TD1 Descriptive Statistics and Probability (Solution)

Exercise 1.

The population Ω is 100 high school students.

The variable X is the number of extracurricular activities by one high school student.

The data are the number of extracurricular activities in which the high school students participate. the data are $2, 5, 7$.

Exercise 2.

Number of children in a family : quantitative variable discrete

Color of eyes : qualitative variable nominal

Socio-professional category : qualitative variable nominal (we can use it ordinal)

City of birth : qualitative variable nominal

Level of education : qualitative variable ordinal

Salary : quantitative variable continuous

Weight : quantitative variable continuous

Gender : qualitative variable nominal

Age : quantitative variable continuous

Mother language : qualitative variable nominal

Type of car: qualitative variable nominal

Height, IQ scores : quantitative variable continuous

.

Exercise 3.

(1)

The population Ω : 40 families in a village,

Variable X : the number of children in one family

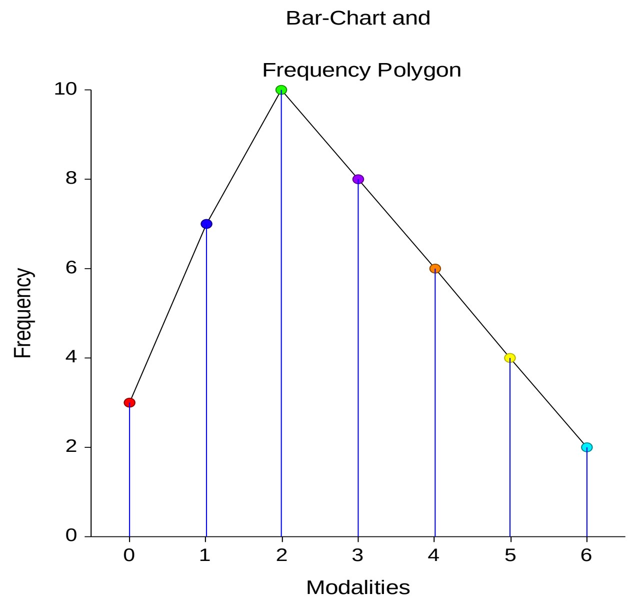
X is a quantitative discrete variable,

$X(\Omega) = \{0, 1, 2, 3, 4, 5, 6\}$.

(2)

Number of childrerns (x_i)	n_i	f_i	N_i	F_i	$n_i x_i$	$n_i x_i^2$
0	3	0,075	3	0,075	0	0
1	7	0,175	10	0,25	7	7
2	10	0,25	20	0,5	20	40
3	8	0,2	28	0,7	24	72
4	6	0,15	34	0,85	24	96
5	4	0,1	38	0,95	20	100
6	2	0,05	40	1	12	72
Total	$n = 40$	1			107	387

(3)



(5) $\overline{X} = \frac{1}{n} \sum_{i=1}^{7} n_n x_i = \frac{1}{40} (107) = 2,675$

1. $Var(X) = \frac{1}{n} \sum_{i=1}^{7} n_n x_i^2 - \left(\overline{X}\right)^2 = \frac{1}{40} (387) - (2,675)^2 = 2,52$

   $\sigma_X = \sqrt{Var X} = 1,6.$

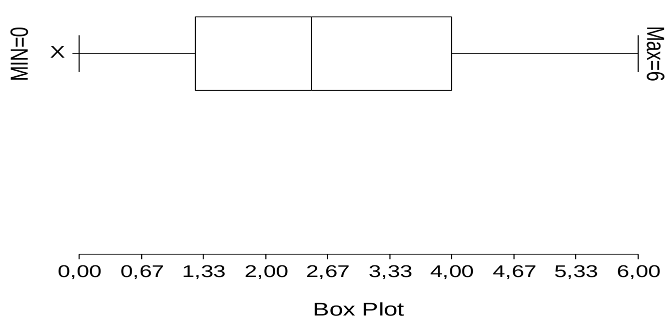
$\frac{n}{4} = 10 \in N_i \Longrightarrow Q_1 = \frac{1+2}{2} = 0,5$

$\frac{n}{2} = 20 \in N_i \Longrightarrow M = \frac{2+3}{2} = 2,5$

$\frac{3n}{4} = 30 \notin N_i \Longrightarrow Q_3 = 4$

Because $M < Mode < \overline{X}$ then the distribution is right skewnss

(6)



Exercise 4.

