

Chapitre 7: Statistiques:

I) Statistiques à une variable:

Exemples:

Série A: 2-3-3-4-5-6-6-7-7-7-8-8-8-8-8-9-9-9-9-9-9-10-10-11-11-11
13-13-15-16

Q_1

Q_3

Notes obtenues à un examen -

Série B:	x_i	Salaire	[900; 1200]	[1200; 1400]	[1400; 1600]	[1600; 1800]	[1800; 2000]	[2000; 2400]	Total
	n_i	Effectif	30	30	60	80	40	40	280

Série C: Adhérents à un club sportif -

17 % jouent au Handball

25 % jouent au rugby

58 % jouent au tennis -

A - Méthodes de représentation.

1) Vocabulaire:

population: l'ensemble des "individus" sur lequel porte l'étude statistique

échantillon: une partie de la population.

caractère: une "propriété" de la population que l'on étudie
(variable)

↳ discret: s'il prend des valeurs isolées (que l'on peut dénombrer)

Ex: nombre de frères et sœurs.

↳ continu: dans le cas contraire

Ex: taille d'un individu.

↳ qualitatif: si différent d'un nombre

Ex: la couleur des yeux

↳ quantitatif: s'il exprime une quantité.

effectif: le nombre d'"individus" associés à une valeur du caractère.

2) Tableaux: On considère une série statistique X à caractère quantitatif dont les p valeurs sont données par les $x_1, \dots, x_p$ d'effectifs associés

$n_1, \dots, n_p$

* à chaque valeur x_i on associe la fréquence $f_i = \frac{n_i}{N}$
avec N l'effectif total ($n_1 + n_2 + \dots + n_p$)

Série A * L'ensemble des fréquences est appelé la distribution en fréquence de X

$X = x_i$	2	3	4	5	6	7	8	9	10	11	13	15	16	Total
n_i	1	2	1	1	2	3	5	6	2	3	2	1	1	30 $N=30$
f_i	$\frac{1}{30}$	$\frac{2}{30}$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{5}{30}$	$\frac{6}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{2}{30}$	$\frac{1}{30}$	$\frac{1}{30}$	1
f_i en %	3	7	3	3	7	10	17	20	7	10	7	3	3	100

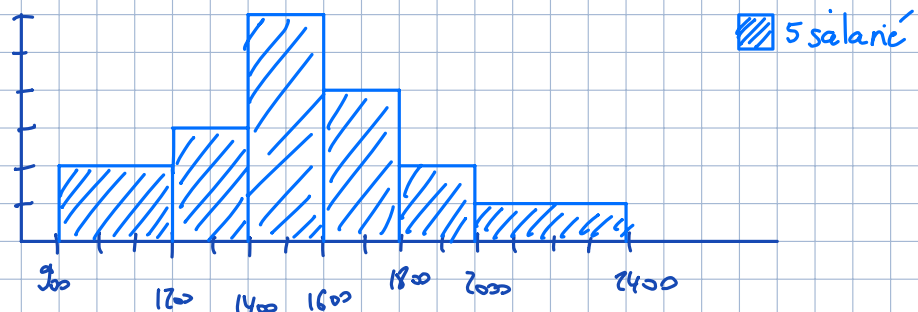
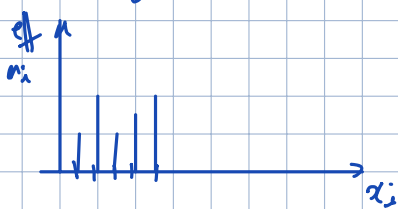
TI → STATS → Editer L_1 | L_2 ... $L_2 : 30 \rightarrow L_3$
 $x_i \rightarrow \begin{matrix} 2 \\ 3 \\ 4 \\ \vdots \end{matrix}$ | $\begin{matrix} 1 \\ 2 \\ 1 \\ \vdots \end{matrix}$ m_i

* Regroupement en classes : Pour une meilleure lecture on peut regrouper les valeurs par intervalles disjoints appelés classes.

X	$[0; 5[$	$[5; 10[$	$[10; 15[$	$[15; 20]$	Total
n_i effectifs	4	17	7	2	30
f_i fréquences	0,13	0,57	0,23	0,07	1

3) Graphiques :

a) Diagramme en bâtons : b) Histogramme Salaire.



c) Diagramme de proportionnalité : Adhérents.

