Sekhar_HW7

Sekhar Mekala

Sunday, November 20, 2016

Problem 9.14 (DES)

Reading and sorting the data into a variable, and dividing the cumulative sum of the data by 100 (distance between the cities is 100 Kilometers).

```
x <- c(88.3,40.7,36.3,27.3,36.8,
       91.7,67.3,7.0,45.2,23.3,
       98.8,90.1,17.2,23.7,97.4,
       32.4,87.8,69.8,62.6,99.7,
       20.6,73.1,21.6,6.0,45.3,
       76.6,73.2,27.3,87.6,87.2)

x <- sort(x)
x
```

```
##  [1]  6.0  7.0 17.2 20.6 21.6 23.3 23.7 27.3 27.3 32.4 36.3 36.8 40.7 45.2
## [15] 45.3 62.6 67.3 69.8 73.1 73.2 76.6 87.2 87.6 87.8 88.3 90.1 91.7 97.4
## [29] 98.8 99.7
```

Let us define the hypothesis:

H_0 : The above values is uniformally distributed

H_a : The above values is NOT uniformally distributed

We have to use a significance level of $\alpha = 0.05$ to test our hypothesis.

```
library(knitr)
y <- vector()
x <- x/100

N <- length(x)

i <- 1:N
A=x
B=i/N
C=((i/N) - x)
D=(x-(i-1)/N)

df <- data.frame(A,B,C,D)

df
```

```
##         A          B            C           D
## 1   0.060 0.03333333 -0.026666667  0.060000000
## 2   0.070 0.06666667 -0.003333333  0.036666667
## 3   0.172 0.10000000 -0.072000000  0.105333333
```

```
## 4  0.206 0.13333333 -0.072666667  0.106000000
## 5  0.216 0.16666667 -0.049333333  0.082666667
## 6  0.233 0.20000000 -0.033000000  0.066333333
## 7  0.237 0.23333333 -0.003666667  0.037000000
## 8  0.273 0.26666667 -0.006333333  0.039666667
## 9  0.273 0.30000000  0.027000000  0.006333333
## 10 0.324 0.33333333  0.009333333  0.024000000
## 11 0.363 0.36666667  0.003666667  0.029666667
## 12 0.368 0.40000000  0.032000000  0.001333333
## 13 0.407 0.43333333  0.026333333  0.007000000
## 14 0.452 0.46666667  0.014666667  0.018666667
## 15 0.453 0.50000000  0.047000000 -0.013666667
## 16 0.626 0.53333333 -0.092666667  0.126000000
## 17 0.673 0.56666667 -0.106333333  0.139666667
## 18 0.698 0.60000000 -0.098000000  0.131333333
## 19 0.731 0.63333333 -0.097666667  0.131000000
## 20 0.732 0.66666667 -0.065333333  0.098666667
## 21 0.766 0.70000000 -0.066000000  0.099333333
## 22 0.872 0.73333333 -0.138666667  0.172000000
## 23 0.876 0.76666667 -0.109333333  0.142666667
## 24 0.878 0.80000000 -0.078000000  0.111333333
## 25 0.883 0.83333333 -0.049666667  0.083000000
## 26 0.901 0.86666667 -0.034333333  0.067666667
## 27 0.917 0.90000000 -0.017000000  0.050333333
## 28 0.974 0.93333333 -0.040666667  0.074000000
## 29 0.988 0.96666667 -0.021333333  0.054666667
## 30 0.997 1.00000000  0.003000000  0.030333333
```

In the above table, "A" represents the distance of the accident divided by the distance between the cities (100 KM). The data is sorted in ascending order. The column "B" represents the value i/N, where i is the sequence of numbers from 1 to N (Number of observations), and N is the number of observations (which is 30). The column "C" represents $i/N - R_{(i)}$, where $R_{(i)}$ is the values in column "A", and the column D represents $R_{(i)} - (i-1)/N$.

The $D^+ = max(column"C") = 0.047$ The $D^- = max(column"D") = 0.172$

Hence $D = max(D^-, D^+) = 0.172$

At 30 degrees of freedon, and 0.05 significance level, the Kolmogorov-Smirnov critical value is $D_\alpha = 0.24$. Since D is less than D_α, we are unable to reject the null hypothesis, and the given data is uniformly distributed.

Problem 9.17 (DES)

Reading the data to a data frame:

```r
Employee <- 1:50

Time <- c(1.88,.54,1.9,.15,.02,2.81,1.5,.53,
          2.62,2.67,3.53,.53,1.8,.79,.21,.8,.26,.63,.36,2.03,
          1.42,1.28,.82,2.16,.05,.04,1.49,.66,2.03,1,.39,.34,.01,
          .1,1.1,.24,.26,.45,.17,4.29,.8,5.5,4.91,.35,.36,.9,1.03,1.73,.38,.48)
#hist(Time)
df <- data.frame(Employee,Time)
df
```

```
##    Employee Time
## 1         1 1.88
## 2         2 0.54
## 3         3 1.90
## 4         4 0.15
## 5         5 0.02
## 6         6 2.81
## 7         7 1.50
## 8         8 0.53
## 9         9 2.62
## 10       10 2.67
## 11       11 3.53
## 12       12 0.53
## 13       13 1.80
## 14       14 0.79
## 15       15 0.21
## 16       16 0.80
## 17       17 0.26
## 18       18 0.63
## 19       19 0.36
## 20       20 2.03
## 21       21 1.42
## 22       22 1.28
## 23       23 0.82
## 24       24 2.16
## 25       25 0.05
## 26       26 0.04
## 27       27 1.49
## 28       28 0.66
## 29       29 2.03
## 30       30 1.00
## 31       31 0.39
## 32       32 0.34
## 33       33 0.01
## 34       34 0.10
## 35       35 1.10
## 36       36 0.24
## 37       37 0.26
## 38       38 0.45
## 39       39 0.17
## 40       40 4.29
## 41       41 0.80
## 42       42 5.50
## 43       43 4.91
## 44       44 0.35
## 45       45 0.36
## 46       46 0.90
## 47       47 1.03
## 48       48 1.73
## 49       49 0.38
## 50       50 0.48
```

