Lab 10

t-test for single mean and t-test for difference of means

The t.test( ) function produces a variety of t-tests. Unlike most statistical packages, the default assumes unequal variance.

```
# independent 2-group t-test
> t.test(y~x)                # where y is numeric and x is a binary factor
# independent 2-group t-test
> t.test(y1,y2)              # where y1 and y2 are numeric
# paired t-test
> t.test(y1,y2,paired=TRUE)  # where y1 & y2 are numeric
# one sample t-test
> t.test(y,mu=3)             # Ho: mu=3
```

We can use the var.equal = TRUE option to specify equal and a pooled variance estimate, use the alternative="less" or alternative="greater" option to specify a one tailed test.

```
>  t.test(len ~ supp, data = ToothGrowth, alt = "greater", var.equal = TRUE)
>  x <- rnorm(13, mean = 2, sd = 3)
> t.test(x, mu = 0, conf.level = 0.9, alternative = "greater")
```

One Sample t-test:-

Comparing the sample mean with a known value, when population variance is not known.

Problem 1 :-

An outbreak of salmonella-related illness was attributed to ice produced at a certain factory. Scientists measured the level of Salmonella in 9 randomly sampled batches ice crean.The levels(in MPN/g) were:

0.593	0.142	0.329	0.691	0.231	0.793	0.519	0.392	0.418

Is there evidence that the mean level pf Salmonella in ice cream greater than 0.3 MPN/g?

Sol:

```
> x=c(0.593,0.142,0.329,0.691,0.231,0.793,0.519,0.392,0.418)
> t.test(x,alternative="greater",mu=0.3)

        One Sample t-test

data:  x
t = 2.2051, df = 8, p-value = 0.02927
alternative hypothesis: true mean is greater than 0.3
95 percent confidence interval:
 0.3245133       Inf
sample estimates:
mean of x
0.4564444
```

.

Inference:-

From the output we see that the p-value = 0.029. Hence, there is moderately strong evidence that the mean Salmonella level in the ice cream is above 0.3 MPN/g.

Problem 2: Suppose that 10 volunteers have taken an intelligence test; here are the results obtained. The average score of the entire population is 75 in the same test. Is there any significant difference (with a significance level of 95%) between the sample and population means, assuming that the variance of the population is not known.

Scores: 65, 78, 88, 55, 48, 95, 66, 57, 79, 81

```
> a = c(65, 78, 88, 55, 48, 95, 66, 57, 79, 81)
> t.test (a, mu=75)

        One Sample t-test

data:  a
t = -0.78303, df = 9, p-value = 0.4537
alternative hypothesis: true mean is not equal to 75
95 percent confidence interval:
 60.22187 82.17813
sample estimates:
mean of x
     71.2
```

```
> qt(0.975, 9)
[1] 2.262157
```

Inference:-

- ➔ *The t-computed value is smaller than t-tabulated, we accept the null hypothesis of equality of the averages.*
- ➔ *Alternatively we could consider the p-value with a significance level of 95%. If **p-value is greater than 0.05** then **we accept the null hypothesis H0,** otherwise **we reject the null .***

t.test usage:

Test a claim about $\mu_1 - \mu_2$ Below, mu is the value of $\mu_1 - \mu_2$ in the null hypothesis.

- Two-Tailed Test: t.test(x, y, mu = ,)
- Right-Tailed Test: t.test(x, y, mu= , alternative="greater")
- Left-Tailed Test: t.test(x, y, mu= ,alternative="less")

Problem 3: Comparing two independent sample means, taken from two populations with unknown variance.The following data shows the heights of individuals of two different countries with unknown population variances. Is there any significant difference b/n the average heights of two groups.

A:	175	168	168	190	156	181	182	175	174	179
B:	185	169	173	173	188	186	175	174	179	180

Solution:-

```
> a = c(175, 168, 168, 190, 156, 181, 182, 175, 174, 179)
> b = c(120, 180, 125, 188, 130, 190, 110, 185, 112, 188)
> t.test(a,b, var.equal=FALSE, paired=FALSE)

        Welch Two Sample t-test

data:  a and b
t = 1.8827, df = 10.224, p-value = 0.08848
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.95955 47.95955
sample estimates:
mean of x mean of y
    174.8     152.8

> qt(0.975, 10.224)
[1] 2.221539
>
```

Inference :-

- ➔ The p-value > 0.05, we conclude that the means of the two groups are significantly similar
- ➔ The value of t is less than the tabulated t-value for 10,224 df, we accept H0.

