Curtin University, Semester 2, 2026

ECOM5005 Tutorial 9 Solution

Hypothesis Testing: Two-sample Tests

Learning Outcomes

1. Conduct hypothesis tests for the means of two independent populations.
2. Conduct hypothesis tests for the means of two related populations.
3. Conduct hypothesis tests for two population proportions.
4. Conduct hypothesis tests for the variances of two independent populations.

PART 1 – Detailed solutions

Question 1

Given $S_1^2 = 4$, $n_1 = 16$; $S_2^2 = 6$, $n_2 = 25$, calculate the pooled-test degrees of freedom and pooled sample variance.

Detailed solution

- For a pooled-variance two-sample t test, the degrees of freedom are $df = n_1 + n_2 - 2$.
- Substitute $n_1 = 16$ and $n_2 = 25$: $df = 16 + 25 - 2 = 39$.
- The pooled variance is a degrees-of-freedom-weighted average of the two sample variances.

$$S_p^2 = [(n_1-1)S_1^2 + (n_2-1)S_2^2] / (n_1+n_2-2)$$

$$= [15(4) + 24(6)] / 39 = 204 / 39 = 5.2308$$

Answer

(1) df = 39 → Option D.

(2) $S_p^2 \approx 5.23$ → Option D.

Question 2. Pooled-variance t test

Female MBA graduates are Population 1 and male MBA graduates are Population 2. The researcher wants to test whether the female population mean starting salary is lower.

	Females	Males
n	18	12
$\bar{x}$	\$48,266.70	\$55,000.00
s	\$13,577.63	\$11,741.29

Detailed solution

1. State the hypotheses: H_0: μfemale ≥ μmale and H_1: μfemale < μmale. Therefore the correct alternative is Option C.

2. Because the exercise specifies a pooled-variance t test, estimate the common variance using both sample variances.

$$S_p^2 = [17(13{,}577.63^2) + 11(11{,}741.29^2)] / 28 = 166{,}086{,}479.24$$

$$S_p = \sqrt{166{,}086{,}479.24} = 12{,}887.45$$

$$SE = S_p \sqrt{(1/18 + 1/12)} = 4{,}802.87$$

$$t = (48{,}266.70 - 55{,}000.00) / 4{,}802.87 = -1.40193$$

- For df = 28, the lower-tail p-value is approximately 0.08596.
- At $\alpha = 0.05$, $p = 0.08596 > 0.05$, so fail to reject H_0.
- There is insufficient evidence at the 5% level to conclude that female MBA graduates have a lower mean starting salary than male MBA graduates.

$$90\%\ CI: (\bar{x}female - \bar{x}male) \pm t_{0.05,28} \times SE$$

$$= -6{,}733.30 \pm 1.70113(4{,}802.87) = (-14{,}903.61, 1{,}437.01)$$

- The 90% interval contains 0, consistent with the one-tail test not rejecting H_0 at $\alpha = 0.05$.
- Do not interpret this as proof that the population means are identical; the sample is simply not strong enough to establish a lower female mean at the chosen significance level.

Answer

(1) Option C: H_1: $\mu female < \mu male$.

(2) t = -1.40193.

(3) Fail to reject H_0 at 5%; insufficient evidence of a lower female mean.

(4) 90% CI for $\mu female - \mu male$: approximately (-$14,903.61, $1,437.01).

Question 3. EV fast-charging software update – paired t test

The same 12 EVs are timed before and after a battery-control update. Test whether mean charging time is reduced.

EV	Before	After	D = Before - After
1	31.2	29.9	1.3
2	33.5	31.7	1.8
3	29.8	29.1	0.7
4	35.1	33.0	2.1
5	32.4	30.8	1.6
6	30.6	29.7	0.9
7	34.0	31.8	2.2
8	36.2	34.1	2.1
9	31.7	30.5	1.2
10	33.1	31.4	1.7
11	29.5	28.8	0.7
12	35.4	33.2	2.2

Detailed solution

- The measurements are paired because each after-update observation belongs to the same vehicle as one before-update observation. The statistical unit for testing is the within-vehicle difference.
- Define D = Before - After. A positive D means the update reduces charging time.
- Hypotheses: H_0: $\mu D \leq 0$ versus H_1: $\mu D > 0$.

