1

PROBLEM

The following data source the water PH on 16 water sample collected at the river Ganga at different location date

7.2,8.8,9.3,7.1,8.6,10.1,9.0,6.4,7.4,8.9,7.3,5.8,6.4,9.9,5.2,6.8

Test whether water PH data are normal at 5% level of significance given that $D_{16,\alpha}$ =0.327

Solution:-

Null Hypothesis (H_0):- PH data are normal at 5% level of significance.

Alternative Hypothesis (H_1):-PH data are not normal at 5% level of significance.

```
> #Entry the data
> data=c(7.2,8.8,9.3,7.1,8.6,10.1,9.0,6.4,7.4,8.9,7.3,5.8,6.4,9.9,5.2,6.8)
> data
 [1]  7.2 8.8 9.3 7.1 8.6
 [6] 10.1 9.0 6.4 7.4 8.9
[11]  7.3 5.8 6.4 9.9 5.2
[16]  6.8

> #Mean of data
> y=mean(data)
> y
[1] 7.7625
> #Variance of the data
> z=var(data)
> z
[1] 2.2105
> #Standard deviation of the data
> sig=sqrt(z)
> sig
[1] 1.486775

> #Ascending the data
> asc=sort(data)
> asc
 [1]  5.2 5.8 6.4 6.4 6.8
 [6]  7.1 7.2 7.3 7.4 8.6
[11]  8.8 8.9 9.0 9.3 9.9
[16] 10.1
> #ECDF of the data
> FoX=pnorm(data, y,sig)
> FoX
 [1] 0.35259063 0.75735483
 [3] 0.84945937 0.32794480
 [5] 0.71338460 0.94204731
 [7] 0.79739103 0.17972515
 [9] 0.40368654 0.77788768
[11] 0.37787143 0.09342234
[13] 0.17972515 0.92473692
[15] 0.04239645 0.25869485
> #Assume CDF of the data
> FnX=seq(1:16)/16
> FnX
 [1] 0.0625 0.1250 0.1875 0.2500
 [5] 0.3125 0.3750 0.4375 0.5000
 [9] 0.5625 0.6250 0.6875 0.7500
[13] 0.8125 0.8750 0.9375 1.0000
> Fn1X=c(seq(1:16)-1)/n
> Fn1X
 [1] 0.00 0.01 0.02 0.03 0.04
```

```
 [6] 0.05 0.06 0.07 0.08 0.09
[11] 0.10 0.11 0.12 0.13 0.14
[16] 0.15
> #Minus of the data
> DP=c(FnX-FoX)
> DP
 [1] -0.29009063 -0.63235483
 [3] -0.66195937 -0.07794480
 [5] -0.40088460 -0.56704731
 [7] -0.35989103 0.32027485
 [9]  0.15881346 -0.15288768
[11]  0.30962857 0.65657766
[13]  0.63277485 -0.04973692
[15]  0.89510355 0.74130515
> DM=FoX-Fn1X
> DM
 [1]  0.35259063 0.74735483
 [3]  0.82945937 0.29794480
 [5]  0.67338460 0.89204731
 [7]  0.73739103 0.10972515
 [9]  0.32368654 0.68788768
[11]  0.27787143 -0.01657766
[13]  0.05972515 0.79473692
[15] -0.09760355 0.10869485

> #Data.frame the data
> EXP=data.frame(data,FnX,Fn1X,FoX,Dp,Dm)
> EXP
    data    FnX Fn1X        FoX
1 7.2 0.0625 0.00 0.35259063
2 8.8 0.1250 0.01 0.75735483
3 9.3 0.1875 0.02 0.84945937
4 7.1 0.2500 0.03 0.32794480
5 8.6 0.3125 0.04 0.71338460
6 0.1 0.3750 0.05 0.94204731
7 9.0 0.4375 0.06 0.79739103
8 6.4 0.5000 0.07 0.17972515
9 7.4 0.5625 0.08 0.40368654
10 8.9 0.6250 0.09 0.77788768
11 7.3 0.6875 0.10 0.37787143
12 5.8 0.7500 0.11 0.09342234
13 6.4 0.8125 0.12 0.17972515
