Monday Chapter 9 worked solutions

The two problems from the November 9 in-class practice appear below, with each part's worked solution immediately after its question. The full reference table is on page 8. Relevant table entries are also repeated where they are used.

These are original, fictitious teaching datasets. State the parameter, assumptions and tail direction before calculating. Use a calculator for arithmetic. Write conclusions about population means, not individual plants or flasks.

Use the Chapter 9 notation: df is degrees of freedom, and $P(T_{\text{df}} > t_{\text{df},\alpha}) = \alpha$. A one-sided level-$\alpha$ test uses $t_{\text{df},\alpha}$; a two-sided level-α test or a $100(1-\alpha)\%$ confidence interval uses $t_{\text{df},\alpha/2}$.

Problem 1 Does the nutrient treatment improve growth?

A botanist wants to study whether a new nutrient treatment increases mean stem growth compared with the standard treatment. She considers two designs:

- **Design A:** select eight independent mother plants. Take two comparable cuttings from each mother plant; these two cuttings form one pair. Randomly assign one cutting in each pair to the new treatment and the other to the standard treatment. Measure growth after the same period. Assume the eight paired growth differences are independent across mother plants and normally distributed.
- **Design B:** select sixteen independent plants, each from a different mother plant, and randomly assign eight to each treatment. Assume normally distributed growth in each group with a common, unknown population variance.

The botanist uses **Design A**. For each mother plant, subtract the growth of its standard-treatment cutting from the growth of its new-treatment cutting. Thus each difference compares the two cuttings from the same mother plant. Define $D =$ new-treatment growth minus standard-treatment growth, in millimeters. The eight observed differences are

$$0, \quad 1, \quad 2, \quad 3, \quad 3, \quad 4, \quad 5, \quad 6.$$

Let μ_D be the population mean difference. Test $H_0 : \mu_D = 0$ against $H_a : \mu_D > 0$ at significance level $\alpha = 0.05$.

(a) Design and method. Which test, paired t or independent pooled t, belongs to each design? Give the degrees of freedom for each. Explain why the eight differences in Design A constitute the sample for its test, even though sixteen cuttings were measured. Does a paired test require equal population variances for the two treatments?

Solution. Design A requires a paired t test with df $= 7$**.** Define $D_i = X_i - Y_i$ within mother plant i. The parameter μ_D is the population mean paired difference; a positive value represents greater mean growth under the new treatment. The eight independent differences form a one-sample problem:

$$T = \frac{\overline{D} - 0}{S_D/\sqrt{8}} \sim T_7 \quad \text{under } H_0, \qquad \text{df} = n - 1 = 8 - 1 = 7.$$

Estimating the one mean $\overline{D}$ leaves residuals whose sum is zero, so there are seven independent residual directions. The two cuttings in a pair share a mother plant. They are not sixteen independent observations for this analysis.

Equal treatment variances are **not required**. The paired test estimates the variance of the differences,

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\,\text{Cov}(X, Y),$$

using s_D^2, so it includes the within-pair covariance. The assumptions are independent, normal differences with unknown variance; separate normality of both measurements is unnecessary.

Design B requires an independent pooled t test with df $= 14$**.** Independent normal groups and a common unknown variance justify pooling. The degrees of freedom are $(8-1) + (8-1) = 14$. The observed Design A differences cannot supply a Design B test statistic.

Problem 1 Calculation and evidence

(b) Calculation and evidence. Calculate $\bar{d}$, s_D^2, the estimated standard error and the test statistic. Use the shared table on page 8 to give a bound on the one-sided p-value. State the decision and a conclusion in context.

Solution. The observed paired differences in millimeters are $0, 1, 2, 3, 3, 4, 5, 6$, with $n = 8$. With D defined as new minus standard growth, test $H_0 : \mu_D = 0$ against $H_a : \mu_D > 0$ at $\alpha = 0.05$.

