T- Test:

```
T-Test will tell the significance difference between the groups

When to use T-Test:
        * Data are Independent
        * Datas are noramlly distributed (check distribution by applying skewness)

Types of T-Test:
        * Paired T-Test (If the groups come from a single population)
        * Two-Sample T-Test (If the groups come from two different)
        * One-Sample T-Test (If there is one group being compared against a standard value)

T-Test:
     * One-tailed T-Test (AIML < Cyber)
     * Two-tailed T-Test (AIML && Cyber)
```

T-Test Table:

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Problem 1:

```
Suppose a botanical wants to know if the mean height of certain species of plant is equal to 15 inches.
Sam collects a random sample of 12 plants and records each of its heights in inches.
```

Answer:

$n = 12 \qquad \bar{x} = 172/12$

$\bar{x} = 14.\overline{333}$

* it is one sample T-test
* level of significance = 0.05

plant height (x_i)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
14	-0.33	0.1089
14	-0.33	0.1089
16	1.67	2.7889
13	-1.33	1.7689
12	-2.33	5.4289
17	2.67	7.1289
15	0.67	0.4489
14	-0.33	0.1089
15	0.67	0.4489
13	-1.33	1.7689
15	0.67	0.4489
14	-0.33	0.1089
	0.04	20.6668

formula of one-sample t-test:

$$T\text{-}test = \frac{m - \mu}{s/\sqrt{n}}$$

$m \to$ mean
$\mu \to$ hypothesis mean
$s \to$ standard deviation
$n \to$ no. of samples.

Standard Deviation $= \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$

$= \sqrt{\frac{20.6668}{12-1}} = \sqrt{1.8788}$

sd $\Rightarrow 1.3706$

T- test $= \frac{14.333 - 15}{\left(\frac{1.3706}{3.464}\right)}$

$\Rightarrow \frac{-0.667}{0.3957} \Rightarrow -1.684$

Degree of Freedom $= n-1$

$= 12-1$

$\Rightarrow 11.$

Now from T-table we need check value of degree of freedom 11 and 0.05

T-Table value $= 1.796$

$-1.684 < 1.798$

T-test value < T-table critical value

$\therefore$ Accept the null hypothesis

Problem 2:

Data given for boys and girls height. Find if there is any significance difference in the boys and girls heright with alpha 0.025

ANSWER:

* Two-sample T-test
* Two-tailed T-test
* level of significance = 0.025

x_1	x_2	$(\bar{x}_1 - x_1)^2$	$(\bar{x}_2 - x_2)^2$
14	8	12.25	0
8	6	6.25	4
7	4	12.25	16
13	14	6.25	36
42	32	37	56

$\bar{x}_1$ = Mean = $42/4$ = 10.5

$\bar{x}_2$ = Mean = $32/4$ = 8

formula for two-way t-test:

$$\Rightarrow \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{n}}}$$

$S \rightarrow$ standard deviation.

$$(S_1)^2 = \frac{\sum |\bar{x}_1 - x_1|^2}{n-1}$$

$= 37/3$

$= 12.33$

$$(S_2)^2 = \frac{\sum |\bar{x}_2 - x_2|^2}{n-1}$$

$= 56/3$

$= 18.66$

formula:

$$\Rightarrow \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}}$$

$$= \frac{|10.5 - 8|}{\sqrt{\frac{12.33}{4} + \frac{18.66}{4}}} = \frac{2.5}{\sqrt{7.74}}$$

$$\Rightarrow 0.89$$

T-test value = 0.89

Degree of freedom $= n_1 + n_2 - 2$

$= 4 + 4 - 2$

$= 6$

T-table value = 2.947

$0.89 < 2.947$

T-test value < T-table critical value

∴ Accept the null hypothesis

Question 3:

A Researcher wants to test a new anti-hunger weight loss pill. They have 10 people, rate their hunger before and after taking the pill. Does the pill do anything?

Answer:

* Two sample special T-test
* level of significance = 0.05

before	after	diff(x)	$[\text{diff}(x)]^2$
9	7	2	4
10	6	4	16
7	5	2	4
5	4	1	1
7	4	3	9
5	6	-1	1
9	7	2	4
6	5	1	1
8	5	3	9
7	7	0	0
		17	49

formula for

$$\Rightarrow \frac{\lambda}{SD/\sqrt{n}}$$

$\lambda \rightarrow$ mean

$\lambda = 17/10$

$\lambda = 1.7.$

$$= \frac{1.7}{1.494/\sqrt{10}} = \frac{1.7}{0.473}$$

$\Rightarrow$ ~~3~~99

$\Rightarrow 3.59$

Standard Deviation:

$$SD = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}}$$

$$= \sqrt{\frac{49 - \frac{(17)^2}{10}}{10-1}}$$

$$= \sqrt{\frac{49 - 289/10}{9}}$$

$$= \sqrt{2.333}$$

$$SD \Rightarrow 1.494$$

Degree of freedom $= n-1$

$= 10-1$

$\Rightarrow 9$

T-table critical value = 2.262

$3.59 > 2.262$

T-test value > T-test critical value

$\therefore$ Reject the null hypothesis

```
Syntax:

    Type:
    Tail:
    alpha(level of significance) :
    solve

    df = Critical value
    Calculated Value < Critical Value
        The Accept the Null Hypothesis
```

