STAT-S 520

RAMPRASAD BOMMAGANTY(RBOMMAGA)

6.4.1::

SOLUTION:

Given,

$$f(x) = \{\begin{matrix} 0 & if x < 1 \\ 2(x-1) & if 1 \leq x \leq 2 \\ 0 & x > 2 \end{matrix}$$

a. Since, the area under the pdf will be 1, the area under triangle with width x-1 and height 2(x-1) will be ½.

   ⇒ area = 0.5 * (q_2-1) * $2(q_2-1)$ = ½
   ⇒ q_2 = 1 + 1 / $\sqrt{2}$

b. To calculate the interquartile range, we need q_3-q_1.

   Calculating q_3,
   0.5 * (q_3-1)* $2(q_3-1)$ = ¾
   q_3 = 1 + $\sqrt{3}$ / 2

   Calculating q_1,
   0.5 * (q_1-1) * $2(q_1-1)$ = ¼
   q_1 = 3/2

   The interquartile range is as follows : iqr(X) = q_3-q_1 = (1 + $\sqrt{3}$ / 2) - 3/2 = ($\sqrt{3}$ -1) / 2 .

6.4.2::

SOLUTION:

a.

We can determine f(x) as follows:

$$f(x) = \{ \begin{array}{cc} 0 & x < 0 \\ cx & x \epsilon [0,1] \\ c & x \epsilon [1,2] \\ c(3-x) & x \epsilon [2,3] \\ 0 & x > 3 \end{array}$$

Since, the area under the pdf is always 1, the value of c can be computed using the area computation of f(x).

0.5 * c * 1 + c * 1 + 0.5 * c * 1 = 1
c = ½

b. Given that a continuous random variable has probability density function f.

$P(1.5 < X < 2.5) = \int_{1.5}^{2} c\, dx + \int_{2}^{2.5} 3c - 3cx\, dx$

$P(1.5 < X < 2.5) = c * (2 - 1.5) + 3c * (2.5 - 2) - (3c/2)[(2.5)^2-(2)^2]$

Substituting, c = ½,

P(1.5 <X < 2.5) = 0.4375

c. Expected value = mean = 1.5, since f(x) is a symmetric function.

d. If F is considered the CDF of f(x), then F(1) :

$F(1) = P(X \leq 1) = 0.5 * c * 1 = 0.5 * 0.5 * 1 = 0.25$

e. 0.90 quantile can be written as the last 0.10 quantile.

0.5 * (3-x) * ½ (3-x) = 0.10
$x = 3 - (2 / \sqrt{10})$

6.4.6 ::

SOLUTION::

Given that X ~ Uniform(5,15)

Mean = EX = 10
Variance = 100/12

a.
```
q1 = qunif(0.25,5,15)
q3 = qunif(0.75,5,15)
iqr = q3- q1
sd = sqrt(100/12)
res = iqr/sd
print(res)
```

