ⵜⴰⴳⵍⴷⵉⵜ ⵏ ⵍⵎⵖⵔⵉⴱ
ⵜⴰⵎⴰⵡⴰⵙⵜ ⵏ ⵓⵙⴳⵎⵉ ⴰⵏⴰⵎⵓⵙ
ⴷ ⵓⵙⵙⵍⵎⴷ ⴰⵎⵣⵡⴰⵔⵓ ⴷ ⵜⵓⵏⵏⵓⵏⵜ

المملكة المغربية
وزارة التربية الوطنية
والتعليم الأولي والرياضة

المركز الجهوي لمهن التربية والتكوين لجهة الدار البيضاء سطات

Centre Régional des Métiers de l'Education et de la Formation de la Région Casablanca-Settat

STATISTIQUE DESCRIPTIVE

PROFESSEUR:
B.BENYOUNES

CHAPITRE I

INTRODUCTION

PLAN

1-Objectifs du module

2-Définitions de la statistique

3-La statistique descriptive et la statistique inférentielle

4-Domaines d'applications de la statistique

5-Histoire de la statistique

6-La démarche statistique

7-importance de la statistique dans le domaine éducatif

8-Choix de l'échantillon

9-Mener une enquête ou une recherche

I-1-Objectifs du module

- Connaitre les notions de base en statistique descriptive.
- Savoir analyser et interpréter les résultats statistiques.
- Pouvoir mener une enquête ou une recherche.

I-2-Définitions de la statistique

- La statistique est une branche des mathématiques appliquées dont les principes découlent de la théorie des probabilités et qui a pour objet le groupement méthodique ainsi que l'étude des séries de faits ou de données numériques. (Larousse)
- La statistique est la science qui a pour objet de recueillir un ensemble de données numériques relatives à tel ou tel phénomène faisant intervenir des variables déterministes et aléatoires et d'exploiter rationnellement ces données pour établir toutes relations de causalité par l'analyse et l'interprétation. (Paul pacè)

I-2-Définitions de la statistique (suite)

Remarque:

Il faut distinguer entre « la statistique » et « les statistiques » qui désigne familièrement l’ensemble des données chiffrés.

I-3-La statistique descriptive et la statistique inférentielle

- La statistique descriptive est la branche des statistiques qui regroupent les nombreuses techniques utilisées pour décrire un ensemble relativement important de données.
- La statistique inférentielle est un ensemble de méthodes permettant de tirer des conclusions fiables à partir de données d’echantillons statistiques

I-4-Domaines d’applications de la statistique

- Démographie (recensement).
- Economie.
- Sociologie.
- Politique.
- Médecine.
- Physique.
- Climatologie.
- Ecologie.

…………………………………………

I-5-Histoire de la statistique

Dans l'évolution de la statistique au cours de l'histoire on peut distinguer trois étapes:

- Dans cette étape qui va de la plus haute antiquité jusqu'au 18 ième siècle, la statistique se réduit à des recensements et inventaires de caractère démographique ou comptable et reste une statistique de constatation.
- Ce n'est qu'au 18 ième siècle que l'on voit apparaitre le rôle prévisionnel des statistiques et c'est à Adolphe Quételet que l'on doit l'idée que la statistique est une science s'appuyant sur les probabilités.

I-5-Histoire de la statistique (suite)

- Sous l’impulsion de grands mathématiciens , la troisième étape est révolutionnaire car elle va permettre désormais d’ouvrir à la statistique des horizons illimités, surtout avec l’essor de l’informatique qui va permettre de traiter et d’analyser un plus grand nombre de données .

I-6-La démarche statistique

La démarche statistique pour une enquête, un sondage, une étude ou un recensement se fait en général en quatre étapes.

- Phase préparatoire:

on détermine les différents caractères que l'on va étudier et sur quel population portera l'étude. Se pose alors le problème de l'échantillon.

- Recueil des données:

La récolte des données peut se faire par des tests, des QCM, des mesures etc.

I-6-La démarche statistique (suite)

- Traitement des données:

  Les résultats d'une étude statistique forment en général une immense série de données brutes difficile à manipuler. Il est don nécessaire de les ranger, de les représenter graphiquement, de calculer certains paramètres significatifs.

- Interprétation et analyse des données:

  Les résultats obtenus lors d'une étude statistique ne constituent en fait qu'un échantillon d'une réalité souvent bien plus vaste.

I-6-La démarche statistique (suite)

Se basant ensuite sur des calculs de probabilités, on essaie de voir si les résultats observés peuvent se généraliser à la population tout entière.

Remarque:

L'ensemble des trois premières étapes entrent dans le cadre de la statistique descriptive. La quatrième intervient dans le cadre de la statistique inférentielle

I-7-importance de la statistique dans le domaine éducatif

Les procédés, méthodes ou techniques statistiques sont nécessaires pour mener n'importe quel étude dans le domaine éducatif, en vue de prendre des décisions objectives. En effet, il permettent au **professeur** dans sa classe ,en tant qu'**évaluateur**, de mieux analyser les instruments de mesure du rendements de ses élèves et de mieux juger ce rendement.(*par exemple, classification des élèves, situation d'un élève dans son groupe, mesure de difficulté d'apprentissage ….).*

I-7-importance de la statistique dans le domaine éducatif (suite)

Donc les statistiques permettent au **professeur** ,en tant que **chercheur** d'analyser les données d'une situation expérimentale, de vérifier ses hypothèses de recherche, d'interpréter ses résultats et de pouvoir les communiquer dans un langage simple et compréhensible, en vue de prendre des décisions objectives et efficaces.

I-8-Choix de l’échantillon

La fiabilité des résultats d’une étude statistique dépend étroitement de l’échantillon choisi. Donc il faut que cet échantillon soit un modèle réduit de la population totale, d'ou l’importance d’une telle décision. Si on connait suffisamment la structure de la population, on peut créer un échantillon représentatif de la population qu’un échantillon pris au hasard.

I-9-Mener une enquête ou une recherche

Pour mener une enquête, généralement on suit les étapes suivantes:

- Choix du thème (justifié)
- Choix de la population étudiée
- Etude bibliographique sur le thème
- Poser la problématique
- Soulever des hypothèses
- Elaborer des QCM, des test, des contrôles…..(visant à confirmer ou infirmer les hypothèses)
- Traitement des données
- Conclusion et suggestions

CHAPITRE II

LES ENQUETES

PLAN

1-Qu’est ce qu’une enquête?

2-sondage et échantillon

3-Les sondages aléatoires simples (SAS)

4-Les sondages stratifiés

5-Sondage à choix raisonnés

6-Autres méthodes empiriques de sondage

7-Le questionnaire

8-Construction du questionnaire

9-Le formatage des questions

II-1-Qu'est ce qu'une enquête?

L'enquête est l'opération technique qui permet la collecte des données sur un ensemble de départ qu'on appelle « population ».

Remarque:

Une enquête demande des moyens importants en temps, en préparation méthodique ,travail de groupe, et en financement.

II-2-sondage et échantillon

En général la plupart des enquêtes que l'on produit sont des enquêtes partielles que l'on appelle aussi

« sondages » et qui consistent à n'interroger qu'une partie de la population mère, la plus représentative possible, appelé « échantillon ».

II-3-Les sondages aléatoires simples (SAS)

On tire de la population mère, de façon aléatoire un échantillon sur lequel portera l'enquête. Evidement, si la population est parfaitement homogène, et parfaitement connue, un tirage de ce genre pourrait être satisfaisant. Dans ce cas la on parle de SAS. Cependant, dans la plupart des cas les choses ne sont pas aussi homogènes que l'on espère.

II-4-Les sondages stratifiés

Ces sondages consistent à découper la population mère en sous-populations plus homogènes, que l'on appelle des «strates »,et sur chacune d'elles sera réaliser un sondage.

II-5-Sondage à choix raisonnés

Il s'agit ici de procédures d'échantillonnage empiriques et non plus probabilistes. Dans ce cas la le principe est de respecter les proportions des principales caractéristiques de la population, pour obtenir une sorte de maquette simplifiée de la réalité. Cette méthode s'appelle aussi « méthode des quotas ».

II-6-Autres méthodes empiriques de sondage

- Le sondage spontané: consiste à lancer un questionnaire auprès d'abonnés d'un journal, ou de clients d'une banque, ou à envoyer par internet, ou même par téléphone, un questionnaire,
- Le sondage sélectif: consiste à interroger par exemple la 5éme personne, puis la 15éme, puis

  la 25éme et ainsi de suite.
- La méthode du panel: utilisée surtout dans les domaines de consommation. Elle consiste à fidéliser un groupe de consommateur pour qu'il répondent à des questionnaires de façon périodique.

II-6-Autres méthodes empiriques de sondage(suite)

Pour conclure, on voit bien qu'il ya place à des champs infinis d'imagination. La recherche de l'homogénéité est l'objectif principale des combinaisons de toutes les méthodes entre elles. Aussi la simple compréhension des méthodes permet de faire fonctionner son imagination pour fabriquer un plan de sondage adapté à une investigation qu'on veut faire sur un problème de terrain.

II-7-Le questionnaire

Le questionnaire est le « support » qui permet à l’enquêteur de passer son enquête .Par ailleurs, il est soit passé en « face à face », soit par téléphone, soit envoyé par poste ou par internet, soit encore déposé à l’entrée d’un office ou en tout endroit stratégique.

II-8-Construction du questionnaire

- Bien préciser l’objet de l’enquête.
- Bien étudier l’ordre des questions.
- Bien choisir les mots et les expressions de façon à ce que tout le monde comprenne la même chose en entendant le même mot.
- Etre court et précis.
- Eviter de poser les doubles questions.
- Mettre le questionnaire à une «pré-enquête».

Cette dernière étape permet de rectifier, reformuler, corriger le questionnaire; et ceci à fin d’optimiser les chances de réussite du questionnaire.

II-9-Le formatage des questions

Dans un questionnaires, les questions peuvent être:

- <u>Fermées</u>.

  <u>exemple</u>: « avez-vous des enfants? »

  oui ☐ non ☐
- <u>Ouvertes</u>.

  <u>exemple</u>:

  « quelles sont vos attentes de cette formation? »
- <u>Semi-ouvertes</u>.

  <u>exemple</u>: « êtes-vous satisfait du service? Argumentez en quelques mots »

  oui ☐ non ☐

II-9-Le formatage des questions (suite)

Remarque:

D’autres types de formatage de questions peuvent exister, parmi ceux-ci existent les formatages de type « échelles ».

exemple: « les maths sont inutiles »

Je n’y crois pas	**0**	**1**	**2**	**3**	**J’en suis totalement persuadé**

Exemple de questionnaire

	Très satisfaisant	satisfaisant	Peu satisfaisant	Très insatisfaisant
1) Qualité du service				
2) Temps de service				
3) Qualité des ingrédients				
4) Qualité de la préparation				
5) Variété du choix des plats				
6) Prix du menu ou de la carte				
7) Qualité des boissons				
8) Propreté de la vaisselle				

CHAPITRE III

VOCABULAIRE STATISTIQUE

PLAN

1-Vocabulaire statistique

2-Modalités-effectifs-fréquences

3-Tableaux statistiques

4-effectif cumulé ↑ -effectif cumulé ↓

5-fréquence cumulée ↑ -fréquence cumulée ↓

III-1-Vocabulaire statistique

1) Population statistique:

« C'est le groupe sur qui portera l'étude statistique »

2) Individu:

« un élément de la population statistique »

3) Echantillon:

«un sous ensemble de la population statistique »

4) Caractère statistique ou variable statistique:

«la propriété sur laquelle porte l'étude statistique »

III-1-Vocabulaire statistique (suite)

On distingue deux types de caractères:

- Caractère qualitatif:

  « C'est un caractère qu'on ne peut pas mesurer »

- Caractère quantitatif:

  « C'est un caractère qu'on peut exprimer numériquement »

III-1-Vocabulaire statistique (suite)

Le caractère quantitatif peut être:

- discret s'il ne prend qu'un nombre fini de valeurs isolées.
- Continue s'il prend toute les valeurs possibles à l'intérieur d'un intervalle.

III-2-Modalités-effectifs-fréquences

- Les valeurs que peut prendre le caractère statistique sont appelées « modalités » qu'on peut noter x_i
- A chaque modalité x_i est associée un effectif n_i qui désigne le nombre d'individus correspondant à cette modalité .
- $f_i = n_i / N$ est la fréquence de la modalité x_i, elle est souvent calculée en pourcentage .

L’ensemble des couples (x_{i}, n_{i}) est appelé série statistique à une variable. on peut s’intéresser à d’autre caractère Y dont les modalités sont y_i . L’ensembles des triplets (x_i ,y_i ,n_{ij}) est appelé série statistique à deux variables.

De même on peut définir une série statistique à n variables.

III-3-Tableaux statistiques

Dans l’étude statistique ,on commence par ranger les données dans un tableau :

caractère	X1	X2			Xp-1	Xp
effectif	n1	n2			np-1	np
fréquence	f1	f2			fp-1	fp

III-4-effectif cumulé ↑ -effectif cumulé ↓

l'effectif cumulé croissant est:

$$N_i \uparrow = \sum_{j=1}^{i} n_j$$

l'effectif cumulé décroissant est:

$$N_i \downarrow = \sum_{j=i}^{p} n_j$$

III-5-fréquence cumulée ↑ -fréquence cumulée ↓

- La fréquence cumulée croissante est:

$$F_i \uparrow = \frac{N_i \uparrow}{N}$$

- La fréquence cumulée décroissante est:

$$F_i \downarrow = \frac{N_i \downarrow}{N}$$

Exemple

A partir d'une enquête portant sur le nombre d'enfants d'un échantillon de familles, on obtient les résultats suivants:

Nbre d'enfants	Effectif	Effectif cum crs	Effectif cum décrs	Fréq	Fréq cum crs	Fréq cum décrs
0	18	18	200	0,09	0,09	1
1	32	50	182	0,16	0,25	0,91
2	66	116	150	0,33	0,58	0,75
3	41	157	84	0,205	0,785	0,42
4	32	189	43	0,16	0,945	0,215
5	9	198	11	0,045	0,99	0,055
6	2	200	2	0,01	1	0,01

Les différentes situations en statistique

- Situation de répartition: lorsque les individus d'une population statistique sont rangés, on parle de situation de répartition.
- Situation de comparaison: lorsqu'on étudie le même caractère sur plusieurs échantillons d'une même population.
- Situation d'évolution ou série chronologique: elle dépend du temps.

CHAPITRE IV

REPRESENTATIONS GRAPHIQUES

Diagrammes statistiques

Le rôle des diagrammes statistiques et des représentations graphiques est de faire passer une information grâce a un effet visuel.

Ils en existent plusieurs types; citons:

- Le diagramme en bâtons.
- Le diagramme en tuyau d'orgues
- Le diagramme en barres
- L'histogramme
- Le diagramme à secteurs angulaires
- Le diagramme en toile d'araignée

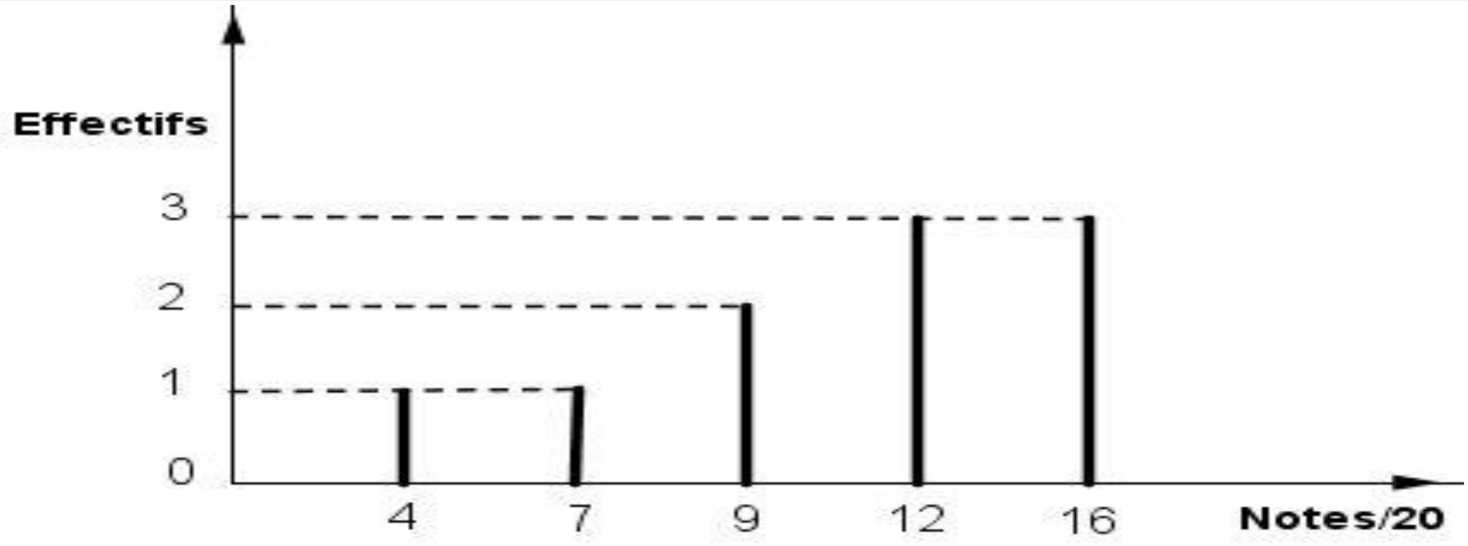
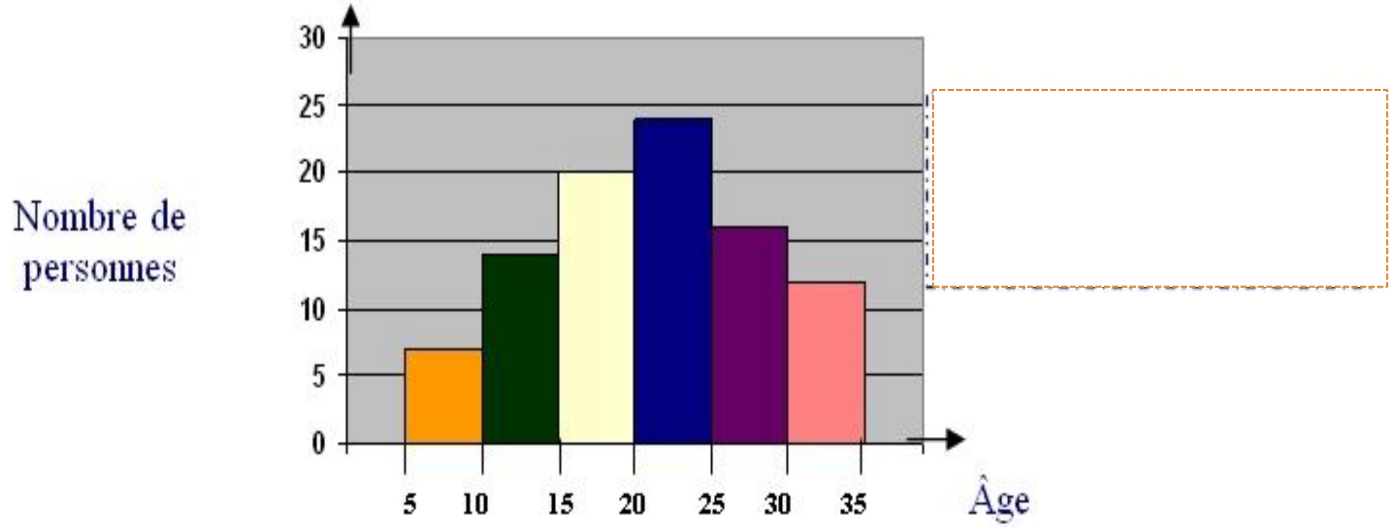