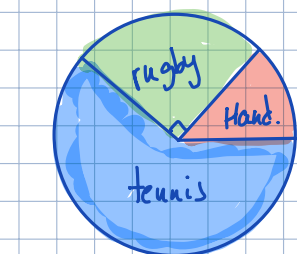
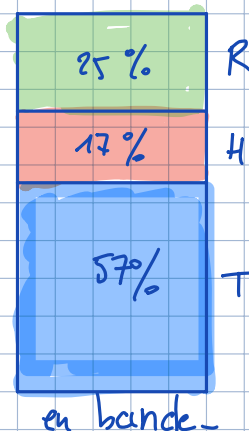
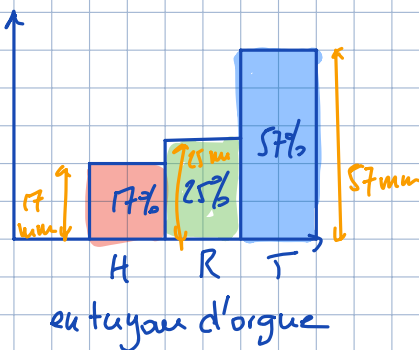


diagramme circulaire



3 - Caractéristique d'une série statistique :

1) Caractéristique de position :

a) Moyenne

Soit X une série statistique de valeurs $x_1, \dots, x_p$ d'effectifs associés

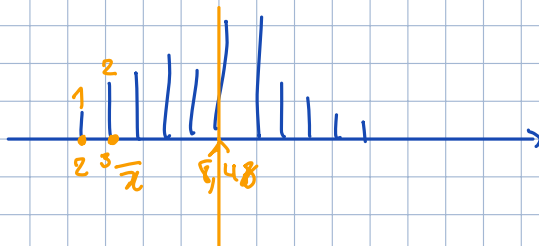
$m_1, \dots, m_p$ on pose $N = m_1 + \dots + m_p$ l'effectif total.

La moyenne pondérée de cette série notée $\bar{x}$

$$\text{et } \bar{x} = \frac{m_1 x_1 + m_2 x_2 + \dots + m_p x_p}{m_1 + m_2 + \dots + m_p} = \frac{1}{N} \sum_{i=1}^p m_i x_i = \sum_{i=1}^p f_i x_i$$

(TI) $\rightarrow x_i \rightarrow L_1, m_i \rightarrow L_2, \text{STATS} \rightarrow \text{Calc} \rightarrow \text{Stat1 1-VAR} \rightarrow X \text{ liste } L_1 \rightarrow \bar{x} \approx 8,47.$
freq liste L_2

(NW) $\rightarrow \text{Statistiques} \rightarrow x_i \rightarrow V_1, m_i \rightarrow N_1 \rightarrow \text{onglet Stats} \rightarrow \bar{x} \approx 8,47.$



Ca particulier du regroupement en classes : on remplace les classes par le milieu des intervalles

Ex: Série B : Le salaire moyen est $\bar{S} \approx 1645$

$$\bar{S} = \frac{30 \times 1050 + 30 \times 1300 + \dots + 40 \times 2200}{280}$$

b) Médiane Soit X une série statistique ordonnée $x_1 \leq x_2 \leq \dots \leq x_p$

La médiane M est le nombre qui permet de diviser la série en deux sous-groupes de même effectif.

* Si N est impair alors la valeur de cette série est celle située au centre de la série - $\left(x_{\frac{N+1}{2}}\right)$

Ex: $\{ \underset{3}{3}; \underset{3}{7}; \underset{3}{8}; \boxed{12}; 12; 13; 15 \}$ $N=7$ impair donc $M=12$.