(a) $\overline{X} = \frac{3+5+2+6+5+9+5+2+8+6}{10} = 5,1.$

Arranging the data $3, 5, 2, 6, 5, 9, 5, 2, 8, 6$ in ascending ordre we obtain : $2, 2, 3, 5, 5, 5, 6, 6, 8, 9$, then $M = \dfrac{5+5}{2} = 5$.

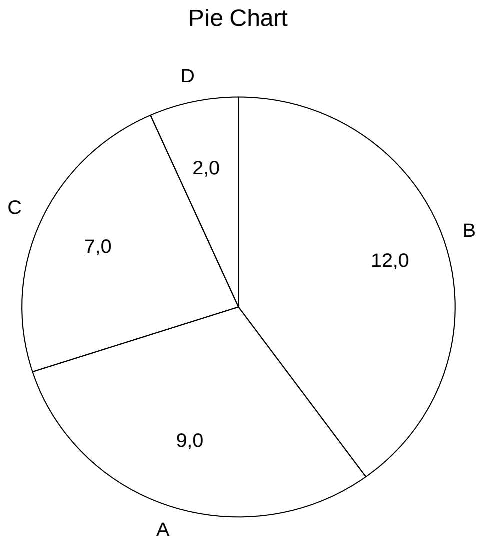
$Mode = 5$.

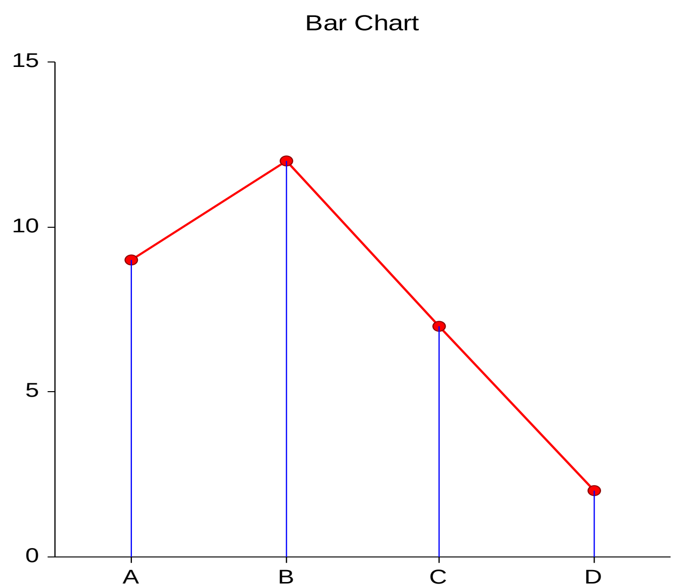
(b) $\overline{X} = \dfrac{51,6 + 48,7 + 50,3 + 49,5 + 48,9.}{5} = 49,72.$

Arranging the data $51,6$; $48,7$; $50,3$; $49,5$; 48.9 in ascending ordre we obtain : $48,7$; 48.9; $49,5$; $50,3$; $51,6$, then $M = 49,5$.

This data has no mode.

Exercise 5.





. **Exercise 7.**

Downtime	Frequencies (n_i)	$Cf\ N_i$	Class mark (x_i)	$f_i x_i$	$f_i x_i^2$
$[0, 10[$	3	3	5	15	75
$[10, 20[$	13	16	15	195	2925
$[20, 30[$	30	46	25	750	18750
$[30, 40[$	25	71	35	875	30625
$[40, 50[$	14	85	45	630	28350
$[50, 60[$	8	93	55	440	24200
$[60, 70[$	4	97	65	260	16900
$[70, 80[$	2	99	75	150	11250
$[80, 90[$	1	100	85	85	7225
Total	$n = 100$			3400	140300

$(a)\quad \overline{X} = \dfrac{1}{n}\sum_{i=1}^{9} n_i x_i = \dfrac{1}{100}(3400) = 34, \quad Var(X) = \dfrac{1}{n}\sum_{i=1}^{9} n_i x_i^2 - \overline{X}^2 = 1403 - 34^2 = 247.$

$(b) \quad \sigma_X = \sqrt{Var(X)} = 15,71.$

(c) Median : we first identify the class containing the median, then apply the following formula

$$M = L_1 + \frac{\frac{n}{2} - C}{f_m} (L_2 - L_1)$$

where

L_1 is the lower bound class containing the median;

n is the total frequency;

C is the cumulative frequency just before the median class;

n_m is the frequency of the median class;

L_2 is the upper bound class containing the median.

