One-sample hypothesis testing

1. The price of a cup of coffee varies from location to location. It has been claimed that the average price for a cup of coffee in Melbourne is \$4.20. To test this hypothesis, 19 coffee stores in Melbourne were sampled. A mean price of \$3.80 was obtained, with a standard deviation of \$1.44. Assume that coffee prices are normally distributed.

   (a) State an appropriate null hypothesis and alternative hypothesis.

   (b) Perform a t-test and find the p-value for your stated hypotheses. Use the fact that $P(T \leqslant 1.21) \approx 0.879$ for $T \sim t_{18}$.

   (c) State an appropriate conclusion, with justification.

   (d) Using the fact that $t_{18,0.975} \approx 2.1$, find a 95% confidence interval for the average coffee price in Melbourne. Does this interval support your conclusion from (c)?

2. Deer mice *Peromyscus maniculatus* are small rodents native to North America. Their adult body lengths (excluding the tail) are known to vary approximately normally, with mean $\mu = 86$mm and standard deviation $\sigma = 8$mm. Deer mice are found in diverse habitats and exhibit different adaptations to their environment. A random sample of 14 deer mice in a rich forest habitat gives an average body length of $\overline{x} = 91.1$ mm.

   Let μ denote the mean body length of adult deer mice in this forest habitat.

   (a) Use an appropriate test to test the hypothesis that the average length of deer mice differs from $\mu = 86$mm at a 5% level of significance. You are given that $P(Z \leqslant 2.385) \approx 0.99$ for $Z \sim N(0, 1)$.

   (b) Now assume that the population standard deviation is *not* known. Consider the following sample data

   $$76, 88, 91, 90, 79, 96, 89, 92, 100, 87.$$

      i. Test the hypothesis that the average length of adult deer mice in the forest habitat differs from $\mu = 86$mm at a 5% level of significance. State an appropriate conclusion, and justify your answer. You are given that $P(T \leqslant 1.242) \approx 0.877$ for $T \sim t_9$.

      ii. Determine a 95% confidence interval for the average length of adult deer mice in the forest habitat. Use the fact that $t_{9,0.975} \approx 2.26$. Does your confidence interval support your conclusion from (i)?

3. A company with a large fleet of cars hopes to keep petrol costs down and sets a goal of attaining a fleet average of at least 26 miles per gallon. To see if the goal is being met, they check the petrol usage for 50 company trips chosen at random, finding a mean of 25.02 miles per gallons and a standard deviation of 4.83 miles per gallons. Is this strong evidence that they have failed to attain their fuel economy goal?

   (a) Write appropriate hypotheses. (Note: this will be a *one-sided* test).

   (b) Are the necessary assumptions to make inferences satisfied?

   (c) Calculate the appropriate test statistic for a t-test.

   (d) Given that $P(T \leqslant -1.43) \approx 0.08$ for $T \sim t_{49}$, find the p-value.

   (e) State an appropriate conclusion, justifying your answer. Assume a significance level of $\alpha = 0.05$.

4. Consider the same scenario from Question 2 of Workshop 7: the company Kraft wanted to see whether they should invest in the training of staff at certain locations to improve efficiency. Of interest is the time required (in hours) for designing advertisements for product promotions. In the initial stages, data from the New York office was collected with a mean design time of 20.19 hours and a sample standard deviation of 3.88. The sample size is $n = 43$. A Kraft employee who visited the New York office believes that the New York design team is slow, claiming that the average design could be 25 hours or more.

   Formulate appropriate hypotheses for testing the employee's claim, and then perform an appropriate test and state your conclusion. Use the `pt` command in R to find the p-value.

Two-sample hypothesis testing

Scientists and engineers frequently wish to compare two different measurement techniques. The table below reports two data on two different measurements of the amount of milk (in mL) ingested by 14 randomly selected infants [1].

The milk consumed by each infant was measured using a deuterium dilution technique ('isotopic') and a test weighing procedure ('test-weighing').

Isotopic	1509	1418	1561	1556	2169	1760	1098	1198	1479	1281	1414	1954	2174	2058
Test.weighing	1498	1254	1336	1565	2000	1318	1410	1129	1342	1124	1468	1604	1722	1518
Difference	11	164	225	−9	169	442	−312	69	137	157	−54	350	452	540

The purpose of the study was to determine whether the novel approach using deuterium had measurements comparable to the conventional test-weighing procedure. Importantly, note that this study is not aiming to determine whether one technique is more accurate than the other. Rather, it is aiming to determine only whether the two techniques are *different*.