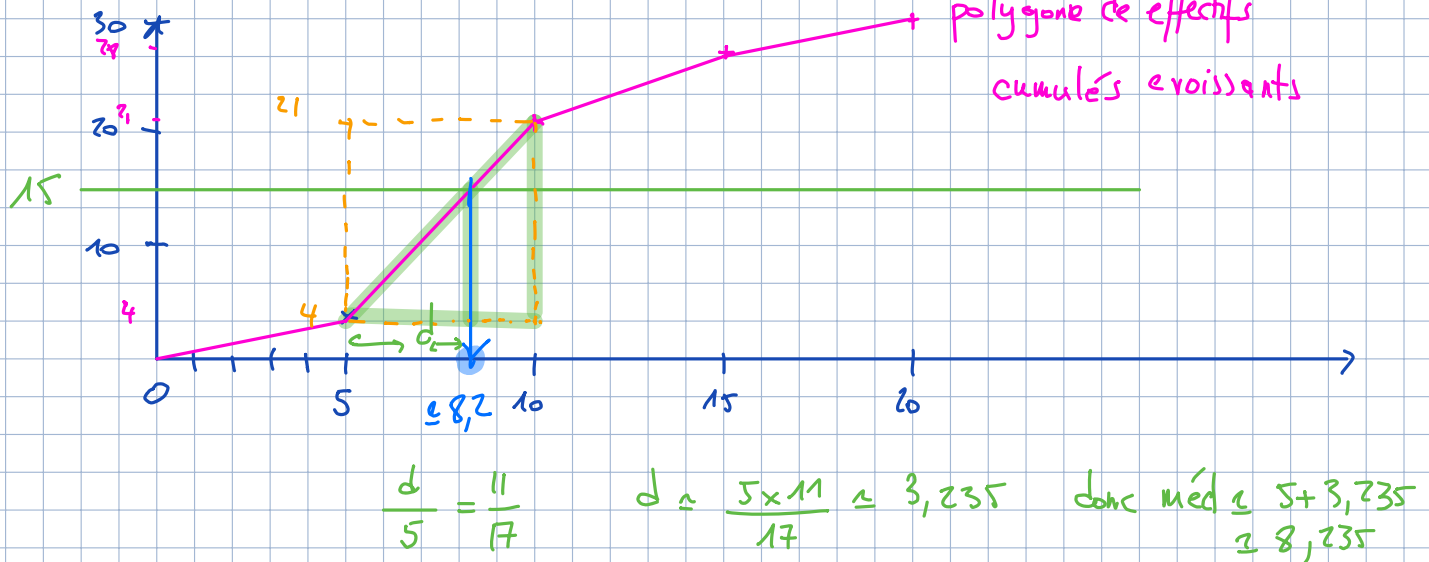
* Si N est pair alors la valeur de la médiane est la moyenne des 2 valeurs centrales $\left(x_{\frac{N}{2}} \text{ et } x_{\frac{N}{2}+1}\right)$

Ex $\{ 11; 13; 15; 18; 19; 20 \}$ $N=6$ donc $M = 16,5$.

$$\frac{15+18}{2} = 16,5$$

Cas particuliers du regroupement en classe:

Note	$[0; 5[$	$[5; 10[$	$[10; 15[$	$[15; 20]$
effectif.	4	17	7	2
effectif cumulés croissants	4	21	28	30



c) Quartiles: Q_1 et Q_3

* le premier quartile Q_1 est la plus petite valeur du caractère x_i telle que l'effectif des valeurs inférieures à x_i représente au minimum 25% de l'effectif total.

* le troisième quartile Q_3 (idem avec 75%)