Quantity	Value
n	12
Mean difference $\bar{D}$	1.5417 min
SD of differences SD	0.5712 min
SE	0.1649 min
df	11

$$t = \bar{D} / (SD/\sqrt{n}) = 1.5417 / 0.1649 = 9.3493$$

- The upper-tail p-value is approximately 0.00000072, far below 0.05. Reject H_0.
- There is very strong evidence that the software update reduces mean charging time.
- For the 95% CI, t* = 2.20099 with df = 11.

$$95\%\ CI\ for\ \mu D = 1.5417 \pm 2.20099(0.1649) = (1.1787, 1.9046)\ minutes$$

- Interpretation: the update is estimated to reduce the mean time to add 40 kWh by about 1.18 to 1.90 minutes.
- For n = 12, the paired differences should be approximately normal with no severe outliers; normality is required for the differences, not separately for before and after measurements.

Answer

Use a paired t test. t = 9.3493, df = 11, one-tail $p < 0.000001$. Reject H_0.

95% CI for the mean reduction: 1.18 to 1.90 minutes.

Question 4. E-commerce checkout redesign – two population proportions

Checkout A: 136/800 conversions. Checkout B: 158/750 conversions. Test whether checkout B has the higher population conversion rate.

Detailed solution

- Let πA and πB be the population conversion rates. Hypotheses: H_0: $\pi B \le \pi A$ versus H_1: $\pi B > \pi A$.
- Sample proportions: pA = 136/800 = 0.1700; pB = 158/750 = 0.2107.
- Under H_0 of equality, use the pooled estimate $\bar{p} = (136+158)/(800+750) = 0.18968$.

$$SE_0 = \sqrt{[\bar{p}(1-\bar{p})(1/800+1/750)]} = 0.01993$$

$$Z = (0.21067 - 0.17000) / 0.01993 = 2.0409$$

- The upper-tail p-value is approximately 0.02063. Since 0.02063 < 0.05, reject H_0.
- There is statistically significant evidence that the redesigned checkout has a higher conversion rate.

$$95\%\ CI\ for\ \pi B - \pi A = 0.04067 \pm 1.96\sqrt{[pA(1-pA)/800 + pB(1-pB)/750]}$$

$$= (0.00156, 0.07977) = (0.16, 7.98)\ percentage\ points$$

- The estimated lift is 4.07 percentage points, but the confidence interval is wide. Product management should consider the economic value of the lift, implementation cost and whether the test was properly randomized.
- Statistical significance answers whether sampling variation plausibly explains the observed lift; practical significance asks whether the size of the lift is commercially meaningful.

Answer

Z = 2.0409, one-tail p = 0.0206. Reject H_0 at 5%.

Estimated conversion-rate improvement: 4.07 percentage points; 95% CI ≈ 0.16 to 7.98 percentage points.

Question 5. Emergency-department triage systems – unequal-variance t test

AI-assisted: $n_1 = 38$, $\bar{x}_1 = 41.6$ min, $s_1 = 12.8$ min. Standard: $n_2 = 34$, $\bar{x}_2 = 49.9$ min, $s_2 = 20.6$ min.

Step 1: Test equality of variances

- Put the larger variance in the numerator so $F \geq 1$: $F = s_2^2/s_1^2 = 20.6^2/12.8^2 = 2.5901$.
- Degrees of freedom: $df_1 = 34\text{-}1 = 33$ and $df_2 = 38\text{-}1 = 37$.
- The two-sided F-test p-value is approximately 0.00559. Reject equal variances at $\alpha = 0.05$.

Step 2: Use Welch's unequal-variance t test

- Hypotheses: H_0: μAI ≥ μStandard versus H_1: μAI < μStandard.
- Because the F test indicates unequal variances, use a separate-variance (Welch) t test.

$$SE = \sqrt{(12.8^2/38 + 20.6^2/34)} = 4.0979$$

$$t = (41.6 - 49.9)/4.0979 = -2.0254$$

- Welch-Satterthwaite df ≈ 53.99. The lower-tail p-value is approximately 0.02389.
- Since $p < 0.05$, reject H_0. There is evidence that the AI-assisted system has a lower mean waiting time.

95% CI for μAI - μStandard ≈ (-16.52, -0.08) minutes

- The entire interval is below zero, supporting a lower mean waiting time under AI-assisted triage.
- The upper endpoint is close to zero, so the estimated improvement is statistically significant at 5% but uncertain in magnitude; operational costs and clinical outcomes should also be considered.

Answer

F = 2.5901, two-tail p = 0.00559 → equal variances are not supported.

Welch t = -2.0254, df ≈ 53.99, one-tail p = 0.0239 → reject H_0 at 5%.

95% CI for μAI - μStandard ≈ (-16.52, -0.08) minutes.