Probability Distribution:

In [5]:

```
x = c(14,14,16,13,12,17,15,14,15,13,15,14)
```

In [6]:

```
length(x)
```

12

In [9]:

```
# t.test = is used for the which sample test
# mu = hypothesis mean value
# conf.level = level of significance

t.test(x, mu=15, conf.level=0.05)
```

```
        One Sample t-test

data:  x
t = -1.6848, df = 11, p-value = 0.1201
alternative hypothesis: true mean is not equal to 15
5 percent confidence interval:
 14.30795 14.35872
sample estimates:
mean of x
 14.33333
```

In [10]:

```
# alternative =  to mention the number of tails
```

```
* In R code it uses the parameter alternative , it will accept 3 value

    * "lesser" ot "greater" is the one tailed T-Test
    * "two.sided" is the two tailed T-Test
```

In [11]:

```
t.test(x, mu=15, conf.level=0.05, alternative = "two.sided")
```

```
        One Sample t-test

data:  x
t = -1.6848, df = 11, p-value = 0.1201
alternative hypothesis: true mean is not equal to 15
5 percent confidence interval:
 14.30795 14.35872
sample estimates:
mean of x
 14.33333
```

In [12]:

```
t.test(x, mu=15, conf.level=0.05, alternative= "greater")
```

```
        One Sample t-test

data:  x
t = -1.6848, df = 11, p-value = 0.9399
alternative hypothesis: true mean is greater than 15
5 percent confidence interval:
 15.04394      Inf
sample estimates:
mean of x 
 14.33333 
```

In [13]:

```
t.test(x, mu=15, conf.level=0.05, alternative = "less")
```

```
        One Sample t-test

data:  x
t = -1.6848, df = 11, p-value = 0.06007
alternative hypothesis: true mean is less than 15
5 percent confidence interval:
     -Inf 13.62273
sample estimates:
mean of x 
 14.33333 
```

In [ ]:

```

```

```
           For Two Sample Cases:

                 *  If its a two sample special use paird = TRUE
                 *  If its a two sample use paird = FALSE
```

In [14]:

```
x1 = c(14,8,7,13)
x2 = c(8,6,4,14)
```

In [23]:

```
t.test(x1, x2, paired= FALSE)
```

```
        Welch Two Sample t-test

data:  x1 and x2
t = 0.89803, df = 5.7596, p-value = 0.4051
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.381432  9.381432
sample estimates:
mean of x mean of y 
     10.5       8.0 
```

In [25]:

```
t.test(x1, x2, paired= FALSE,  conf.level=0.01)
```

```
        Welch Two Sample t-test

data:  x1 and x2
t = 0.89803, df = 5.7596, p-value = 0.4051
alternative hypothesis: true difference in means is not equal to 0
1 percent confidence interval:
 2.463568 2.536432
sample estimates:
mean of x mean of y 
     10.5       8.0 
```

In [21]:

```
b = c(9,10,7,5,7,5,9,6,8,7)
a = c(7,6,5,4,4,6,7,5,5,7)
x = c(2,4,2,1,3,-1,2,1,3,0)
```

In [26]:

```
t.test(b,a,paired=TRUE, conf.level=0.05)
```

```
        Paired t-test

data:  b and a
t = 3.5973, df = 9, p-value = 0.005773
alternative hypothesis: true difference in means is not equal to 0
5 percent confidence interval:
 1.669529 1.730471
sample estimates:
mean of the differences
                    1.7
```

Binomial Distribution:

```
* A Binomial Distribution is a discrete probability distribution
* It describe the outcome of n independent trials in an experiment
```

$$P_x = \binom{n}{x} p^x q^{n-x}$$

```
x = 0,1,2,3,......
n = independent trial
p = probability of successfull trial
x = sucessful outcomes in an experiment of "n"
q = (1 - p)
```

In [27]:

```
# pbinom = is used for binomial distribution
# pbinom(successful outcome, size, probability)

pbinom(4, size = 12, prob = 0.2)  # probability = 1/5 = 0.2
```

0.92744450048

Poison Distribution:

```
Poison is a probability ditribution of independent of event occurance in an interval
```

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

```
x = 0,1,2,3,.......
λ = Mean Occurance per interval
e = Euler's Constant ~ 2.71828
```

In [28]:

```
# ppois = is used for poison distribution
# ppois( ,lambda)

ppois(16, lambda = s)
```

```
Error in ppois(16, lambda = s): object 's' not found
Traceback:

1. ppois(16, lambda = s)
```

Continuous Uniform Distribution:

Continuous Uniform Distribution is a probability distribution of the random numbers selection from the continuous interval between A to B

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

In [29]:

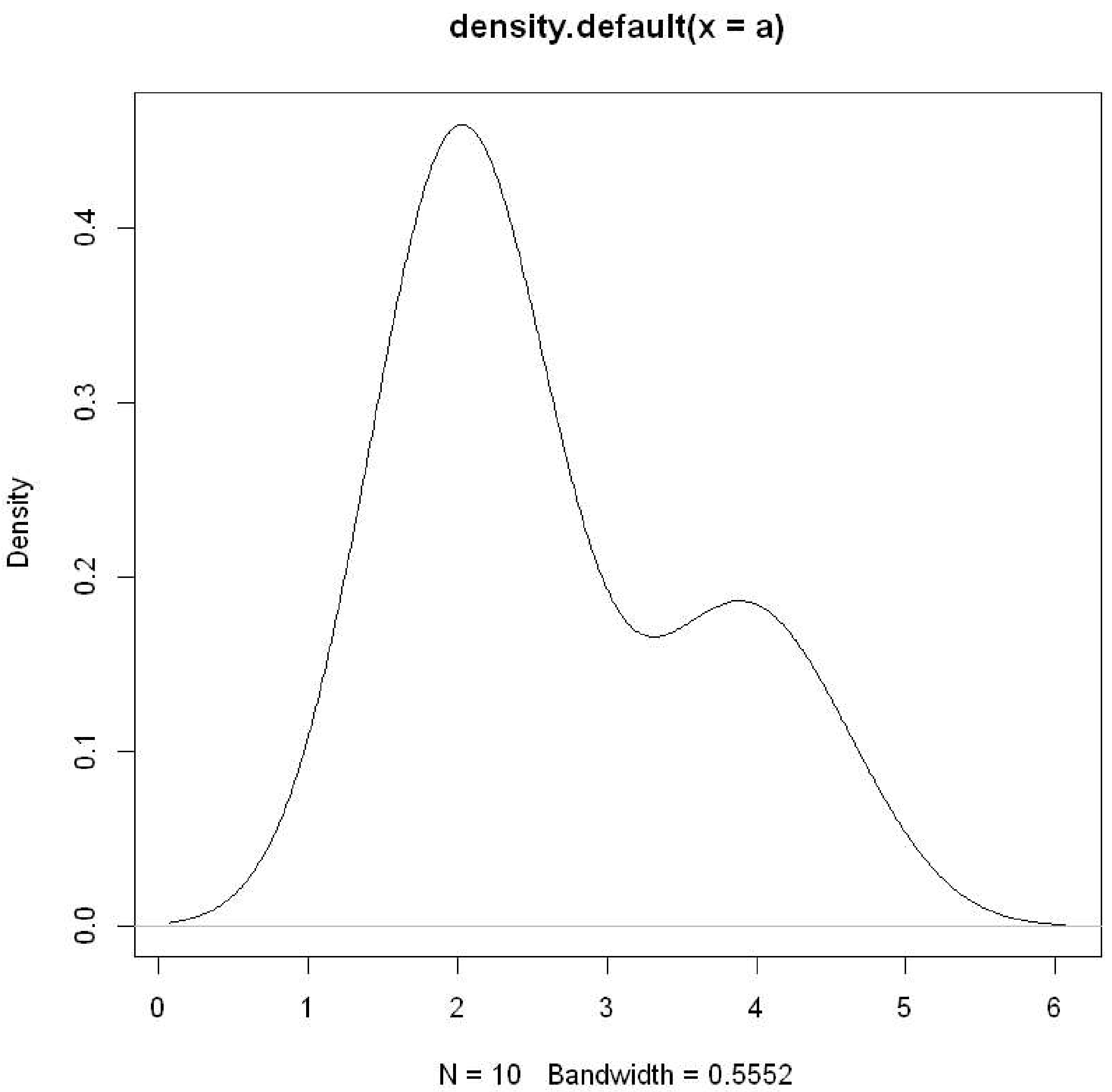
```
# runif = is used for continuous Uniform Distribution
# runif(number of numbers needed, minimum , maximum)

runif(10, min=1, max =5)
```

3.16520885378122 2.24853338487446 1.85274896863848 3.91063662618399 3.43775863293558 4.23783768713474 4.69304928742349 1.7086231932044 4.58909172657877 4.32584514189512

In [30]:

```
a = runif(10, min=1, max =5)
plot(density(a))
```



Exponential Distribution:

Exponential Distribution describe the arrival time of a randomly reccuring independent event sequence

1/(μ) e^(-x/u)

μ = mean waiting time for the next event
e = Euler's Constant ~ 2.71828
x = random variable

In [31]:

```
# pexp() = is used for exponential distribution
# pexp(value, rate)

pexp(2, rate = 1/3)
```

0.486582880967408

Normal Distribution:

Normal Distribution definied by the following probablity density function, where

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

```
* σ = Variance
* μ = Population Mean
* x = Value of the variable or data
*
```

In [33]:

```
# pnorm() = is used for normal distribution
# pnorm(value, mean , standard deviation)

pnorm(84, mean=72, sd=15.2)
```

0.785082397688728

In [ ]:

```

```