Ex: Série A: $N=30$ 25% de $30 = 7,5 \approx 8$ par excès

donc Q_1 est la 8^{ème} valeur de la série

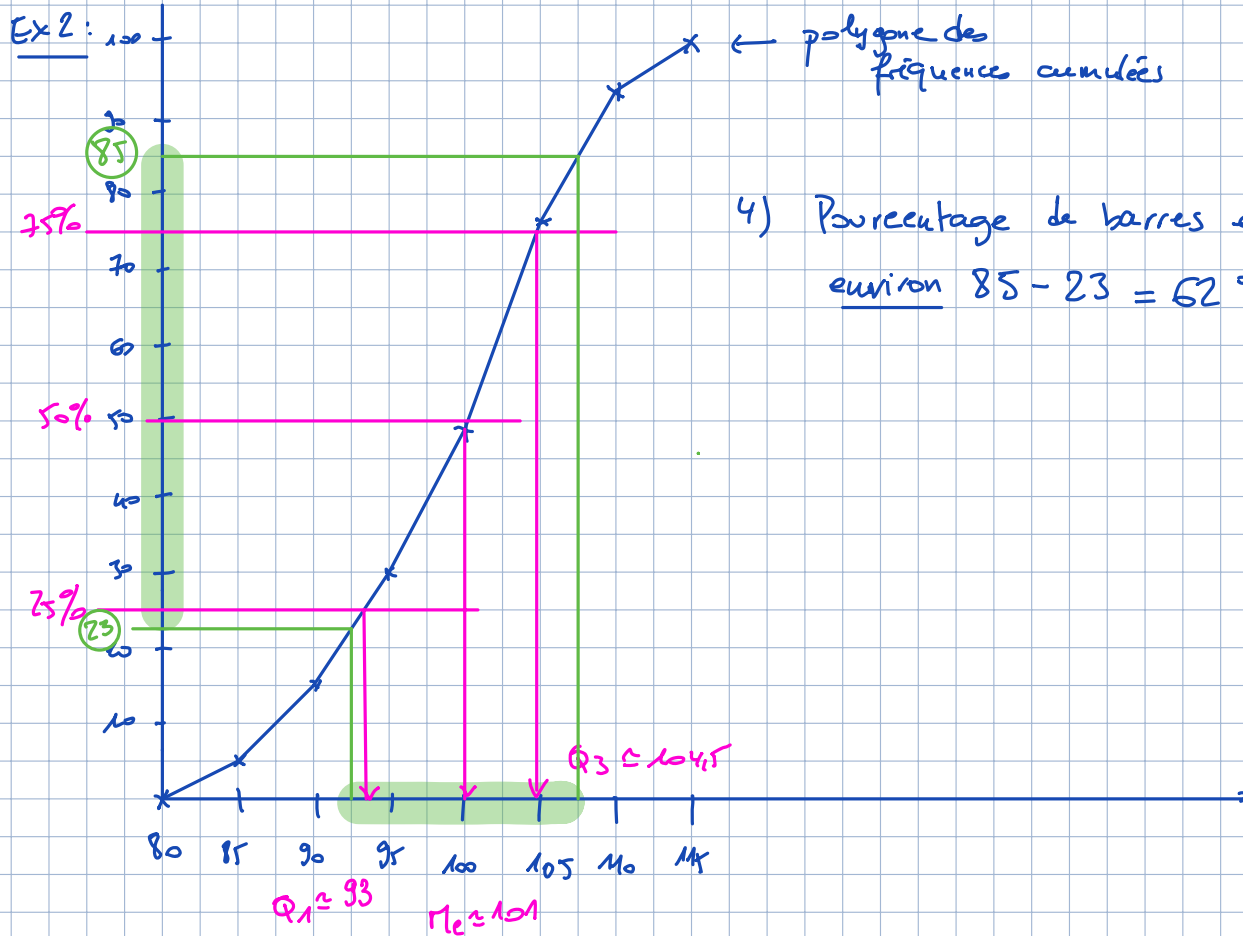
$$Q_1 = 7$$

et 75% de $30 = 22,5 \approx 23$ par excès

Q_3 est la 23^{ème} valeur de la série.

$$Q_3 = 10$$

* Cas particulières du regroupement en classe: On utilise les effectifs cumulés



2) Caractéristiques de Dispersion.

a) Étendue: $V_{\max} - V_{\min}$.

Ex Série A: $16 - 2 = 14$ est l'étendue de la série A

b) Écart-type et variance:

Soit X une série statistique quantitative, discrète.
on appelle variance de X notée $V(X)$

la moyenne des écarts entre la moyenne $\bar{x}$ et les valeurs de la série au carré.

$$V(X) = \frac{1}{N} \sum_{i=1}^p (\bar{x} - x_i)^2 \times m_i$$

on appelle écart-type de X noté $\sigma(X)$ la racine carrée de $V(X)$

$$\sigma(X) = \sqrt{V(X)}$$

Exemple: cf. Ex 1: Notes et effectifs.

