class_01_discrete_variables

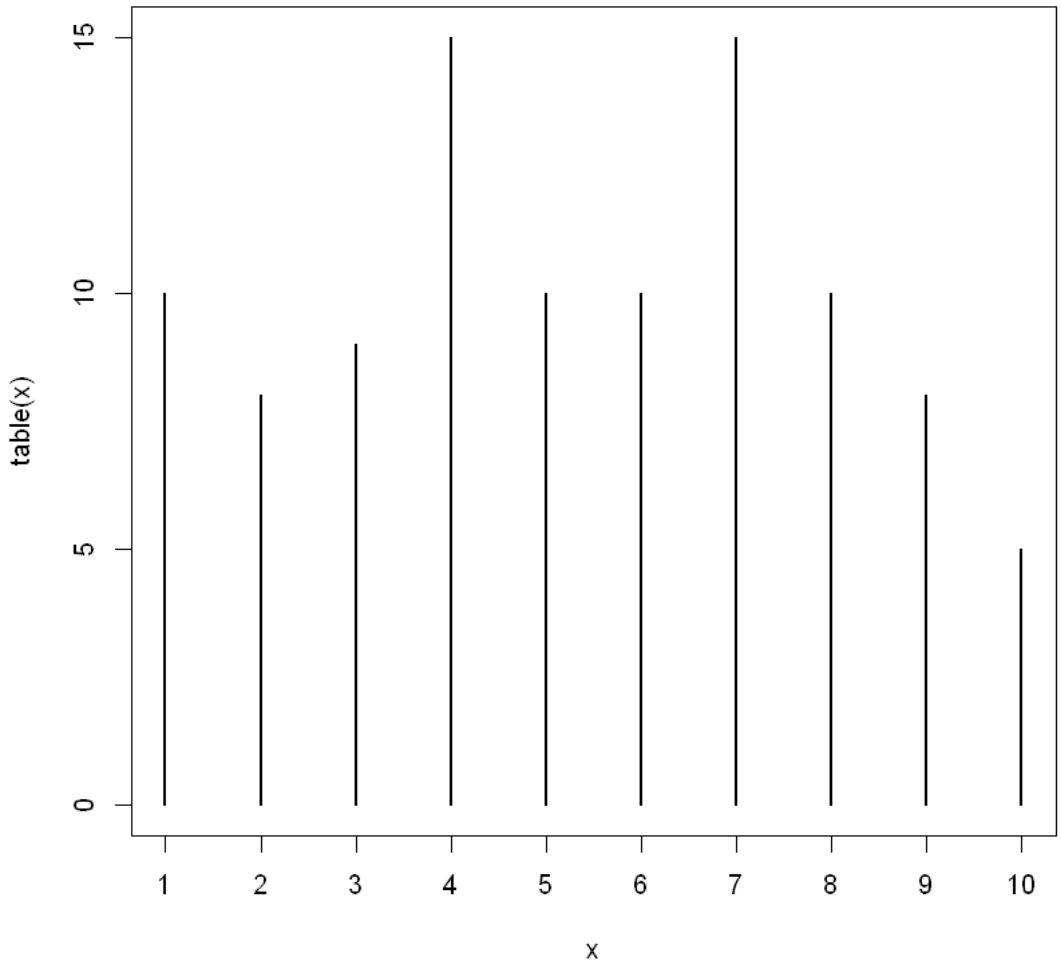
September 20, 2025

[1]:
```r
# Sampling
n <- 100
x <- sample(1:10, n, replace = TRUE)

summary(x)
mean(x)
plot(table(x))
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    1.0     3.0     5.0     5.3     7.0    10.0 
```

5.3



[16]:
```
median(x)
quantile(x)
```

6

0\% 1 **25\%** 3 **50\%** 6 **75\%** 8 **100\%** 10

Let the variance be: (population variance)

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

[17]:
```
sum((x - mean(x))^2) / n
# there's no R builtin function for population variance
```

7.6064

Let the sample variance be:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

[18]:
```
sum((x - mean(x))^2) / (n - 1)
var(x) # R builtin function for sample variance
```

7.68323232323232

7.68323232323232

Population variance from sample variance

$$\sigma^2 = \frac{n-1}{n} \times s^2$$

$$\sigma^2 = \frac{n-1}{n} \times \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

