ST 551 Homework 5

Luis Garcia-Lamas

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.

First, let's start with Unequal variance (Welch's) two-sample t-test:

$$t_U(\delta_0) = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}}$$

And when the two groups sample size are the same, $m = n$:

$$t_U(\delta_0) = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{n}}} = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\frac{s_X^2 + s_Y^2}{n}}}$$

We know that Equal Variance two-sample t-test:

$$t_E(\delta_0) = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{s_p^2(\frac{1}{m} + \frac{1}{n})}}$$

And when the two groups sample size are the same, $m = n$:

$$t_E(\delta_0) = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{s_p^2(\frac{1}{n} + \frac{1}{n})}}$$

$$t_E(\delta_0) = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{s_p^2(\frac{2}{n})}}$$

Now, what does $s_p^2 =$? As seen in lecture:

$$s_p^2 = \frac{(m-1)s_X^2 + (n-1)s_Y^2}{m + n - 2}$$

And when the two groups sample size are the same, $m = n$:

$$s_p^2 = \frac{(n-1)s_X^2 + (n-1)s_Y^2}{n + n - 2} = \frac{(n-1)(s_X^2 + s_Y^2)}{2(n-1)} = \frac{(s_X^2 + s_Y^2)}{2}$$

And plugging it back in:

$$t_E(\delta_0) = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\frac{(s_X^2 + s_Y^2)}{2}(\frac{2}{n})}} = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\frac{s_X^2 + s_Y^2}{n}}}$$

Hence, it was shown that when the 2 groups have equal sample size, the t statistic will be the same for both equal variance and unequal varince two-sample t-test.

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), $n = 30$	.0085	.0481
Exp(1), $m = 30$	Normal(1,4), $n = 10$	.1696	.0571
$\chi^2(5)$, $m = 20$	Normal(5,1), $n = 50$	.1863	.0608
t(5), $m = 30$	t(3), $n = 10$	.066	.0383

```r
N <- 10000

p_values_equal <- rep(0,N)
p_values_unequal <- rep(0,N)
count1 <- 0
count2 <- 0

for (i in 1:N) {


  sampleA <- rt(30,5)
  sampleB <- rt(10,3)

  p_values_equal[i] <- t.test(sampleA, sampleB, var.equal = TRUE)$p.value

  if(p_values_equal[i] <= .05) {count1 <- count1 +1}
  p_values_unequal[i] <- t.test(sampleA, sampleB, var.equal = FALSE)$p.value
  if(p_values_unequal[i] <= .05) {count2 <- count2 +1}

}

count1/N
```

```
## [1] 0.0713
```

```r
count2/N
```

```
## [1] 0.0444
```

We typically reject a null hypothesis when the p-value $< .05$. Therefore, I sampled from sample 1 distribution, and sample 2 distribution with their respective parameters and did a t test to compare the samples (10,000 times). I did an equal and unequal variance t test and counted the number of p values. This is how I compared rejection rates between tests.

Comparing the first two distributions, N(0,1) and N(0,4), here we know that N(0,1) would have a smaller variance. It also has a smaller sample size here of 10. We can see in the results, that we reject more infrequently in the equal variance t-test at least compared to the unequal variance t-test.

Comparing the second pair of distributions, Exp(1) and N(1,4). In this case, the exponential distribution can have a calculated variance of 1. The normal distribution clearly has a variance of 4. However, in this case the normal distribution has a larger variance and a smaller sample size for the t-test. As observed, in this case we actually reject more frequently for equal variance than unequal variance.

Looking at the third pair, $\chi^2(5)$ and N(5,1). In this case, the chi-squared distribution has a calculated variance of 10 while the normal distribution has a variance of 1. The normal distribution has more samples than the chi-squared. We can observe that for the equal variance t-test we reject far more often than the unequal test. Similar to the last result!

And finally, looking at t distribution with df = 5 and df = 3 for their respective sample size of 30 and 10. For the first sample, we have a calculated variance of 5/3 and 3 for sample 2. t(5) has a larger sample size in this case and smaller variance in this case. For both unequal and equal t-test the values seem close to .05 but the equal test seems to be ever so slightly farther from .05 on average. Overall, very similar but you can see the sample size and variance trade off with the 2 tests.

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.

We know that the t-score formula is:

$$t_U(\delta_0) = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}}$$

And the denomanitor contains sample variance and we use it to estimate the true population variance. When we try to look at the equation for $Var(\bar{X} - \bar{Y})$:

$$Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(\bar{Y}) - 2Cov(\bar{X}, \bar{Y}) = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{n} - 2\frac{\sigma_{XY}}{n}$$

We can assume m and n to be the same size for convenience of covariance. Since the covariance between X and Y is negative:

$$Var(\bar{X} - \bar{Y}) = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{n} + 2\frac{\sigma_{XY}}{n}$$

where $\sigma_{XY} > 0$

This means that because the t-score is not taking into account the negative covariance, it will be an over estimate of the true statistic. Therefore, a t-score without the negative covariance $t_{U|regular}$ will more often than not be larger than $t_{U|negative}$

$t_{U|regular}(\delta_0) > t_\alpha$ is more likely than $t_{U|negative}(\delta_0) > t_\alpha$

hence, the test would reject more often than need be.

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), $n = 10$	.0523	.0523
$\chi^2(5)$, $m = 30$	Normal(5,4), $n = 30$	.532	.0546
Normal(0,1), $m = 10$	Normal(2,4), $n = 10$	.7106	.7392
$\chi^2(5)$, $m = 30$	Normal(3,4), $n = 30$	.8446	.8579

```
N <- 10000
p_values_pair <- rep(0,N)
p_values_samp <- rep(0,N)
count1 <- 0
count2 <- 0
for (i in 1:N) {
```

```
  sampleA <- rchisq(30, 5)
  sampleB <- rnorm(30,3,2)

  p_values_pair[i] <- t.test(sampleA, sampleB, paired = TRUE)$p.value
  if(p_values_pair[i] <= .05) {count1 <- count1+1}
  p_values_samp[i] <- t.test(sampleA, sampleB, var.equal = FALSE)$p.value
  if(p_values_samp[i] <= .05) {count2 <- count2+1}


}

count1/N
```

```
## [1] 0.8352
```

```
count2/N
```

```
## [1] 0.8461
```

We can see that both test are properly calibrated for the first two pairs of distributions we look at. This makes sense because they have the same means, therefore they should both be following a similar rejection rate of type 1 error. Then we can see when the means are actually different for the last two pairs of distirbutions, the power we get is high because they are clearly not the same mean. Also, the power for an unequal variance t-test has a slightly higher value but not a huge amount.