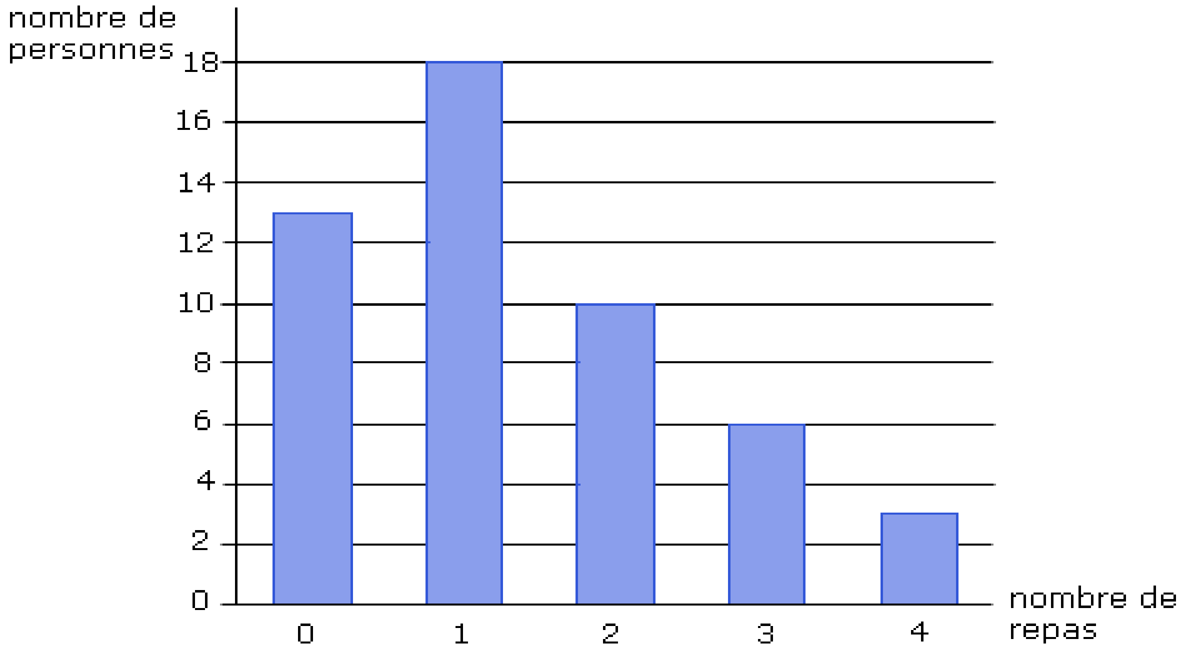
diagramme en bâton

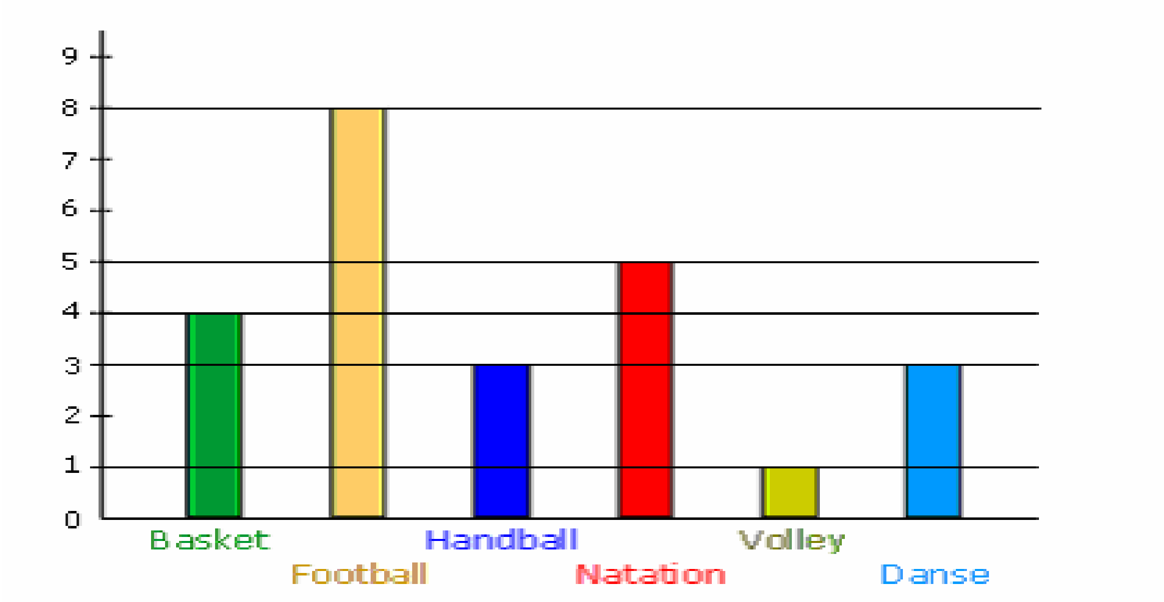
Nombre de personnes assistant à un concert selon leur âge

L'histogramme



Le diagramme en barres





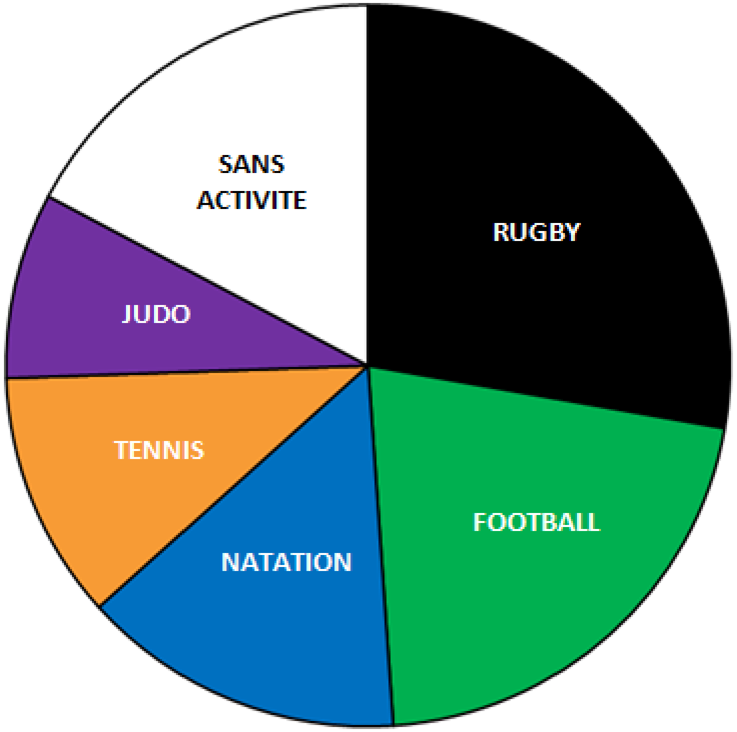
Le diagramme en tuyau d'orgues

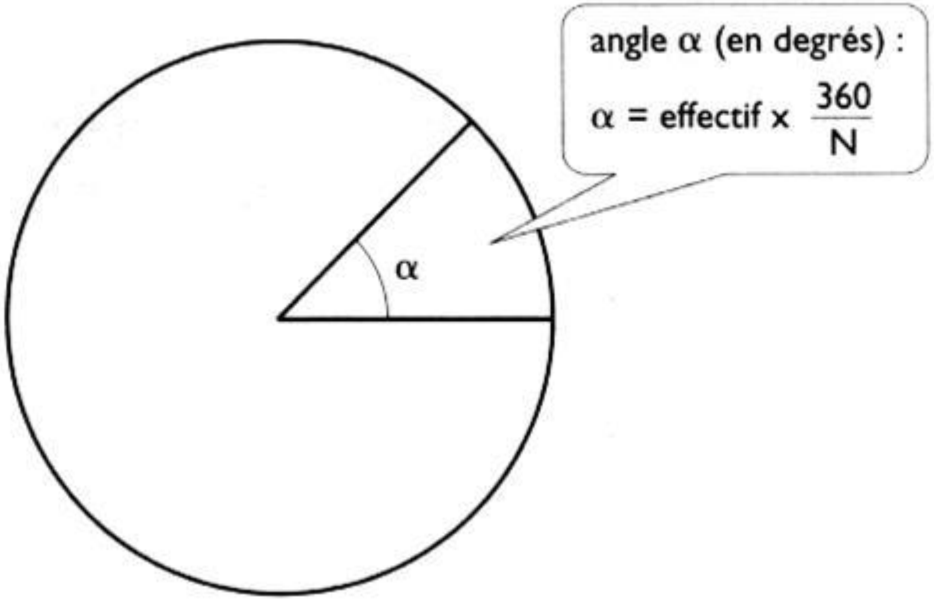
Impact du 11 septembre 2001 sur le tourisme dans le monde

Dans le graphique en tuyaux d'orgue, on laisse un espace entre chaque série composée de 2 ou de plusieurs items.

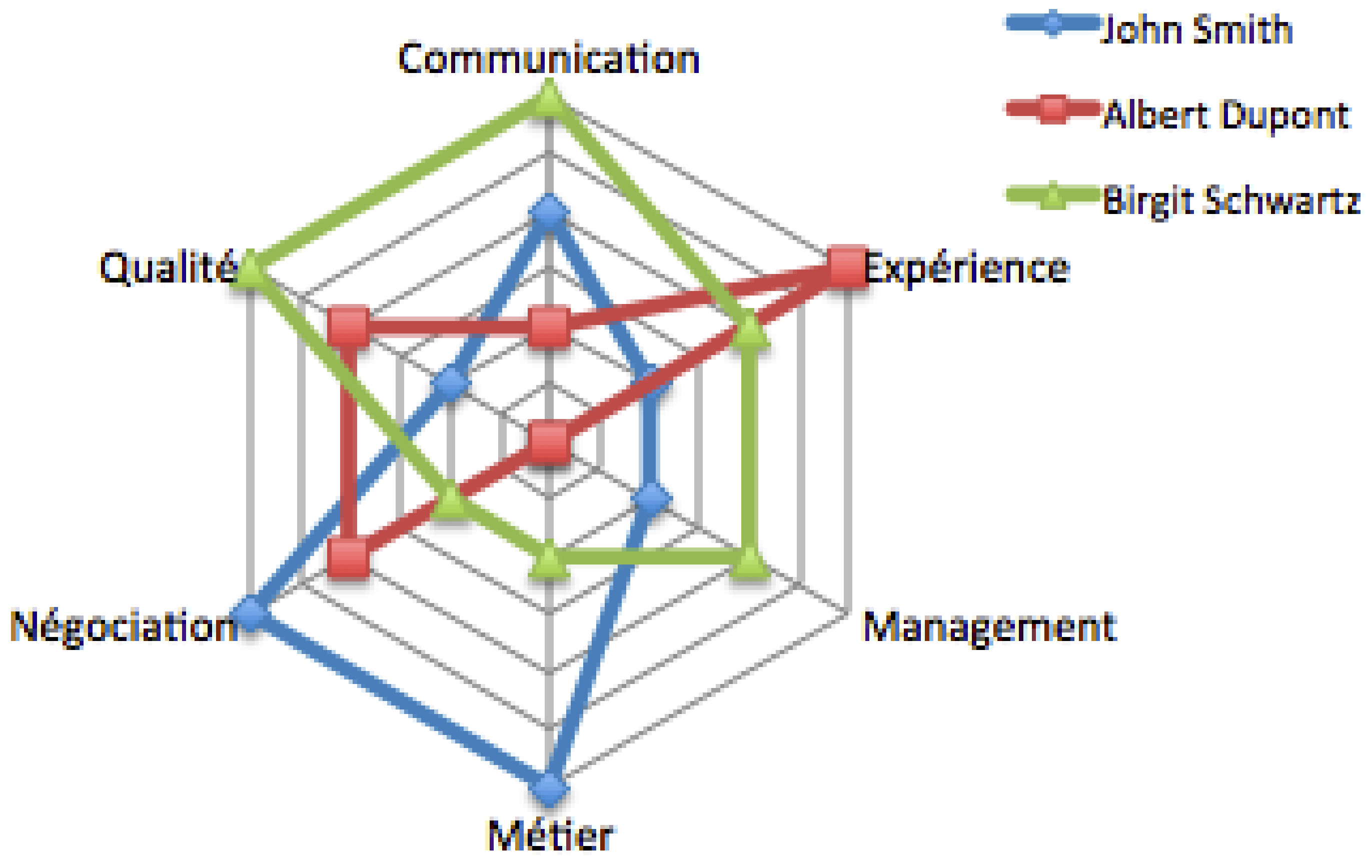
<u>Usage</u> : Pour comparer la croissance individuelle de deux ou plusieurs séries.

Le diagramme à secteurs angulaires





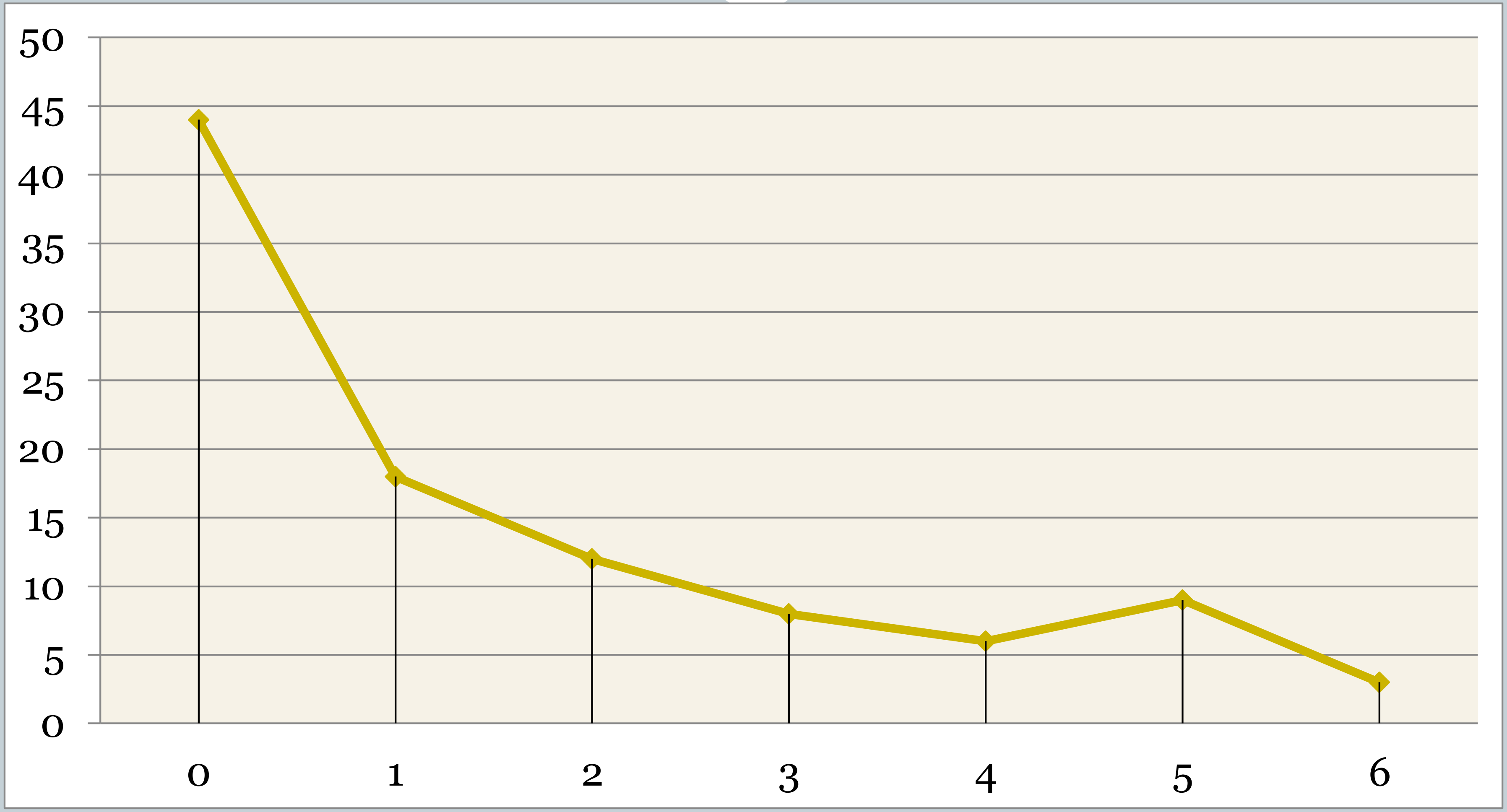
Le diagramme en toile d'araignée , radar, polaire ou Kiviat,



DANS UNE ENTREPRISE, LE CONTRÔLEUR DE FABRICATION A COMPTÉ LE NOMBRE DE PIÈCES DÉFECTUEUSES DANS LA PRODUCTION JOURNALIÈRE LES OBSERVATIONS FAITES SUR 100 JOURS ONT DONNÉ LES RESULTATS:

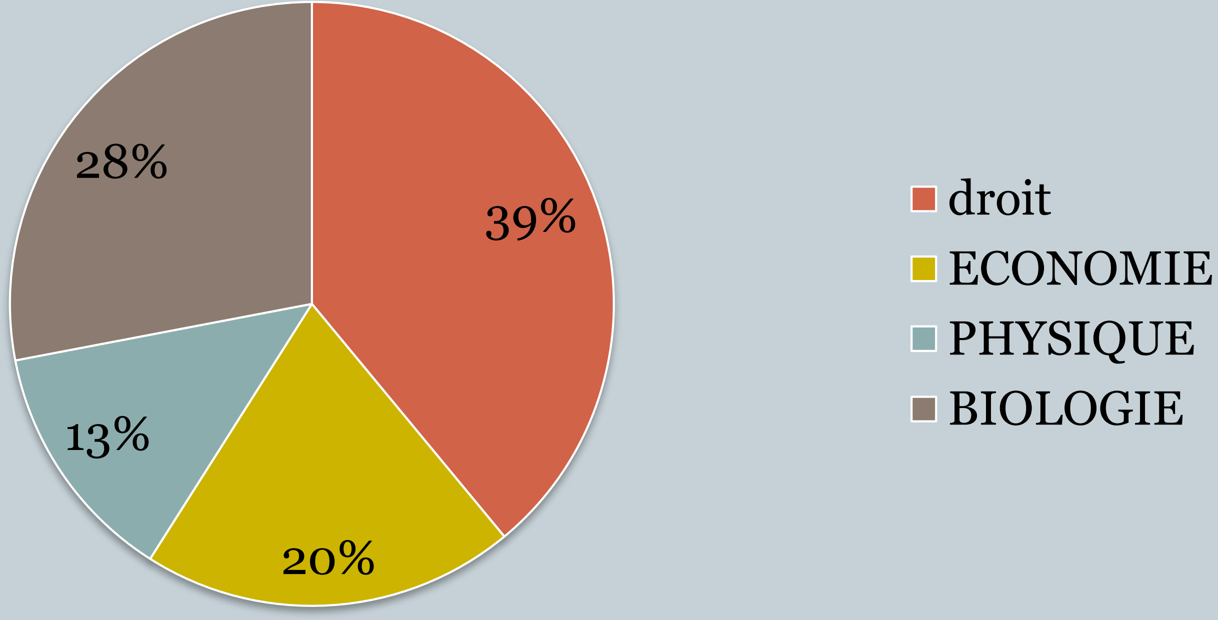
NOMBRE DE PIECES DEFECTUEUSES	NOMBRE DE JOURS
0	44
1	18
2	12
3	8
4	6
5	9
6	3

Diagramme des effectifs



Diagrammes a secteurs circulaires

- Dans l’université XX, on compte 39% d’étudiants en droit, 20% en économie, 13% en physique et 28% en biologie.
- Représenter cette série statistique par un diagramme à secteurs circulaires.



