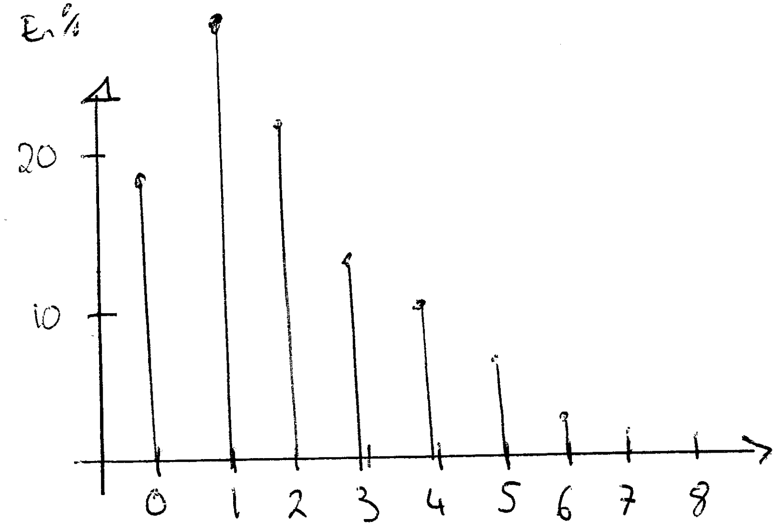
PSSD, Feuille 1

Exo 1

①



En fréquences :

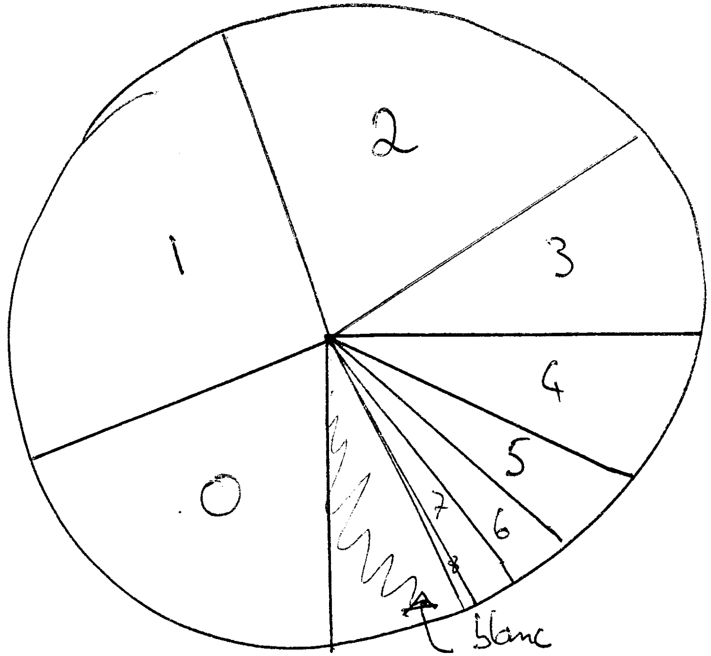
(n = 500)

0	1	2	3	4	5	6	7	8
18,2	29,2	20,8	12,6	9,4	6,6	2	0,8	0,4
18,2	47,4	68,2	80,8	90,2	96,8	98,8	99,6	100

Cumulées :

④ Médiane : 2 ; $q_{25\%}$: 1, $q_{75\%}$: 3.

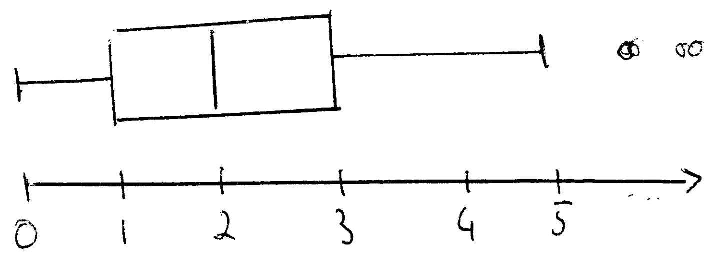
②



③ Moyenne observée : ~~0,47× 1 +~~ 0,29×1 + 0,2×2 + 0,12×3 + 0,09×4 + 0,07×5 + 0,02×6 + 0,01×7 + 0,004×8.

$\simeq 2$.