STATISTIQUES		
Données	Graphique	Stats
V1/N1		
Effectif total	n	210
Minimum	Min	5
Maximum	Max	14
Etendue	E	9
Moyenne	$\bar{x}$	7.971429
Ecart type	σ	2.561462
Variance	σ^2	6.561088
Premier quartile	Q1	5
Troisième quartile	Q3	10

TI Stats $\rightarrow$ Calc $\rightarrow$ Stats 1 VAR
 $\downarrow$
 $\sigma_x \approx 2,561$

$\rightarrow \sigma(X) \approx 2,561$
 $\rightarrow V(X) \approx 6,561$

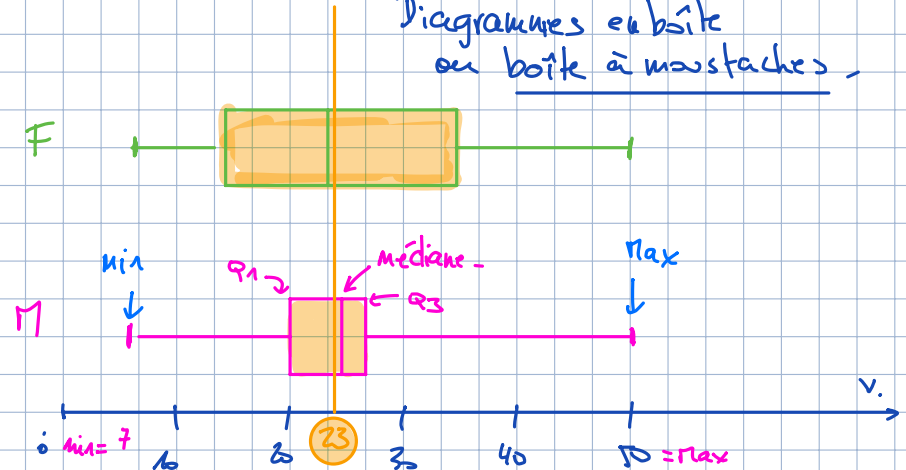
Ex3

Vitesse du vent (en noeuds)	7	14	16	18	20	22	24	26	27	30	44	50
Effectif (en jours)	1	2	1	1	4	5	3	4	4	2	1	2
Effectif cumulés croissants	1	3	4	5	9	14	17	21	25	27	28	30

STATISTIQUES		
Données	Graphique	Stats
V1/N1		
Maximum	Max	50
Etendue	E	43
Moyenne	$\bar{x}$	24.9
Ecart type	σ	9.173694
Variance	σ^2	84.15667
Premier quartile	Q1	20
Troisième quartile	Q3	27
Médiane	Med	24
Ecart interquartile	EI	7

5.

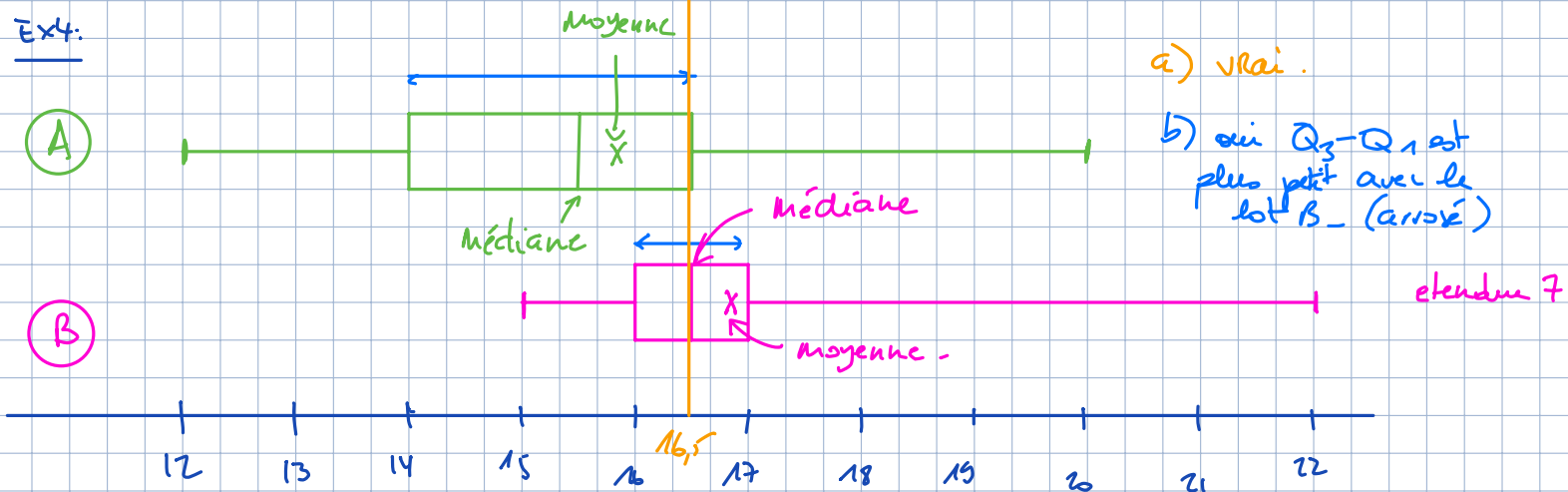
Diagrammes en boîte
ou boîte à moustaches