L’histogramme

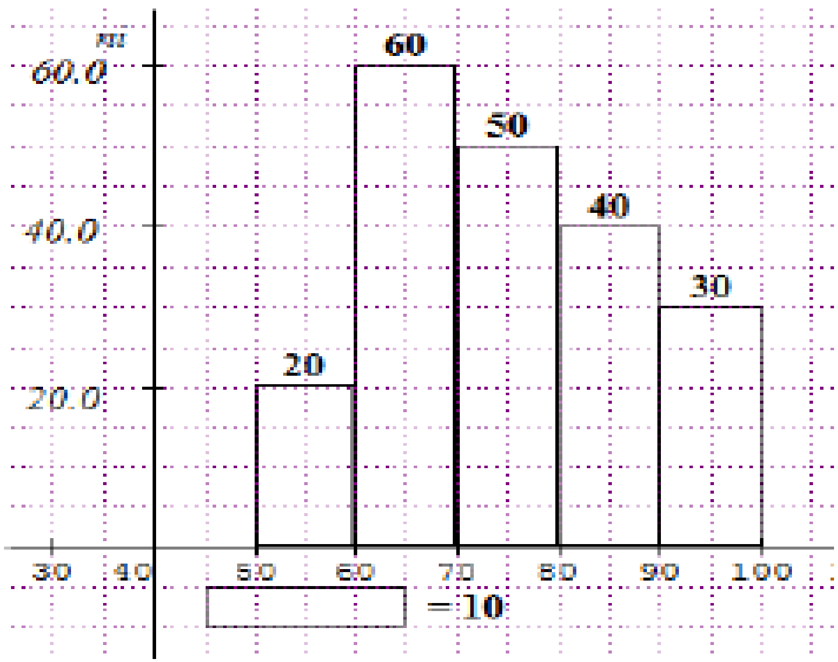
L’histogramme donne une image pour laquelle, on repère la surface des rectangles, et non leurs hauteurs, comme pour les bâtons ou les tuyaux d’orgues.

- **Cas ou l'amplitude des classes est la même**

Soit le tableau suivant :

classes	Effectifs n_i
[50- 60[	20
[60- 70[	**60**
[70- 80[	50
[80- 90[	40
[90- 100[	30

Tab.1 : données statistiques réparties par classes d'amplitudes égales



histogramme pour des classes d'amplitudes égales

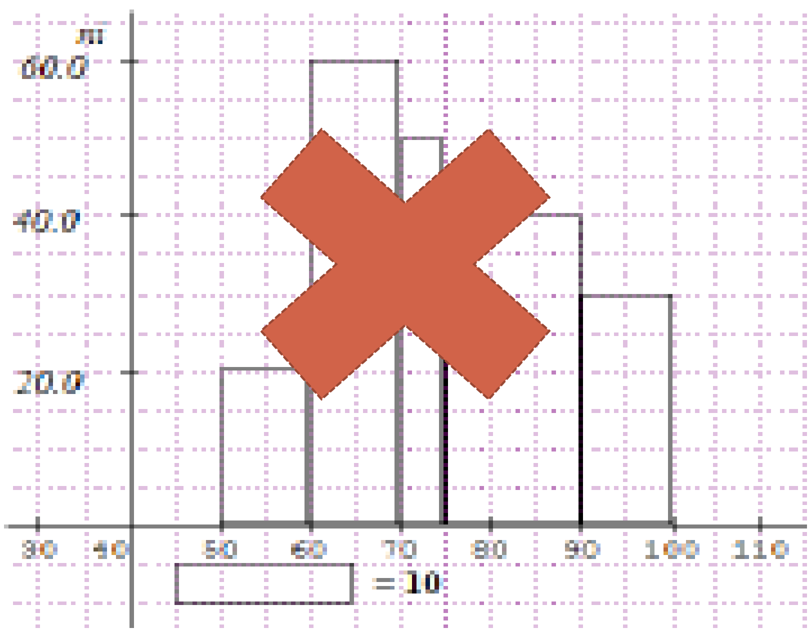
classes d'amplitudes différentes

Une correction d'effectifs s'impose pour que l'air reste proportionnelle à l'effectif total .En effet si on choisit l'intervalle qui a la plus petite amplitude comme unité et si un intervalle (classe) d'effectif n_i et k fois plus grand que l'intervalle unité, la correction de cet intervalle sera n_i/k.

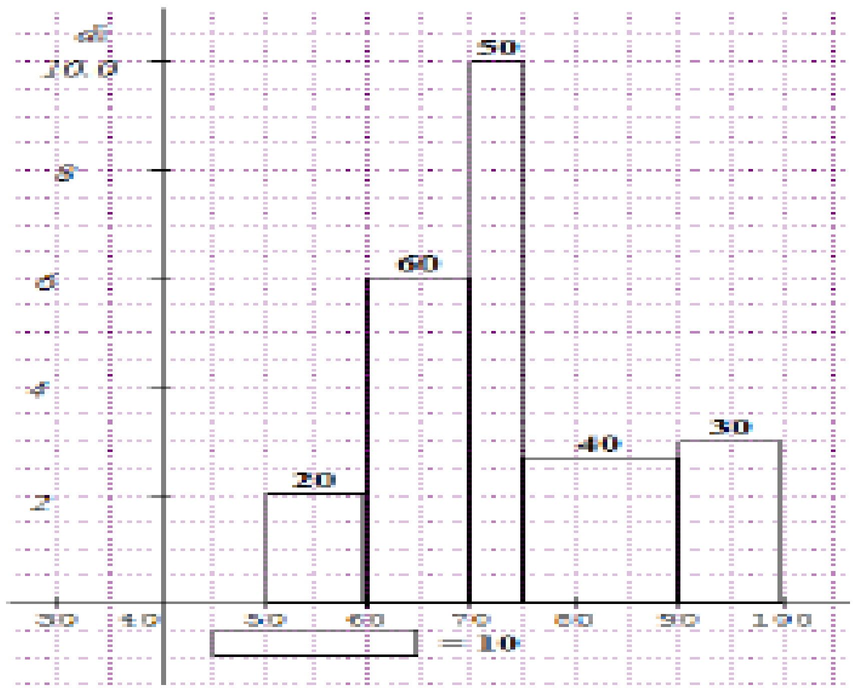
- **Cas ou l'amplitude des classes est différente**

classes	Effectifs n_i	densité d_i
[50- 60[	20	2
[60- 70[	60	6
[70- 75[	**50**	**10**
[75- 90[	**40**	**2.67**
[90- 100[	30	3

Tab.2 : données statistiques réparties par classes d'amplitudes différen



graphe a



graphe b

Exemples

Remplissez le tableau suivant, puis représentez cette série par un histogramme.

Classe Ci	**[300,400[**	**[400,500[**	**[500,600[**	**[600,700[**	**[700,800[**
Effectifs	25	90	145	90	50

Le tableau suivant donne la répartition des âges de 110 personnes

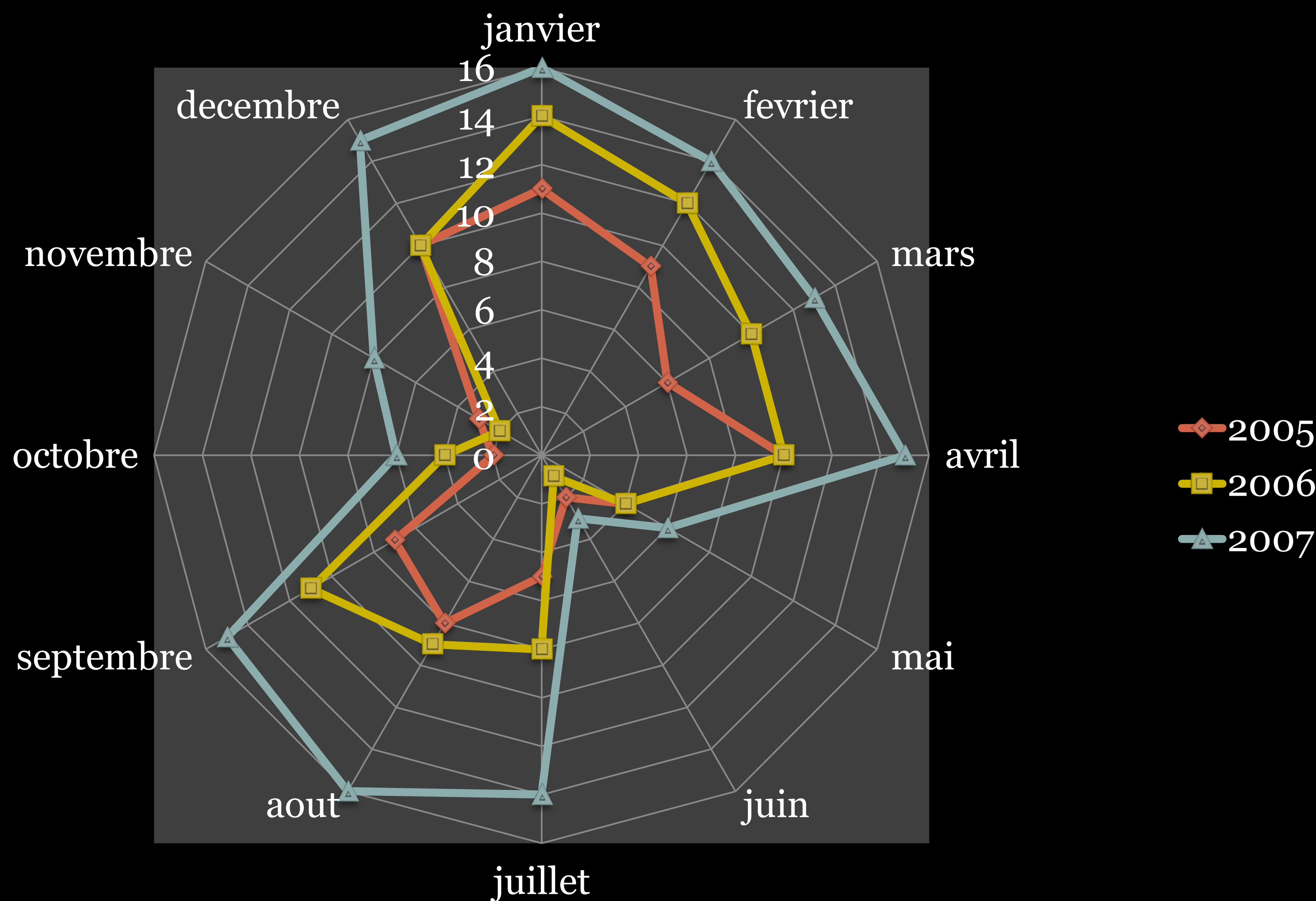
Représentez cette série statistique par un histogramme.

âges	effectifs
[15,20[	10
[20,25[	30
[25,30[	40
[30,35[	20
[35,45[	10

Diagramme en toile d'araignée

Chiffres d'affaires du commerce X (en milliers d'euros)

	2005	2006	2007
Janvier	11	14	16
Février	9	12	14
Mars	6	10	13
Avril	10	10	15
Mai	4	4	6
Juin	2	1	3
Juillet	5	8	14
Aout	8	9	16
Septembre	7	11	15
Octobre	2	4	6
Novembre	3	2	8
Décembre	10	10	15


janvier
fevrier
mars
avril
mai
juin
juillet
aout
septembre
octobre
novembre
decembre
16
14
12
10
8
6
4
2
0
2005
2006
2007

CHAPITRE V

PARAMETRES DE POSITION ET DE DISPERTION

V-1 PARAMETRES DE POSITION

Le rôle des paramètres de position ou de dispersion est de transmettre, par un calcul, une information liée à une réalité statistique .

V-1-1 LE MODE

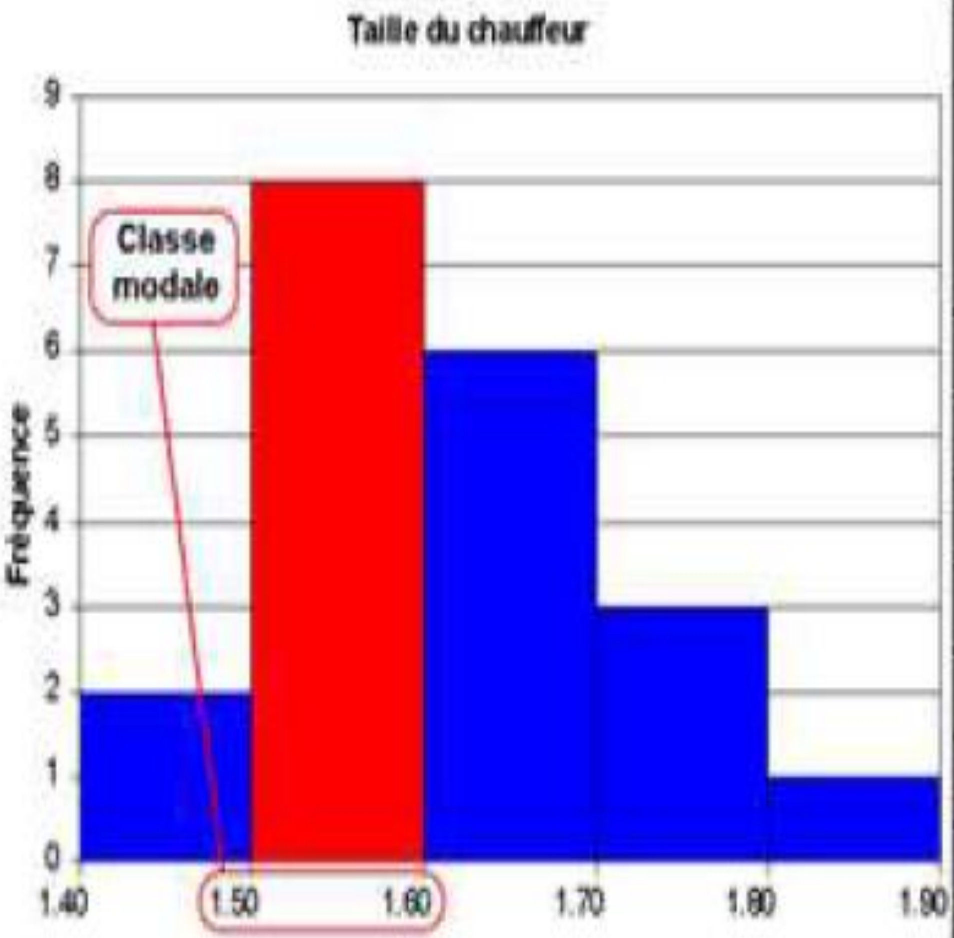
Définition:

Le mode (classe modale) est la valeur (classe) ayant le plus grand effectif.

Remarques:

- Le mode n'est pas unique.
- Le mode peut aussi être attribuer à une série statistique à caractère qualitatif.

Classe modale d'une variable quantitative continue



Taille du conducteur (m)		
Valeurs de la variable x_i	Fréquence absolue n_i	Fréquence relative f_i
[1.40 , 1.50 [	2	10 %
[1.50 , 1.60 [	8	40 %
[1.60 , 1.70 [	6	30 %
[1.70 , 1.80 [	3	15 %
[1.80 , 1.90 [	1	5 %
Nbre observ.	20	100 %

les étapes pour bien calculer le mode

1- Calculer la densité des classes $d_i = n_i/a_i$

2- Trouver la classe la plus dense, c'est la classe où se trouve le mode

3- Appliquer la formule suivante :

$$Mo = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} a_i$$

l_i : Limite inférieure de la classe modale

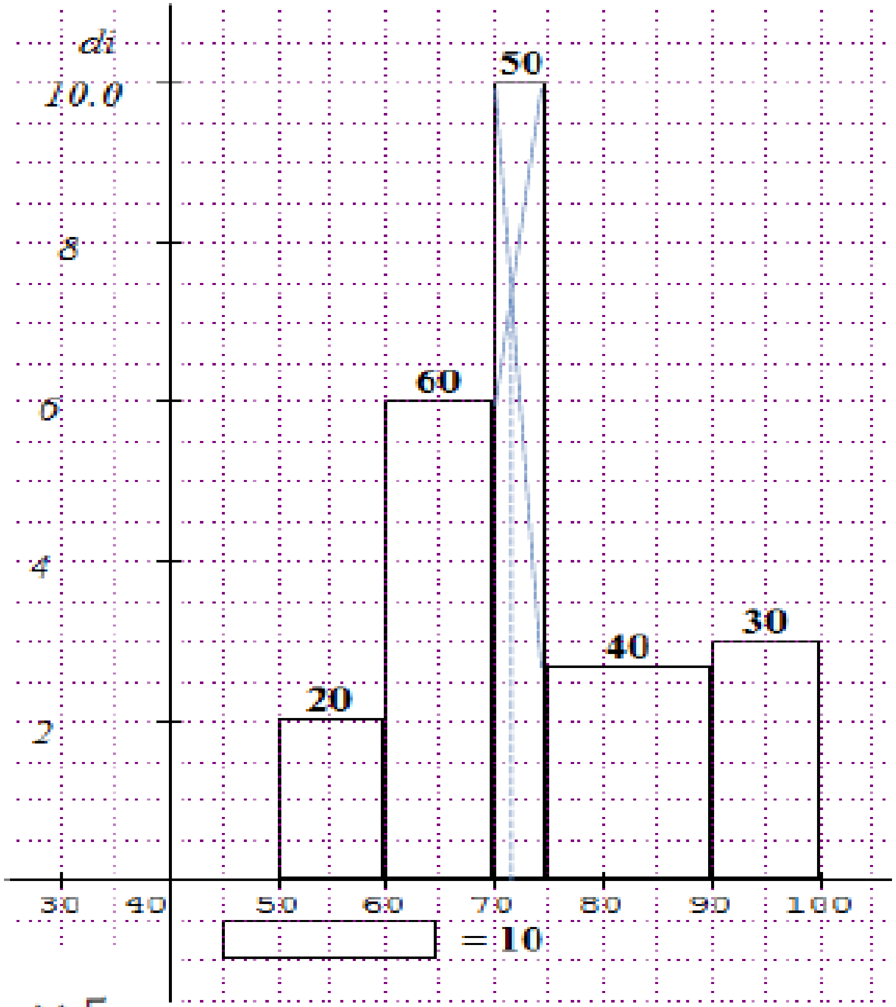
Δ_1 : La différence en densité entre la classe modele est la classe avant

Δ_2 : La différence en densité entre la classe modale est la classe suivante

a_i : L'amplitude de la classe modale

- classe d'amplitudes différentes

classes	Effectifs n_i	densité d_i
[50- 60[	20	2
[60- 70[	60	6
[70- 75[	50	10
[75- 90[	40	2.67
[90- 100[	30	3



$$Mo = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} a_i$$

$$Mo = 70 + \frac{(10-6)}{(10-6)+(10-2.67)} \times 5$$

$$= 70 + \frac{4}{4+7.33} \times 5 = 71.76$$

V-1-2 LA MEDIANE

Définition:

La médiane qu'on note Me est la valeur de la variable statistique, telle que l'effectif des valeurs inférieures soit égal à l'effectif des valeurs supérieures.

Remarque:

- La médiane est la valeur à partir de laquelle la fréquence cumulée croissante dépasse 50%.

CALCUL DE LA MEDIANE

Pour calculer la médiane on ordonne la distribution par ordre croissant.

<u>Cas discret</u>:

- Si le nombre total des valeurs du caractère est paire, soit 2n, la médiane est comprise entre le terme de rang n est celui de rang n+1.On pourra prendre Me comme demi-somme des deux.
- Si le nombre total des valeurs du caractère est impaire, soit 2n+1, la médiane Me est le terme de rang n+1.
- C'est aussi la plus petite valeur du caractère dont l'effectif cumulé croissant est supérieur ou égal à N/2.