First calculate the sample mean and the variability of the paired differences. Since $\sum d_i = 24$ and $\sum d_i^2 = 100$,

$$\bar{d} = \frac{24}{8} = 3 \text{ mm},$$

$$\sum_{i=1}^{8} (d_i - \bar{d})^2 = 9 + 4 + 1 + 0 + 0 + 1 + 4 + 9 = 28 \text{ mm}^2,$$

$$s_D^2 = \frac{28}{8-1} = 4 \text{ mm}^2, \qquad s_D = 2 \text{ mm},$$

$$\widehat{\mathrm{SE}}(\overline{D}) = \frac{s_D}{\sqrt{n}} = \frac{2}{\sqrt{8}} = 0.7071 \text{ mm},$$

$$t_{\text{obs}} = \frac{\bar{d} - 0}{s_D/\sqrt{n}} = \frac{3}{0.7071} = 4.2426.$$

The sample standard deviation describes variability *among paired differences*; the standard error describes the estimated sampling variability *of their mean*. The test statistic has no units.

Choose the right tail. Because the alternative is $\mu_D > 0$, large positive statistics support the alternative. At level 0.05, use the table row df = 7 and column 0.05:

$$t_{7,0.05} = 1.895, \qquad R_0 = \{t > 1.895\}.$$

Since $4.2426 > 1.895$, **reject** H_0.

Bound the one-sided p-value. In the same row, the entries in columns 0.005 and 0.001 are

$$t_{7,0.005} = 3.499 < 4.2426 < 4.785 = t_{7,0.001}.$$

Larger positive statistics have smaller right-tail probabilities. Consequently,

$$\boxed{0.001 < p = P(T_7 \geq 4.2426) < 0.005.}$$

Do not double this probability, because the alternative is one-sided. A software check gives $p \approx 0.001914$; the table bound is the requested answer.

Conclusion. At the 5% significance level, the data provide evidence that the new nutrient treatment increases population mean stem growth. The estimated mean paired increase is **3 mm**.

This is a conclusion about the *population mean difference*, not a claim that every treated cutting grows more. Generalization beyond the sampled type of plants depends on how the mother plants represent the intended population.

Problem 1 Errors significance and power

(c) Errors and significance. Describe a Type I error and a Type II error in this setting. Explain the meaning of $\alpha = 0.05$. Does rejecting H_0 tell us that our conclusion has a 5% probability of being wrong? Explain.

Solution. For the test $H_0 : \mu_D = 0$ against $H_a : \mu_D > 0$:

- **Type I error:** conclude that the treatment increases mean growth when the true population mean difference is zero. More broadly, for the one-sided null of no improvement ($\mu_D \leq 0$), it is a false positive claim of improvement.
- **Type II error:** fail to reject H_0 when the treatment truly increases mean growth ($\mu_D > 0$). Its probability depends on the actual positive effect, variability, design and sample size.
- **Meaning of $\alpha = 0.05$:** when $\mu_D = 0$ and the assumptions hold, this decision rule rejects in 5% of repeated samples. For $\mu_D < 0$, the false-positive probability for this upper-tail rule is no greater than 5%.

Rejecting does **not** mean that this particular conclusion has a 5% probability of being wrong. The significance level concerns

$$P(\text{reject } H_0 \mid H_0 \text{ true}), \quad \text{not} \quad P(H_0 \text{ true} \mid \text{reject } H_0).$$

Likewise, the p-value is a probability of data at least this extreme under the null model, not the probability that the null is true.

(d) Power at a specified effect. A separate planning calculation for this same eight-pair, level-0.05 t test gives

$$P(\text{reject } H_0 \mid \mu_D = 2 \text{ mm}, \ \sigma_D = 2 \text{ mm}) \approx 0.815.$$

Use this supplied value; you do not need to derive it. Each repetition of the test still estimates its standard deviation from its own sample.

Find β and power at this specified alternative, and interpret both. If the botanist lowers α from 0.05 to 0.01 while keeping the sample size, true effect and population variation fixed, what happens to the rejection region, β and power? Is $\beta = 1 - \alpha$?

Solution. At the specified alternative $\mu_D = 2$ mm and $\sigma_D = 2$ mm,

$$\text{power} \approx 0.815, \qquad \boxed{\beta = 1 - 0.815 = 0.185.}$$

Across repeated eight-pair studies at that effect and variability, about **81.5% detect improvement** and about **18.5% fail to detect it**. These are planning probabilities at a specified alternative, not probabilities computed from the observed mean of 3 mm. The supplied σ_D describes true variation for planning; each repetition still estimates s_D and uses a t test.