Ex4:

A

B



a) vrai.

b) oui $Q_3 - Q_1$ est plus petit avec le lot B. (arrêté)

Ex5: 1) $\bar{x} = 4,946$ $\sigma \approx 0,386$

2) $4,9 < 4,946 < 5,1$
 $0,386 < 0,4$

$[\bar{x} - 2\sigma; \bar{x} + 2\sigma] = [4,174; 5,718] \approx [4,7; 5,7]$

Où selon ces critères le prod° est bonne.

$90\% < \frac{4+14+25+18+15+9+7+3}{100} \approx 95\%$

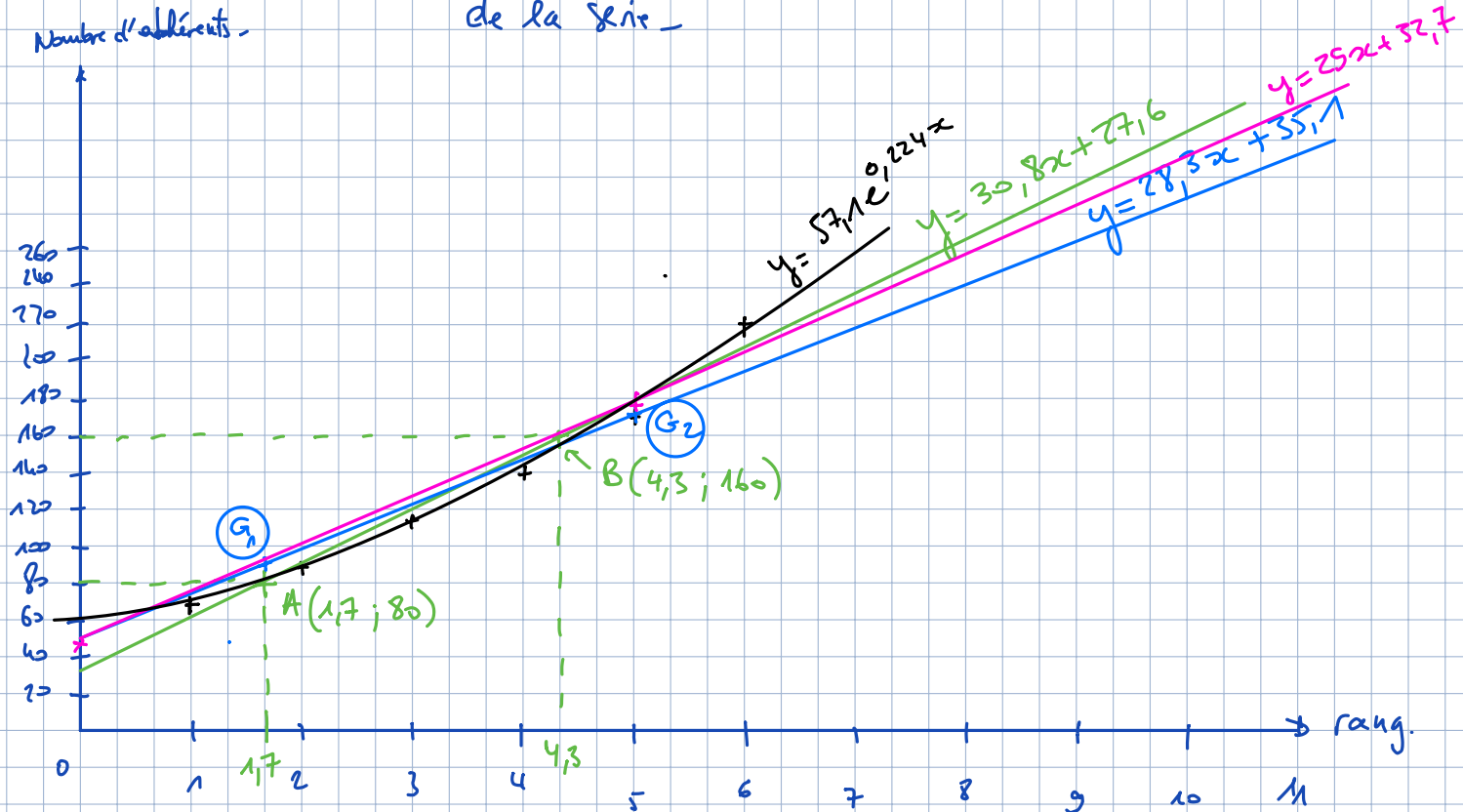
II) Statistiques à 2 variables:

Exemples: L'évolution des adhérents d'un club de rugby et donné dans le tableau suivant.

Année	2020	2021	2022	2023	2024	2025
rang: x_i	1	2	3	4	5	6
no. d'adhérents y_i	70	90	115	140	170	220

N_1 N_2

1) Nuage de points: L'ensemble des points $M_i(x_i; y_i)$ est appelé nuage de points de la série.



2) Régression

a) Régression linéaire:

i) ajustement à la règle:

$A(1,7; 80)$ $B(4,3; 160)$

$$a = \frac{y_B - y_A}{x_B - x_A} = \frac{160 - 80}{4,3 - 1,7} \approx 30,8$$

$$b = y_A - ax_A \approx 80 - 30,8 \times 1,7 \approx 27,6$$

$$y = 30,8x + 27,6$$

grâce à cette régression on a en 2030:

$$y = 30,8 \times 11 + 27,6 \approx \underline{366} \text{ adhérents prévus.}$$

ii) ajustement de Mayer

on utilise les points moyens de 2 sous-muages d'effectifs similaires pour déterminer la droite de régression.

$$N_1 : \{ (\underline{1}; 70) ; (\underline{2}; 90) ; (\underline{3}; 115) \} \quad G_1(\bar{x}_1; \bar{y}_1) = (2; 91,7)$$

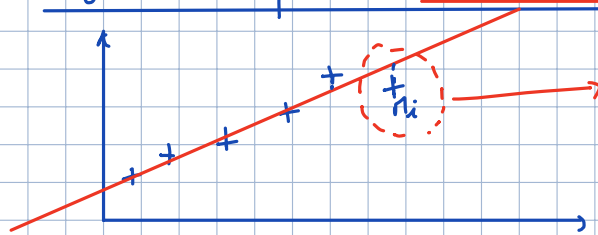
$$N_2 : \{ (4; 140) ; (5; 170) ; (6; 220) \} \quad G_2(\bar{x}_2; \bar{y}_2) = (5; 176,7)$$

$$a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{176,7 - 91,7}{5 - 2} \approx 28,3$$

$$b = y_{G_1} - ax_{G_1} = 91,7 - 28,3 \times 2 \approx 35,1$$

$$\left. \begin{array}{l} y = 28,3x + 35,1 \\ \text{en 2030:} \\ 28,3 \times 11 + 35,1 \\ \approx 346 \text{ adhérents.} \end{array} \right\}$$

iii) ajustement par la méthode des moindres carrés



$$\begin{array}{c} Q_i \\ \updownarrow \\ M_i \end{array} \quad \begin{array}{c} Q_i \\ \updownarrow \\ M_i \end{array} \quad \begin{array}{c} Q_i \\ \updownarrow \\ M_i \end{array} \quad \begin{array}{c} Q_i \\ \updownarrow \\ M_i \end{array} \quad \begin{array}{c} Q_i \\ \updownarrow \\ M_i \end{array}$$

$$\sum M_i Q_i^2 \text{ minimal.}$$

à la calculatrice: (NW)