EXEMPLES

- Soit la série statistique suivante (effectif total pair):

6 – 6 – 7 – 7 – 7 - 7 – 8 – 9 – 9 – 10 – 10 – 11 – 12 –
12 – 13 – 13 – 14 – 14 – 14 – 15 – 16 – 16 – 17 – 18

$$N = 24 \Rightarrow \frac{N}{2} = 12 \Rightarrow M_e = \frac{11+12}{2} = 11,5$$

- Soit la série statistique (effectif total impair) :

1 – 2 – 2 – 3 – 3 – 3 – **4** – 4 - 5 – 5 – 6 – 7 – 7

$$N = 13 \Rightarrow M_e = 4$$

CALCUL DE LA MEDIANE

Cas continue:

- La médiane Me peut être approché graphiquement comme abscisse du point d'intersection du polygone des effectifs croissants et décroissants.
- On détermine d'abord la classe médiane $\left[a_{k-1}; a_k\right[$ tel que : $N_{k-1} \leq \frac{N}{2} < N_k$ puis en utilisant une interpolation affine, on calcule Me par la formule: $$\frac{Me - a_{k-1}}{\frac{N}{2} - N_{k-1}} = \frac{a_k - a_{k-1}}{N_k - N_{k-1}}$$

$$\frac{M_e - a_{k-1}}{a_k - a_{k-1}} = \frac{\frac{N}{2} - N_{k-1}}{N_k - N_{k-1}} \quad \Rightarrow \quad M_e = (a_k - a_{k-1})\frac{\frac{N}{2} - N_{k-1}}{N_k - N_{k-1}} + a_{k-1}$$

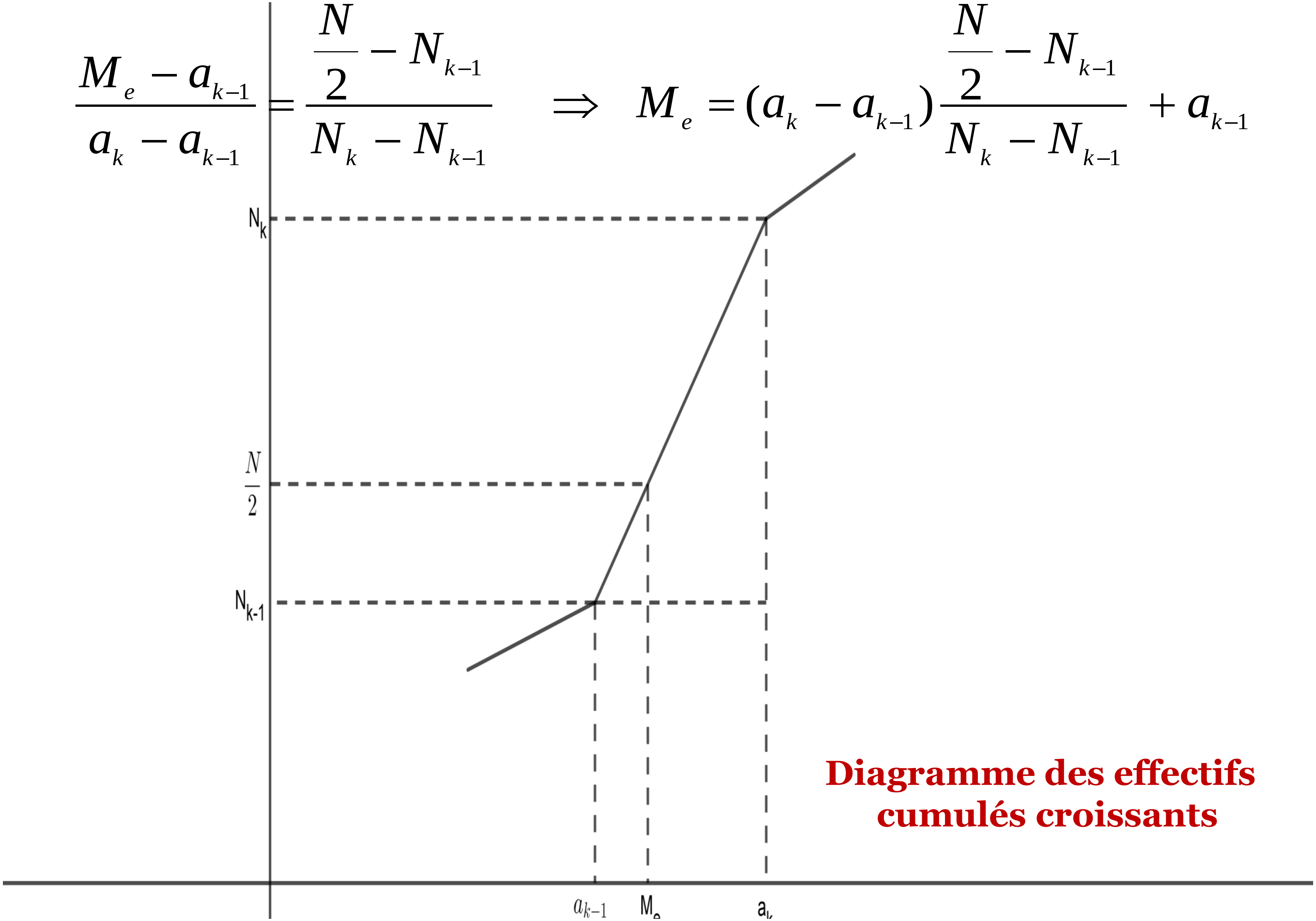


Diagramme des effectifs cumulés croissants

EXEMPLE 1

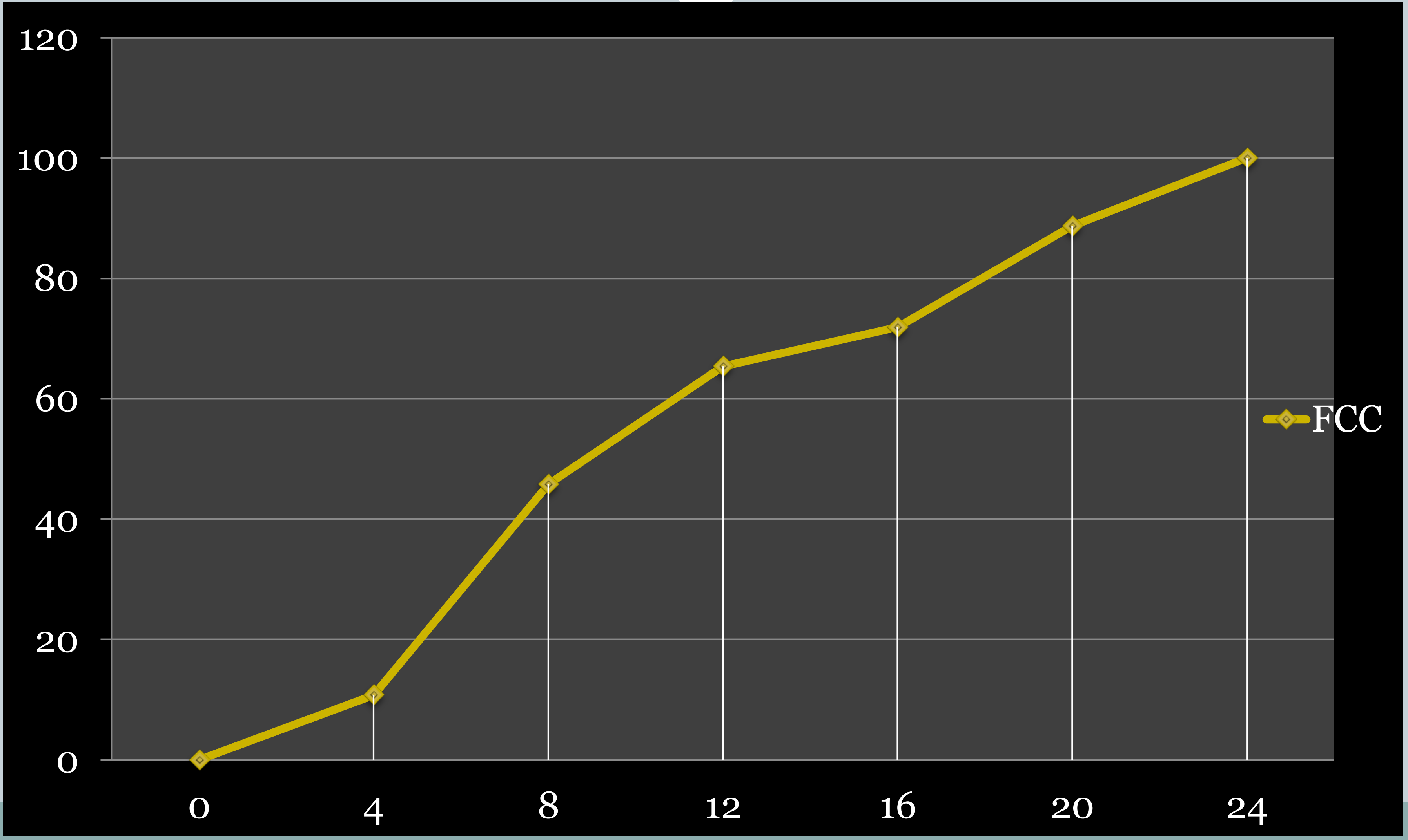
On compte le nombre de voitures à la sortie d'une ville un jour:

heure	$[0,4[$	$[4,8[$	$[8,12[$	$[12,16[$	$[20,24[$	$[20,24[$
Nbre de voitures	1100	3800	2100	700	1800	1200

On construit la courbe des fréquences cumulées croissantes :

heure	0	4	8	12	16	20	24
FCC en %	0	10,3	45,8	65,4	71,9	88,7	100

Polygone des fréquences cumulées croissantes



EXEMPLE 2

Sur le tableau ci-dessous on donne la répartition des dépenses mensuelles D'un échantillon de foyers.

On a: N=210

Dépenses en Dhs	Nombre de foyers	Effectifs cumuculés croissants
[3000,4000[	5	5
[4000,5000[	60	65
[5000,6000[	15	80
[6000,7000[	95	175
[7000,8000[	30	205
[8000,10000[	5	210

La classe médiane

$\frac{N}{2} = 105$

- Détermination de la médiane par interpolation:

$$\frac{M_e - 6000}{7000 - 6000} = \frac{105 - 80}{175 - 80} \quad \Rightarrow \quad M_e \simeq 6236 \; Dhs$$

Remarque: La médiane sépare l'histogramme en deux parties de même aire

V-1- 3 LES QUANTILES

Définition:

Les quantiles sont les valeurs du caractère qui partagent la distribution en *n* parties comprenant le même effectif égal à $\frac{1}{n}$ de l'effectif total.

Remarque:

- Si n=4 ,on parle de quartiles.
- Si n=10,on parle de déciles.
- Si n=100,on parle de centiles.
- Les quantiles servent à mesurer la symétrie d'une série statistique mais aussi la concentration du caractère étudié.

CALCUL DES QUARTILES

Cas discret:

Pour déterminer les quartiles Q_1 et Q_3 ;on range d'abord les valeurs de la série statistique dans l'ordre croissant.

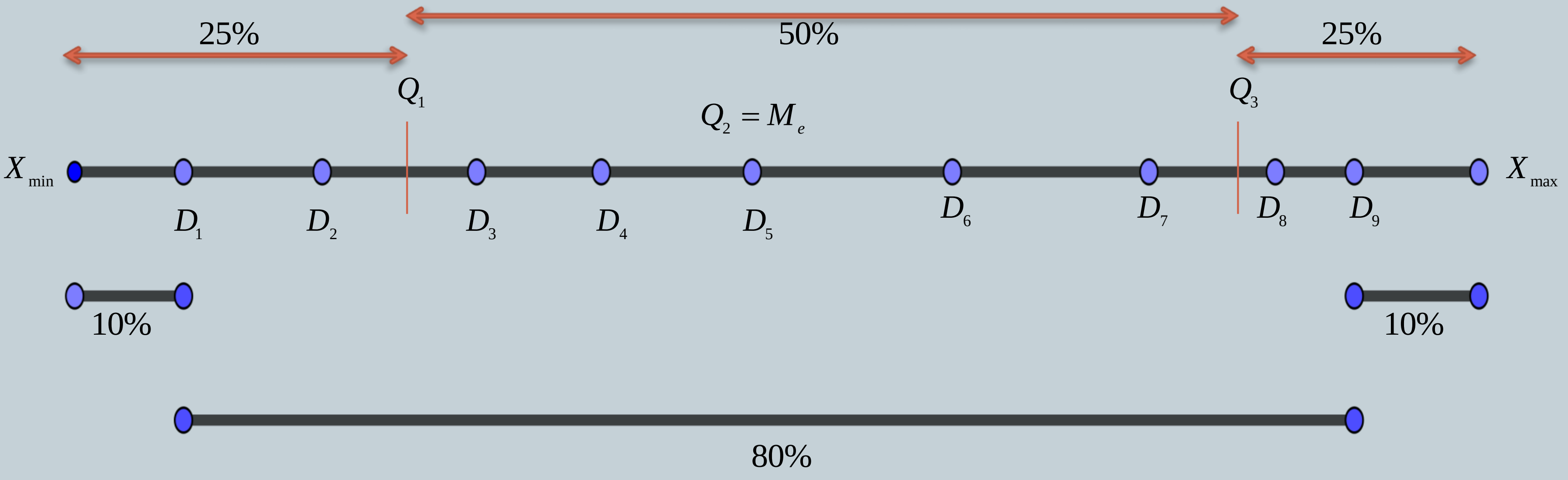
- Q_1 est la valeur classé $\frac{N}{4}$ arrondie par excès.
- Q_3 est la valeur classé $\frac{3N}{4}$ arrondie par excès.
- Q_2 =Médiane

CALCUL DES QUARTILES

<u>Cas continue</u>:

- Les quartiles Q_1 et Q_3 sont approximativement les abscisses respectifs des points du polygone des effectifs cumulés croissant qui correspondent à $\frac{N}{4}$ et à $\frac{3N}{4}$.
- On peut aussi déterminer Q_1 et Q_3 par interpolation affine comme dans le cas de la médiane.

Répartition des déciles et des quartiles



V-1- 4 LA MOYENNE ARITHMETIQUE

Définition:

La moyenne arithmétique d'une série statistique $(x_i, n_i)_{1 \le i \le p}$ est le nombre réel:

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{p} n_i x_i$$

Remarque:

- Dans le cas d'une série statistique à caractère continue $([a_i, a_{i+1}[; n_i)_{1 \le i \le p}$, la moyenne arithmétique est : $\overline{x} = \frac{1}{N} \sum_{i=1}^{p} n_i c_i$ ou $c_i = \frac{a_i + a_{i+1}}{2}$

QUELQUES PROPRIETES DE LA M.A

- La moyenne arithmétique est le paramètre le plus utilisé. Il a l'inconvénient d'être sensible à des valeurs aberrantes, ce qui le rend moins significatif dans certains cas.
- Exemple:

La moyenne d'âges de cinq personnes âgées respectivement: 22,23,24,25,26 ans est égale à 24 ans.

A ce groupe de personnes vient s'ajouter une personne âgée 57 ans. La moyenne d'âges des six personnes devient alors égale à 29,5 ans.

QUELQUES PROPRIETES DE LA M.A

Théorème 1:

Soit $(x_i, n_i)_{1\le i\le p}$ une série statistique de moyenne arithmétique $\overline{x}$ .On partage la population en deux groupes A et B d'effectifs respectifs N_A et N_B et de moyenne arithmétique $\overline{x}_A$ et $\overline{x}_B$.

alors:

$$\overline{x} = \frac{N_A\overline{x}_A + N_B\overline{x}_B}{N_A + N_B}$$

QUELQUES PROPRIETES DE LA M.A

- Exemple:

Apres avoir corrigé les copies des trois classes de terminales scientifique d'effectifs respectifs 32,35 et 39 élèves , le professeur calcule les moyennes de chaque classe et trouve:

$$\overline{x}_1 = 12,2 \quad ; \quad \overline{x}_2 = 10,7 \quad ; \quad \overline{x}_3 = 9,5$$

La moyenne de l'ensemble des élèves des trois classes égale à :

$$\overline{x} = \frac{32 \times 12,2 + 35 \times 10,7 + 39 \times 9,5}{32 + 35 + 39} = 10,71$$

QUELQUES PROPRIETES DE LA M.A

Théorème 2:

Soit la série statistique $(x_i, n_i)_{1 \leq i \leq p}$ de moyenne arithmétique $\overline{x}$. Si on considère la série statistique $(y_i, n_i)_{1 \leq i \leq p}$ tel que $y_i = a.x_i + b$ (a et b sont deux réels)

alors: $\overline{y} = a\overline{x} + b$

EXERCICE

Le professeur X corrige les copies du devoir qu'il a donné à ses élèves. Les notes obtenues sont:

6-4-7-9-11-8-5-12-3-7-1-13-5-7-9-2-8-10-10-5

1) Calculer la moyenne arithmétique.
2) Conscient de la faiblesse des résultats, le prof décide de relever les notes en les multipliant par un coefficient a puis ajoutant 1 à toute la classe à fin d'obtenir une moyenne égale à 11,65.

a) Déterminer a.
b) Déterminer les notes ainsi obtenues.

AUTRES TYPES DE MOYENNES

A-La moyenne harmonique:

Définition:

La moyenne harmonique d'une série statistique $\left(x_i, n_i\right)_{1 \le i \le p}$ est le nombre :

$$\overline{x}_H = \frac{N}{\sum_{i=1}^{p} n_i \cdot \frac{1}{x_i}}$$

AUTRES TYPES DE MOYENNES

Remarques:

- La moyenne harmonique est l'inverse de la moyenne arithmétique des inverses des valeurs de la série statistique.
- La moyenne harmonique n'est utilisée que si les inverses des données ont vraiment une signification.

AUTRES TYPES DE MOYENNES

B-La moyenne géométrique:

Définition: La moyenne géométrique d'une série statistique $\left(x_i, n_i\right)_{1 \leq i \leq p}$ est le nombre :

$$\overline{x}_G = \sqrt[N]{x_1^{\;n_1} x_2^{\;n_2} \ldots\ldots x_p^{\;n_p}}$$

AUTRES TYPES DE MOYENNES

Remarques:

- On peut calculer aisément la moyenne géométrique en utilisant le logarithme, à savoir:

$$\log(\overline{x}_G) = \frac{\sum_{i=1}^{p} n_i \log(x_i)}{N}$$

AUTRES TYPES DE MOYENNES

- La moyenne géométrique est rarement utilisée en statistique. En effet, elle n'a de signification réelle que lorsqu'elle s'applique à des valeurs qui suivent sensiblement une progression géométrique.

  C'est le cas de certaines données telles que les valeurs acquises à intérêt composés par un capital donné, si on considère ces valeurs à la fin des périodes de placement.

AUTRES TYPES DE MOYENNES

C-La moyenne quadratique:

Définition:

La moyenne quadratique d'une série statistique $\left(x_i, n_i\right)_{1\leq i\leq p}$ est le nombre:

$$\overline{x}_Q = \sqrt{\frac{1}{N}\sum_{i=1}^{p} n_i {x_i}^2}$$

EXERCICE

Montrer que:

$$\overline{x}_H < \overline{x}_G < \overline{x} < \overline{x}_Q$$

V-2 PARAMETRES DE DISPERTION

En général, les paramètres de position ne suffisent pas à caractériser une série statistique. Par ailleurs, on peut concevoir qu'il existe des séries statistiques qui ont même médiane, même mode et même moyenne. De ce fait, pour distinguer ces séries, il nous faut trouver des valeurs qui caractérisent la manière de distribution des données autour d'une valeur centrale telle que la médiane ou la moyenne.

V-2-1 L'ETENDUE

Définition:

L'étendue d'une série statistique, c'est la différence entre la plus grande et la plus petite des données de la série statistique, c-à-d:

$$e = x_{max} - x_{min}$$

V-2-2 L'ECART INTERQUARTILE

Définition:

C'est la différence entre le troisième et le premier quartile, c-à-d: $Q_3 - Q_1$

Remarque:

- L'intervalle interquartile ne contient que 50% des observations privées de 25% des valeurs extrêmes de chaque coté.
- Pour pouvoir comparer des séries statistiques exprimées par des unités différentes, on utilise l'interquartile relatif $\frac{Q_3 - Q_1}{Q_2}$.

V-2- 3 L'ECART INTERDECILE

Définition:

C'est la différence entre le neuvième et le premier décile c-à-d: $D_9 - D_1$

Remarque:

L'intervalle inter décile contient 80% de la population étudiée et ne rend pas compte de 10% des valeurs extrêmes des deux cotés.

