Small Sample Test

size less than thirty

Small Sample Test

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.

Syntax: one sample – single mean

- t.test(y, mu=3, alt = "greater"/ "lesser")

Problem

- 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?

R- Code & Interpretation

```
> 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
```

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 0.3MPN/g.

Problem

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

R- Code & Interpretation

```
> 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
```

*the p-value with a significance level of 95%. If **p-value is lesser than 0.05** hence w**e reject the null hypothesis***

T- test for two samples (independent)

- 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

Comparing two independent sample means, taken from two populations with unknown variances. The following data shows the heights of individuals of two different countries with unknown population variances. Is there any significant difference between 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

R- Code & Interpretation

```
> x=c(175,168,168,190,156,181,182,175,174,179)
> y=c(120,180,125,188,130,190,110,185,112,188)
> t.test(x,y)

        Welch Two Sample t-test

data:  x and y
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
```

The p-value > 0.05, we conclude that the means of the two groups are significantly similar

Problem

- 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?

R- Code & Interpretation

```
> 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")

        Welch Two Sample t-test

data:  x and y
t = -0.53311, df = 16.245, p-value = 0.3006
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 2.044664
sample estimates:
mean of x mean of y
     11.4      12.3
```

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

Paired t-test (Dependent Sample)

```
# paired t-test

> t.test(y1,y2,paired=TRUE)        # where y1 & y2 are numeric
```

Problem

- *A school athletics has taken a new instructor, and want to test the effectiveness of the new type of training proposed by the new instructor comparing the average times of 10 runners in the 100 meters. The results are given below(time in seconds)*

Before training	12.9	13.5	12.8	15.6	17.2	19.2	12.6	15.3	14.4	11.3
After training	12.7	13.6	12.0	15.2	16.8	20.0	12.0	15.9	16.0	11.1

- *Solu:*
- *In this case we have two sets of paired samples, since the measurements were made on the same athletes before and after the workout. To see if there was an improvement, deterioration, or if the means of times have remained substantially the same (hypothesis H0), we need to make a Student's t-test for paired samples, proceeding in this way*

R- Code & Inference

```
> before = c(12.9, 13.5, 12.8, 15.6, 17.2, 19.2, 12.6, 15.3, 14.4, 11.3)
> after = c(12.7, 13.6, 12.0, 15.2, 16.8, 20.0, 12.0, 15.9, 16.0, 11.1)
> t.test(before,after, paired=TRUE)

        Paired t-test

data:  before and after
t = -0.21331, df = 9, p-value = 0.8358
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.5802549  0.4802549
sample estimates:
mean of the differences
                  -0.05
```

Interpretation :-

The p-value is greater than 0.05, then we do not reject the hypothesis H_0 of equality of the averages and conclude that the new training has not made any significant improvement to the team of athletes.

Problem

Suppose now that the manager of the team (given the results obtained) fired the coach who has not made any improvement, and take another, more promising. We report the times of athletes after the second training:

Before training:	12.9	13.5	12.8	15.6	17.2	19.2	12.6	15.3	14.4	11.3
After the second training:	12.0	12.2	11.2	13.0	15.0	15.8	12.2	13.4	12.9	11.0

Solu:

Now we check if there was actually an improvement, ie perform a t-test for paired data, specifying in R to test the alternative hypothesis H1 of improvement in times. To do this simply add the syntax alt = "less" when you call the t-test

R- Code & Inference

```
> before=c(12.9, 13.5, 12.8, 15.6, 17.2, 19.2, 12.6, 15.3, 14.4, 11.3)
> after = c(12.0, 12.2, 11.2, 13.0, 15.0, 15.8, 12.2, 13.4, 12.9, 11.0)
> t.test(before,after, paired=TRUE, alt="less")

        Paired t-test

data:  before and after
t = 5.2671, df = 9, p-value = 0.9997
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 2.170325
sample estimates:
mean of the differences
                   1.61
```

In response, we obtained a p-value well above 0.05, which leads us to conclude that we can reject the null hypothesis H_0 in favour of the alternative hypothesis H1: the new training has made substantial improvements to the team

Problem

- *Consider the paired data below that represents cholesterol levels on 10 men before and after a certain medication Test the claim that, on average, the drug lowers cholesterol in all men. i.e., test the claim that* $\mu_d > 0$*. Test this at the 0.05 significance level*

Before(x)	237	289	257	228	303	275	262	304	244	233
After(y)	194	240	230	186	265	222	242	281	240	212

R- Code and Interpretation

```
> before=c(237,289,257,228,303,275,262,304,244,233)
> after=c(194,240,230,186,265,222,242,281,240,212)
> t.test(before,after,paired=TRUE,alternative="greater",mu=0)

        Paired t-test

data:  before and after
t = 6.5594, df = 9, p-value = 5.202e-05
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 23.05711      Inf
sample estimates:
mean of the differences
                     32
```

We can reject the null hypothesis and support the claim because the P-value ($5.2\ X10^{5}$) is less than the significance level

F- Test (Variance Ration Test)

- Syntax:

```
var.test(x, y)
```

Problem

- *Five Measurements of the output of two units have given the following results (in kilograms of material per one hour of operation) .Assume that both samples have been obtained from normal populations, test at 10% significance level if two populations have the same variance*

Unit A	14.1	10.1	14.7	13.7	14.0
Unit B	14.0	14.5	13.7	12.7	14.1

$H_0: S_1^2 = S_2^2$

$H_1: S_1^2 \neq S_2^2$

R- Code and Inference

```
> Unit_A=c(14.1,10.1,14.7,13.7,14.0)
> Unit_B=c(14.0,14.5,13.7,12.7,14.1)
> var.test(Unit_A,Unit_B)

        F test to compare two variances

data:  Unit_A and Unit_B
F = 7.3304, num df = 4, denom df = 4, p-value = 0.07954
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.7632268 70.4053799
sample estimates:
ratio of variances
          7.330435
```

Here p valuve >0.05 ,then there is no evidence to reject the null hypothesis

Practice Problems

- 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.
- 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

Practice Problems cont…

- 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
- 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				