⑤



I

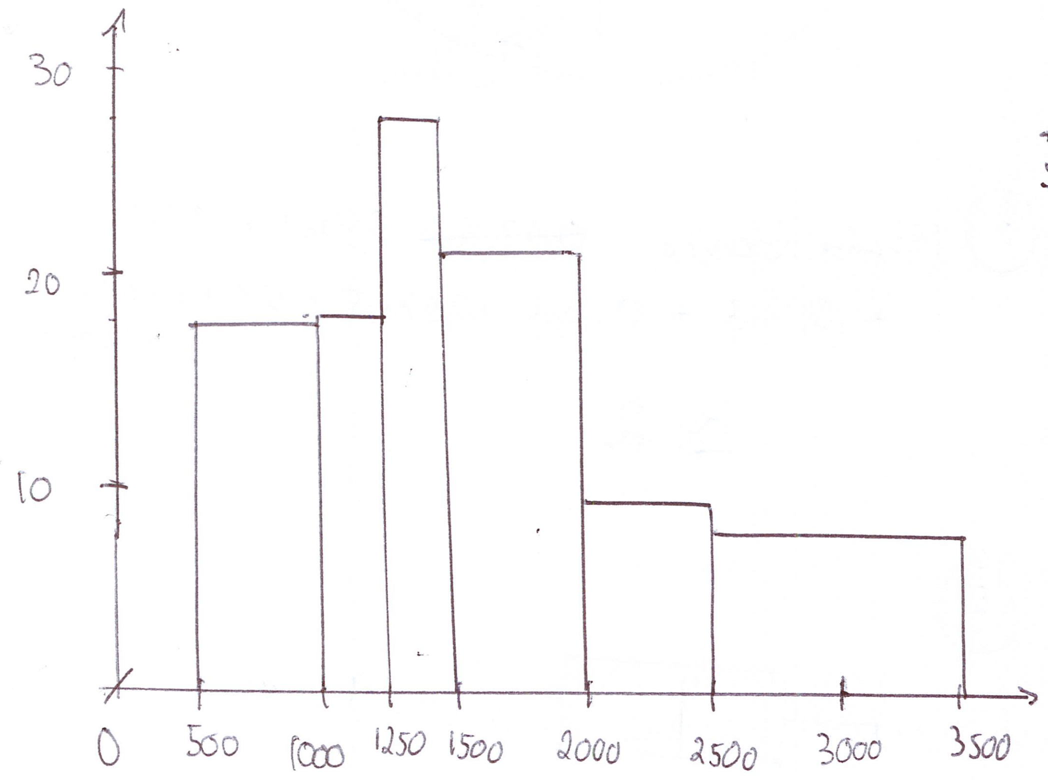
Exo 2 |

Freq / Freq cumulées

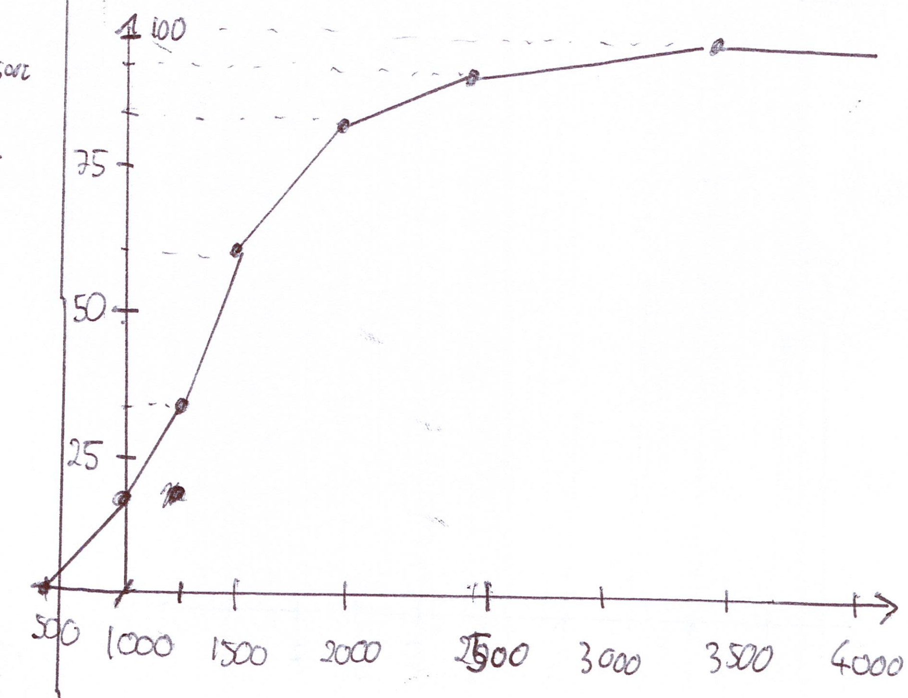
①

[500, 1000[	[1000, 1250[	[1250, 1500[	[1500, 2000[	[2000, 2500[	[2500, 3500[
17,1	17,5	27,4	20,3	9,6	8,1
17,1	34,6	62	82,3	91,9	100

