mercredi 12 octobre 2022 18:47

Statistique descriptive à une dimension

Exercice 1h: Le responsable de la sécurité d'une entreprise a relevé le nombre d'accidents de travail observés durant une période de 100 jours. Les résultats apparaissent dans le tableau suivant:

Nombre d'accidents observées en une journée																								
0	0	0	1	1	2	0	0	0	0	0	3	1	0	2	2	0	0	1	0	2	1	1	0	3
1	0	0	0	1	1	4	2	1	1	1	0	0	0	4	3	0	0	0	1	1	1	3	0	0
1	0	0	2	4	1	0	0	1	1	1	1	3	3	0	0	1	0	1	1	2	0	2	1	2
3	4	2	1	0	0	1	1	2	0	1	1	0	0	2	1	1	1	3	0	0	1	1	0	2

a) Identifier la population et donner sa taille.
b) Identifier la variable statistique et préciser sa nature.
c) Dépouiller la série statistique.
d) Représenter avec un graphique approprié la distribution de cette variable.
e) Calculer les fréquences cumulées ascendantes et descendantes.
f) Tracer les courbes des fréquences cumulées ascendantes et descendantes.
g) Sur la période observée,
   1. quel pourcentage de jours a-t-on observé 3 accidents ou moins en une journée ?
   2. quel pourcentage de jours a-t-on observé 3 accidents au moins en une journée ?
   3. quel pourcentage de jours a-t-on observé moins de 3 accidents en une journée ?
   4. quel pourcentage de jours a-t-on observé 2 accidents ou plus en une journée ?
   5. quel pourcentage de jours a-t-on observé 2 accidents au plus en une journée ?
   6. quel pourcentage de jours a-t-on observé plus de 2 accidents en une journée ?
h) Déterminer l'étendue, la moyenne, la variance, l'écart-type, les écarts interquartile et interdécile et la dominante.

Étude statistique

N = Taille = nbr des individus

Population → Êtres humains ; des objets → voitures ; abstraits → jours

Echantillon

Variable statistique → quantitative ; qualitative

quantitative → discrète ; continue

discrète → $\{0, 1, 2, \dots, 8\}$

continue → $[155, 175[$, $[175, 179[$

$X : P \longrightarrow E$

individu ⟶ = une voiture ⟶ état mécanique $\in \{m, bm, b, e\}$

Corrections

① / ② La population : Les jours
La variable statistique : le nbr d'accidents / jour
C'est une variable quantitative discrète

③

x_i	n_i	N_i	f_i	F_i / f_{cc}	f_{cd}	$f_i \times x_i$
0	40	40	0,4	0,4	1	0
1	35	75	0,35	0,75	0,6	0,35
2	13	88	0,13	0,88	0,25	0,26
3	8	96	0,08	0,96	0,12	0,24
4	4	100	0,04	1 / 100 %	0,04	0,16
			1			

$\frac{n}{N} \times 100$

25 %

$X : P \longrightarrow \mathbb{N}$

ind ⟶ valeur

$X(P) = \{0, 1, 2, 3, 4\}$ → x_1, x_2, x_3, x_4, x_5

un accident ou moins → 0

au plus un accident → 1 → n_2 + 0 → n_1

④ Variable statistique quant discrète : Graphe en bâtons

$\longrightarrow \circ \longrightarrow n_1$ $\quad x_1\ x_2\ x_3\ x_4\ x_5$

④ Variable statistique quant discrète : Graphe en bâtons.

⑤ $G : x \longrightarrow$ nbr des individus tq la valeur de la V.S st $\leq x$.

50 famille

$\hookrightarrow \{0, 1, 2, \dots\dots, 8\}$

$\hookrightarrow G(4)$.

$\{0, 1, 2, 3, \underset{x_i}{4}, 6, 7, 7\}$ $\longrightarrow N_5$

$\hookrightarrow G(5) \longrightarrow G(4)$

Cas discret (Fonctions de répartition)

$\hookrightarrow x_i \leq x < x_{i+1}$

$G(x) =$ nbs des ind tq Var $\leq x$

$=$ nbr " " Var $\leq x_i = N_i$

$F(x) \Longrightarrow F_5$ (4 enfants?)