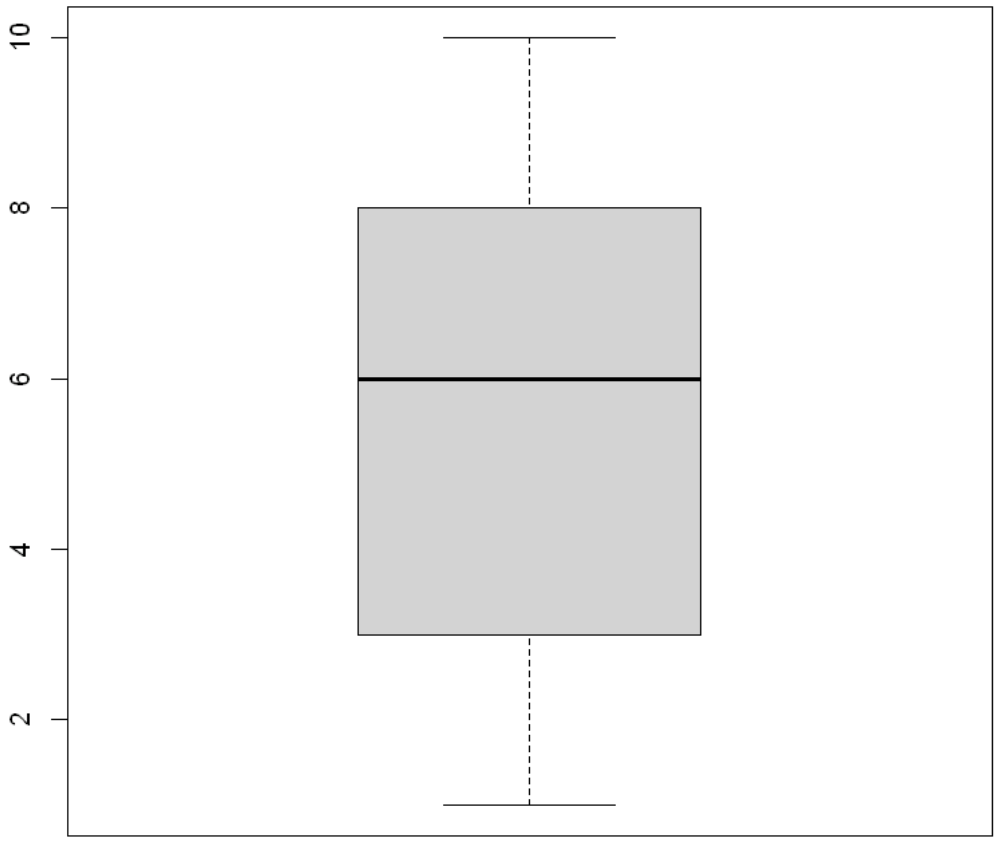
$$\sigma^2 = \times\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

[19]:
```
var(x) * (n - 1) / n
```

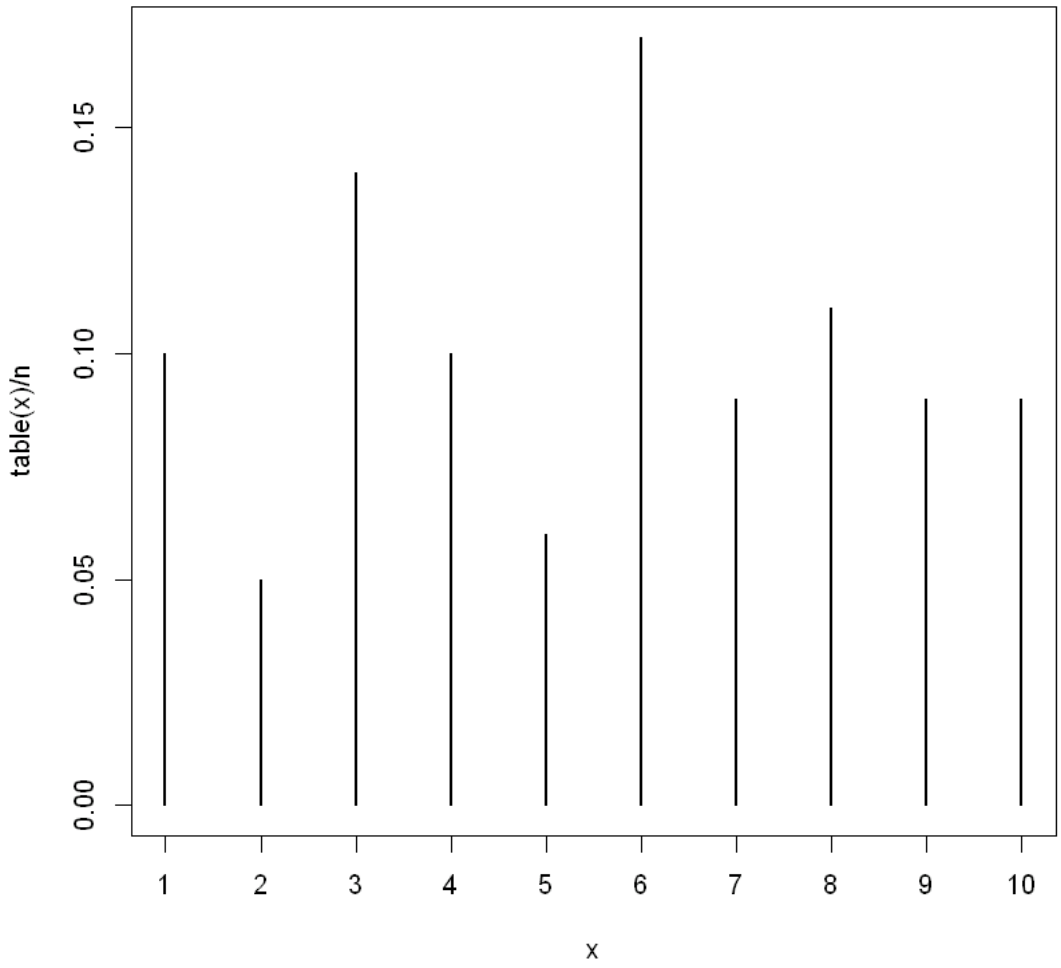
7.6064

[20]:
```
table(x)
boxplot(x)
```

```
x
 1  2  3  4  5  6  7  8  9 10
10  5 14 10  6 17  9 11  9  9
```



[23]:
```
# Relative frequencies
plot(table(x) / n)
```



[24]:
```
# which sum up to 1
fi <- table(x) / n
fi
sum(fi)
cumsum(fi)
```

```
x
   1    2    3    4    5    6    7    8    9   10
0.10 0.05 0.14 0.10 0.06 0.17 0.09 0.11 0.09 0.09
```

1

1 0.1 **2** 0.15 **3** 0.29 **4** 0.39 **5** 0.45 **6** 0.62 **7** 0.71 **8** 0.82 **9** 0.91 **10** 1

[25]:
```
plot.ecdf(x, main = "Empirical Distribution Function")
```