1. The department chair wanted to compare the final examination percentage (**Final Score**: measured in %) of introductory statistics students between two new instructors (Dr. Jekyll and Mr. Hyde). A random sample of 32 students was drawn in Dr. Jekyll's class and the average final score was 71.1% and a standard deviation of 7.6%. A random sample of 24 students was drawn in Mr. Hyde's class and the average final score was 67.2% and a standard deviation of 9.0%. It is reasonable to assume that the Final Score follows a Normal distribution, ~~and that the equal variance assumption is valid~~.

   a) Provide a description of the parameter of interest. **[2 marks]**

   Define $\mu_1 - \mu_2$ as the difference in average final score from Dr. Jekyll's and Dr. Hyde's class respectively.

   b) Set up the null hypothesis and the alternative hypothesis. **[2 marks]**

   $H_0: \mu_1 - \mu_2 = 0$ **vs.** $H_a: \mu_1 - \mu_2 \neq 0$

   c) Can we safely assume the "equal variance" assumption if valid? Briefly justify your answer. **[0+2 marks]**

   $s_1^2 = 7.6^2 = 57.8$
   $S_2^2 = 9^2 = 81$
   -> The equal variance assumption is good as 81 is not more than twice as big as 57.8

   d) Calculate the test statistic and find the p-value. **[4 marks]**

   $$s_p^2 = \frac{(n_1-1)\times s_1^2+(n_2-1)\times s_2^2}{n_1+n_2-2} = \frac{(32-1)\times 57.8+(24-1)\times 81}{32+24-2} = 67.68$$
   $$TS1 = \frac{\bar{X}_1-\bar{X}_2}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{71.1-67.2}{\sqrt{(67.68)\left(\frac{1}{32}+\frac{1}{24}\right)}} = \frac{3.9}{2.221} = 1.756$$

   $DF = n_1 + n_2 - 2 = 32 + 24 - 2 = 54$
   From the t-table, we cannot find DF of 54. We usually down to the next biggest degree of freedom, i.e. DF = 60.
   ~~**Note:**~~ $\cancel{t_{60,0.20} = 0.8477}$
   ~~**-> right-hand area (or the area to the right of TS) is bigger than 0.20.**~~
   ~~**-> p-value > twice of 0.20 or 0.40 (because of a two-tailed test)**~~
   Note: $t_{60,0.05} = 1.6706$ **and** $t_{60,0.025} = 2.0003$
   -> right-hand area (or the area to the right of the TS) is between 0.025 and 0.05
   -> p-value will be twice of that, i.e. p-value between 0.05 and 0.10

   e) Draw an appropriate conclusion with 1% significance level. **[2 marks]**

   Since the p-value is bigger than the 1% significance level, we have enough statistical evidence to reject the null hypothesis and conclude that the difference in average final score from Dr. Jekyll's and Dr. Hyde's class is significantly different from zero. In other words, the average final scores between the two teachers are significantly different.

2. A subscription-based e-commerce company wanted to test if the new version of the product browsing page (Version 2) will generate higher sales (in dollars) compared to the current version (Version 1). A random sample of 50 customers were randomly assigned to the two versions during a 7-day period. The results are summarized in the data “SalesComparison.xlsx”. Use R function “t.test()” to answer the following questions.

a) Suppose we want to see if the current version (Version 1) generates sales less than $50. Use t.test() to generate the results and make an appropriate conclusion. Use $\alpha = 0.05$. **[2+4 marks]**

R code	R output

`t.test(Data$V1,`
`alternative="less",`
`mu = 50)`
	`One Sample t-test`
	
	`data:  Data$V1`
	`t = -0.85029, df = 24, p-value = 0.2018`
	`alternative hypothesis: true mean is less than 50`
	`95 percent confidence interval:`
	`-Inf 51.56675`
	`sample estimates:`
	`mean of x`
	`48.452`

Define μ as the average sales (in $) of customers/transactions using the current version.
$H_0: \mu \geq 50$ vs. $H_a: \mu < 50$
From the output, p-value = 0.2018
Since p-value is bigger than the 5% significance level, we do not have enough evidence to reject the null hypothesis and conclude that the average sales of customers/transactions using the current version is not significantly less than $50.

b) Suppose we want to see if the new version (Version 2) generates more sales than the current version (Version 1). Can we safely make the assumption that the equal variance assumption is valid? Use t.test() to generate the results and make an appropriate conclusion. Use $\alpha = 0.05$. **[2+2+4 marks]**

R code	R output

`t.test(Data$V1,`
`Data$V2,`
`alternative="less",`
`mu = 0,`
`var.equal = TRUE)`
	`Two Sample t-test`
	
	`data:  Data$V1 and Data$V2`
	`t = -2.6519, df = 48, p-value = 0.005407`
	`alternative hypothesis: true difference in means is less than 0`
	`95 percent confidence interval:`
	`-Inf -2.503732`
	`sample estimates:`
	`mean of x mean of y`
	`48.452    55.264`

Define $\mu_1 - \mu_2$ as the difference average sales (in $) of customers/transactions between using the current version and the new version.
$H_0: \mu_1 - \mu_2 \geq 0$ vs. $H_a: \mu_1 - \mu_2 < 0$
From the output, p-value = 0.0054
Since p-value is smaller than the 5% significance level, we have enough evidence to reject the null hypothesis and conclude that average sales (in $) of customers/transactions between using the current version and the new version are significantly different from each other.