ST 551 Homework 5

Student's Name Goes Here

Fall 2022

Instructions

This assignment is due by 11:59 PM, November 14th on Canvas via Gradescope. **You should submit your assignment as a typed PDF which you can compile using the provide .Rmd (R Markdown) template.** Include your code in your solutions and indicate where the solutions for individual problems are located when uploading into Gradescope. You should also use complete, grammatically correct sentences for your solutions.

Question 1 (13 Points) Two-sample t-test comparions

Part A. (5 points) Show (algebraically) that when the sample sizes in the two groups are equal, the equal-variance t-statistic and the unequal-variance t-statistic are equal.

Part B. (8 Points) Recall from class that the two-sample t-test assuming equal variance (i.e. with a pooled variance estimate) rejects too infrequently when the smaller sample comes from the population with the smaller variance. Simulate 10,000 datasets consisting of two samples under each of the following scenarios to confirm. Fill out the table below with the rejection rates for the equal variance and unequal variance t-tests (Let $\alpha = 0.05$), and discuss your results:

Sample 1 Dist	Sample 2 Dist	Equal Variance t-test	Unequal Variance t-test
Normal(0,1), $m = 10$	Normal(0,4), $n = 30$		
Exp(1), $m = 30$	Normal(1,4), $n = 10$		
$\chi^2(5)$, $m = 20$	Normal(5,1), $n = 50$		
t(5), $m = 30$	t(3), $n = 10$		

Question 2 (13 Points) Paired t-test

Part A. (5 Points) If the true population covariance between X and Y is *negative*, use analytical/algebraic arguments to show that the regular, unequal variance, two-sample t-test will reject too often (more than the nominal/target level α). Hint: Remember that when computing the variance of two random variables that are not independent that a covariance term pops up.

Part B. (8 Points) Simulate data sets with two independent samples X and Y in the following scenarios to show that the paired t-test is properly calibrated (i.e. rejects at the correct level α). Also show that the two-sample t-test has better power. How much better power? Fill out the table below with the rejection rates for the paired t-test and unequal variance t-test (Let $\alpha = 0.05$), and discuss your results (Notice that in some of the scenarios, the null hypothesis is true while in others it's false):

Sample 1 Dist	Sample 2 Dist	Paired t-test	Unequal Variance t-test
Normal(0,1), $m = 10$	Normal(0,4), $n = 10$		
$\chi^2(5)$, $m = 30$	Normal(5,4), $n = 30$		
Normal(0,1), $m = 10$	Normal(2,4), $n = 10$		
$\chi^2(5)$, $m = 30$	Normal(3,4), $n = 30$		