Histogramme



Polygone Freq cumulées



② Classe modale : [1250, 1500[

Intervalle médian : [1250, 1500[

③ Non, il faudrait pour cela la répartit° des salaires au sein de chaque classe.

En supposant cette répartit° uniforme (cela revient à prendre comme valeur le milieu dans le calcul de la moyenne),

Salaire moyen $\simeq 0{,}17 \times 750 + 0{,}17 \times 1125 + 0{,}27 \times 1375$
$+ 0{,}2 \times 1750 + 0{,}1 \times 2250 + 0{,}8 \times 3000$
$\simeq 1515{,}36$.

II

Exo 3

(i) D'après le premier tableau, il vaut mieux choisir le Lycée B : bien que le nombre total de réussites soit moins élevé, le <u>taux</u> d'échec y est plus faible (estime $\mathbb{P}(\text{échec} \mid \text{Lycée B})$).

(i) Cela dépendra de votre classe sociale : on a

$$\hat{\mathbb{P}}(\text{échec} \mid \text{Favorisé} \cap \text{Lycée A}) \doteq 0{,}01$$
$$< \hat{\mathbb{P}}(\text{échec} \mid \text{Favorisé} \cap \text{Lycée B}) \doteq 0{,}013 .$$

Si vous êtes favorisé, il vaudra mieux choisir le Lycée A (encore que au vu de l'écart ce n'est pas une recommandation forte).

À l'inverse, si vous êtes défavorisé il vaudra mieux choisir le lycée B (à pas grand chose encore).

(i) Ce n'est pas vraiment un paradoxe : à classe sociale fixée les taux d'échecs des deux lycées sont sensiblement les mêmes, mais le taux d'échec des défavorisés est environ 2,5 fois plus élevé que celui des favorisés.

Le Lycée A comportant 7 fois plus d'élèves défavorisés que le B, cela pénalise artificiellement son score global (Tableau 1).

D'un point de vue maths : si votre classe sociale était tirée au hasard au sein du Lycée, il faudrait choisir le B ($\hat{\mathbb{P}}(\text{échec} \mid \text{Lycée B})$ plus petit).

Comme elle est 'fixe', il faut regarder $\hat{\mathbb{P}}(\text{échec} \mid \text{Lycée} \cap \text{Classe})$.

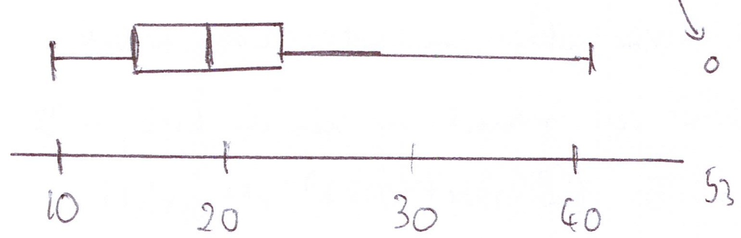
Exo 4

(I) La série étant triée, ça se lit directement:

- Médiane empirique : 19
- $q_{25\%}$: 15 (celui du 8 ième)
  $q_{75\%}$: 23 (—— 23 ième).