V-2-3 L'ECART ABSOLU MOYEN

Définition:

L'écart absolu moyen d'une série statistique $(x_i, n_i)_{1\leq i\leq p}$ est le nombre:

$$e_m = \frac{1}{N}\sum_{i=1}^{p} n_i \left| x_i - \overline{x} \right|$$

Remarque:

Lorsqu'il s'agit d'une série statistique à caractère continue $\left(\left[a_i, a_{i+1} \right[, n_i \right)_{1\leq i\leq p}$;on remplace les x_i par $c_i = \frac{a_i + a_{i+1}}{2}$

V-2-4 LA VARIANCE ET L'ECART-TYPE

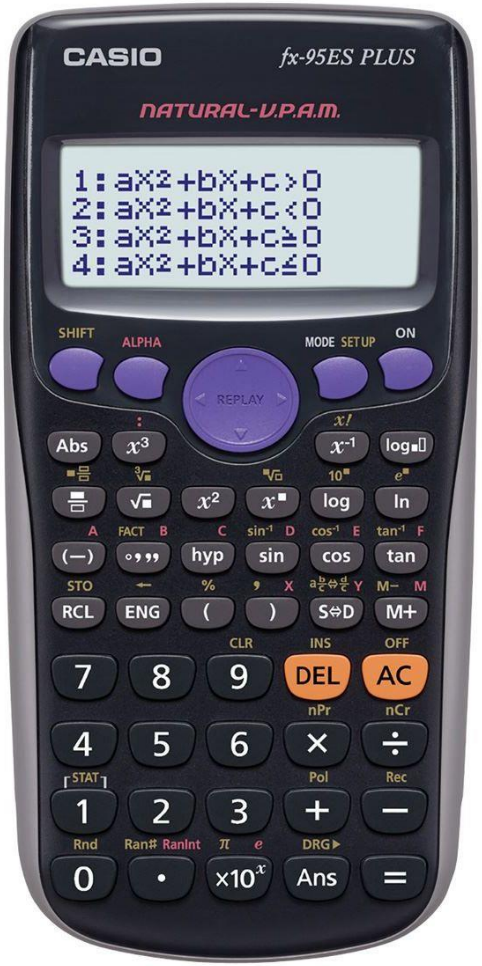
Définition:

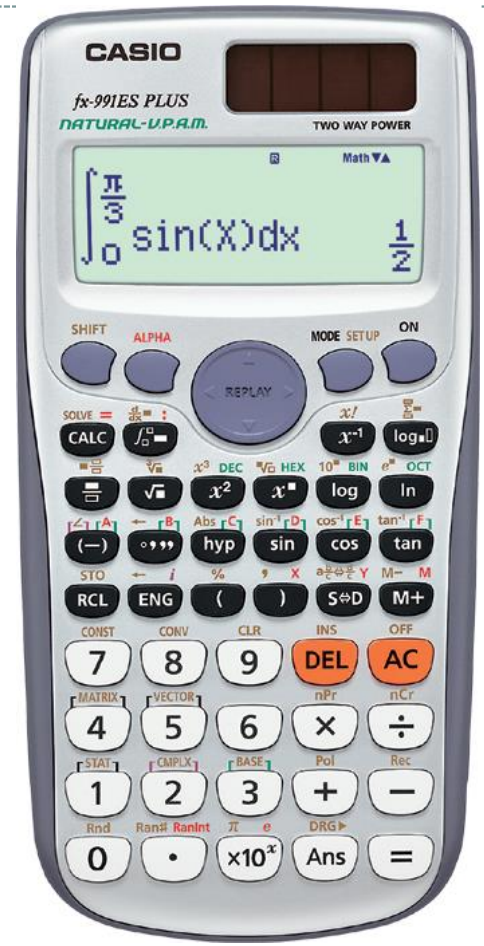
- La variance d'une série statistique $\left(x_i, n_i\right)_{1 \le i \le p}$ est le nombre:

$$V = \frac{1}{N} \sum_{i=1}^{p} n_i \left(x_i - \overline{x}\right)^2$$

- L'écart-type est le nombre: $\sigma = \sqrt{V}$

Usage de la calculatrice scientifique





Casio fx-95ES PLUS

Casio fx-991ES PLUS

Les calculatrices scientifiques sans affichage graphique sont destinées aux élèves et aux enseignants des classes scientifiques et mathématiques du collège et lycée. Toutes dotées de plusieurs options qui permettent de réaliser la plupart des fonctions scientifiques abordées dans les classes :

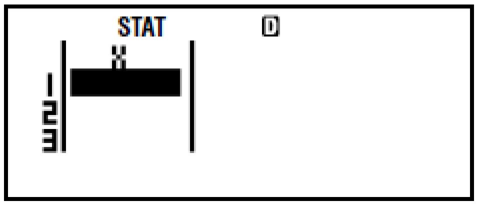
- fonctions mathématiques
- fonctions trigonométriques
- calcul algébrique
- fonctions statistiques

Calculs statistiques sur la calculatrice Casio fx-991ES PLUS

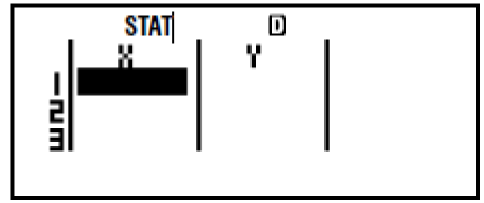
Tous les types de calculs mentionnés ici s'effectuent dans le mode STAT ($\boxed{\text{MODE}}$ $\boxed{3}$), et voilà le menu affiché :

Variable unique (X)	$\boxed{1}$ (1-VAR)
Variable double (X, Y), régression linéaire (y = A + Bx)	$\boxed{2}$ (A+BX)
Variable double (X, Y), régression quadratique (y = A + Bx + Cx^2)	$\boxed{3}$ (_+CX²)
Variable double (X, Y), régression logarithmique (y = A + Bln$x$)	$\boxed{4}$ (ln X)
Variable double (X, Y), e régression exponentielle ($y = Ae^{Bx}$)	$\boxed{5}$ (e^X)
Variable double (X, Y), ab régression exponentielle ($y = AB^x$)	$\boxed{6}$ (A•B^X)
Variable double (X, Y), régression de puissance ($y = Ax^B$)	$\boxed{7}$ (A•X^B)
Variable double (X, Y), régression inverse (y = A + B/x)	$\boxed{8}$ (1/X)

○ Ecran de l'éditeur STAT:



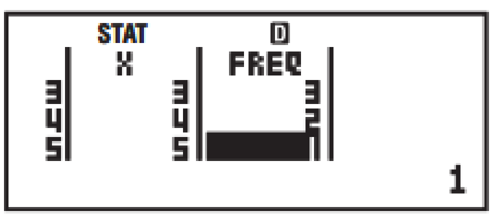
Statistiques à une variable

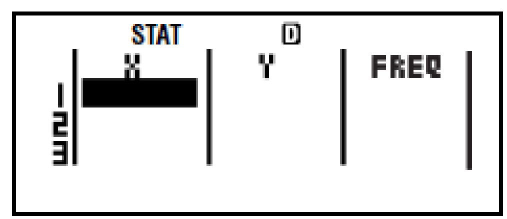


Statistiques à deux variables

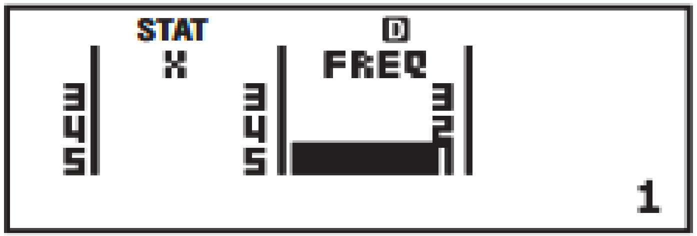
Remarque : Pour afficher la colonne intitulée « FREQ » (effectif) sur l'écran de l'éditeur STAT, on fait un réglage sur l'écran de paramétrage de la calculatrice en appuyant sur les touches suivantes :

SHIFT MODE ▼ 4 (STAT) 1 (ON)





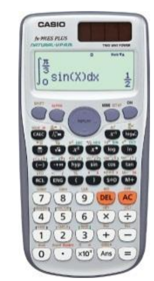
- Etape 1 :La saisie des données se fait sur l’écran de l’éditeur, ils sont insérées dans la cellule où se trouve le curseur. Utilisez les touches de curseur pour déplacer le curseur sur les cellules.



- Etape 2 : Après la saisie des données et lorsque l’écran de l’éditeur STAT ou l’écran de calcul STAT est affiché, on appuie sur SHIFT 1 (STAT) pour afficher le menu STAT, le contenu du menu est différent selon qu’une variable ou deux variables sont utilisées pour le calcul statistique actuellement sélectionné.

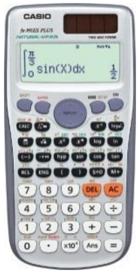
Statistiques à une variable

Statistiques à deux variables



○ Menu STAT: (statistique à une variable)

Sélectionnez cet élément du menu :	Pour effectuer cette opération :
[1] Type	Afficher l'écran de sélection du type de calcul statistique
[2] Data	Afficher l'écran de l'éditeur STAT
[3] Edit	Afficher le sous-menu Edit pour l'édition du contenu de l'écran de l'éditeur STAT
[4] Sum	Afficher le sous-menu de commandes Sum pour le calcul de sommes
[5] Var	Afficher le sous-menu de commandes Var pour le calcul de la moyenne, de l'écart-type, etc.
[6] MinMax	Afficher le sous-menu de commandes MinMax pour obtenir les valeurs maximales et minimales



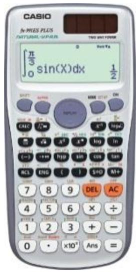
○ Menu STAT :

Sélectionnez cet élément du menu :	Pour effectuer cette opération :
[1] Type	Afficher l'écran de sélection du type de calcul statistique
[2] Data	Affich… iteur STAT
[3] Edit	Affich… Edit pour l'édition du contenu de l'é… STAT
[4] Sum	Afficher le sous-menu de comm… le calcul de sommes
[5] Var	Afficher le sous-menu de comma… calcul … de l'écart-type, etc.
[6] MinMax	Affiche… u de commandes MinMax pour o… s maximales et minimales

[1] Σx^2
[2] Σx

[1] n
[2] $\bar{x}$
[3] $x\sigma n$
[4] $x\sigma n{-}1$

[1] minX
[2] maxX



○ Menu STAT: (statistique à deux variable)

Sélectionnez cet élément du menu :	Pour effectuer cette opération :
[1] Type	Affic … type de calcul stati…
[2] Data	Affic…
[3] Edit	Affic… …dition du contenu de l'écran de…
[4] Sum	Afficher le so… …n pour le calcul de s…
[5] Var	Afficher le so… …our le calcul de la mo… etc.
[6] MinMax	Afficher le sou… …des MinMax pour obtenir les… …t minimales
[7] Reg	Afficher le sous-… …eg pour les calculs de rég…

[1] Σx^2 [3] Σy^2 [5] Σxy
[2] Σx [4] Σy [7] Σx^2y
…

[1] n [2] $\bar{x}$ [3] $x\sigma n$ [4] $x\sigma n{-}1$
[5] $\bar{y}$ [6] $y\sigma n$ [7] $y\sigma n{-}1$

[1] minX [3] minY
[2] maxX [4] maxY

[1] A [2] B [3] r
[4] $\hat{x}$ [5] $\hat{y}$

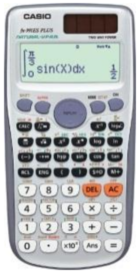
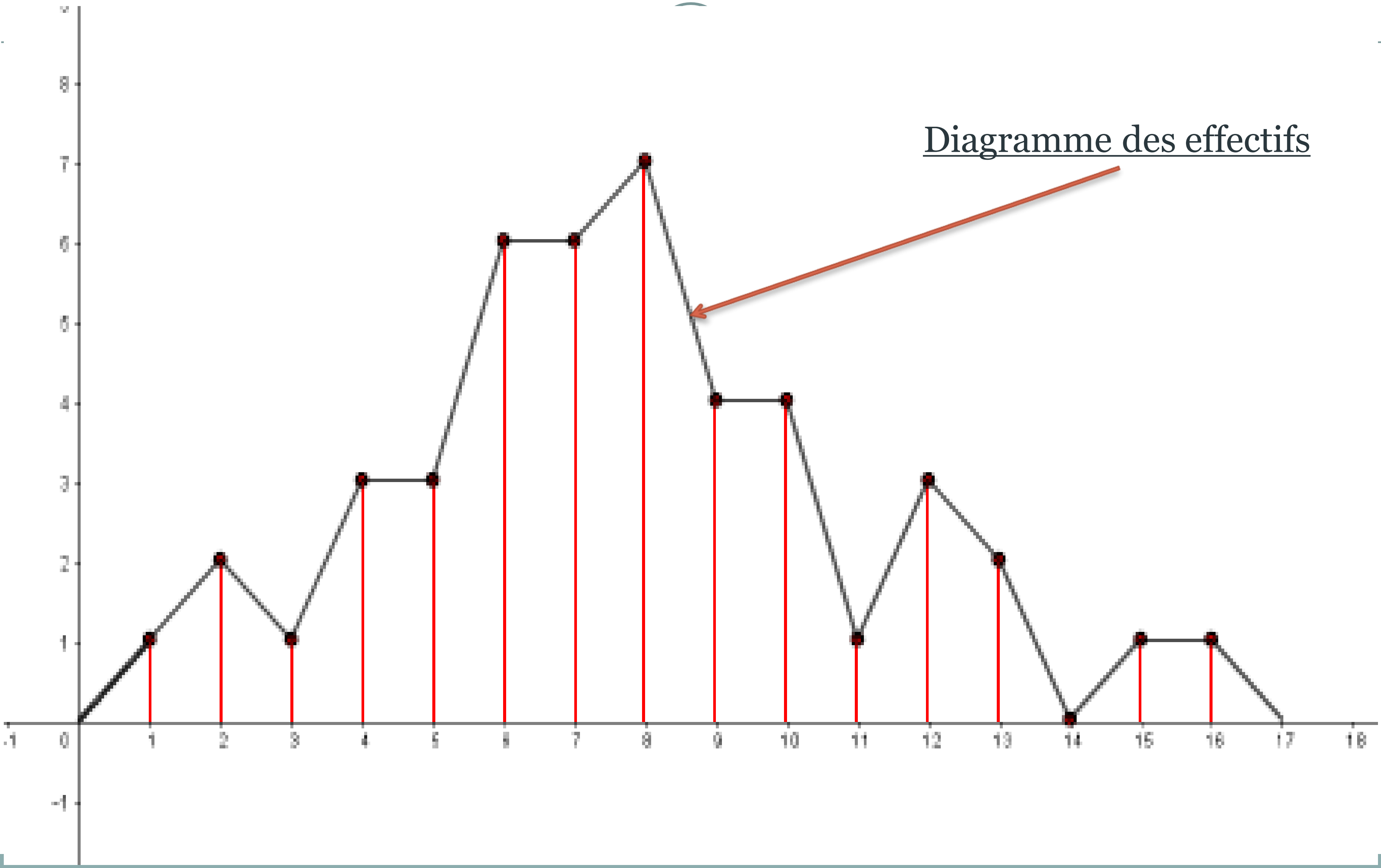
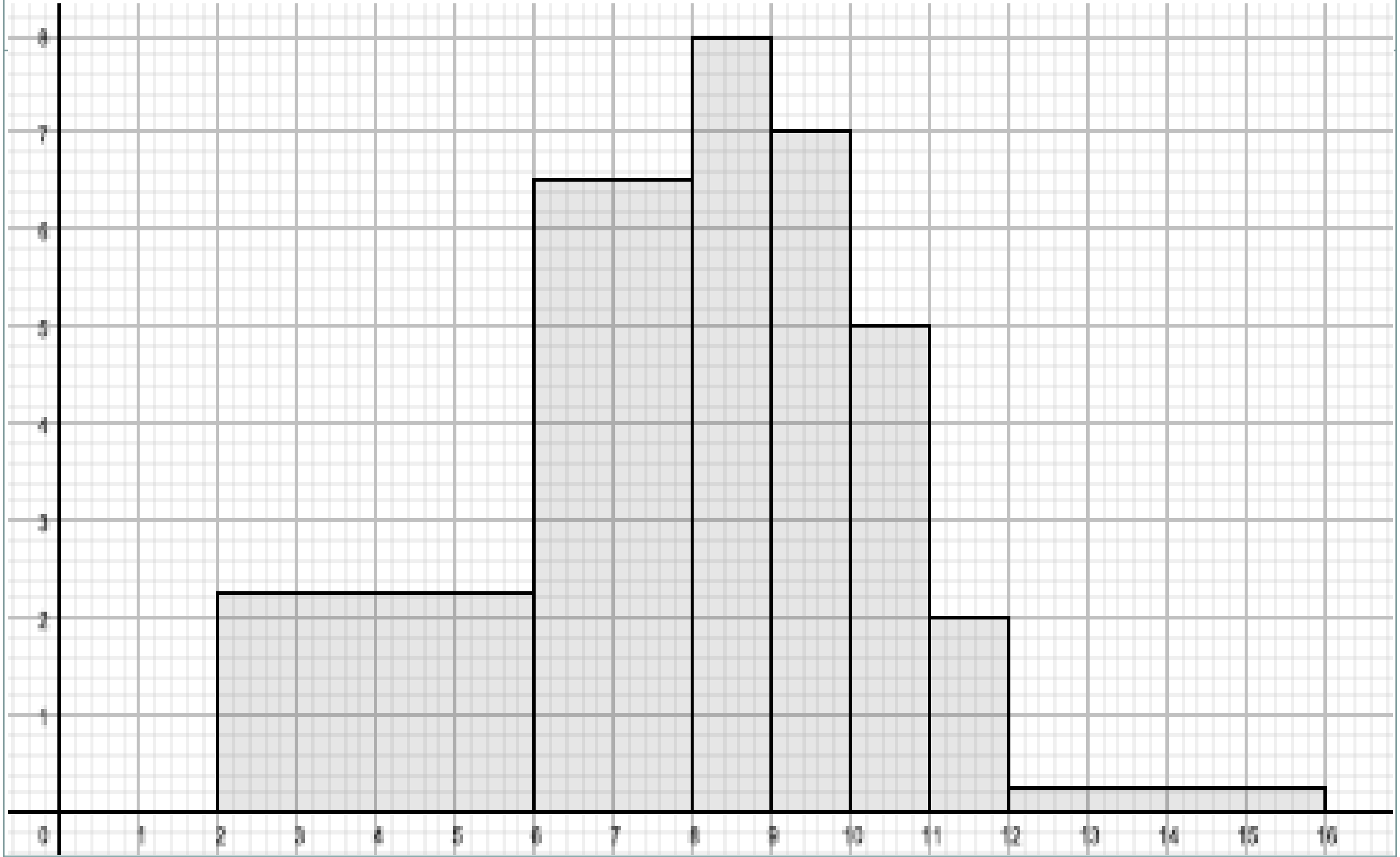


Diagramme en bâtons (Ex3-serie1)



Histogramme (Ex4-serie1)



CHAPITRE VI

SERIES STATISTIQUES A DEUX CARACTERES

VI-1 GENERALITES

On considère deux variables statistiques X et Y d'une population Ω.

Soient x_1,x_2 …….et x_p les valeurs prises par X

et y_1,y_2,……….y_q les valeurs prises par Y

Deux présentations de la série statistique associée au couple (X,Y) sont possibles:

LES DONNEES SONT NON GROUPEES

X_i	X_1	X_2	X_3	…………….	X_n
Y_i	Y_1	Y_2	Y_3	……………….	Y_n

Dans ce cas il ya autant de valeurs que d'individu de la population étudiée.

