean coffee prices differ significantly from \$4.20.

   (d) The 95% confidence interval is given by
$$\overline{x} \pm t_{18,0.975} \times \frac{s}{\sqrt{n}} = 3.8 \pm 2.1 \times \frac{1.44}{\sqrt{19}},$$
giving $(3.1, 4.5)$.
This confidence interval includes 4.2, indicating that we do not reject the null hypothesis. This is consistent with the answer in (c).

2. (a) We are testing the hypotheses
$$H_0\colon \mu = 86 \text{ versus } H_1\colon \mu \neq 86.$$
Since the population standard deviation is known, we use a z-test. The test statistic is
$$z = \frac{91.1 - 86}{8/\sqrt{14}} = 2.385.$$
Using $Z \sim N(0, 1)$, the p-value is
$$p = 2 \times P(Z \geqslant |2.385|) = 2\,(1 - P(Z \leqslant 2.385)) = 2(1 - 0.999) = 0.002.$$
Since $p < 0.05$, we reject the null hypothesis. The data suggests that the mean length of the mice in the forest habitat differs significantly from 86mm.

   (b) i. We are testing the hypotheses
$$H_0\colon \mu = 86 \text{ versus } H_1\colon \mu \neq 86.$$
The sample mean is
$$\overline{x} = \frac{1}{10}(76 + 88 + 91 + 90 + 79 + 96 + 89 + 92 + 100 + 87) = 88.8\text{mm}.$$
The sample variance is
$$\begin{aligned} s^2 &= \tfrac{1}{9}\big((76 - 88.8)^2 + (88 - 88.8)^2 + (91 - 88.8)^2 + (90 - 88.8)^2 + (79 - 88.8)^2 \\ &\qquad + (96 - 88.8)^2 + (89 - 88.8)^2 + (92 - 88.8)^2 + (100 - 88.8)^2 + (87 - 88.8)^2\big) \\ &= \tfrac{1}{9}\big((-12.8)^2 + (-0.8)^2 + (2.2)^2 + (1.2)^2 + (-9.8)^2 + (7.2)^2 + (0.2)^2 + (3.2)^2 + (11.2)^2 + (-1.8)^2\big) \\ &= 50.84 \end{aligned}$$
So $s = \sqrt{50.84} = 7.13$.
Then $\text{SE} = \frac{7.1}{\sqrt{10}} = 2.255$, and the test statistic is
$$t = \frac{88.8 - 86}{2.255} = 1.242.$$
Using $T \sim t_9$, this gives a p-value of
$$p = 2 \times P(T \geqslant |1.242|) = 2 \times (1 - P(T \leqslant 1.242)) = 2 \times (1 - 0.877) = 0.246.$$
Since p is greater than the significance level, we do not reject the null hypothesis. The data does not suggest that the length of deer mice in the forest habitat differs significantly from 86mm.

ii. The 95% confidence interval is

$$\overline{x} \pm t_{9,0.975} \times \text{SE} = 88.8 \pm 2.26 \times 2.255,$$

giving an interval of $(83.7, 93.9)$. The interval includes the possibility of $\mu = 86$, so we do not reject the null hypothesis. This is consistent with our conclusion in (i).

3. (a) Let μ denote the mean fuel economy for the fleeet (in miles per gallon). We are testing

$$H_0\colon \mu \leqslant 26 \text{ versus } H_1\colon \mu > 26$$

(b) The sample was chosen randomly, and the sample size is large (greater than 30) so the necessary assumptions are met.

(c) The test statistic is

$$t = \frac{25.02 - 26}{4.83/\sqrt{50}} = -1.43.$$

(d) Using $T \sim t_{49}$, the p-value is

$$p = P(T \geqslant -1.43) = 1 - P(T \leqslant -1.43) = 1 - 0.92 = 0.08.$$

(Note that we have used the formula for a one-sided test here).

(e) The p-value is greater than 0.05, so we do not reject the null hypothesis. We conclude that there is not strong evidence that the company has failed to meet their fuel economy goal.

4. Let μ denote the mean design time (in hours). We are testing

$$H_0\colon \mu \leqslant 25 \text{ versus } H_1\colon \mu > 25.$$

The test statistic is

$$t = \frac{20.19 - 25}{3.88/\sqrt{43}} = -8.13.$$

Using $T \sim t_{42}$, the p-value is then given by

$$p = P(T \geqslant -8.13),$$

and using `pt(-8.13, df=42, lower.tail=FALSE)` or `1 - pt(-8.13, df=42)` in R gives a p-value of 1 (this precise value of 1 is due to rounding —to 10 decimal places, the p-value is 0.9999999998). This is much greater than 0.05, so we do not reject the null hypothesis. In other words, the data does not support the claim that the average design is 25 hours or more.

5. (a) It represents the mean difference in the measurement obtained by the deuterium dilution technique versus the test weighing technique.