$x_i \leq x < x_{i+1} \Rightarrow \underline{G(x)} = \underline{N_i}$

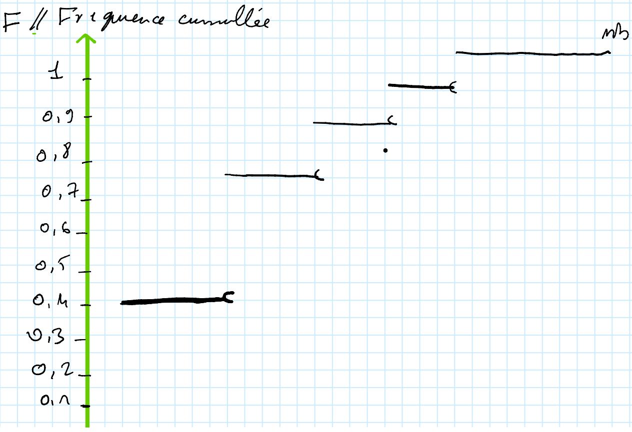
$x_1 \qquad x_2 \longrightarrow N_1$

$x_2 \qquad x_3 \longrightarrow N_2$

$\vdots$

$x_{n-1} \qquad x_n$

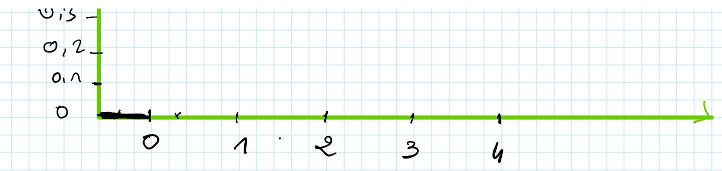
F // Fréquence cumullée



nb $\leq x < 0$

% des jours tq nb $\leq x$ / $x \geq 0$

$= \frac{0}{40}$ $\quad 0,5$



$1 \leqslant x < 2$

$\leqslant x \longrightarrow \leqslant 1$

$F(1) \Rightarrow$ % nb d'acc $\leqslant 1$

$x \geqslant 4 \rightarrow F(n) \Rightarrow$ % nbr d'acc $\leqslant x$

$\leqslant 4$

$F(n) \Rightarrow$ % des j / nbr d'acc $\leqslant x$.

(3) * $f_{cc}(3) = 0{,}96 = 96\%$ — 3 acci ou moins

* au moins 3 accidents

$\hookrightarrow f_{cd}(3) = 0{,}12 = 12\%$

* $f_{cc}(2) = 88\%$ — 2 acci au plus

* $f_{cd}(2) = 25\%$ — 2 acc au plus

* $f_{cd}(3)$ — 3/4 acci

* f_3

(4) $e = x_{max} - x_{min} = 4 - 0 = 4$. // L'étendue

Moyenne :

20 élèves → 5 → 6 ; 7 → 5 ; 4 → 8 ; 4 → 9

× 6	5
× 5	7
× 8	4
× 9	4
	20

$$\frac{5 \times 6 + 7 \times 5 + 4 \times 8 + 4 \times 9}{20}$$

4 → mauvais
10 → moy
7 → bonne
3 → excellent

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{k} n_i \times x_i = \sum_{i=1}^{k} \underbrace{\frac{n_i}{N}}_{f_i} \times x_i = \sum_{i=1}^{k} f_i \times x_i$$

$\overline{X} = 1{,}01$ accidents / jours

* Variance : $V(X) = \sigma^2(X)$

$= \overline{X^2} - \overline{X}^2$

1.203

$(f_1 \times 0^2 + f_2 \times 1^2 + f_3 \times 2^2 + f_4 \times 3^2 + f_5 \times 4^2) - (0{,}01)^2$

$= \overline{X^2} - \overline{X}^2$

$1{,}20\beta \to 1{,}20$

$= (f_1 \times 0^2 + f_2 \times 1^2 + f_3 \times 2^2 + f_4 \times 3^2 + f_5 \times 4^2) - (0{,}01)^2$

$= \underbrace{1{,}209}_{1} \longrightarrow 1{,}21.$

* Ecart-type : $\sigma(X) = \sqrt{V(X)}$
$= 1{,}10.$

* Mode : 0 accidents.

$\frac{N}{2} =$

Exercice 3: On considère les données suivantes:

11582	5231	35756	8916	5852	5354	11417	10705	4588	2025	11096
8167	5788	4627	5678	2921	5771	4672	16948	9015	20217	1766

a) Dépouiller ces données par classes de même amplitude.
b) Donner un découpage en classes plus approprié d'amplitudes inégales.