LES DONNEES SONT GROUPEES

Y / X	**Y1**	**........**	**Yj**	**.............**	**Yq**	**Marginale X**
X1	N11	....	N1i	....	N1q	N1.
. .	. .	. .	. .	. .	. .	. .
Xi	Ni1	....	Nij	....	Niq	Ni.
. .	. .	. .	. .	. .	. .	. .
Xp	Np1	....	Npj	....	Npq	Np.
Marginale Y	N.1	....	N.j	....	N.q	N

DEFINITIONS

- Chaque couple (x_i,y_j) est associé à l'effectif n_{ij} .
- La série obtenue est une série à 2 variables notée (x_i,y_j,n_{ij}) attachée au couple (X ,Y).
- L'effectif total est $N=\sum\sum n_{ij} =\sum n_{i.} = \sum n_{.j}$
- La fréquence de (x_i,y_j) est $f_{ij}= n_{ij}/N$
- Le pourcentage $p_{ij}=f_{ij}\bullet 100$
- La série peut se noter également par (x_i,y_j,f_{ij})
- Les deux séries X et Y s'appellent les séries marginales de la série (x_i,y_j,n_{ij})

EXEMPLE

- A un examen, chaque candidat est évalue sur deux disciplines A(noté Xi) et B(noté Yi) à l'aide d'une échelle de 0 à 10.

Xi \ Yi	2	4	5	6	7	8	9	MARGINALE X
3				1	6		1	8
4			3	1	4		4	12
5			2	3	5	1		11
6	1		6	4	2	3	3	19
7		2	3		1	4		10
MARGINALE Y	1	2	14	9	18	8	8	60

VI-2 NUAGE STATISTIQUE

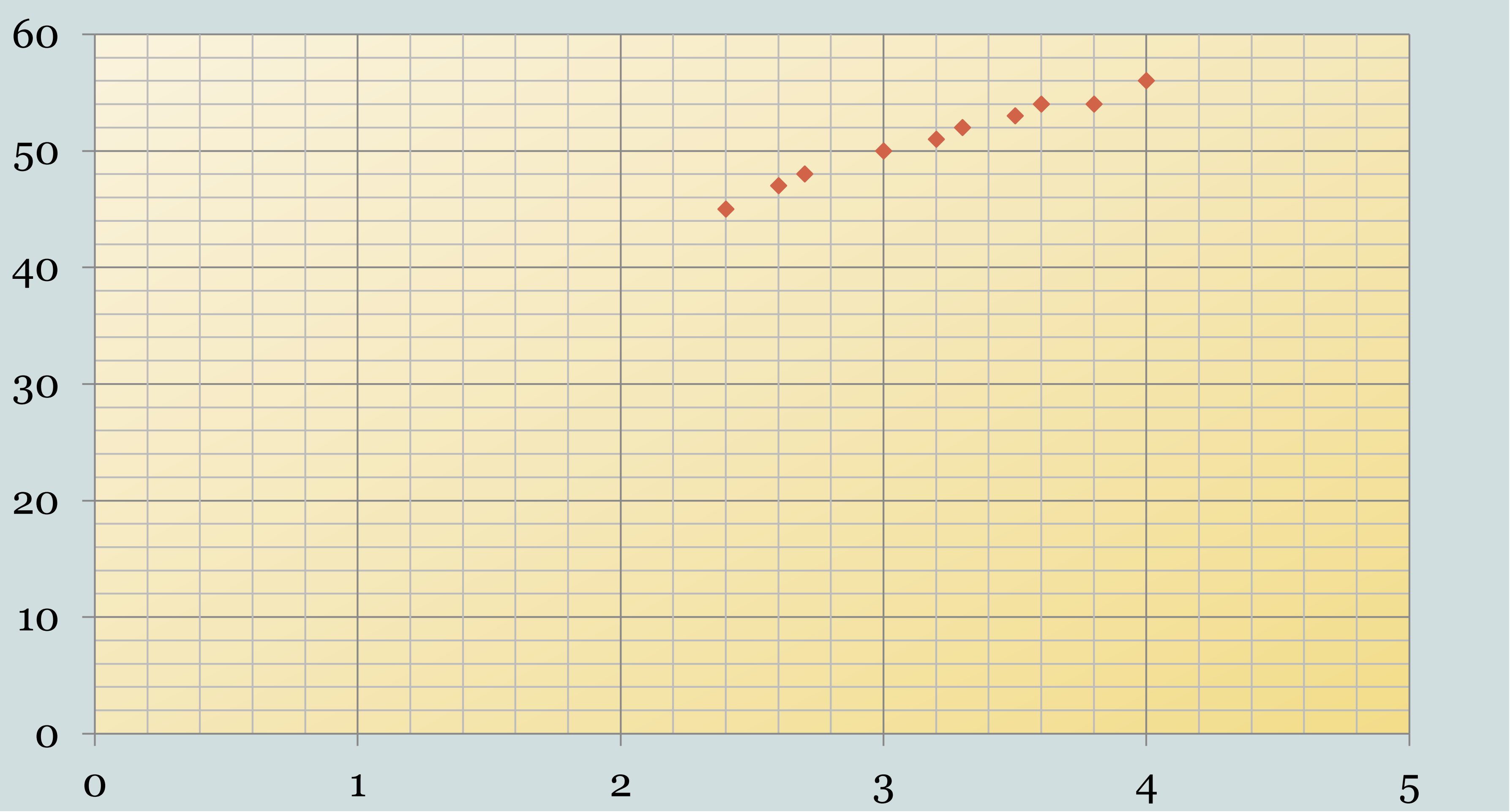
Dans un repère cartésien ,on considère les points $M_{ij}(x_i,y_j)$ .L'ensemble des points M_{ij} s'appelle nuage statistique attaché à la série (x_i,y_j,n_{ij}).

EXEMPLE 1

- L' étude statistique suivante porte sur une population de nouveau-nés .Deux caractères sont étudiés:

enfant	1	2	3	4	5	6	7	8	9	10
Masse en kg	2,4	2,6	2,7	3	3,2	3,3	3,5	3,6	3,8	4
Taille en cm	45	47	48	50	51	52	53	54	54	56

REPRESESENTATION GRAPHIQUE



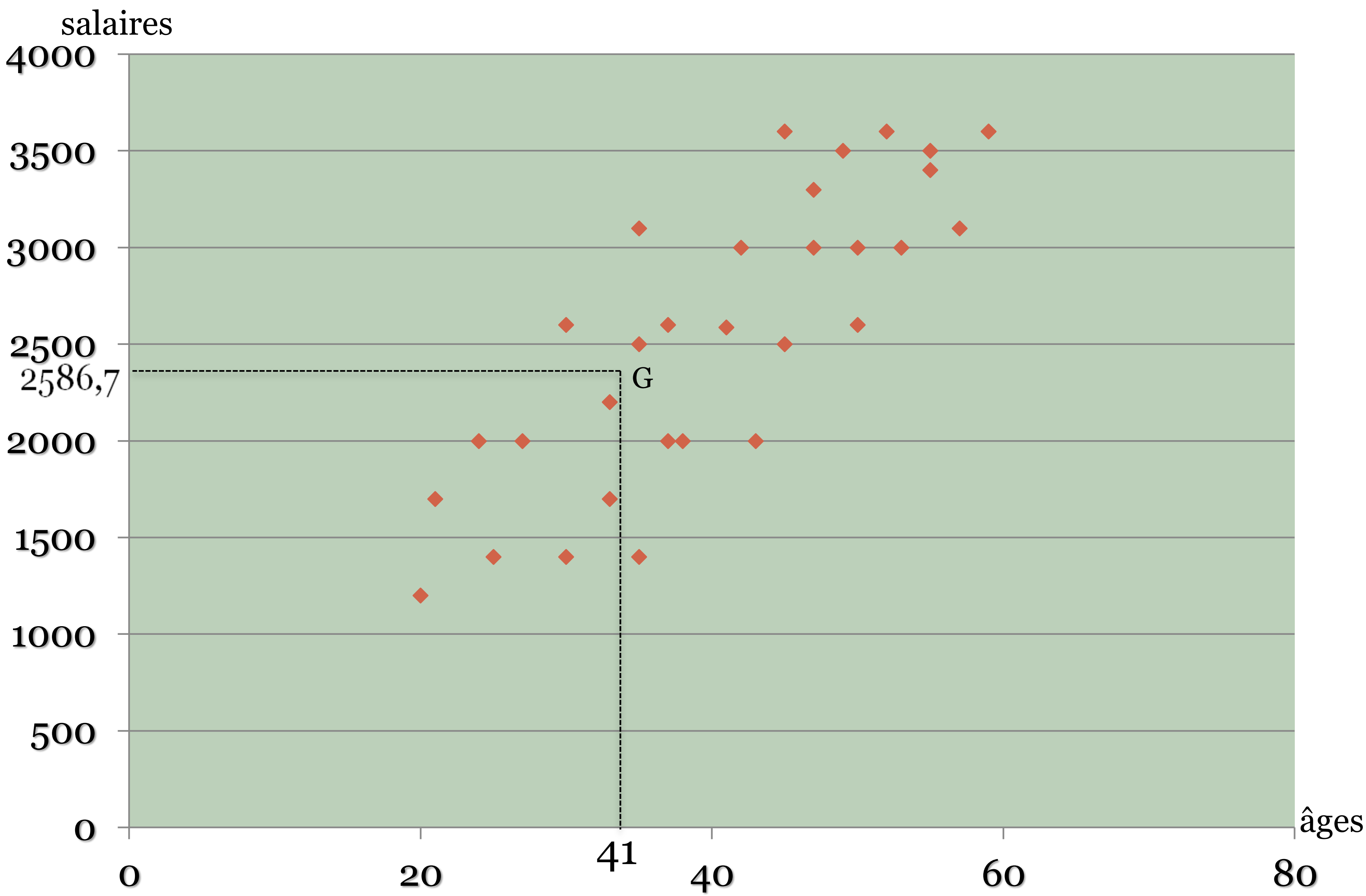
VI-3 POINT MOYEN

- Le point G($\overline{x}$, $\overline{y}$) s'appelle « point moyen » du nuage statistique.
- Dans le cas de l'exemple, on a: $\overline{x} = 3{,}19$

$$\overline{y} = 51$$

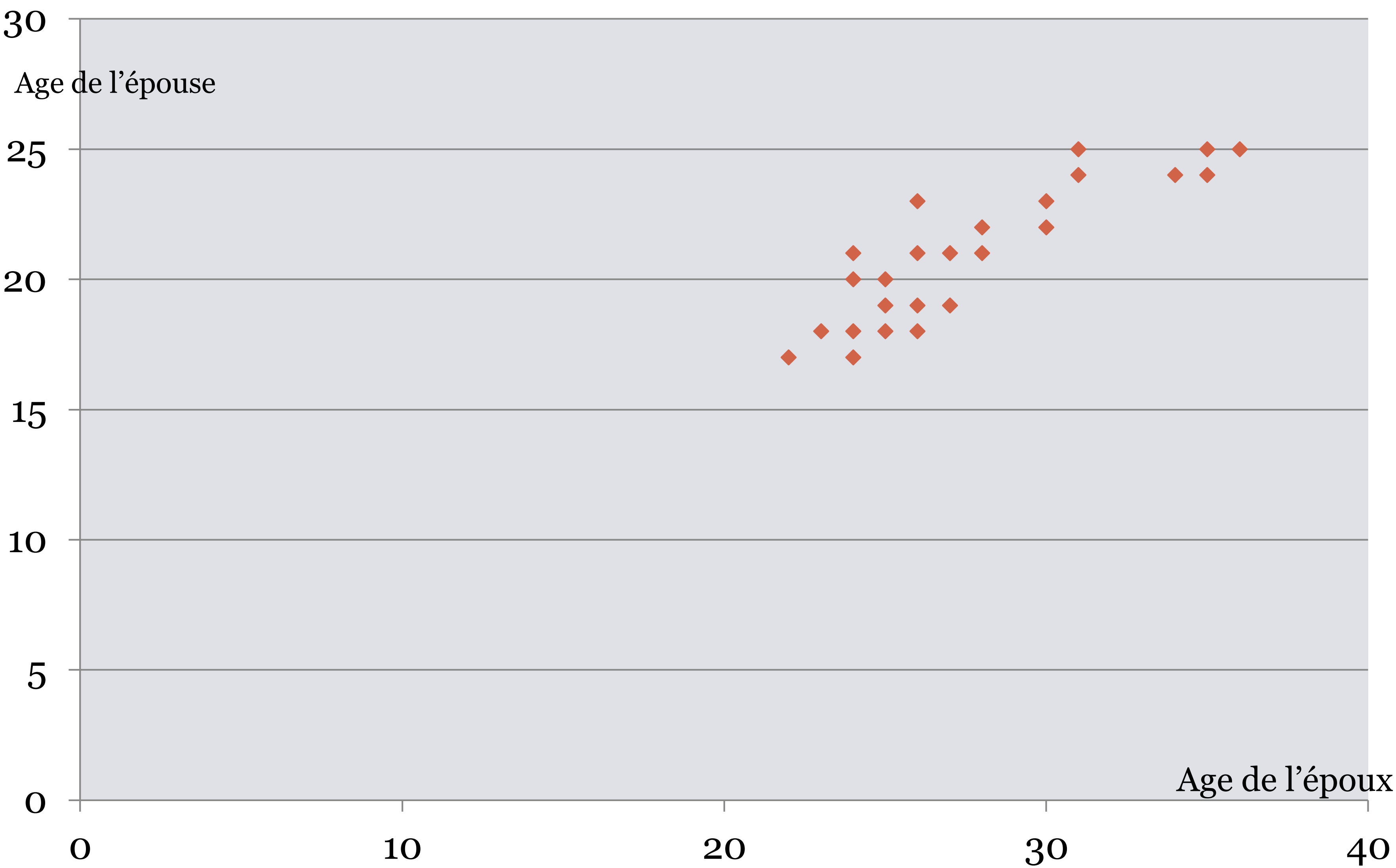
EXEMPLE 2

- On recense les salaires mensuels nets en euros des 30 salariés d'une certaine entreprise et leurs âges; on obtient : (20,1200)-(21,1700)-(25,1400)-(24,2000)-(30,2600)-(27,2000)-(33,1700)-(33,2200)-(35,1400)-(35,2500)-(35,3100)-(37,2000)-(37,2600)-(38,2000)-(43,2000)-(42,3000)-(45,3600)-(45,2500)-(47,3300)-(47,3000)-(50,2600)-(50,3000)-(49,3500)-(52,3600)-(53,3000)-(55,3500)-(55,3400)-(57,3100)-(59,3600).


salaires
4000
3500
3000
2500
2586,7
G
2000
1500
1000
500
0
0
20
41
40
60
80
âges

EXEMPLE 3

- Pour une étude démographique locale concernant l'âge des époux et des épouses au moment du mariage, on prélève les données suivantes:(24,17)-(23,18)-(22,17)-(24,18)-(24,20)-(24,21)-(25,18)-(25,19)-(25,20)-(26,18)-(26,19)-(26,21)-(26,23)-(27,19)-(27,21)-(28,21)-(28,22)-(30,22)-(30,23)-(31,24)-(31,25)-(34,24)-(35,24)-(36,25)-(35,25).