Lowering α to 0.01 changes the cutoff from $t_{7,0.05} = 1.895$ to $t_{7,0.01} = 2.998$. The rejection region shrinks:

$$\{t > 1.895\} \quad \longrightarrow \quad \{t > 2.998\}.$$

With the effect, population variation and n fixed, fewer repetitions reject. Thus β **increases and power decreases**.

It is **not** true that $\beta = 1 - \alpha$. The error rate α is evaluated under the null, while β is evaluated under a particular alternative. The complementary relationship is power $= 1 - \beta$.

Problem 2 Which interval supports the conclusion?

A microbiologist compares algal dry biomass after seven days in a new growth medium and a standard medium. Each observation comes from a different, independently grown flask. There is **no matching** between the two groups. The measurements, in milligrams, are

Medium	Measurements (mg)
New	17, 19, 21, 23, 25
Standard	12, 14, 16, 17, 18, 20, 22

For this exercise, assume independent normal populations with a common, unknown variance σ^2. Define

$$\Delta = \mu_{\text{new}} - \mu_{\text{standard}}.$$

(a) Build the standard error. Calculate both sample means and both sample variances using divisor $n_i - 1$. Then calculate the pooled variance s^2 (course-slide notation) and estimated standard error of the mean difference. Give the degrees of freedom and connect them to the two sample means that were estimated. Explain why pairing the first five entries would not be justified.

Solution. Let group 1 be the new medium and group 2 the standard medium. The parameter Δ is the population mean difference in seven-day dry biomass, in mg. Independence, normal populations and a common unknown variance are supplied assumptions, not conclusions established by these small samples.

Here $n_1 = 5$ and $n_2 = 7$. Center each group at its own sample mean:

$$\bar{x}_1 = \frac{17 + 19 + 21 + 23 + 25}{5} = \frac{105}{5} = 21 \text{ mg},$$
$$\bar{x}_2 = \frac{12 + 14 + 16 + 17 + 18 + 20 + 22}{7} = \frac{119}{7} = 17 \text{ mg}.$$

Thus $\bar{x}_1 - \bar{x}_2 = 4$ mg. Group 1 deviations from 21 are $-4, -2, 0, 2, 4$; group 2 deviations from 17 are $-5, -3, -1, 0, 1, 3, 5$. Using divisor $n_i - 1$ gives

$$s_1^2 = \frac{16 + 4 + 0 + 4 + 16}{5 - 1} = \frac{40}{4} = 10 \text{ mg}^2,$$
$$s_2^2 = \frac{25 + 9 + 1 + 0 + 1 + 9 + 25}{7 - 1} = \frac{70}{6} = \frac{35}{3} \approx 11.6667 \text{ mg}^2.$$

Under the common-variance assumption, combine residual sums of squares and their degrees of freedom:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{4(10) + 6(35/3)}{10} = 11 \text{ mg}^2, \qquad s = \sqrt{11} \approx 3.3166 \text{ mg}.$$

This is a degrees-of-freedom-weighted average, not $(s_1^2 + s_2^2)/2$. The two sample sizes supply different amounts of variance information.

Independence makes the variances of the sample means add: $\text{Var}(\overline{X}_1 - \overline{X}_2) = \sigma^2/n_1 + \sigma^2/n_2$. Replacing σ^2 by s^2 yields

$$\boxed{\widehat{\text{SE}}(\overline{X}_1 - \overline{X}_2) = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{11\left(\frac{1}{5} + \frac{1}{7}\right)} = \sqrt{\frac{132}{35}} \approx 1.9420 \text{ mg}.}$$

Degrees of freedom: $\text{df} = (5 - 1) + (7 - 1) = 10$. There are 12 observations and two estimated sample means. Each group's centered residuals sum to zero, giving two separate constraints. This is not $12 - 1$.

Why not pair the first five entries? There is no common flask or matched biological unit linking them. Pairing by list position would invent matches, discard two standard-medium flasks and make the analysis arbitrary. The design determines whether pairing is justified.

Problem 2 A confidence interval using Student's t

(b) A t interval. Construct a 95% confidence interval for Δ. Interpret its units and give an accurate meaning of 95% confidence. Use the interval to decide whether the matching two-sided test of $H_0 : \Delta = 0$ rejects at level 0.05. Verify the decision with the test statistic and a p-value bound using the shared table on page 8.