(b) $H_0\colon \mu_d = 0$ versus $H_1\colon \mu_d \neq 0$

(c) i. The estimate of the mean difference is 167, so on average the deuterium dilution technique will give a measurement 167mL higher than the test weighing technique.

ii. The interval is $(35.44928, 299.97929)$. This interval does not contain zero, so we are 95% confident that there is a positive true mean difference between approximately 35.45 and 299.98 mL.

iii. The p-value is 0.01681. So, with a level of significance of $\alpha = 0.05$, we reject the null hypothesis since the p-value is smaller than α.

iv. The data provides evidence that, on average, the two techniques give different measurements. We are 95% confident that the true mean difference is between 35.45 and 299.98, with the deuterium dilution technique typically providing higher measurements on average.

(d) i. The estimated mean difference is $\hat{\mu}_1 - \hat{\mu}_2 = 167.2143$, so the test statistic is

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\text{SE}} = \frac{167.2143}{\sqrt{\frac{352.9697^2}{14} + \frac{234.042^2}{14}}} \approx 1.4773.$$

ii. The p-value is $2P(T \geqslant |1.4773|)$, where $T \sim t_{22.579}$. So, the p-value is

$$p = 2(1 - P(T \leqslant 1.4773)) \approx 2(1 - 0.9233) = 0.1534.$$

The 95% confidence interval is

$$\hat{\mu}_1 - \hat{\mu}_2 \pm t_{0.975,22.579} \times \text{SE} = 167.2143 \pm 2.07 \times \sqrt{\frac{352.9697^2}{14} + \frac{234.042^2}{14}},$$

giving an interval of $(-67.09, 401.51)$.

iii. This time, the p-value is greater than 0.05, and the confidence interval does contain zero. Hence the evidence is not sufficient to suggest that there is a statistically significant difference in the measurements obtained by the two techniques.
Of course, this is the *incorrect* conclusion to make, as the correct test to apply would be a paired t-test.

6. (a) A paired t-test should be used, because the samples are paired.

(b) The data we are using is

Isotopic	1509	1418	1561	1556
Test.weighing	1498	1254	1336	1565
Difference	11	164	225	−9

The estimated mean difference is

$$\hat{\mu}_d = \frac{1}{4}(11 + 164 + 225 - 9) = 97.75$$

The sample variance is

$$s^2 = \frac{1}{3}\left((11 - 97.75)^2 + (164 - 97.75)^2 + (225 - 97.75)^2 + (-9 - 97.75)^2\right) = 13167.58,$$

giving a sample standard deviation of $s = \sqrt{13167.58} = 114.75$. Hence the test statistic is

$$t = \frac{\hat{\mu}_d}{s/\sqrt{n}} = \frac{97.75}{114.75/\sqrt{4}} \approx 1.7037.$$

(c) For the paired t-test, the degrees of freedom is $\text{df} = n - 1 = 3$. So the p-value is $2 \times P(T \geqslant |1.7037|)$, where $T \sim t_3$. This gives $p = 2(1 - 0.9065) = 0.187$. The confidence interval is

$$97.75 \pm 3.18 \times \frac{114.75}{\sqrt{4}},$$

giving a 95% confidence interval of $(-84.84, 280.34)$.

(d) The estimated mean difference is 97.75, but this difference was not significant ($p = 0.187$). We are 95% confident that the true mean is between approximately -84 and 280mL.

(e) We have a much smaller sample size of $n = 4$, which means less confidence and more variability in our estimates.

(f) In Question 1, we rejected H_0 based on evidence from $n = 14$ observations. However, if the study stopped at four observations, then the conclusion would have been different due to less confidence in our estimates. We should therefore avoid making statements such as "H_0 is proven".

7. (a) The approximate 95% confidence interval is $17.56 \pm 1.96 \times 3.70/\sqrt{36} = (16.35,\ 18.77)$.

(b) The estimated mean difference is $20.19 - 17.56 = 2.63$ hours.

(c) We have

$$\text{SE} = \sqrt{\frac{3.88^2}{43} + \frac{3.70^2}{36}} \approx 0.855.$$

(d) The test statistic is $t = 2.63/0.855 \approx 3.08$.

(e) The approximate p-value is $2 \times (1 - 0.999) = 0.002$.

(f) Since our p-value is less than 0.05 we reject that the means are the same. In fact, the p-value is very small so the evidence for there being a difference is strong.

(g) $(20.19 - 17.56) \pm 1.99 \times \text{SE} = (0.93,\ 4.33)$. This interval suggests that the New York office has a higher mean design time.

(h) Based on our collected data, we conclude that there is a difference in mean design times between the New York and Los Angeles offices ($p = 0.002$). We estimate that the New York office takes, on average, somewhere between 0.95 hours and 4.31 hours longer than the Los Angeles office.