(II) Moyenne empirique : 21,1

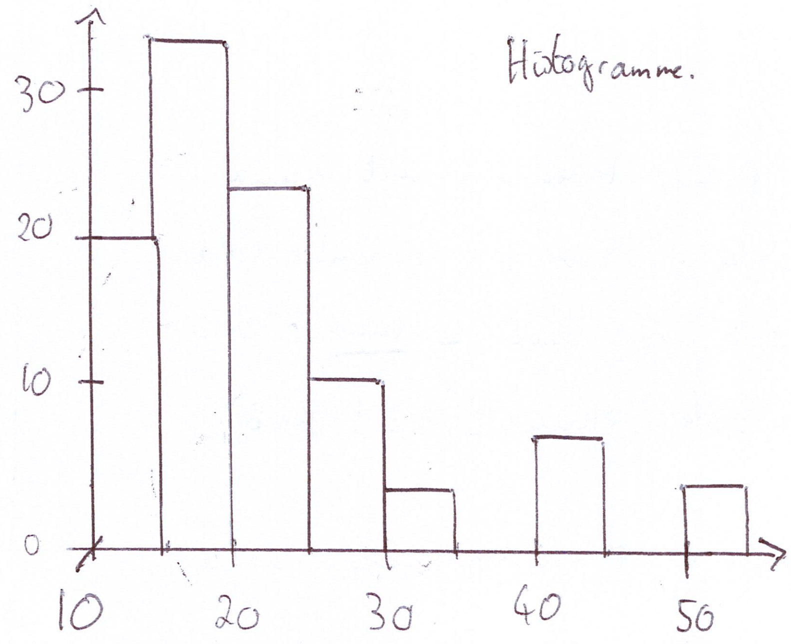
(III)



(IV)

[10,15[	[15,20[	[20,25[	[25,30[	[30,35[	[35,40[	[40,45[	[	[50,55[
6	10	7	3	1	0	2	0	1
20	33,3	23,3	10	3,3	0	6,7	0	3,3

Histogramme.



Moyenne approchée: $\dfrac{6\times 12{,}5 + 10\times 17{,}5 + 7\times 22{,}5 + 3\times 27{,}5 + 32{,}5 + 2\times 42{,}5 + 52{,}5}{30}$

$= 22$

Exo 5

Verso maths : ~~$\Omega = [1,13]$~~ $X \sim \mathcal{B}(p)$, p taux de mortalité.

Y à valeurs dans $[\![1,k]\!]$ (k tranches d'âge)

(X,Y) : pour un individu tiré au hasard,
(décès dans l'année, tranche d'âge).

Z : pays d'origine.

La deuxième assertion se traduit par

$$\forall j \in [\![1,k]\!] \quad \mathbb{P}(X=1 \mid Z=F \cap Y=j) \leqslant \mathbb{P}(X=1 \mid Z=I \cap Y=j)$$

La première par

$$\mathbb{P}(X=1 \mid Z=I) \leqslant \mathbb{P}(X=1 \mid Z=F).$$

Comme

$$\mathbb{P}(X=1 \mid Z=I) = \sum_{j=1}^{k} \mathbb{P}(X=1 \mid Z=I \cap Y=j)\, \mathbb{P}(Y=j \mid Z=I)$$

$$\mathbb{P}(X=1 \mid Z=F) = \sum_{j=1}^{k} \mathbb{P}(X=1 \mid Z=F \cap Y=j)\, \mathbb{P}(Y=j \mid Z=F),$$

les deux assertions ne sont pas incompatibles.

Comme pour les lycées il suffit que $\mathbb{P}(Y=j \mid Z=I) \gg \mathbb{P}(Y=j \mid Z=F)$ pour les classes d'âge à faible mortalité (typiquement les jeunes). On peut subodorer qu'il y a beaucoup plus de jeunes en Inde.

Ex6

1) $\overline{(x-a)} = \frac{1}{n}\left(\sum_{i=1}^{n}(x_i-a)\right) = \bar{x}-a$

$\overline{(x/a)} = \bar{x}/a$

$s_{(x-a)} = \left(\frac{1}{n}\sum_{i=1}^{n}\left[(x_i-a)-(\overline{x-a})\right]^2\right)^{\frac{1}{2}} = s_x$

$s_{x/a} = \left(\frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i}{a}-\frac{\bar{x}}{a}\right]^2\right)^{\frac{1}{2}} = \frac{s_x}{a}.$

2) $\overline{\frac{x-\bar{x}}{s_x}} = \frac{\bar{x}-\bar{x}}{s_x} = 0$

$s_{\frac{x-\bar{x}}{s_x}} = \frac{s_{x-\bar{x}}}{s_x} = \frac{s_x}{s_x} = 1.$

Ex7

1) On a $\bar{x} = \frac{1}{n_y+n_z}\sum_{u\in x} u = \frac{1}{n_y+n_z}\left(\sum_{u\in y} u + \sum_{u \in z} u\right)$

$= \frac{n_y}{n_y+n_z}\bar{y} + \frac{n_z}{n_y+n_z}\bar{z}.$

$= p_y\bar{y} + p_z\bar{z}.$

V

(2)

$$s_x^2 = \frac{1}{n_y+n_z} \sum_{u \in x} (u-\bar{x})^2$$

$$\cancel{= \frac{1}{n_y+n_z} \left(\sum_{u \in y} (u - p_y\bar{y} - p_z\bar{z})^2 + \sum_{u \in z} (u - p_y\bar{y} - p_z\bar{z})^2 \right)}$$