① on rentre le tableau $X_1; Y_1$

② graphique → régression → affine

③ Stats on lit a et b.

rad REGRESSIONS		
Données	Graphique	Stats
chantillon s	X1	Y1
de points N	1.870829	54.99242
Covariance cov		6
produits $\sum xy$		84.58333
corrélation r		3325
Régression y		0.9865728
Efficient a		y=a·x+b
Efficient b		29
Elimination r2		32.66667
		0.973326

$$y = 29x + 32,7$$

en 2030:

$$29 \times 11 + 32,7 \approx 352 \text{ adhérents.}$$

$$\textcircled{TI} \text{ Stats} \rightarrow \text{calc} \rightarrow \text{Reglin}(ax+b) L_1, L_2 \Rightarrow \begin{array}{l} a: 29 \\ b: 32,7 \end{array} \left. \vphantom{\begin{array}{l} a: 29 \\ b: 32,7 \end{array}} \right\} y = 29x + 32,7$$

$$(0; 32,7) \text{ et } (5; 5 \times 29 + 32,7 = 177,7)$$

Finalement ces 3 méthodes donnent: en 2030 : 366 ; 346 ; 352 adhérents.

b) Ajustements exponentiel:

x_i	1	2	3	4	5	6
y_i	70	90	115	140	170	220
$z_i = \ln(y_i)$	4,248	4,5	4,745	4,942	5,136	5,394

$$\begin{cases} z = \ln(y) \\ z = 0,224x + 4,044 \end{cases}$$

donc $\ln(y) = 0,224x + 4,044$

exp $\hookrightarrow y = e^{0,224x + 4,044}$

$$y = e^{4,044} \times e^{0,224x}$$

$$y = 57,1 \times e^{0,224x}$$

en 2030: $57,1 \times e^{0,224 \times 11} \approx 671$ adhérents.

Regression linéaire de z en x

on obtient

$$z = 0,224x + 4,044$$

Ex1:

t_i	0	1	2	3	4	5	6	7
v_i	215	140	85	57	36	29	27	22
$m_i = \ln(v_i - 15)$	5,3	4,83	4,25	3,74	3,04	2,64	2,48	1,95

(TI) $\ln(L_2 - 15) \xrightarrow{\text{sto}} L_3$

2) 3) $m = -0,485t + 5,224$

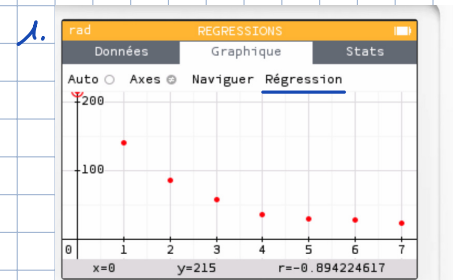
4) on a $\begin{cases} m = \ln(v - 15) \\ m = -0,485t + 5,224 \end{cases}$

exp $\hookrightarrow v - 15 = e^{-0,485t + 5,224}$

$v - 15 = e^{5,224} \times e^{-0,485t}$

$v = e^{5,224} \times e^{-0,485t} + 15$

$v = 185,7 \times e^{-0,485t} + 15$

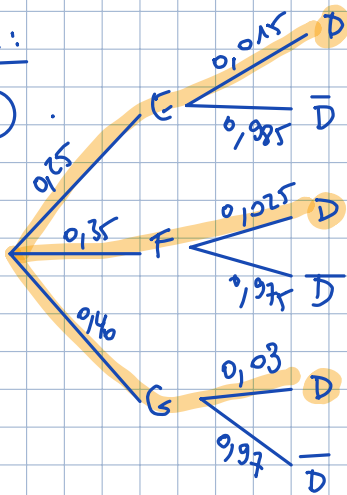


le nuage ne ressemble pas à une droite.

donc $\alpha = 185,7$ $\beta = -0,485$ $\gamma = 15$

Ex 2:

(A)



$$P(D) = 0,25 \times 0,015 + 0,35 \times 0,025 + 0,4 \times 0,03$$

$$P(D) = 0,0245$$

(B)

1) $y = 0,292x + 0,775$

2) $0,292 \times 11 + 0,775 = 3,937 \approx 3,9\% - \hat{=} 0,1\% \text{ pr\%}$