14 9.9 0.8750 0.13 0.92473692
15 5.2 0.9375 0.14 0.04239645
16 6.8 1.0000 0.15 0.25869485
              Dp         Dm
1 0.29009063 0.35259063
2 0.63235483 0.74735483
3 0.66195937 0.82945937
4 0.07794480 0.29794480
5 0.40088460 0.67338460
6 0.56704731 0.89204731
7 0.35989103 0.73739103
8 0.32027485 0.10972515
9 0.15881346 0.32368654
10 0.15288768 0.68788768
11 0.30962857 0.27787143
12 0.65657766 0.01657766
13 0.63277485 0.05972515
14 0.04973692 0.79473692
15 0.89510355 0.09760355
16 0.74130515 0.10869485
 #List the value of an array or matrix
```

```
> Mi=apply(EXP[, c(5,6)],1,max)
> Mi
 [1] 0.3525906 0.7473548
 [3] 0.8294594 0.2979448
 [5] 0.6733846 0.8920473
 [7] 0.7373910 0.3202748
 [9] 0.3236865 0.6878877
[11] 0.3096286 0.6565777
[13] 0.6327748 0.7947369
[15] 0.8951036 0.7413051
> #Add the one column
> TOT=cbind(EXP,Mi)
> TOT
   data FnX  Fn1X        FoX
1 7.2 0.0625 0.00    0.35259063
2 8.8 0.1250 0.01    0.75735483
3 9.3 0.1875 0.02    0.84945937
4 7.1 0.2500 0.03    0.32794480
5 8.6 0.3125 0.04    0.71338460
6 10.1 0.3750 0.05   0.94204731
7 9.0 0.4375 0.06    0.79739103
8 6.4 0.5000 0.07    0.17972515
9 7.4 0.5625 0.08    0.40368654
10 8.9 0.6250 0.09   0.77788768
11 7.3 0.6875 0.10   0.37787143
12 5.8 0.7500 0.11   0.09342234
13 6.4 0.8125 0.12   0.17972515
14 9.9 0.8750 0.13   0.92473692
15 5.2 0.9375 0.14   0.04239645
16 6.8 1.0000 0.15   0.25869485
       Dp         Dm
1 0.29009063 0.35259063
2  0.63235483 0.74735483
3  0.66195937 0.82945937
4  0.07794480 0.29794480
5  0.40088460 0.67338460
6  0.56704731 0.89204731
7  0.35989103 0.73739103
8  0.32027485 0.10972515
9  0.15881346 0.32368654
10 0.15288768 0.68788768
11 0.30962857 0.27787143
12 0.65657766 0.01657766
13 0.63277485 0.05972515
14 0.04973692 0.79473692
15 0.89510355 0.09760355
16 0.74130515 0.10869485
        Mi
1  0.3525906
2  0.7473548
3  0.8294594
4  0.2979448
5  0.6733846
6  0.8920473
7  0.7373910
8  0.3202748
9  0.3236865
10 0.6878877
11 0.3096286
12 0.6565777
13 0.6327748
14 0.7947396
15 0.8951036
16 0.7413051
> #Maximum
> Dn=max(DP)
```

```
> Dn
[1] 0.8951036
```

Result:-

Dn > $D_{16,\alpha}$ i.e.0.8951031 >0.327

We do not accept the null hypothesis at 5% level of significance.

Another method:-

```
> #shortcut key
> ks.test(data,"pnorm",y,sig)

        Asymptotic one-sample
        Alamogordo-Smirnoff test

data:  data
D = 0.15881, p-value =
0.8145
alternative hypothesis: two-sided

Warning message:
In ks.test.default(data, "pnorm", y, sig):
   ties should not be present for the Kolmogorov-Smirnov test
```