~~Regardons le 1er terme~~ :

$$\cancel{\sum_{u \in y} (u - p_y\bar{y} - p_z\bar{z})^2 = \sum_{u \in y} (u - p_y\bar{y}}$$

$$= \frac{1}{n_y+n_z} \sum_{u \in y} (u - \bar{y} + \bar{y} - \bar{x})^2 + \frac{1}{n_y+n_z} \sum_{u \in z} (u - \bar{z} + \bar{z} - \bar{x})^2$$

Regardons le premier terme.

$$\frac{1}{n_y+n_z} \sum_{u \in y} (u-\bar{y}+\bar{y}-\bar{x})^2 = \frac{1}{n_y+n_z} \sum_{u \in y} (u-\bar{y})^2$$

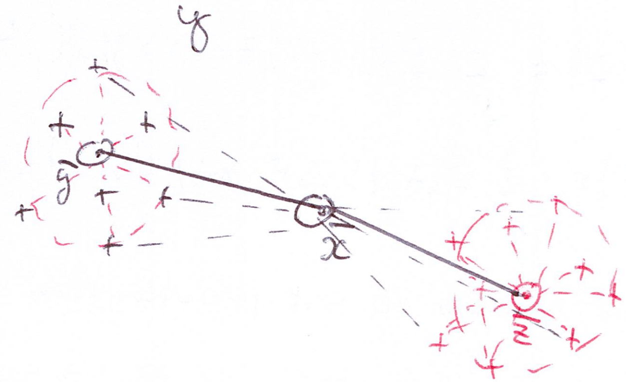
$$+ \frac{1}{n_y+n_z} \times n_y (\bar{y}-\bar{x})^2 + \frac{1}{n_y+n_z} (\bar{y}-\bar{x}) \underbrace{\sum_{u \in y} (u-\bar{y})}_{0}$$

$$= p_y s_y^2 + p_y (\bar{y}-\bar{x})^2.$$

Le deuxième terme se traite pareillement. On a alors

$$s_x^2 = p_y s_y^2 + p_z s_z^2 + \left(p_y (\bar{y}-\bar{x})^2 + p_z (\bar{z}-\bar{x})^2\right)$$

Dessin



s_x^2 : - -

variance inter : —

variance intra :

Ex 8

Pour $y \in \mathbb{R}$, on note $N_y = \{i \mid x_i \leq y\}$, et

$$G(y) = \sum_{i=1}^{n} |x_i - y|.$$

On a alors

$$G(y) = \sum_{i \in N_y} (y - x_i) + \sum_{i \in N_y^c} (x_i - y).$$

On remarque que G est affine par morceaux, dérivable sur $\mathbb{R} \setminus x_{1:n}$, et convexe.

Pour $y \notin x_{1:n}$, on a $G'(y) = |N_y| - |N_y^c|$.

S'il existe $y \notin x_{1:n}$ tq $|N_y| = |N_y^c|$, alors il minimise G, et c'est une médiane.

Sinon, comme $y \mapsto |N_y| - |N_y^c|$ est $\nearrow$, ~~(de −n à n)~~ vaut $-n$ pour $y < \min(x_i)$, n pour $y \geq \min(x_i)$, constante par morceaux, il existe x_{j_0} tq

$$\begin{cases} |N_y| - |N_y^c| < 0 & \text{si } y < x_{j_0} \\ |N_y| - |N_y^c| > 0 & \text{si } y > x_{j_0} \end{cases}$$

$|N_y|$ étant continue à droite, on en déduit

$$|N_{x_{j_0}}| \geq |N_{x_{j_0}}^c|$$

Donc $|\{i \mid x_i \leq x_{j_0}\}| \geq \frac{n}{2}$.

Pour y assez proche de x_{j_0}, $y < x_{j_0}$, on a

$$|N_y| = |\{i \mid x_i < x_{j_0}\}|$$

$$|N_y^c| = |\{i \mid x_i \geq x_{j_0}\}|$$

et $|N_y| - |N_y^c| < 0$ donne

$$|\{i \mid x_i \geq x_{j_0}\}| \geq \frac{n}{2}.$$

x_{j_0} est alors une médiane, et par convexité de G minimise bien G.