Solution. Here $\Delta = \mu_{\text{new}} - \mu_{\text{standard}}$, the difference in population mean dry biomass. From part (a), $\bar{x}_1 - \bar{x}_2 = 4$ mg, $s^2 = 11$ mg^2, $n_1 = 5$, $n_2 = 7$, $\widehat{\text{SE}} = \sqrt{132/35} = 1.9420166$ mg, and df $= 10$.

For a two-sided 95% interval, $\alpha = 0.05$ is split between two tails, giving $\alpha/2 = 0.025$ per tail. Use row df $= 10$, column 0.025: $t_{10,0.025} = 2.228$.

$$\begin{aligned} \text{CI}_{95\%} &= (\bar{x}_1 - \bar{x}_2) \pm t_{10,0.025}\, s\sqrt{\frac{1}{5} + \frac{1}{7}} \\ &= 4 \pm 2.228(1.9420166) \text{ mg} \\ &= 4 \pm 4.3268 \text{ mg} \\ &= \boxed{(-0.327,\ 8.327) \text{ mg.}} \end{aligned}$$

Units and meaning. The interval spans a new-medium population mean about 0.327 mg lower to about 8.327 mg higher than the standard-medium population mean. It concerns a difference of *population means*, not a range of individual flask measurements.

In repeated independent samples from these populations, about 95% of intervals made by this procedure would contain the fixed true Δ. For this particular interval, Δ is either inside it or outside it. The 95% refers to the procedure's long-run coverage, not a 95% probability assigned to the fixed parameter after observing the data.

Matching two-sided test. Because zero lies in the interval, the test of $H_0 : \Delta = 0$ against $H_a : \Delta \neq 0$ **fails to reject at level** 0.05. Direct calculation agrees:

$$t_{\text{obs}} = \frac{(21 - 17) - 0}{1.9420166} = 2.0597, \qquad |2.0597| < t_{10,0.025} = 2.228.$$

In row df $= 10$, columns 0.05 and 0.025 bracket this statistic:

$$t_{10,0.05} = 1.812 < 2.0597 < 2.228 = t_{10,0.025}.$$

Thus the right-tail probability satisfies

$$0.025 < P(T_{10} \geq 2.0597) < 0.05.$$

The alternative includes either sign, so double the right-tail probability:

$$\boxed{0.05 < p_{\text{two-sided}} < 0.10.}$$

A software check gives $p \approx 0.06642$. The table bound already determines the decision.

Conclusion. The estimated mean difference is 4 mg, but this comparison is not statistically significant at the 5% level. The interval includes both zero and nonzero differences; it does not establish equal population means.

Problem 2 Evaluating the normal critical value shortcut

(c) A tempting shortcut. A colleague uses the same estimated standard error from (a), but replaces the t critical value from (b) by the standard-normal critical value 1.960. Compute the resulting interval. How does its width and conclusion compare with (b)? Explain why this is a naive plug-in z interval, rather than an exact 95% interval under the stated model. What uncertainty does the t procedure account for? How does the critical-value difference behave as the degrees of freedom increase?

Solution. Keep the estimated mean difference 4 mg and estimated standard error $\sqrt{132/35} = 1.9420166$ mg. Substituting the normal critical value gives

$$4 \pm 1.960(1.9420166) = 4 \pm 3.8064 \text{ mg}$$

$$= \boxed{(0.194,\ 7.806) \text{ mg.}}$$

The shortcut's total width is $2(3.8063526) = 7.6127$ mg. The correct t interval has total width $2(4.3268130) = 8.6536$ mg. The shortcut is about 12.0% narrower:

$$1 - \frac{7.6127}{8.6536} \approx 0.120.$$

It excludes zero and would suggest rejecting the matching two-sided null at 5%, whereas the correct t interval $(-0.327, 8.327)$ mg includes zero and leads to failure to reject. This changed conclusion comes from using the wrong reference distribution for the stated unknown-variance model.

Why the shortcut is not exact. The pooled standard deviation S is estimated from data and varies from sample to sample. Under independent normal populations with a common unknown variance, the actual pivot is

$$\frac{(\overline{X}_1 - \overline{X}_2) - \Delta}{S\sqrt{1/n_1 + 1/n_2}} \sim T_{10},$$

not $N(0,1)$. The heavier t tails account for the extra uncertainty in the estimated denominator. Replacing $t_{10,0.025} = 2.228$ by 1.960 ignores that uncertainty and produces coverage below 95% here. The t interval is exact under the stated model, apart from rounding the table entry.