Age de l’épouse
Age de l’époux
30
25
20
15
10
5
0
0
10
20
30
40

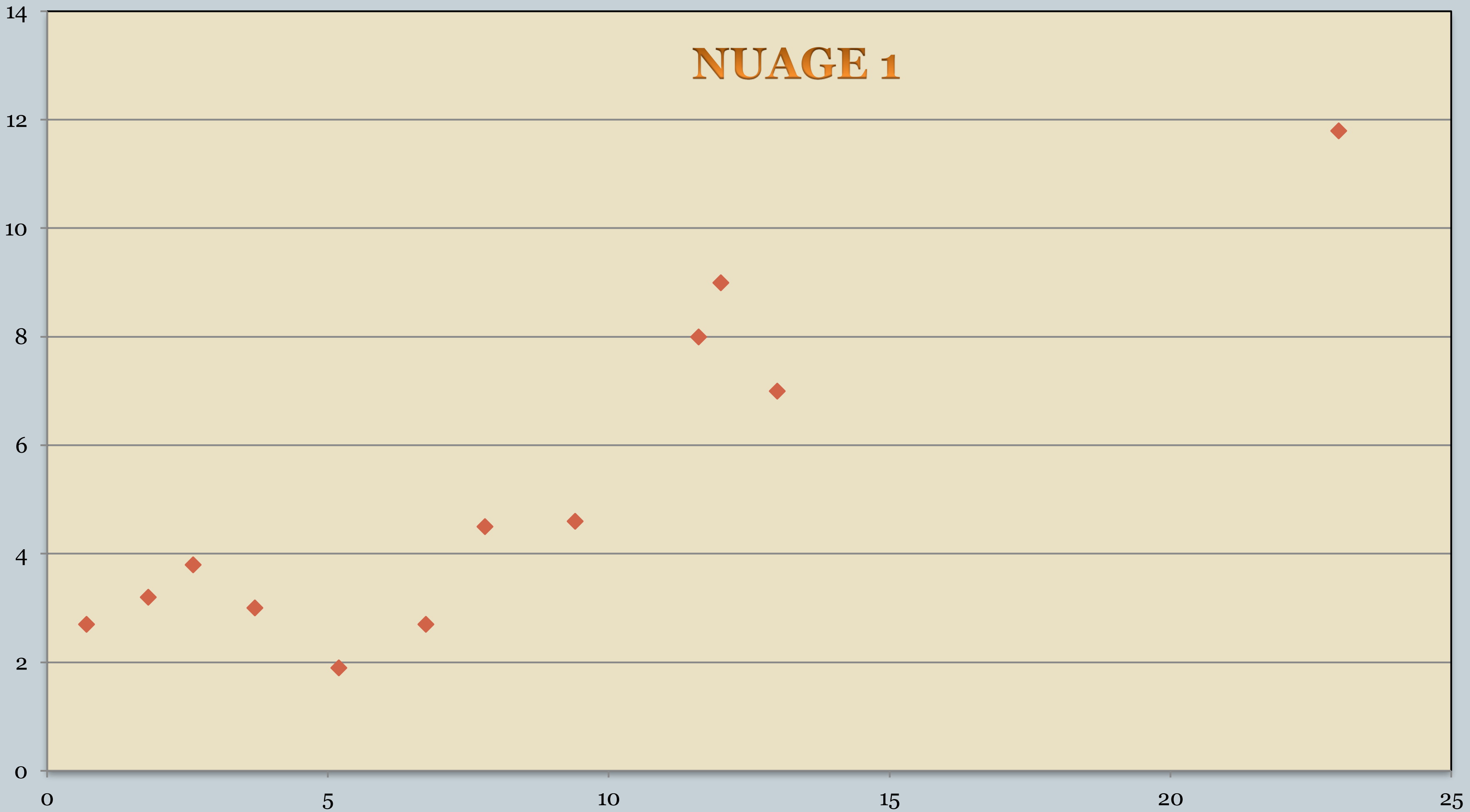
L'ajustement

La corrélation

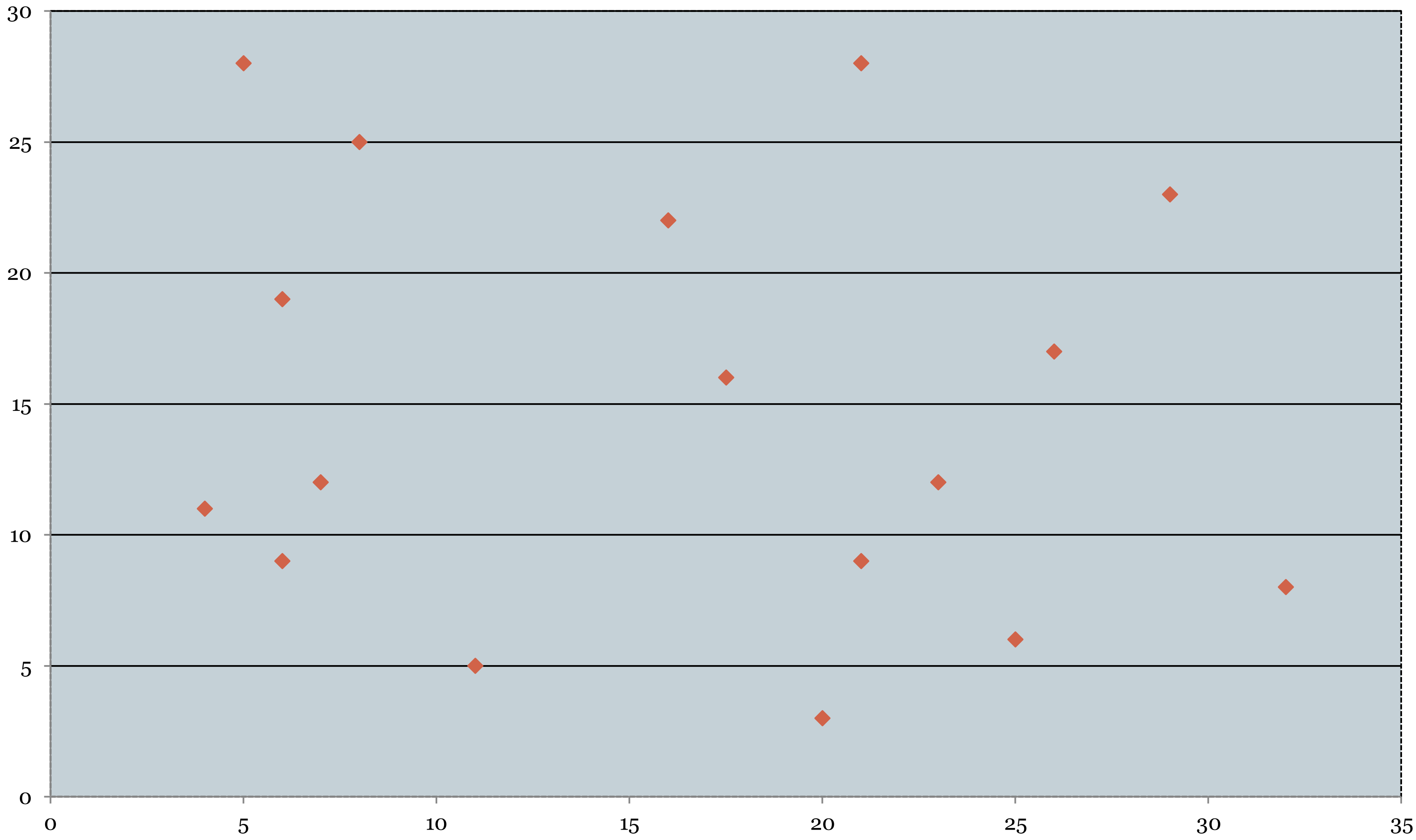
VI-4 L' AJUSTEMENT

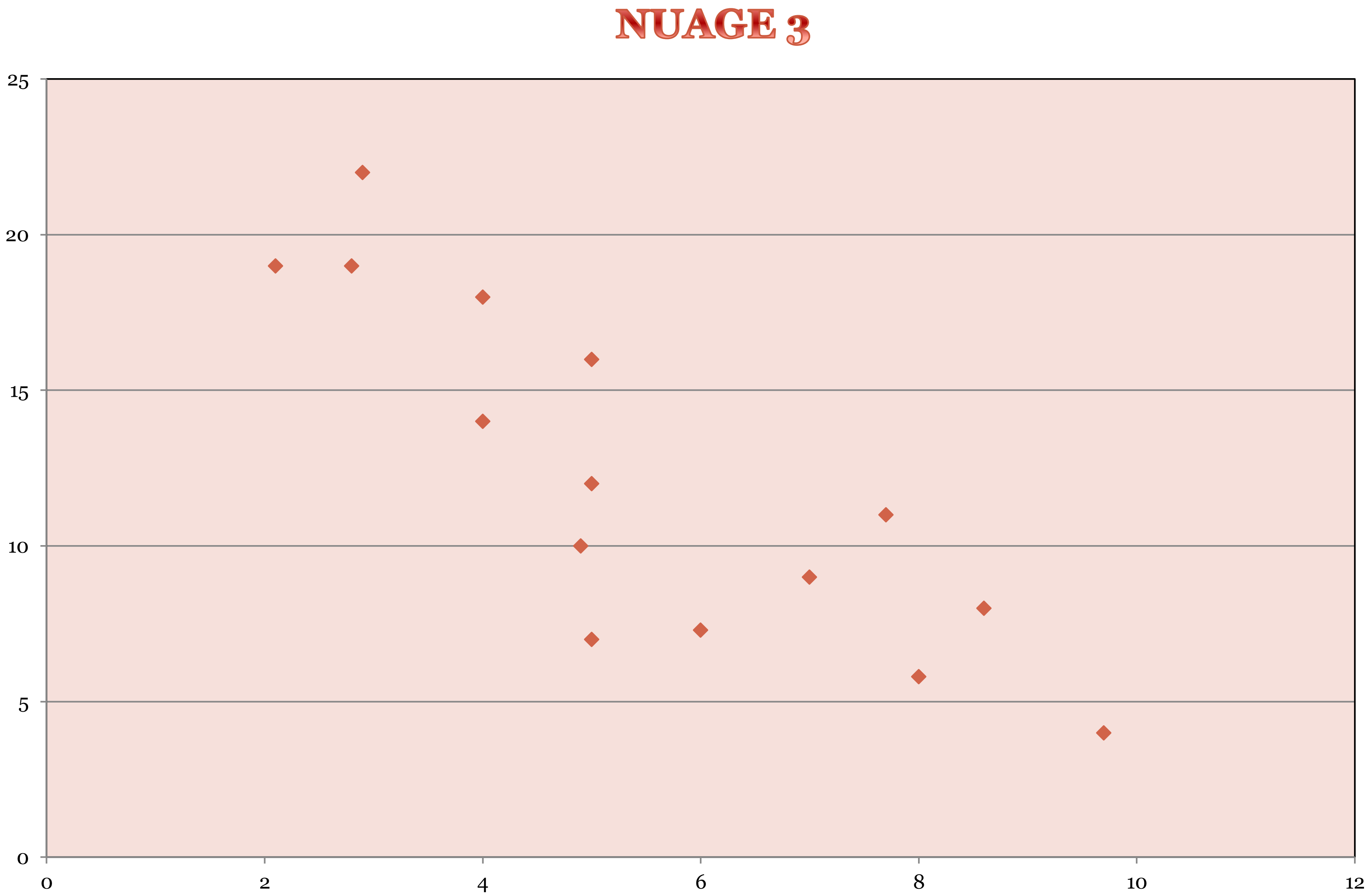
- A partir du nuage de points représentant la série statistique, on cherche à trouver un lien éventuel entre les deux caractères .Mathématiquement, il s'agit de déterminer la fonction analytique y=f(x) qui parait la mieux représentatif pour l'ensemble du nuage.
- Cet ajustement pourrait servir pour procéder à des interpolations ou à des extrapolations.

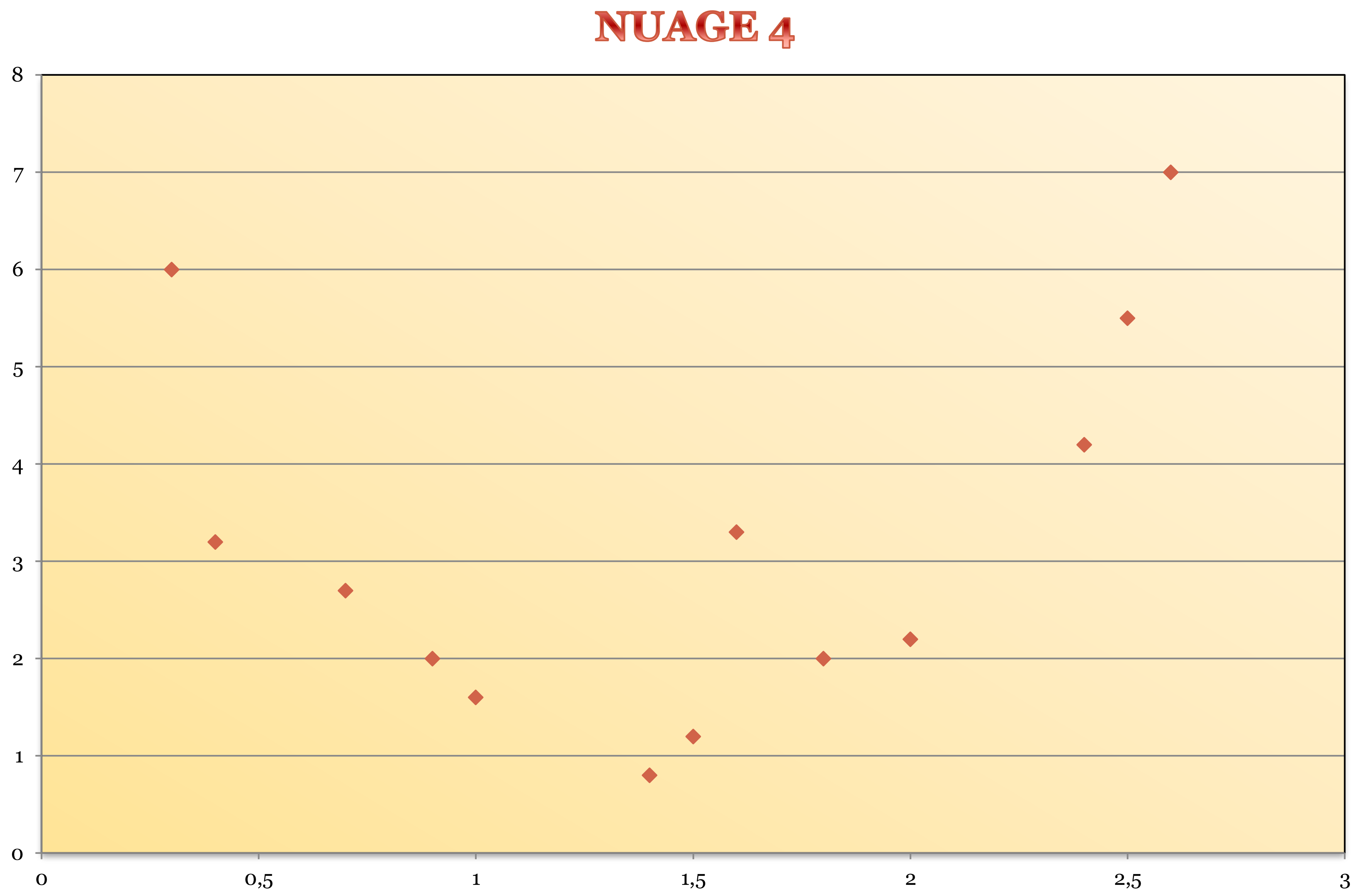
Exemples de nuages de points



NUAGE 2




NUAGE 3
25
20
15
10
5
0
0
2
4
6
8
10
12


NUAGE 4
8
7
6
5
4
3
2
1
0
0
0,5
1
1,5
2
2,5
3

VI-5 L’AJUSTEMENT LINEAIRE

- L’idée est de déterminer la droite qui sera la plus proche de tous les points du nuage et donnera les meilleurs résultats. Ceci est le mieux adapté quand le nuage a un aspect rectiligne.

- Plusieurs méthodes sont possibles:
  - La méthode de Mayer.
  - La méthode des moindres carrés.

VI-5-1 METHODE DE MAYER

Cette méthode consiste à :

- Fractionner le nuage de points en deux nuages dont les effectifs sont égaux ou différent de 1 si l'effectif total est impair.
- Calculer les coordonnées des points moyens G_1 et G_2 des deux nuages.
- Déterminer l'équation de la droite (G_1G_2).

Exemple

- Le mur d'une habitation est constitué d'une couche de béton et une couche de polystyrène d'épaisseur variable x. On a mesuré la résistance thermique R de ce mur pour diverses valeurs de x, et on a obtenu les résultats suivants:

x (cm)	2	4	6	8	10	12	15	20
R	0,83	1,34	1,63	2,29	2,44	2,93	4,06	4,48

La résistance est exprimée en $(m^2.dC^0/W)$

1) Calculer les coordonnées du point moyen G_1 associé aux points du nuage ayant les quatre plus petites abscisses et

   les coordonnées du point moyen G_2 associé aux points du nuage ayant les quatre autres points du nuage.

2) On choisit la droite (G_1G_2) comme droite d'ajustement. En déterminer une équation.

3) Quel résistance thermique peut-on obtenir avec une épaisseur de polystyrène de 25 cm.

Solution

1) $G_1(5;1.52)$; $G_2(14.25;3.48)$

2) $G(9.625;2.5)$

$$(G_1G_2):\ R = ax + b \quad avec \quad a = \frac{R_2 - R_1}{x_2 - x} \quad et \quad b = R_G - ax_G.$$

$$(G_1G_2):\ R = 0.21x + 0.48$$

3) $x = 25cm \Rightarrow R \approx 5.73 m^2.d^0C/W$

VI-5- 2 METHODE DES MOINDRES CARRES

Cette méthode consiste à :

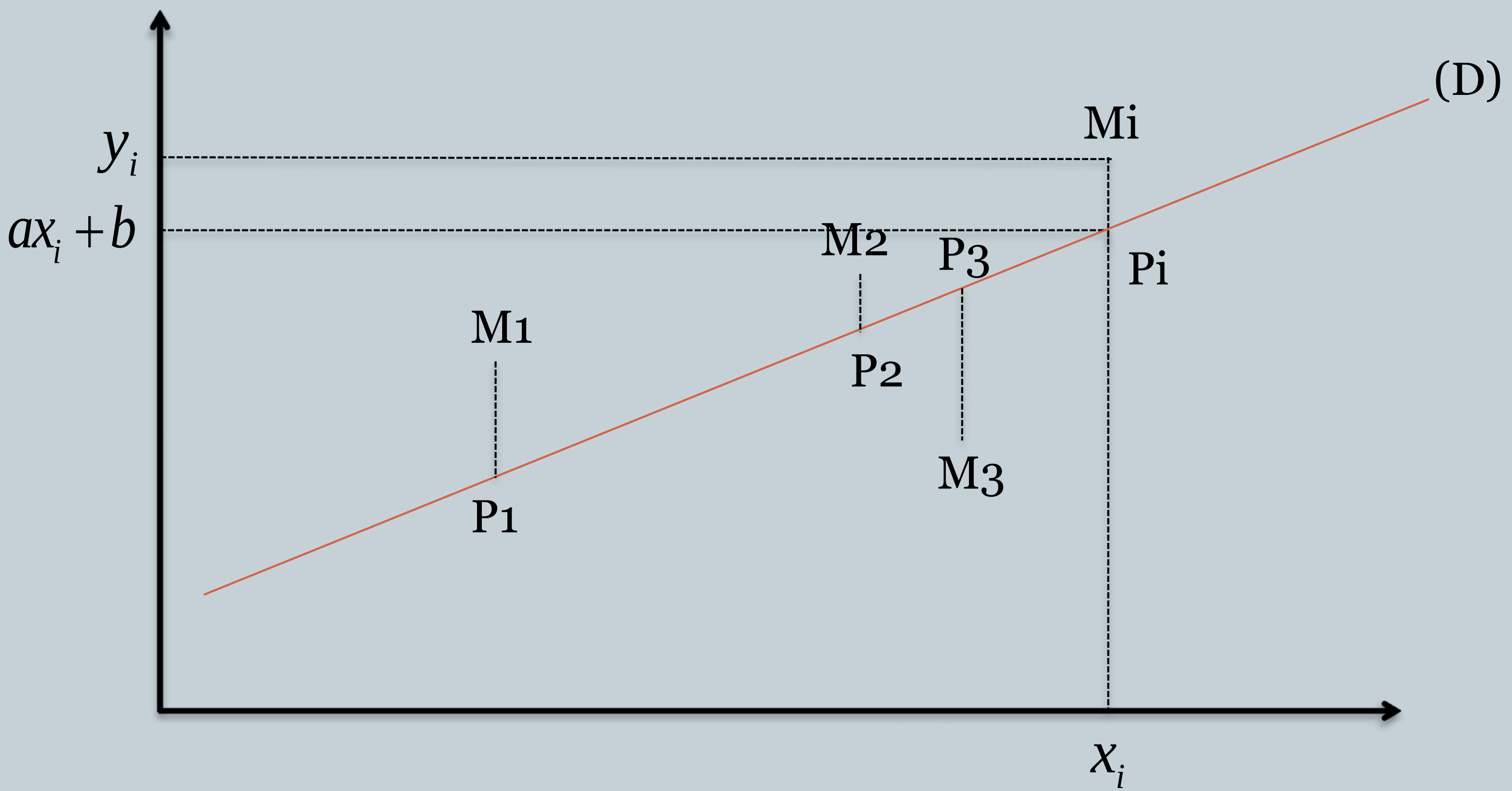
- Déterminer la droite (D) qui minimise la somme des carrés des écarts des points du nuage à cette droite.

On peut procéder de deux manières suivant que l'on prend les écarts suivant la direction de l'axe des abscisses ou de l'axe des ordonnées.

Le problème c'est de déterminer la droite (D) : $y=ax+b$ de telle sorte que

$$f(a,b)= \sum(a\, x_i+b - y_i)^2 \text{ soit minimale.}$$

VI-5- 2 METHODE DES MOINDRES CARRES



$$Soit \quad S(a,b)=\sum_{i=1}^{N}(y_i-ax_i-b)^2. \; On\ cherche\ a\ et\ b\ tq\ S(a,b)\ \ soit\ \min.$$

$$On\ a: S(a,b)=\sum_{i=1}^{N}b^2-2b\sum_{i=1}^{N}(y_i-ax_i)+\sum_{i=1}^{N}(y_i-ax_i)^2$$

$$=Nb^2-2b\sum_{i=1}^{N}(y_i-ax_i)+\sum_{i=1}^{N}(y_i-ax_i)^2.$$

$$\frac{\partial S}{\partial b}=0 \iff 2Nb-2\sum_{i=1}^{N}(y_i-ax_i)=0$$

$$\iff Nb-\sum_{i=1}^{N}y_i-a\sum_{i=1}^{N}x_i=0$$

$$\iff \overline{y}=a\overline{x}+b \;\Rightarrow G(\overline{x},\overline{y})\in(D)$$

$$On\ \ pose : \begin{cases} X = x - \overline{x} \\ Y = y - \overline{y} \end{cases} \Rightarrow \begin{cases} X_i = x_i - \overline{x} \\ Y_i = y_i - \overline{y} \end{cases}$$

$$\Rightarrow Y_i - aX_i = y_i - ax_i - b$$

$$\Rightarrow S(a,b) = \sum (Y_i - aX_i)^2$$

$$\Rightarrow S(a,b) = a^2 \sum Y_i^2 - 2a \sum X_i Y_i + \sum Y_i^2$$

$$\frac{\partial S}{\partial a} = 0 \Leftrightarrow a = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{\frac{1}{N}\sum (x_i - \overline{x})(y_i - \overline{y})}{\frac{1}{N}\sum (x_i - \overline{x})^2} = \frac{\operatorname{cov}(X,Y)}{V(X)}$$

$Conclusion :$

$$(D)_{y\,en\;x} : y - \overline{y} = a(x - \overline{x})$$

$$a = \frac{\operatorname{cov}(X,Y)}{V(X)}$$

DROITE DE REGRESSION DE Y EN X

La droite d'équation y=ax+b qui rend minimale la somme des résidus est la droite :

- qui passe par le point moyen G.
- qui a pour coefficient directeur a= $\frac{\operatorname{cov}(X,Y)}{V(X)}$