Output : 1.732051

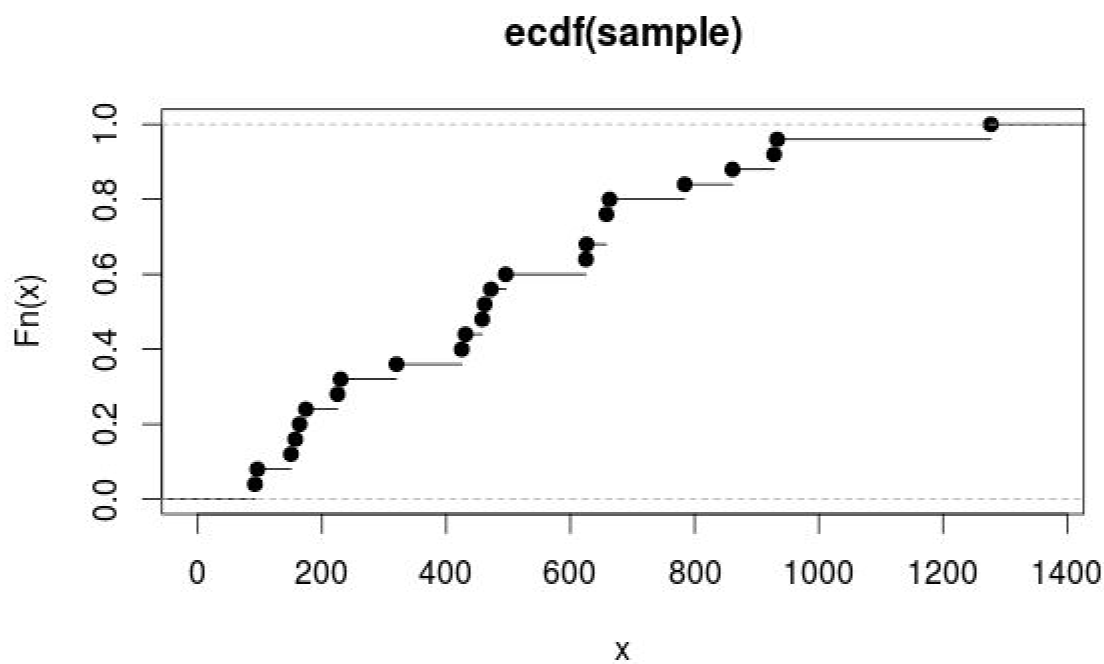
b.
```
q1 = qnorm(0.25, 10, sqrt(100/12))
q3 = qnorm(0.75,10,sqrt(100/12))
iqr = q3 - q1
sd =  sqrt(100/12)
res = iqr/sd
print(res)
```

Output : 1.34898

7.7.1::
SOLUTION:

a.
```
sample =
c(462,425,164,784,625,472,658,658,663,928,92,230,96,626,1277,225,150,320,496,
157,458,933,861,174,431)

plot(ecdf(sample))
```
The empirical cdf is as given below:



b. mean(sample) = 494.6
var(sample) = 94873.67

c. median = quantile(sample,0.5)
print(median)

Output : 462 (50%)

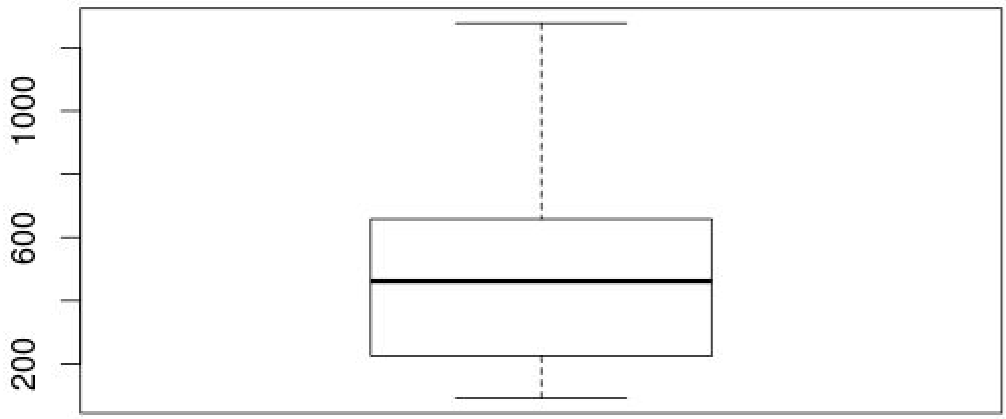
q1 = quantile(sample,0.25)
q3 = quantile(sample,0.75)
iqr = q3 - q1
print(iqr)
Output: 433(75%)

d. ratio = 433/ sqrt(94873.67)
print(ratio)

Output : 1.405773

e. boxplot(sample)

The boxplot is as under:



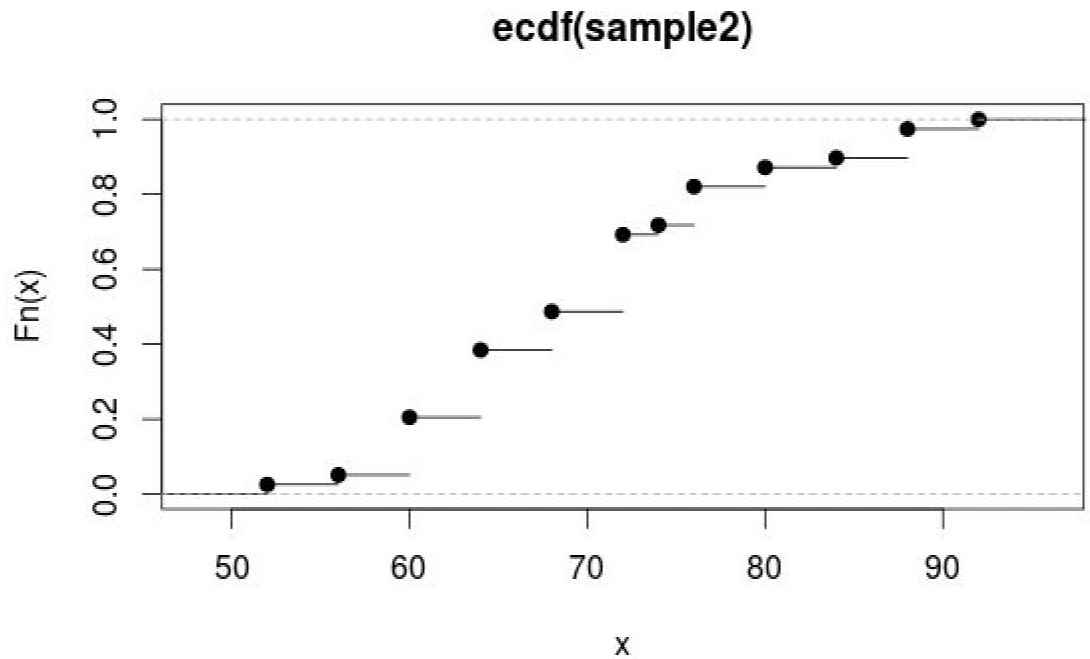
7.7.2::

SOLUTION::

a. sample2 = c(88,76,84,64,60,64,60,64,68,74,68,68,72,76,72,52,72,64,60,56,72,88,80,76,64,72,60,76,88,72,64,60,60,72,92,80,72,64,68)

plot(ecdf(sample2))

The plot is as given below:



b.

sample2 = c(88,76,84,64,60,64,60,64,68,74,68,68,72,76,72,52,72,64,60,56,72,88,80,76,64,72,60,76,88,72,64,60,60,72,92,80,72,64,68)

print(mean(sample2))

Output : 70.30769

sample2 = c(88,76,84,64,60,64,60,64,68,74,68,68,72,76,72,52,72,64,60,56,72,88,80,76,64,72,60,76,88,72,64,60,60,72,92,80,72,64,68)

print(var(sample2))

Output: 90.21862

c.

median = quantile(sample2,0.5)
print(median)

Output : 72 (50%)

```
q1 = quantile(sample2,0.25)
q3 = quantile(sample2,0.75)
iqr  = q3 - q1
print(iqr)
```

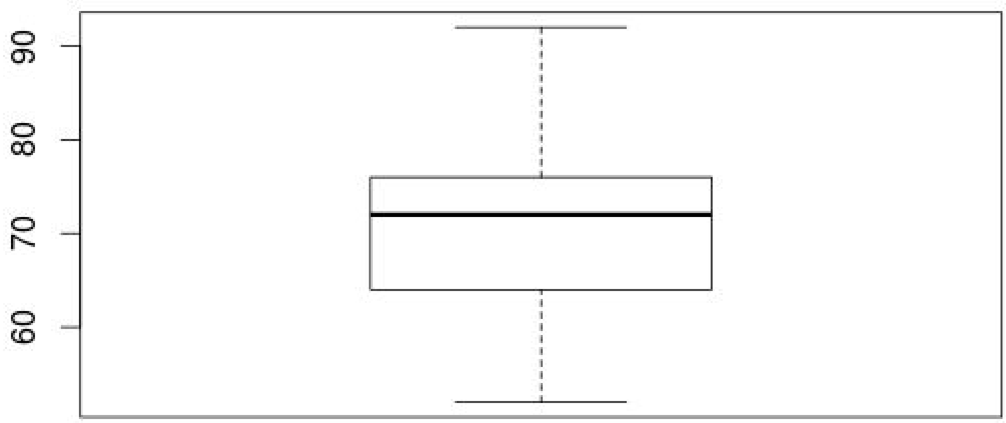
Output : 12 (75%)