Let us define the hypothesis:

H_0 : The Time data is exponentially distributed

H_a : The above values is NOT exponentially distributed

Let $k = 6$, then each interval will have probability p as 1/6.

The end point of each interval is calculated as:

$$a_i = -(1/\lambda)ln(1 - ip)$$

where $i = 0, 1, 2, 3, 4, 5, 6$

$$\lambda = 1/mean(Time) = 0.8291874$$

Using the above formulae, we can get the following table:

```
lambda <- 1/mean(Time)
a <- (-1/lambda)*(log(1-(0:6)/6))

Class_Interval <- paste(a[1:5],"-",a[2:6],sep=" ")
Class_Interval <- c(Class_Interval,paste(a[6]," - Infinity",sep=""))

Observed <- vector()

for(i in 2:7){
Observed[i-1] <- sum(Time>=a[i-1] & Time < a[i])
}

df <- data.frame(Class_Interval,Observed_Freq=Observed,Expected_Freq=50/6)

df$Chi_sq <- ((df$Observed_Freq - df$Expected_Freq)^2/df$Expected_Freq)
df
```

```
##                       Class_Interval Observed_Freq Expected_Freq
## 1                 0 - 0.219879797493509             8      8.333333
## 2 0.219879797493509 - 0.488990920378446            11      8.333333
## 3 0.488990920378446 - 0.835935499755294             9      8.333333
## 4  0.835935499755294 - 1.32492642013374             5      8.333333
## 5   1.32492642013374 - 2.16086191988903            10      8.333333
## 6          2.16086191988903 - Infinity              7      8.333333
##       Chi_sq
## 1 0.01333333
## 2 0.85333333
## 3 0.05333333
## 4 1.33333333
## 5 0.33333333
## 6 0.21333333
```

```
sum(df$Chi_sq)
```

```
## [1] 2.8
```

The degrees of freedom = 6 - 1 -1 = 4. The $\chi^2_{0.05}$= 9.488. Since the calculated χ^2 (2.8) is less than $\chi^2_{0.05}$(9.488), we are unable to reject the null hypothesis. Hence the Time variable is exponentially distributed.

Problem 10.1(A) (DES)

Given the following data:

```
x <- c(18.9,22,19.4,22.1,19.8,21.9,20.2)
x
```

```
## [1] 18.9 22.0 19.4 22.1 19.8 21.9 20.2
```

The average value of the sample is $\bar{x} = 20.6142857$ and the Standard deviation of the sample $S = 1.3557637$.

Let us define the hypothesis:

$$H_0 : \mu = 22.5$$

$$H_a : \mu \neq 22.5$$

The desired significance level is $\alpha = 0.05$

We have to perform the two side hypothesis test, with the sample mean $\bar{x} = 20.6142857$ and the standard error $S = 0.5124305$.

The confidence interval at 0.05 level of significance is [21.4956546, 23.5043454].

Since the sample mean (20.6142857) does not lie in the confidence interval, we reject the null hypothesis.

Problem 10.1(B) (DES)

Let us compute the Type-II error.

$$\delta = |20.61429 - 22.5|/1.355764 = 1.390884$$

Using the operating characteristic curves given in Table A.10 (Appendix A) in DES book, we can identify that the probability of accepting null hypothesis, when it is false is approximately 0.1, hence the probability of rejecting the null hypothesis when it is indeed fasle is 0.9 (which is the power of the test). To make the power of test as atleast 0.8, the sample size should be greater than or equal to 6.

Problem 11.13 (DES)

See the attached simio file for the detailed experiment results.

- Job 1 average system time is 117.82 hours. The 95% Conf. Interval is [115, 123] hours
- Job 2 average system time is 97.5542 hours. The 95% Conf. Interval is [94.13, 102.5] hours
- Job 3 average system time is 85.1304 hours. The 95% Conf. Interval is [83, 87] hours
- Job 4 average system time is 47.1026 hours. The 95% Conf. Interval is [40, 50] hours
- Station 1 utilization is 99%. The 95% Conf. Interval is [97.8167, 100] %
- Station 2 utilization is 95.5%. The 95% Conf. Interval is [91, 99.5] %
- Station 3 utilization is 100%. The 95% Conf. Interval is [100, 100] %
- Station 4 utilization is 97.65%. The 95% Conf. Interval is [90.93, 100] %