

Problem 1 (6 points, 1 point for each part)

Indicate if a statement is **true** if it is always or **false** otherwise. You do not need to explain why it is true or false.

- (1) One of the assumptions underlying the use of the pooled estimate of population variance in the two sample t -test is that the samples are drawn from populations having equal variances.
- (2) In the two sample t -test, the number of degrees of freedom for the test statistic increases as sample sizes increase.
- (3) If every observation is multiplied by 2, then the t statistic is multiplied by 2.
- (4) The standard normal distribution can be used for inferences concerning proportions of success from binomial populations.
- (5) The pooled variance estimate is used when comparing means of two populations using independent samples.
- (6) It is not necessary to have equal sample sizes for the paired t -test.

Problem 2 (22 points)

The manager of a large office building needs to buy a large shipment of light bulbs. After reviewing specifications and prices from a number of suppliers, the choice is narrowed to two brands whose specifications with respect to price and quality appear identical. The manager intends to buy the bulbs with a longer mean life. To test the life of bulbs of each brand, he purchases some bulbs of each brand and subjects them to an accelerated life test, recording hours to burnout, the following are the data obtained from the manager's experiment:

Brand A: sample size = 38; sample mean = 1532; sample variance = 128008

Brand B: sample size = 40; sample mean = 1390; sample variance = 92564

Based on the above data, the manager consults you about which brand of bulbs he should buy. As a student taking a statistical course, you try to answer his question with the following calculations:

- (1) **(13 points)** You test if the mean life of bulbs from two brands differs with a significance level of 0.05 and the equal variance assumption. Clearly specify 5 steps with the p -value approach. You need to provide sufficient details about your calculations.

- (2) **(5 points)** Construct a 95% confidence interval on the mean difference in the life of bulbs for two brands. Here you still assume the equal variance of life of bulbs from two brands. Based on this confidence interval, do you reject the null hypothesis in (1)?
- (3) **(4 points)** The complete data can be found in data file `bulb.csv`. Use R function `t.test()` to find the corresponding test statistic, the p -value, and the 95% confidence interval and report them in your solution file. You can use `bulb$life[bulb$brand == "a"]` and `bulb$life[bulb$brand == "b"]` to get the life of brand A and brand B bulbs. Once you create a data named “bulb” with R function `read.csv()`.

Problem 3 (22 points)

A maintenance manager must test if a new repair method can increase the expected time between repairs. For each machine used in the study, she recorded the last time between failures prior to using the new method, which she called “Current”, and the first time between failures after using the new method, which she called “New”. These are the times (in hours):

Machine	1	2	3	4	5	6	7	8	9	10
Current	155	222	346	287	115	389	183	451	140	252
New	211	345	419	274	244	420	319	505	396	222
Difference	56	123	73	-13	129	31	136	54	256	-30

Sample means: Current = 254.0; New = 333.5; Difference = 81.5

Sample variances: Current = 12752.667; New = 9711.89; Difference = 6976.722

- (1) **(11 points)** Conduct the most appropriate hypothesis test using a 0.05 significance level and the p -value approach. You must clearly specify 5 steps for your test and provide sufficient details about your calculations.
- (2) **(5 points)** Construct a 95% two-sided confidence interval for the difference of time between repairs before and after using the new repair method. You must provide sufficient details about your calculations.
- (3) **(3 points)** Use R function `t.test()` based on the data of “Current” and “New” to find the test statistic, the p -value, and confidence interval for the test from (1) and report them in your solution file. In other words, you need to perform a two sample t -test with R function `t.test()` and the option “paired = TRUE”.

- (4) **(3 points)** Use R function `t.test()` based on the data of “Difference” to find the test statistic, the p -value, and confidence interval for the test from (1) and report them in your solution file.