Homework 9 — Inference for Two Samples

STAT 515 · Ott & Longnecker §6.2, §6.4 · Lecture 16

(Due: Wednesday, Nov 18, 2026, before class)

Please submit a hard copy of your homework before class and email your R code, if applicable, to Alex Goebel at AJGOEBEL@email.sc.edu.

This assignment covers **Lecture 16: Inference for Two Samples** — the two-sample z, the pooled t, Welch's t, and paired differences. Show your work throughout.

Useful values: $z^* = 1.96$ (95%), 2.576 (99%). For df = 14: $t^*_{0.025} = 2.145$ and $t^*_{0.05} = 1.761$. For df = 7: $t^*_{0.025} = 2.365$ and $t^*_{0.05} = 1.895$.

Problem 1: Large-Sample Inference for $\mu_1 - \mu_2$ (15 points)

A transit planner compares morning commute times (minutes) on two routes into downtown. Independent random samples of weekday trips give

Route	n	Mean	Std. dev.
A	45	34.2	7.5
B	52	30.8	6.4

(a) Define μ_1 and μ_2 in words, state which route you are calling population 1, and write H_0 and H_a for a two-sided test.

(b) Verify that the conditions for the large-sample (z) procedure are met. State each condition and say why no assumption about the shape of either population is needed.

(c) Compute the standard error $\sqrt{s_1^2/n_1 + s_2^2/n_2}$ and the test statistic.

(d) Report the p-value and your decision at $\alpha = .05$. Write the conclusion in one sentence, in context.

(e) Construct a 95% confidence interval for $\mu_1 - \mu_2$. State its sign pattern and say what that pattern tells you.

(f) Now construct a 99% confidence interval. It covers 0, even though the 95% interval did not. Explain why this is not a contradiction, and state what your decision would have been at $\alpha = .01$.

Problem 2: Small Samples and the Pooled t (12 points)

Two fertilizer formulations are compared in a greenhouse trial. Eight plants are randomly assigned to each formulation and the dry weight of the yield (grams) is recorded.

Formulation A: 21, 24, 19, 26, 22, 20, 25, 23

Formulation B: 27, 25, 30, 24, 28, 26, 31, 29

(a) Compute $\bar{x}_A$, s_A, $\bar{x}_B$, and s_B.

(b) Compute the pooled sample variance s_p^2, and state its degrees of freedom. Explain in one sentence what quantity s_p^2 is estimating and why the two samples may be combined to estimate it.

(c) Test at the .05 level whether the two formulations differ in mean yield. Report the test statistic, the rejection region, and your decision.

(d) Construct a 95% confidence interval for $\mu_A - \mu_B$ and interpret it in context.

(e) List the three conditions required for this procedure to be valid, and comment on how plausible each is here.

Problem 3: A Paired Design (10 points)

Eight volunteers perform a reaction-time task (milliseconds) before and after a short training session. Each volunteer is measured twice.

Subject	1	2	3	4	5	6	7	8
Before	212	285	243	198	267	231	255	220
After	205	271	236	195	254	222	247	214

(a) Compute the eight differences $d_i = \text{Before} - \text{After}$, then $\bar{d}$ and s_d.

(b) Test at the .05 level whether training reduces mean reaction time. State H_0 and H_a in terms of μ_d, give the test statistic and its degrees of freedom, and state your decision.

(c) Construct a 95% confidence interval for μ_d and interpret it.

(d) Someone ignores the pairing and runs a two-sample pooled t test on the same 16 numbers, obtaining $s_p = 27.72$, $t = 0.604$ on 14 degrees of freedom, and a one-sided p-value of .278. Verify the standard error they must have used, and state the conclusion their analysis reaches.

(e) The two analyses disagree completely. Explain why, referring to the size of the between-subject variation (the "Before" values run from 198 to 285) compared with the size of the differences.

Problem 4: Choosing and Justifying a Procedure (10 points)

(a) Starting from $\mathrm{Var}(aX) = a^2\,\mathrm{Var}(X)$ and the fact that variances of independent random variables add, show that

$$\mathrm{Var}(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Identify the step at which the minus sign disappears.

(b) Two independent samples have $SE(\bar{x}_1) = 6$ and $SE(\bar{x}_2) = 8$. Compute $SE(\bar{x}_1 - \bar{x}_2)$. Explain in one sentence why the answer is neither 2 nor 14.

(c) A classmate reports two separate 95% confidence intervals, one for each group mean, notes that they overlap, and concludes that the means do not differ significantly. State what is wrong with this reasoning and what should be reported instead.

(d) For each study below, state whether the samples are independent or paired, and name the procedure you would use:

(i) 30 patients have cholesterol measured before and after eight weeks on a diet.

(ii) 40 students are randomly assigned to an online section and 38 to an in-person section; final exam scores are compared.

(iii) Twelve pairs of twins are split, one twin per teaching method, and test scores are compared.

(e) Two independent samples have $n_1 = 9$ with $s_1 = 2.4$ and $n_2 = 31$ with $s_2 = 11.0$, drawn from normal populations. State whether you would use the pooled t or Welch's t, and justify your choice in one or two sentences.

Problem 5: Two-Sample Procedures in R (8 points)

(a) Explain what each line computes, and state which of the three procedures from this assignment the final line carries out:

```r
A <- c(21, 24, 19, 26, 22, 20, 25, 23)
B <- c(27, 25, 30, 24, 28, 26, 31, 29)
t.test(A, B, var.equal = TRUE)
```

(b) Write the R code that reproduces the paired analysis in Problem 3(b)–(c), and state what changes in the output if `paired = TRUE` is omitted.