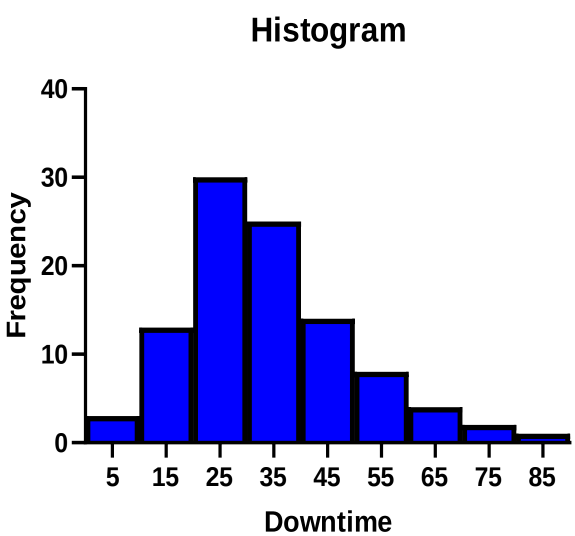
We proceed similarly with Q_1 and Q_3, by replacing $\frac{n}{2}$ by $\frac{n}{4}$ and $\frac{3n}{4}$ in the formula.
We have $M \in [30, 40[$, $Q_1 \in [20, 30[$ and $Q_3 \in [40, 50[$

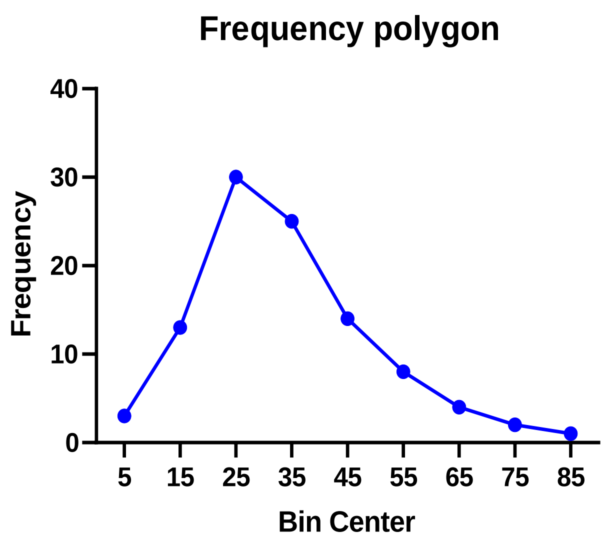
$$(c) \quad M = 30 + \frac{50 - 46}{25} \times 10 = 31.6,$$

$$(d) \quad Q_1 = 20 + \frac{25 - 16}{30} \times 10 = 23, \quad Q_3 = 40 + \frac{75 - 71}{14} \times 10 = 42.85$$

(e) The range $= 90 - 0 = 90$, (f) The coefficient of variation $CV = \frac{\sigma_X}{\overline{X}} \times 100\% = 47\%$, (g) The interquartile range $IQ = Q_3 - Q_1 = 42,85 - 23 = 19.85$

The mean is greater than the median, the distribution is right-skewed as is shown in frequency polygon below.





(3) The number of Downtime which are within $\overline{X}$ and $\overline{X} + \sigma_X$ is $N(\overline{X} + \sigma_X) - N(\overline{X})$ where N is the cumulative frequency function in its form defined as follows

$$N(x) = N_{i-1} + \frac{n_i}{l_i}(x - a_{i-1}), \quad x \in [a_{i-1}, a_i[$$

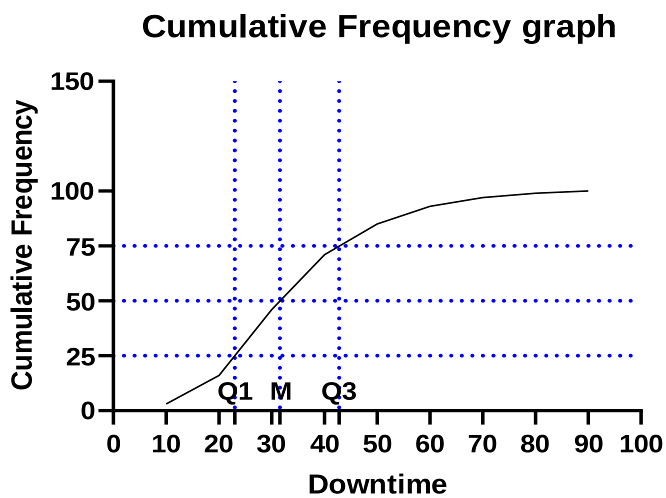
where $[a_{i-1}, a_i[$ is the class such that $x \in [a_{i-1}, a_i[$, N_{i-1} is the cumulative frequency just before $[a_{i-1}, a_i[$, n_i and l_i are the frequency and the width of $[a_{i-1}, a_i[$.