PART 2 – Palmer Penguins data exercise: Detailed Excel solution

Dataset context

- The workbook contains 344 penguins from three species observed in the Palmer Archipelago, Antarctica. Data were collected by Dr. Kristen Gorman and Palmer Station LTER.
- For the main analysis, the observational unit is one penguin, the numerical response is body_mass_g, and the grouping variable is sex within the Adelie species.
- The prepared Adelie_Comparison sheet contains 73 male and 73 female Adelie penguins with recorded sex and body mass.

Source: https://allisonhorst.github.io/palmerpenguins/ ; Gorman, Williams & Fraser (2014), PLOS ONE 9(3): e90081.

Before starting: turn on Excel Analysis ToolPak

1. Open Excel. Choose File → Options → Add-ins.
2. At the bottom, next to Manage: Excel Add-ins, click Go.
3. Tick Analysis ToolPak and click OK.
4. Return to the Data tab. You should now see Data Analysis on the right side of the ribbon.

Part 2(a)-(b): Descriptive statistics

1. Open ECOM5005_Tutorial_9_Palmer_Penguins_Data.xlsx and go to Adelie_Comparison.
2. Columns A and B contain male and female Adelie body mass in grams. Row 3 contains the labels; rows 4:76 contain 73 observations in each group.
3. Choose Data → Data Analysis → Descriptive Statistics.
4. Set Input Range to $A$3:$B$76, choose Grouped By: Columns, tick Labels in first row and Summary statistics.
5. Choose an Output Range or New Worksheet Ply, then click OK.

Statistic	Male Adelie	Female Adelie
n	73	73
Mean (g)	4,043.49	3,368.84
Median (g)	4,000	3,400
Standard deviation (g)	346.81	269.38
Sample variance	120,278.25	72,565.64
Minimum (g)	3,325	2,850
Maximum (g)	4,775	3,900

- The sample mean difference is 4,043.49 - 3,368.84 = 674.66 g, with males heavier on average.
- The male sample also has a larger standard deviation, suggesting more variability in male body mass.

Part 2(c): Box-and-whisker chart

1. Select A3:B76 on Adelie_Comparison.
2. Choose Insert → Insert Statistic Chart → Box & Whisker.
3. Change the chart title to “Adelie Penguin Body Mass by Sex”.
4. Add the vertical-axis title “Body mass (g)”.

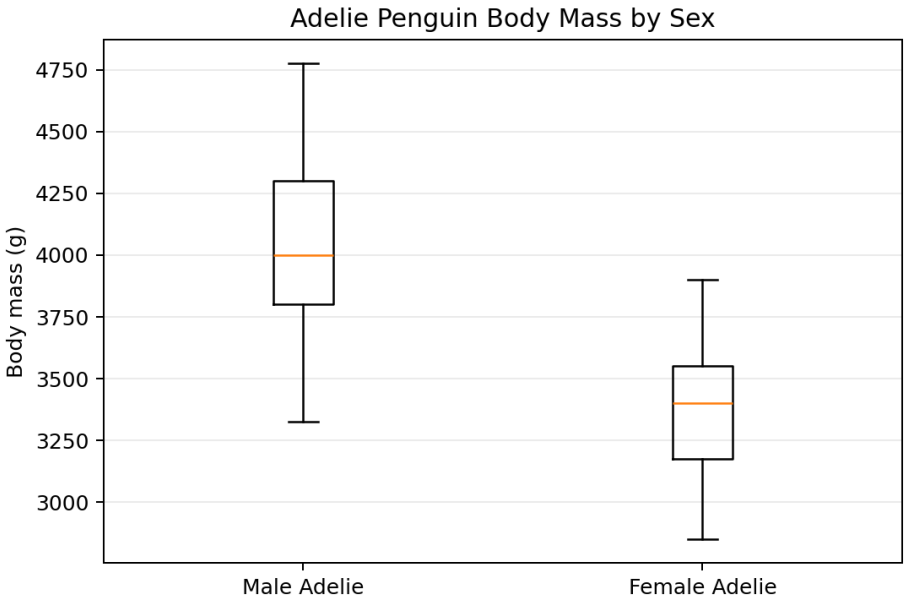


Figure 1: Final box-and-whisker comparison for the Adelie body-mass data

- The male distribution is shifted upward relative to the female distribution.
- There is some overlap, but the separation in the centres is large relative to the within-group spread.

Part 2(d): Pooled-variance t test for mean body mass

- We first follow the question's equal-variance assumption. Let Population 1 = male Adelie and Population 2 = female Adelie.
- Hypotheses: H_0: μmale ≤ μfemale versus H_1: μmale > μfemale.