La règle de sturges

$$k = \boxed{1 + \frac{10}{3} \log_{10}(N)}$$

$$k = 1 + \log_2(N) = 1 + \frac{\log_{10}(N)}{\log_{10}(2)} = 1 + \frac{10}{3} \times \log_{10}(N)$$

$$N = 22 \Rightarrow k = 1 + \frac{3}{10} \log_{10}(22)$$

$$= 5{,}47 \longrightarrow 6 \text{ classes.}$$

$$a = \frac{E(X)}{k} = \frac{\max X - \min X}{k} \simeq 6\,000.$$

X	n_i
[0, 6000[	x
[6000, 12000[	
[12000, 18000[	
[18000, 24000[	
[24000, 30000[	
[30000, 36000[	

Exercice 4g: La répartition des 9100 exploitations agricoles d'une région selon leurs superficies en hectares est donnée dans le tableau suivant:

Superficie en ha des exploitations	% de la région	Superficie totale en ha
0 – 5	12.7	3 658
5 – 10	25.6	18 432
10 – 25	21.2	30 740
25 – 50	14.5	44 950
50 – 75	12.3	75 399
75 – 100	13.7	109 600
	100%	

a) Préciser la population. Identifier le caractère et préciser sa nature
b) Tracer l'histogramme et le polygone des fréquences.
c) Tracer, sur le même graphique, les polygones des effectifs cumulés.
d) Calculer l'écart interquartile.
e) Une entreprise spécialisée dans la vente de matériels agricoles, désire toucher les 30% exploitations les plus importantes. Déterminer la taille minimale des exploitations contactées.

a) Population : exploitation
V.S : Surface

$S_i = h_i \times a_i = n_i / f_i$

$h_i \times a_i = f_i$

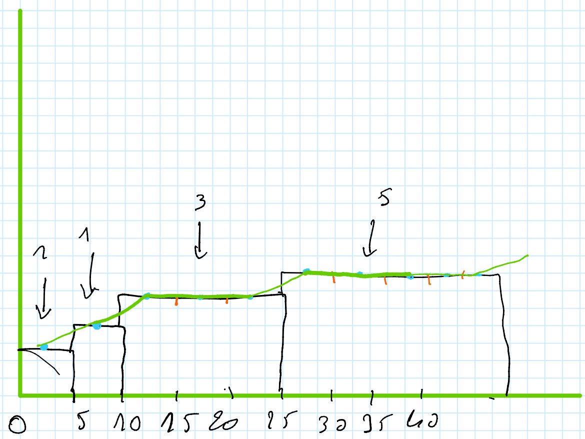
$h_i = \frac{f_i}{a_i}$

	X	f_i	$h_i = \frac{f_i}{a_i}$ → fréquence corrigé
5	[0 , 5[		
5	[5 , 10[		
15	[10 , 25[		
25	[25 , 50[		
25	[50 , 75[		
25	[75 , 100		

$S_i = C f_i$

$a_i \times h_i = C f_i$

$h_i = C \times \frac{f_i}{a_i}$



Exercice 2b: On a effectué une étude sur le taux de compactage de sols de remblai, et des couches de fondation, en mesurant les valeurs de densité maximale sur un site constitué de sols fins homogènes. 50 observations ont été effectuées et sont présentées ci-dessous:

85.9	86.3	88.1	88.6	88.9	89	89.4	89.5	89.9	90	90.3	90.4	90.4	90.6	90.7	90.9	90.9
91	91.1	91.1	91.6	91.7	91.8	91.9	92.1	92.3	92.3	92.3	92.7	92.7	92.8	93.3	93.4	93.7
93.8	93.8	93.8	94.1	94.2	94.3	94.4	94.7	95.1	95.3	95.6	96.1	96.5	96.6	97.3	98.2	

a) Préciser la population. Identifier le caractère, préciser sa nature et déterminer son étendue.
b) Dépouiller les données du site suivant une distribution des effectifs avec 85 comme limite inférieure de la première classe et une amplitude égale 2.
c) Tracer l'histogramme et le polygone des effectifs de la distribution.
d) Dresser le tableau des fréquences cumulées croissantes et décroissantes et tracer sur le même graphique les courbes correspondantes.
e) Quel pourcentage d'observations ont un taux inférieur à 94? supérieur ou égal à 92? compris dans l'intervalle [88 , 96[?
f) Dans quel intervalle de classe se situe le plus grand nombre d'observations?
g) Déterminer la valeur autour duquel le taux de compactage a tendance à se grouper.
h) Calculer la dispersion et le coefficient de variation du taux de compactage.
i) Déterminer, graphiquement et par le calcul, la valeur médiane du taux de compactage.
j) Quel pourcentage de mesures a un taux inférieur à la valeur moyenne ?