Problem 4: Suppose the recovery time for patients taking a new drug is measured (in days). A placebo group is also used to avoid the placebo effect. The data are as follows

with drug	: 15 10 13 7 9 8 21 9 14 8
placebo	: 15 14 12 8 14 7 16 10 15 2

Is there any significant difference between the average effect of these two drugs?

Solution:-

```
> x = c(15, 10, 13, 7, 9, 8, 21, 9, 14, 8)
> y = c(15, 14, 12, 8, 14, 7, 16, 10, 15, 12)
> t.test(x,y,alt="less",var.equal=TRUE)

        Two Sample t-test

data:  x and y
t = -0.53311, df = 18, p-value = 0.3002
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 2.027436
sample estimates:
mean of x mean of y
     11.4      12.3
```

Inference :-

P value(0.3002) > 0.05 then there is no evidence to reject our Null hypothesis.

Problem 5: Six subjects were given a drug (trearment group) and an additional 6 subjects a placebo(control group).Their reaction time to stimulus was measured(in ms).We want to perform a two sample t-test for comparing the means of the treatment and control groups.

Control	**91**	**87**	**99**	**77**	**88**	**91**
Treatment	**101**	**110**	**103**	**93**	**99**	**104**

Rcode:-

```
> control=c(91,87,99,77,88,91)
> Treat=c(101,110,103,93,99,104)
> t.test(control,Treat,alternative="less",var.equal=TRUE)

        Two Sample t-test

data:  control and Treat
t = -3.4456, df = 10, p-value = 0.003136
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -6.082744
sample estimates:
mean of x mean of y
 88.83333 101.66667
```

```
> control=c(91,87,99,77,88,91)
> Treat=c(101,110,103,93,99,104)
> t.test(control,Treat,alternative="less",var.equal=FALSE)

        Welch Two Sample t-test

data:  control and Treat
t = -3.4456, df = 9.4797, p-value = 0.003391
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -6.044949
sample estimates:
mean of x mean of y
 88.83333 101.66667
```

Inference:-

From both the output we see that the p-value = 0.003136(equal) and 0.003391(Unequal). Therefore, it infers that there is different between treatment and control group.

Practice Problems :-

1. A certain stimulus administered to each of the 13 patients resulted in the following increase of blood pressure: 5, 2, 8,-1, 3, 0, -2, 1, 5, 0, 4, 6, 8. Can it be concluded that the stimulus, in general, be accompanied by an increase in the blood pressure?

2. The manufacturer of a certain make of electric bulbs claims that his bulbs have a mean life of 25 months with a standard deviation of 5 months. Random samples of 6 such bulbs have the following values: Life of bulbs in months: 24, 20, 30, 20, 20, and 18. Can you regard the producer's claim to valid at 1% level of significance?

3. the life time of electric bulbs for a random sample of 10 from a large consignment gave the following data: 4.2, 4.6, 3.9, 4.1, 5.2, 3.8, 3.9, 4.3, 4.4, 5.6 (in '000 hours). Can we accept the hypothesis that the average life time of bulbs is 4, 000 hours

4. Data on weight (grams) of two treatments of NMU (nistroso- methyl urea) are recorded. Find out whether these two treatments have identical effects by using t test for sample means at 5% level of significance.

Sample	1	2	3	4	5	6	7	8	9	10	11	12
Treatments 0.2 %	2.0	2.7	2.9	1.9	2.1	2.6	2.7	2.9	3.0	2.6	2.6	2.7
0.4%	3.2	3.6	3.7	3.5	2.9	2.6	2.5	2.7				

5. Hypothesis Tests for Two Means :Independent Data:Here we test for a difference in means for the following data

No Drug(x1)	237	289	257	228	303	275	262	304	244	233
Drug(x2)	194	240	230	186	265	222	242	281	240	212