1. Choose Data → Data Analysis → t-Test: Two-Sample Assuming Equal Variances.
2. Variable 1 Range: $A$3:$A$76; Variable 2 Range: $B$3:$B$76.
3. Hypothesized Mean Difference: 0. Tick Labels. Alpha: 0.05.
4. Choose an output location and click OK. Because the alternative is μmale > μfemale, use the one-tail p-value.

Output	Value
Male mean	4,043.493
Female mean	3,368.836
Pooled variance	96,421.946
df	144
t Stat	13.1263
P(T≤t) one-tail	$\approx 1.11 \times 10^{-26}$
t Critical one-tail	≈ 1.655

- The p-value is effectively zero at normal reporting precision. Reject H_0.
- There is extremely strong statistical evidence that male Adelie penguins have a higher population mean body mass than female Adelie penguins.

Part 2(e): 95% confidence interval for μmale - μfemale

$$\text{Difference} = 4{,}043.493 - 3{,}368.836 = 674.658 \text{ g}$$

$$SE_{pooled} = \sqrt{[96{,}421.946(1/73 + 1/73)]} = 51.397 \text{ g}$$

$$t^* = \text{T.INV.2T}(0.05,144) = 1.9766$$

$$95\% \text{ CI} = 674.658 \pm 1.9766(51.397) = (573.07, 776.25) \text{ g}$$

- Interpretation: we are 95% confident that male Adelie penguins in the represented population are, on average, about 573 to 776 grams heavier than female Adelie penguins.
- The confidence interval does not contain zero, consistent with rejecting equality of means.

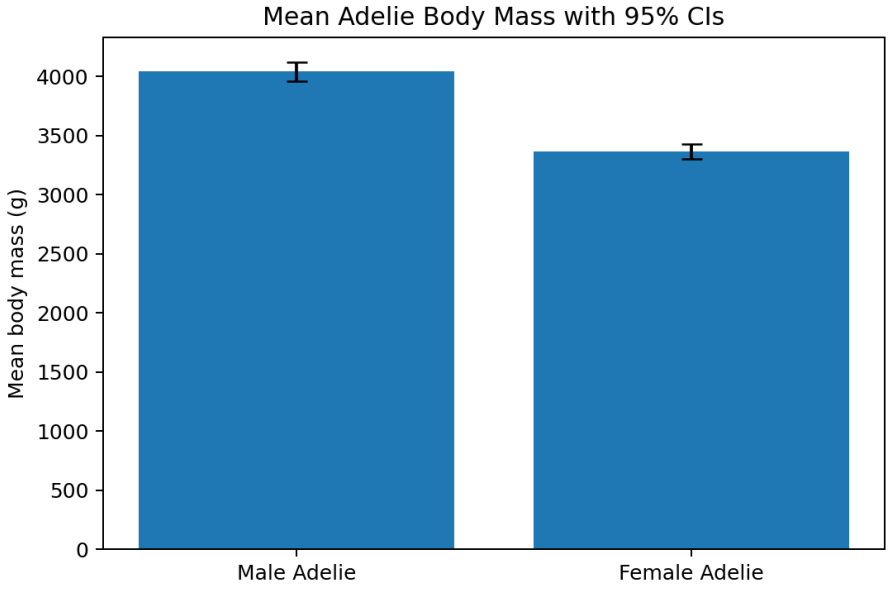


Figure 2: Sample mean body mass with separate 95% confidence intervals for each group mean

Part 2(f): F test for equal variances

1. Choose Data → Data Analysis → F-Test Two-Sample for Variances.
2. Variable 1 Range: male A3:A76; Variable 2 Range: female B3:B76; tick Labels and set Alpha = 0.05.
3. For a direct two-sided p-value, Excel can also use =F.TEST(A4:A76,B4:B76).

$$F = 120{,}278.253 / 72{,}565.639 = 1.6575$$

- The two-sided p-value is approximately 0.03356. Since $p < 0.05$, reject the equal-variance null hypothesis.
- This means the pooled-variance assumption is questionable for these data, so the unequal-variance (Welch) t test is preferred as a robustness check.

Repeat with Welch's unequal-variance t test

1. Choose Data → Data Analysis → t-Test: Two-Sample Assuming Unequal Variances.
2. Use the same male and female ranges and Hypothesized Mean Difference = 0.
3. For a one-tail Excel formula you can also use =T.TEST(A4:A76,B4:B76,1,3).