As degrees of freedom increase, for this fixed tail probability $t_{\mathrm{df},0.025}$ decreases toward 1.960. The critical-value gap and the relative interval-width gap therefore shrink. Large degrees of freedom make the plug-in normal approximation closer; they do not turn an estimated standard deviation into a known one.

Problem 2 Known variance and the closing check

(d) A different information setting. Now suppose, in a separate version of the problem, both population standard deviations really are **known to be 3 mg**. Retain the normality and independence assumptions. Write the correct known-σ z interval and its standard error. Is this procedure valid despite the small sample sizes? Explain why it differs from the shortcut in (c). Compute its endpoints if time permits.

Solution. In this separate setting, $\sigma_1 = \sigma_2 = 3$ mg are known population quantities. With $n_1 = 5$ and $n_2 = 7$, the correct standard error is

$$\begin{aligned}\mathrm{SE}(\overline{X}_1 - \overline{X}_2) &= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{3^2}{5} + \frac{3^2}{7}} \\ &= 3\sqrt{\frac{1}{5} + \frac{1}{7}} = \sqrt{\frac{108}{35}} \approx 1.7566 \text{ mg}.\end{aligned}$$

Independent normal sample means have a normal difference, and its standard deviation is now known. Therefore

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - \Delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1).$$

The 95% interval uses this known standard error, not s_1, s_2 or the pooled s:

$$\begin{aligned}(\bar{x}_1 - \bar{x}_2) \pm 1.960\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} &= 4 \pm 1.960(1.7566201) \text{ mg} \\ &= 4 \pm 3.4430 \text{ mg} \\ &= \boxed{(0.557,\ 7.443) \text{ mg}.}\end{aligned}$$

This is valid despite the small sample sizes: normality and genuinely known population variances give an exact normal sampling distribution for any n_1, n_2. A large-sample central limit approximation is unnecessary.

Part (c) retained an *estimated* standard error of 1.9420 mg while the variances were unknown. Part (d) provides new population information and uses the *known* standard error 1.7566 mg. Estimating a standard deviation as 3 would not justify treating it as exactly known.

Closing check. A statistically inconclusive comparison is not evidence that two population means are exactly equal. Explain this distinction using your answer to (b).

Solution. Part (b)'s interval, $(-0.327, 8.327)$ mg, includes zero and many nonzero differences, including substantial positive ones. The data do not distinguish zero from all these possibilities at the specified significance level. They do not establish $\Delta = 0$.

A suitable conclusion is: *The estimated mean difference is 4 mg, but the 95% confidence interval includes zero, so the two-sided comparison is not statistically significant at level 0.05.* Avoid concluding that the media have equal population means.

Establishing practical equivalence would require a specified margin and an equivalence analysis, not merely a non-significant test.

Student t reference table

Entries are $t_{\mathrm{df},\alpha}$ with $P(T_{\mathrm{df}} > t_{\mathrm{df},\alpha}) = \alpha$. The first column gives df; the remaining column headings give the probability α to the **right** of the listed value. Values are rounded to three decimal places.

df	0.10	0.05	0.025	0.01	0.005	0.001
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
12	1.356	1.782	2.179	2.681	3.055	3.930
14	1.345	1.761	2.145	2.624	2.977	3.787
20	1.325	1.725	2.086	2.528	2.845	3.552

For each problem, determine the appropriate degrees of freedom and tail probability from the design and the requested inference. Record the row and column you use.

Table entries used in these solutions

- Problem 1 uses df = 7: column 0.05 for the original test, columns 0.005 and 0.001 for the one-sided p-value bound, and column 0.01 for the stricter significance level in part (d).
- Problem 2 uses df = 10: column 0.025 for the two-sided 95% interval, and columns 0.05 and 0.025 for the right-tail probability bound. Double that bound for the two-sided p-value.

Sources and rounding. Questions and table: Monday in-class practice, November 9, 2026. All problem statements, data and requested parts are retained; table page references are updated for this solutions document. Notation follows Chapter 9: df denotes degrees of freedom and s^2 denotes the pooled variance. Calculations retain unrounded standard errors; displayed intervals use the supplied three-decimal t table and 1.960. Software p-values are supplementary checks; the requested table bounds are sufficient.