5. Let μ_1 denote the mean measurement using the isotopic procedure and let μ_2 denote the mean measurement using the test-weighing procedure.
   (a) What does the mean difference $\mu_1 - \mu_2$ represent?
   (b) State appropriate hypotheses for the study.
   (c) Now suppose that the vectors `Isotopic` and `Test.weighing` contain the data from each row in the table above. The R command below was used to perform a paired t-test.

```
t.test(Isotopic, Test.weighing, paired=TRUE)
```

   which resulted in the following output:

```
        Paired t-test

data:  Isotopic and Test.weighing
t = 2.7416, df = 13, p-value = 0.01681
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  35.44928 298.97929
sample estimates:
mean of the differences
       167.2143
```

   Based on the R output above, answer the following questions.
      i. Provide the estimate for the mean difference. Interpret it in context.
      ii. What is the 95% confidence interval for the mean difference? Does the interval provide evidence that the true mean difference is not equal to zero? Explain.
      iii. Give the p-value for the paired t-test, and state an appropriate conclusion.
      iv. Summarise your findings of this paired t-test.
   (d) Suppose instead that an *unpaired* t-test was (incorrectly) used for the study.
      i. Using the information from the R output below, calculate the test statistic for the unpaired t-test.

```
> sd(Isotopic)
[1] 352.9697
> sd(Test.weighing)
[1] 234.042
```

      Hint: the mean difference from part (c) will make an appearance in the calculation.
      ii. You are given that df = 22.579 for the unpaired t-test in this case. Based on the R output below, determine the p-value and the 95% confidence interval for the mean difference.

```
> pt(1.4773, df=22.579)
[1] 0.9232956
> qt(0.975, df=22.579)
[1] 2.070794
```

      iii. How does your conclusion differ in this instance?

6. You are now going to conduct the paired t-test yourself. However, this time suppose that only the first four observations (pairs) have been collected for testing.
   (a) Should a paired or unpaired test be used? Explain.
   (b) Calculate the test statistic for the test.
   (c) Some R commands and corresponding output are shown below.

```
> pt(1.7037, df=3)
[1] 0.9065067
> pt(1.2388, df=4.2285)
[1] 0.860129
> qt(0.975, df=3)
[1] 3.182446
> qt(0.975, df=4.2285)
[1] 2.718264
```

   Use the appropriate lines from above to calculate the p-value and the 95% confidence interval.
   (d) Summarise the findings.
   (e) This time, you should have found that you could not reject the null hypothesis. Why do you think this is?
   (f) Explain why this question highlights why it is important to realise that failing to reject the null hypothesis is not the same as proving that the null hypothesis is true. Make sure your summary in (b) does not make such a claim.

7. Once again we will consider the scenario from Workshop 7, where Kraft wanted to see whether they should invest in the training of staff at certain locations to improve efficiency with regards to design time of promotional products. Now suppose that they are interested in comparing the mean design times between two offices. Based on a sample size of 43, the sample mean and sample deviation for the New York office was found to be 20.19 and 3.88 respectively. For the Los Angeles office they were 17.56 and 3.70 based on 36 observations. You may use the fact that for this test, df $\approx$ 76.
   (a) In Lab 7, we found that an approximate 95% confidence interval for the mean design time for the New York office was (19.03, 21.35). Similarly, calculate an approximate 95% confidence interval for the mean design time for the Los Angeles office.
   (b) What is the estimated mean difference between the two offices?
   (c) Calculate the standard error for this estimated mean difference.
   (d) Calculate the test statistic for the t-test for comparing the two means.
   (e) Given that $P(T \leqslant 3.08) \approx 0.999$ for $T \sim t_{76}$, calculate an approximate p-value for the test.
   (f) Choosing a level of significance of $\alpha = 0.05$, do you reject that the two means are the same? Explain.
   (g) Now, calculate a 95% confidence interval for the difference in two means. You may use the fact that $t_{76,0.975} = 1.99$. Does the interval suggest a difference in the mean design times?
   (h) Write a simple statement that summarises your findings.

References

[1] N F Butte, C Garza, E O Smith, and B L Nichols. Evaluation of the deuterium dilution technique against the test-weighing procedure for the determination of breast milk intake. *The American Journal of Clinical Nutrition*, 37(6):996–1003, June 1983.