Welch result	Value
t Stat	13.1263
Approx. df	135.69
One-tail p-value	$\approx 3.20 \times 10^{-26}$
Decision	Reject H_0

- The substantive conclusion is unchanged: the evidence of a higher male mean body mass is overwhelming.
- This is a useful example of why checking assumptions matters even when the final inference is robust.

Part 2(g): Extension – compare male proportions for Adelie and Gentoo

Species	Male	Female	Known sex	Male proportion
Adelie	73	73	146	0.5000
Gentoo	61	58	119	0.5126

- Hypotheses: H_0: πAdelie = πGentoo versus H_1: πAdelie ≠ πGentoo.
- Pooled proportion $\bar{p}$ = (73+61)/(146+119) = 0.50566.
- All success/failure counts are far above 5, so the normal approximation is appropriate.

$$Z = (0.5000 - 0.5126) / \sqrt{[0.50566(1-0.50566)(1/146+1/119)]} = -0.2041$$

$$\text{Two-tail p-value} = 0.8382$$

- Fail to reject H_0 at α = 0.05. The sample does not provide evidence that the recorded male proportions differ between Adelie and Gentoo penguins.
- Failing to reject is not proof that the proportions are exactly equal; it means the observed difference is entirely consistent with ordinary sampling variation.

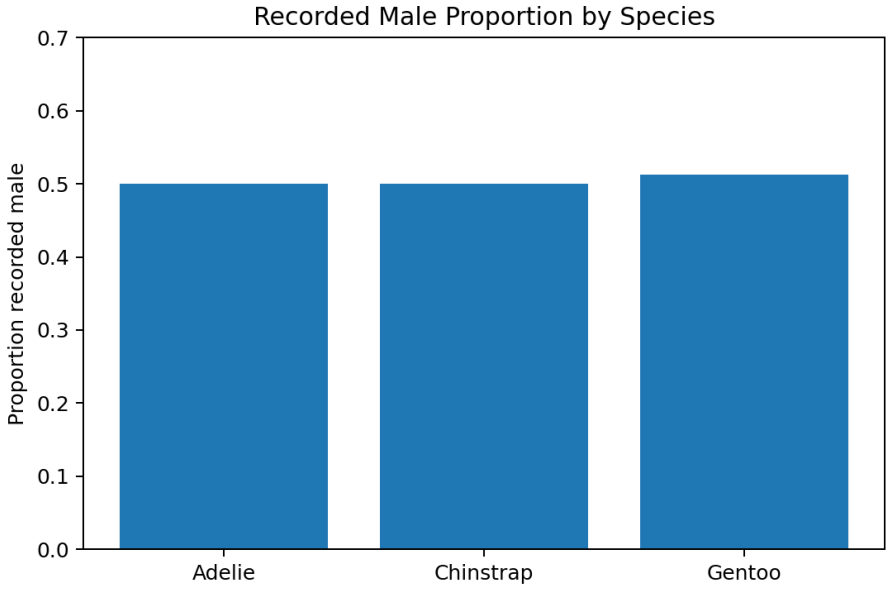


Figure 3: Recorded male proportions for the three penguin species

Part 2(h): Example 100-word statistical report

Among 73 male and 73 female Adelie penguins with recorded body mass, males were heavier on average (4,043.5 g versus 3,368.8 g). The pooled two-sample t test gave t = 13.13 with a one-tail p-value well below 0.001. A 95% confidence interval estimated the male-minus-female mean difference at about 573 to 776 g. An F test indicated unequal variances (p = 0.0336), but Welch's unequal-variance test gave the same inference. The difference is therefore statistically very strong. Biological interpretation should still consider species ecology, sampling design and whether these sampled adults represent the wider Adelie population.

Final checks for two-sample testing

- Independent samples: observations in one group must not determine observations in the other group.
- Paired samples: analyze within-pair differences, not the two raw groups as if they were independent.
- Pooled t test: use only when equal population variances are a defensible assumption.
- Welch t test: appropriate when variances are not assumed equal.
- Two proportions: under H_0 of equal proportions, use a pooled proportion in the test standard error.
- Always report the statistical decision in the context of the real-world question, and do not equate "not significant" with "exactly equal".

PART 3 – Pearson MyLab

1. Please ensure that you have finished the MyLab Homework #4.

Note: If you are having issues with obtaining access to MyLab, please re-read the instructions (available in the welcome topic). If you are still unsuccessful, please send an email explaining the issue. It is important that you complete MyLab tasks on a weekly basis since the final e-test in this unit will be completed using MyLab.