d. `ratio = 12 /  sqrt(90.21862)`

Output : 1.263378

e. `boxplot(sample2)`

The box plot is as under:



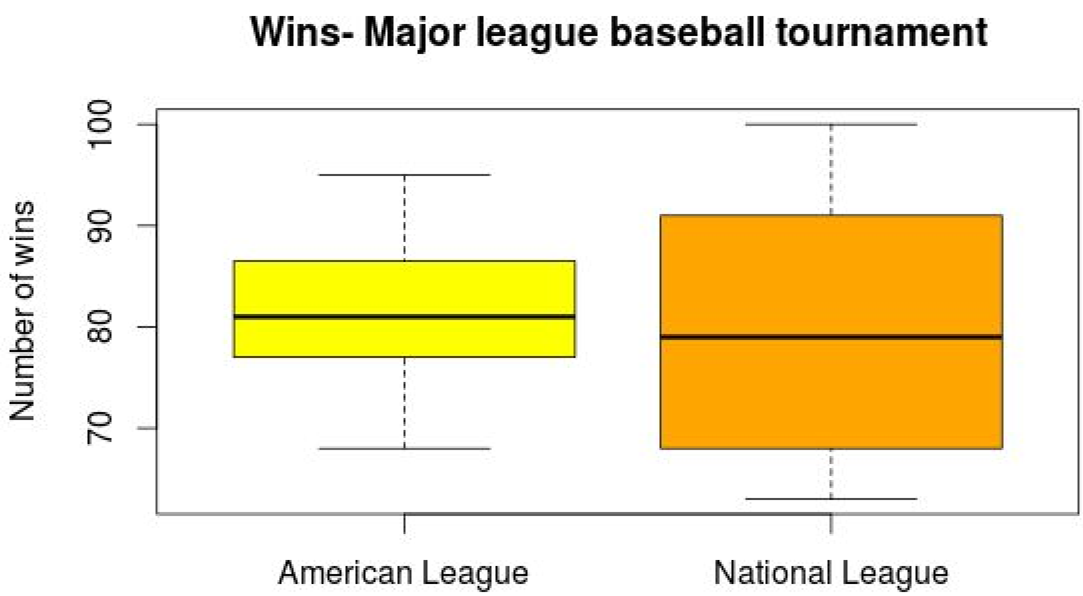
6 SOLUTION::

a. The code for the box plot is as given under:

```
AL = c(93,87,81,80,78,95,83,81,76,74,88,86,85,76,68)
NL = c(90,83,71,67,63,100,98,97,68,64,92,84,79,74,68)
```

```
boxplot(AL,NL,main = "Wins- Major league baseball tournament", ylab = "Number of wins", names = c("American League","National League"), col = c("yellow","orange"))
```

The box plot for the above code is :



b. Upon running the summary command in R, we get the following results for the distributions:

<u>American League:</u>

Minimum number of wins - 68
First Quartile - 77
Median - 81
Mean - 82.07
Third Quartile - 86.50
Maximum - 95

<u>National League:</u>

Minimum number of wins - 63
First Quartile - 68

Median - 79
Mean - 79.87
Third Quartile - 91
Maximum - 100

From the box plot and the above results, it is evident that the American League has tougher competition between teams as opposed to a much easier competition in the National League. The disparity between teams is more in National League in comparison to American League.

DISCUSSION: This assignment was discussed with Sairam Rakshith Bhyravabhotla (bsairamr).