PRACTICAL 8: Practical of Hypothesis Testing

THEORY:

There are three types of *t*-test:

- **One-sample *t*-test**
  Used to compare a sample mean with a known population mean or some other meaningful, fixed value
- **Independent samples *t*-test**
  Used to compare two means from independent groups
- **Paired samples *t*-test**
  1. Used to compare two means that are repeated measures for the same participants - scores might be repeated across different measures or across time.
  2. Used also to compare paired samples, as in a two treatment randomized block design.

==

1. **one sampled t-test**
   Note:
   - You can perform a one-sample t-test with the t.test function. To compare a sample mean with a constant value mu0, use the command:
     t.test(bottles$Volume, mu=500)
   - By default, R performs a two-tailed test. To perform a one-tailed test, set the alternative argument to "greater" or "less", as shown below.
     t.test(bottles$Volume, alternative = "less", mu=500)

   - Q: A bottle filling machine is set to fill bottles with soft drink to a volume of 500 ml. The actual volume is known to follow a normal distribution. The manufacturer believes the machine is under-filling bottles. A sample of 20 bottles is taken and the volume of liquid inside is measured.
   - Suppose you want to use a one-sample t-test to determine whether the bottles are being consistently under filled, or whether the low mean volume for the sample is purely the result of random variation. A one-sided test is suitable because the manufacturer is specifically interested in knowing whether the volume is *less* than 500 ml. The test has the null hypothesis that the mean filling volume is equal to 500 ml, and the alternative hypothesis that the mean filling volume is less than 500 ml. A significance level of 0.01 is to be used.

```
> bottles <- read.csv(file.choose(), sep=",", header=T)
> summary(bottles)
     Volume
 Min.   :449.1
 1st Qu.:473.3
 Median :491.5
 Mean   :491.6
 3rd Qu.:505.1
 Max.   :538.1
```

```
> t.test(bottles$Volume, alternative = "less", mu=500)

        One Sample t-test

data:  bottles$Volume
t = -1.5205, df = 19, p-value = 0.07243
alternative hypothesis: true mean is less than 500
95 percent confidence interval:
     -Inf 501.1569
sample estimates:
mean of x
 491.5705
```

- **Analysis:** From the output, we can see that the mean bottle volume for the sample is 491.6 ml. The one-sided 95% confidence interval tells us that mean filling volume is likely to be less than 501.2 ml. The p-value of 0.07243 tells us that if the mean filling volume of the machine were 500 ml, the probability of selecting a sample with a mean volume less than or equal to this one would be approximately 7%. Since the p-value is not less than the significance level of 0.01, we cannot reject the null hypothesis that the mean filling volume is equal to 500 ml. This means that there is no evidence that the bottles are being under-filled.

==

2. **paired t-test**

```
> data1 <- read.csv(file.choose(), sep=",", header=T)
> summary(data1)
       BS            Fasting            pp
 Min.   : 56.60   Min.   : 80.00   Min.   :120.0
 1st Qu.: 74.67   1st Qu.: 89.25   1st Qu.:127.0
 Median : 96.62   Median : 93.00   Median :137.5
 Mean   :101.09   Mean   : 92.70   Mean   :136.6
 3rd Qu.:100.49   3rd Qu.: 96.75   3rd Qu.:143.8
 Max.   :194.00   Max.   :100.00   Max.   :160.0
> t.test(data1$pp, data1$Fasting, alternative = "greater", paired = T)

        Paired t-test

data:  data1$pp and data1$Fasting
t = 11.88, df = 9, p-value = 4.193e-07
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 37.1263      Inf
sample estimates:
mean of the differences
                   43.9
```

t test for variance

```
> var<-read.csv(file.choose(),sep="," , header = T)
> summary(var)
       BS             Fasting             pp       
 Min.   : 56.60   Min.   : 80.00   Min.   :120.0  
 1st Qu.: 74.67   1st Qu.: 89.25   1st Qu.:127.0  
 Median : 96.62   Median : 93.00   Median :137.5  
 Mean   :101.09   Mean   : 92.70   Mean   :136.6  
 3rd Qu.:100.49   3rd Qu.: 96.75   3rd Qu.:143.8  
 Max.   :194.00   Max.   :100.00   Max.   :160.0  
> var.test(var$Fasting,var$pp,alternative = "two.sided")

	F test to compare two variances

data:  var$Fasting and var$pp
F = 0.217, num df = 9, denom df = 9, p-value = 0.0327
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.05390025 0.87364915
sample estimates:
ratio of variances 
         0.2170021 
```

As the p-value is less than the level of significance (0.05) we reject the null hypothesis. And we see that the pp sugar values are higher than fasting sugar values.

===

3. **t test for correlation**

We take a dataset having aptitude test score (out of 25) and job proficiency (on a scale of 10) for employees. We use t test to find out the correlation between these two. That is whether aptitude test score is correlated to job proficiency.

If the p-value is < 5%, then the correlation between them is significant.

```
> corr<-read.csv(file.choose(),sep="," , header = T)
> summary(corr)
     emp_id      aptitude_test_score job_proficiency apti_test_score_2
 Min.   : 1.00   Min.   :12.34       Min.   :5.6     Min.   :13.55    
 1st Qu.: 5.75   1st Qu.:18.55       1st Qu.:6.8     1st Qu.:19.36    
 Median :10.50   Median :20.11       Median :7.9     Median :21.61    
 Mean   :10.50   Mean   :19.88       Mean   :7.9     Mean   :20.96    
 3rd Qu.:15.25   3rd Qu.:22.54       3rd Qu.:8.9     3rd Qu.:24.36    
 Max.   :20.00   Max.   :24.33       Max.   :9.8     Max.   :24.99    
> cor.test(corr$aptitude_test_score, corr$job_proficiency,alternative =
+            "two.sided",method= "pearson")

	Pearson's product-moment correlation

data:  corr$aptitude_test_score and corr$job_proficiency
t = 3.9612, df = 18, p-value = 0.0009156
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3437382 0.8640342
sample estimates:
      cor 
0.6824505 
```

here p value is very less than 0.05 hence correlation between job proficiency and aptitude score is significant.