We have $\overline{X} = 34 \in [30, 40[$, $N_{i-1} = 46$, $n_i = 25$, $l_i = 10$ and $\overline{X} + \sigma_X = 49.71 \in [40, 50[$, $N_{i-1} = 71$, $n_i = 14$, $l_i = 10$, then

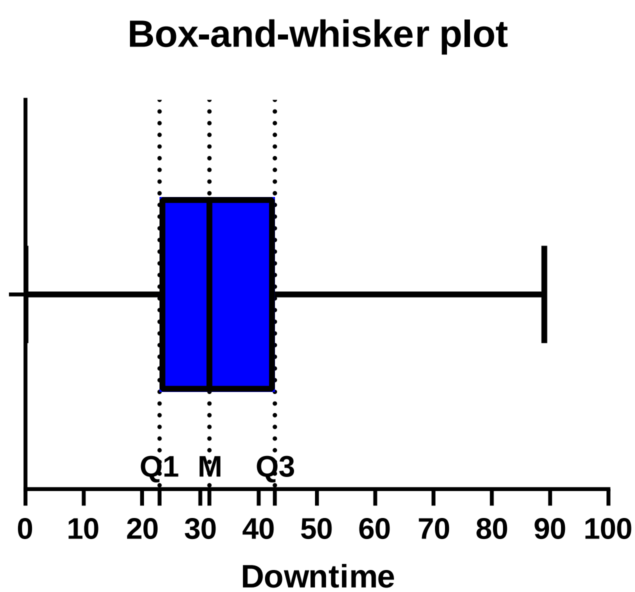
$$N(\overline{X}) = 46 + 2.5(34 - 30) = 56 \text{ and } N(\overline{X} + \sigma_X) = 71 + 1,4(49.71 - 40) = 84,59$$

Finally $N(\overline{X} + \sigma_X) - N(\overline{X}) = 29$ Downtimes.

(4)



(5)



Exercise 8.

Weight (grams)	Frequencies (n_i)	N_i	Class mark (x_i)	$n_i x_i$	$n_i x_i^2$
$[5,00;5,01[$	4	4	5.005	20.02	100.200
$[5,01,5,02[$	18	22	5.015	90.27	452.704
$[5,02;5,03[$	25	47	5.025	125.625	631.265
$[5,03;5,04[$	36	83	5.035	181.26	912.644
$[5,04;5,05[$	30	113	5.045	151.35	763.560
$[5,05;5,06[$	22	135	5.055	111.21	562.166
$[5,06;5,07[$	11	146	5.065	55.715	282.196
$[5,07;5,08[$	3	149	5.075	15.225	77.267
$[5,08;5,09[$	1	150	5.085	5.085	25.857
Total	$n = 150$			755.76	3807.859

(a)

$$\overline{X} = \frac{\sum_{i=1}^{9} n_i x_i}{n} = \frac{755.76}{150} = 5,038, \quad \sigma_X^2 = \frac{\sum_{i=1}^{9} n_i x_i^2}{n} - \overline{X}^2 = 0,004 \quad \sigma_X = \sqrt{0.004} = 0,063$$

(b) the number of bolts which are within one standard deviation of the mean : $N\left(\overline{X} + \sigma_X\right) - N\left(\overline{X}\right) = N(5,101) - N(5,038) = 150 - N(5,038) = 150 - \left(25 + \frac{36}{0,01} \times 0,008\right) = 96.$

(c) Let Y present the nut, we have $\overline{Y} = 2.043$ and we know that $\overline{X} = 5.038$, then $\overline{X+Y} = \overline{X} + \overline{Y} = 7.081.$

$Var(X+Y) = VarX + VarY = VarX + \sigma_Y^2 = 0,004 + (0,008)^2 = 0,004064.$

$\sigma_{X+Y} = \sqrt{Var(X+Y)} = 0,063.$

Exercise 9.

We have $\overline{X}$ =25,000 and σ_X =2000, all salaries are increased by \$2500, then $Y = X + 2500$ has $\overline{Y} = \overline{X} + 2500 = 27500$ and $\sigma_Y = |1| \times \sigma_X$ =2000.