Donc l'équation de la droite de régression de y en x a pour équation:

$$y-\overline{y}=\frac{\operatorname{cov}(X,Y)}{V(X)}(x-\overline{x})$$

DROITE DE REGRESSION DE X EN Y

En procédant de la même manière, mais en prenant les écarts suivant la direction de l'axe des abscisses, on trouve l'équation de la droite de régression de x en y:

$$x - \bar{x} = \frac{\operatorname{cov}(X, Y)}{V(Y)}(y - \bar{y})$$

DROITES DE REGRESSION

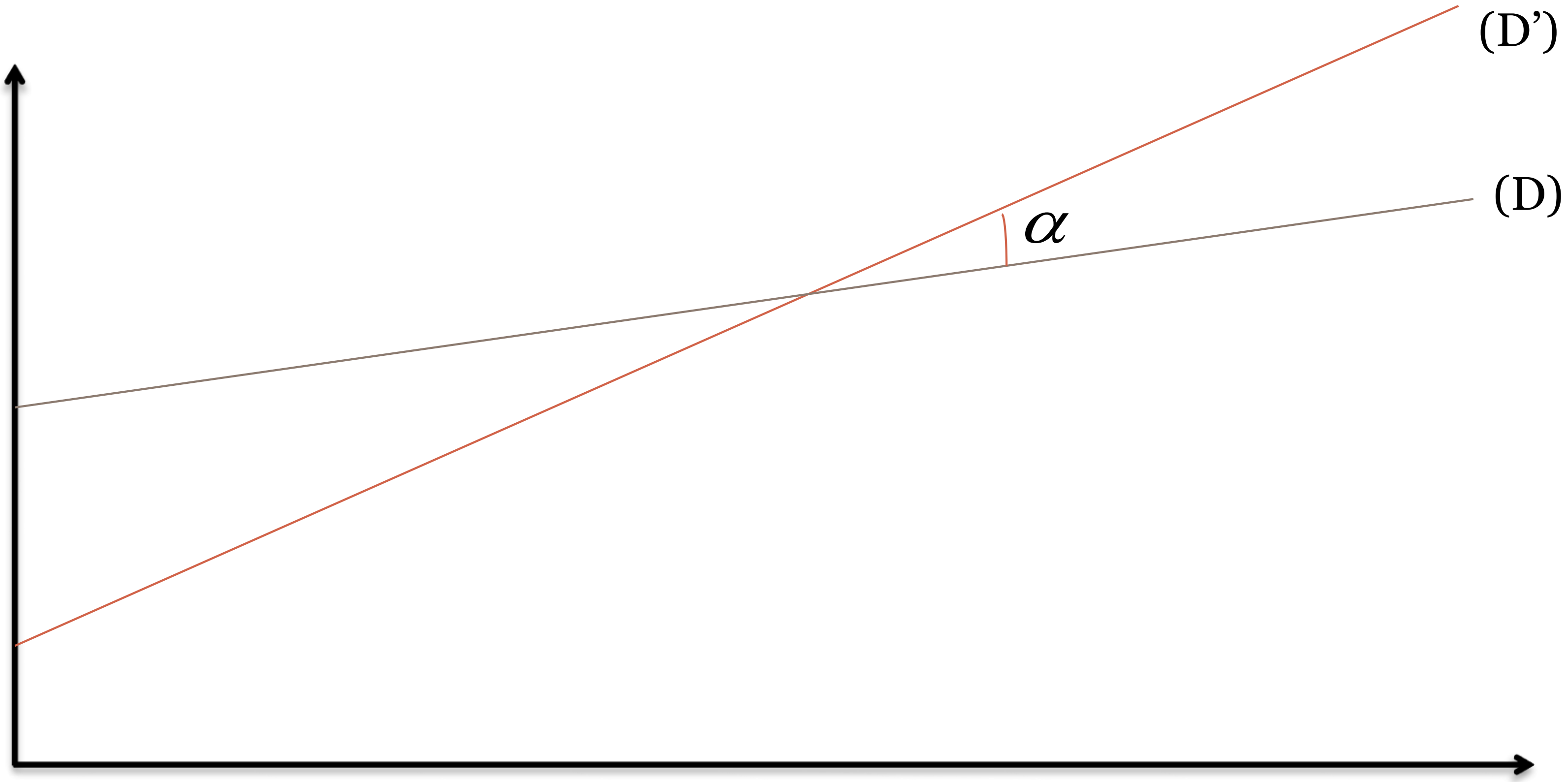
- Le coefficient directeur de la droite de régression de y en x est :

$$a = \frac{\operatorname{cov}(X,Y)}{V(X)}$$

- Le coefficient directeur de la droite de régression de x en y est :

$$\frac{1}{a'} = \frac{V(Y)}{\operatorname{cov}(X,Y)}$$

Les deux droites de régression sont toutes les deux croissantes ou toutes les deux décroissantes.



VI-6 LA CORRELATION

- La décision d'ajuster un nuage par une droite ne se prend pas uniquement à la seule vue du nuage, suivant s'il est allongé ou non. A cet égard, les statisticiens ont éprouvé le besoin de quantifier cette décision par un nombre ; c'est le coefficient de corrélation.

VI-6 LA CORRELATION

Définition:

le nombre $r(X,Y)=\dfrac{\text{cov}(X,Y)}{\sigma(X)\sigma(Y)}$ s'appelle coefficient

de corrélation linéaire.

Remarque:

$$a.a' = \left[r(X,Y)\right]^2$$

Proposition:

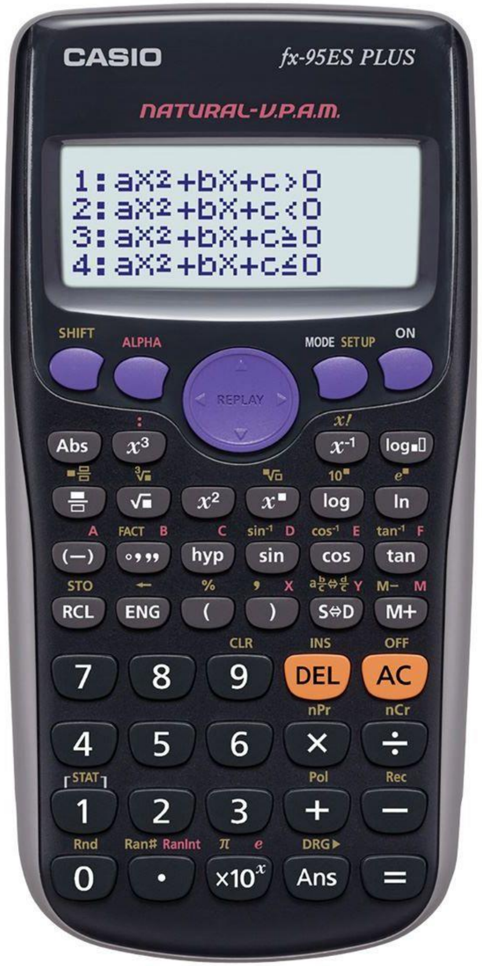
$$-1 \leq r(X,Y) \leq 1$$

VI-6 LA CORRELATION

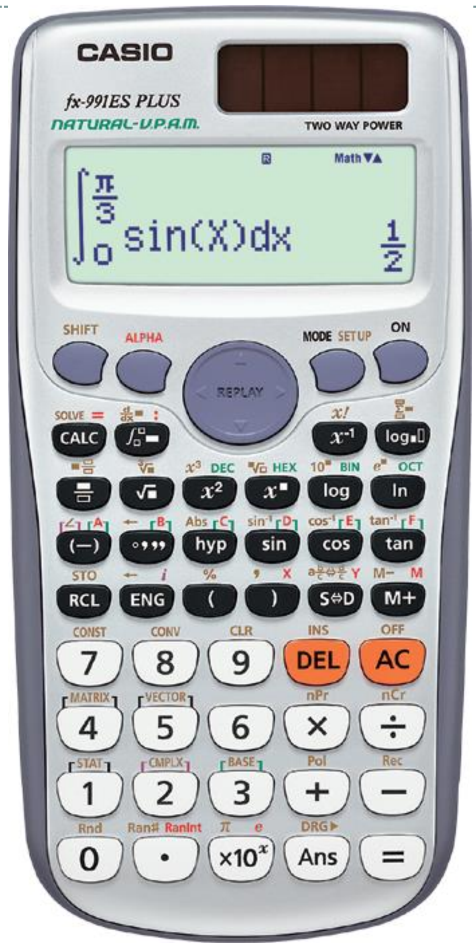
Remarques:

- La corrélation est d'autant meilleure que $|r|$ est proche de 1.
- De façon calculée, on estime que la corrélation est bonne lorsque $|r| \geq 0,87$

Usage de la calculatrice scientifique

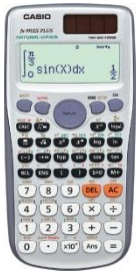


Casio fx-95ES PLUS



Casio fx-991ES PLUS

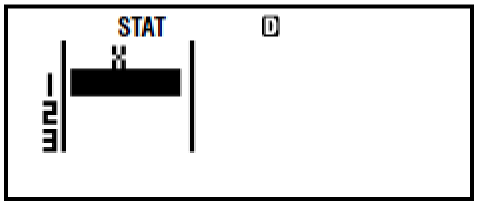
Calculs statistiques sur la calculatrice Casio fx-991ES PLUS



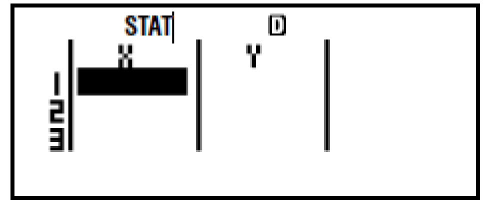
Tous les types de calculs mentionnés ici s'effectuent dans le mode STAT (MODE 3), et voilà le menu affiché :

Variable unique (X)	1 (1-VAR)
Variable double (X, Y), régression linéaire (y = A + Bx)	2 (A+BX)
Variable double (X, Y), régression quadratique (y = A + Bx + Cx^2)	3 (_+CX²)
Variable double (X, Y), régression logarithmique (y = A + Bln$x$)	4 (ln X)
Variable double (X, Y), e régression exponentielle ($y = Ae^{Bx}$)	5 (e^X)
Variable double (X, Y), ab régression exponentielle ($y = AB^x$)	6 (A•B^X)
Variable double (X, Y), régression de puissance ($y = Ax^B$)	7 (A•X^B)
Variable double (X, Y), régression inverse (y = A + B/x)	8 (1/X)

- **Ecran de l'éditeur STAT:**



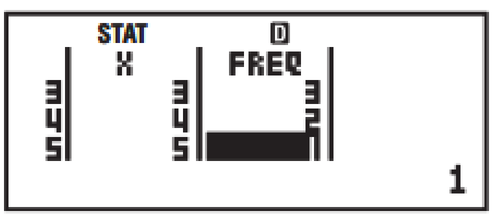
Statistiques à une variable

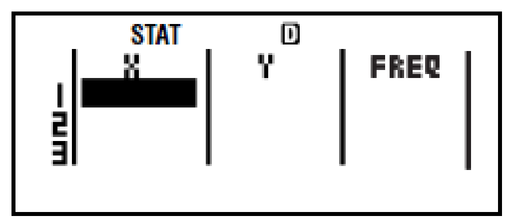


Statistiques à deux variables

Remarque : Pour afficher la colonne intitulée « FREQ » (effectif) sur l'écran de l'éditeur STAT, on fait un réglage sur l'écran de paramétrage de la calculatrice en appuyant sur les touches suivantes :

SHIFT MODE ▼ 4 (STAT) 1 (ON)





Menu STAT: (statistique à deux variable)

Sélectionnez cet élément du menu :	Pour effectuer cette opération :
$\boxed{1}$ Type	Affic… type de calcul stati…
$\boxed{2}$ Data	Affic…
$\boxed{3}$ Edit	Affic… dition du contenu de l'écran de …
$\boxed{4}$ Sum	Afficher le so… pour le calcul de s…
$\boxed{5}$ Var	Afficher le so… our le calcul de la mo… etc.
$\boxed{6}$ MinMax	Afficher le sou… des MinMax pour obtenir les … minimales
$\boxed{7}$ Reg	Afficher le sous-… eg pour les calculs de rég…

$\boxed{1}\,\Sigma x^2$ $\boxed{3}\,\Sigma y^2$ $\boxed{5}\,\Sigma xy$

$\boxed{2}\,\Sigma x$ $\boxed{4}\,\Sigma y$ $\boxed{7}\,\Sigma x^2 y$

…

$\boxed{1}\,n$ $\boxed{2}\,\bar{x}$ $\boxed{3}\,x\sigma n$ $\boxed{4}\,x\sigma n{-}1$

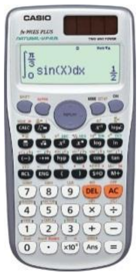
$\boxed{5}\,\bar{y}$ $\boxed{6}\,y\sigma n$ $\boxed{7}\,y\sigma n{-}1$

$\boxed{1}$ minX $\boxed{3}$ minY

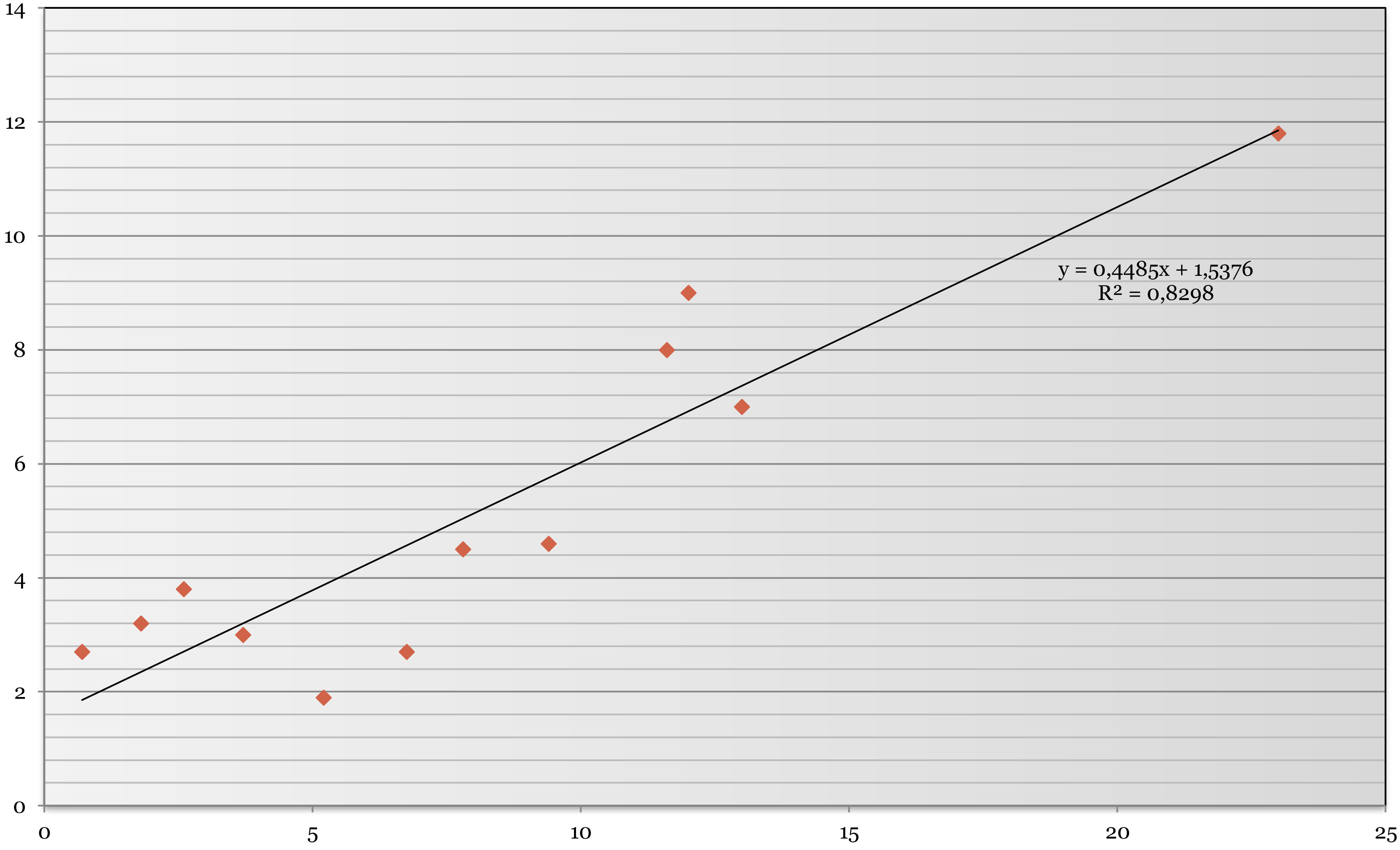
$\boxed{2}$ maxX $\boxed{4}$ maxY

$\boxed{1}\,A$ $\boxed{2}\,B$ $\boxed{3}\,r$

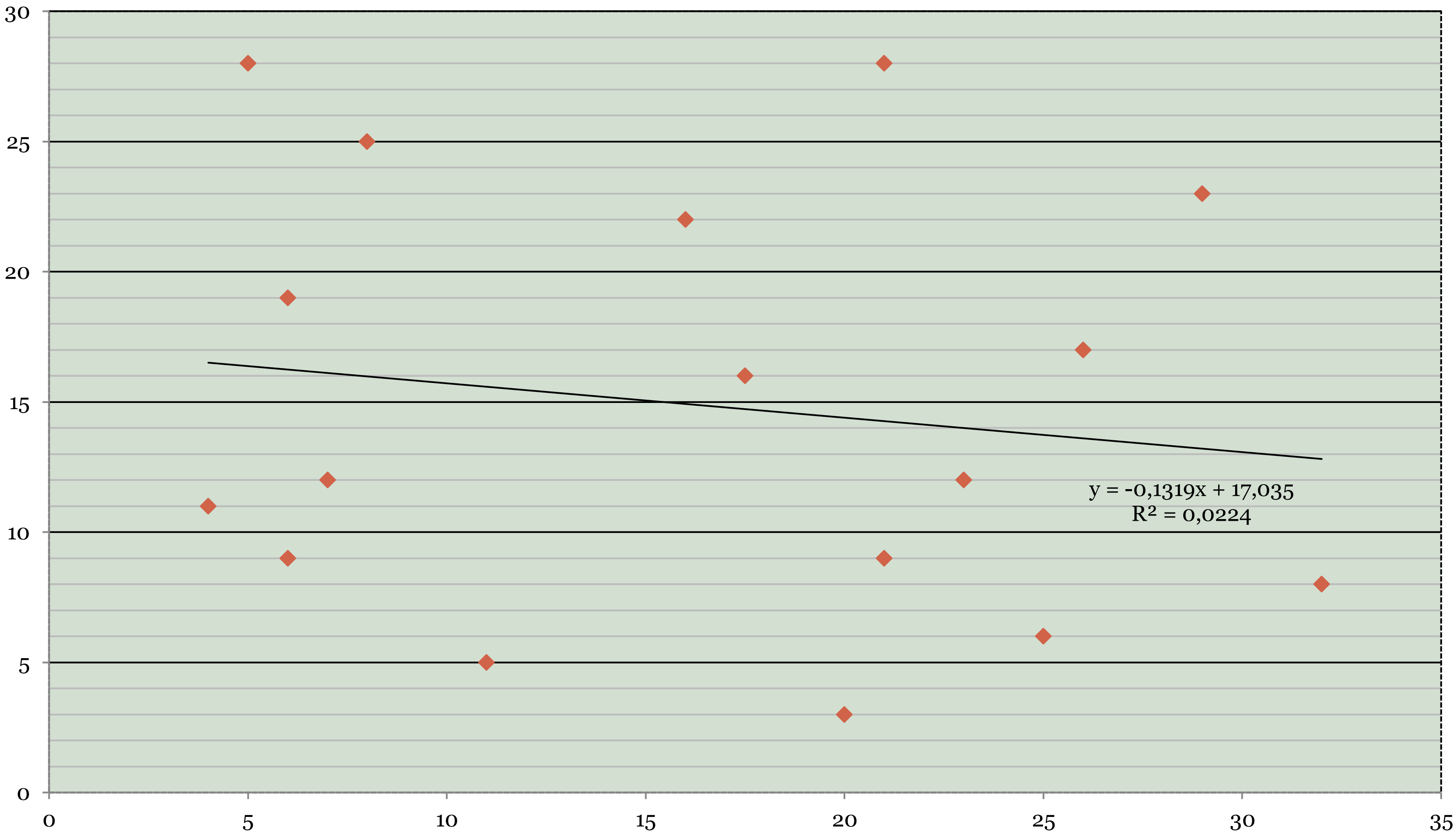
$\boxed{4}\,\hat{x}$ $\boxed{5}\,\hat{y}$



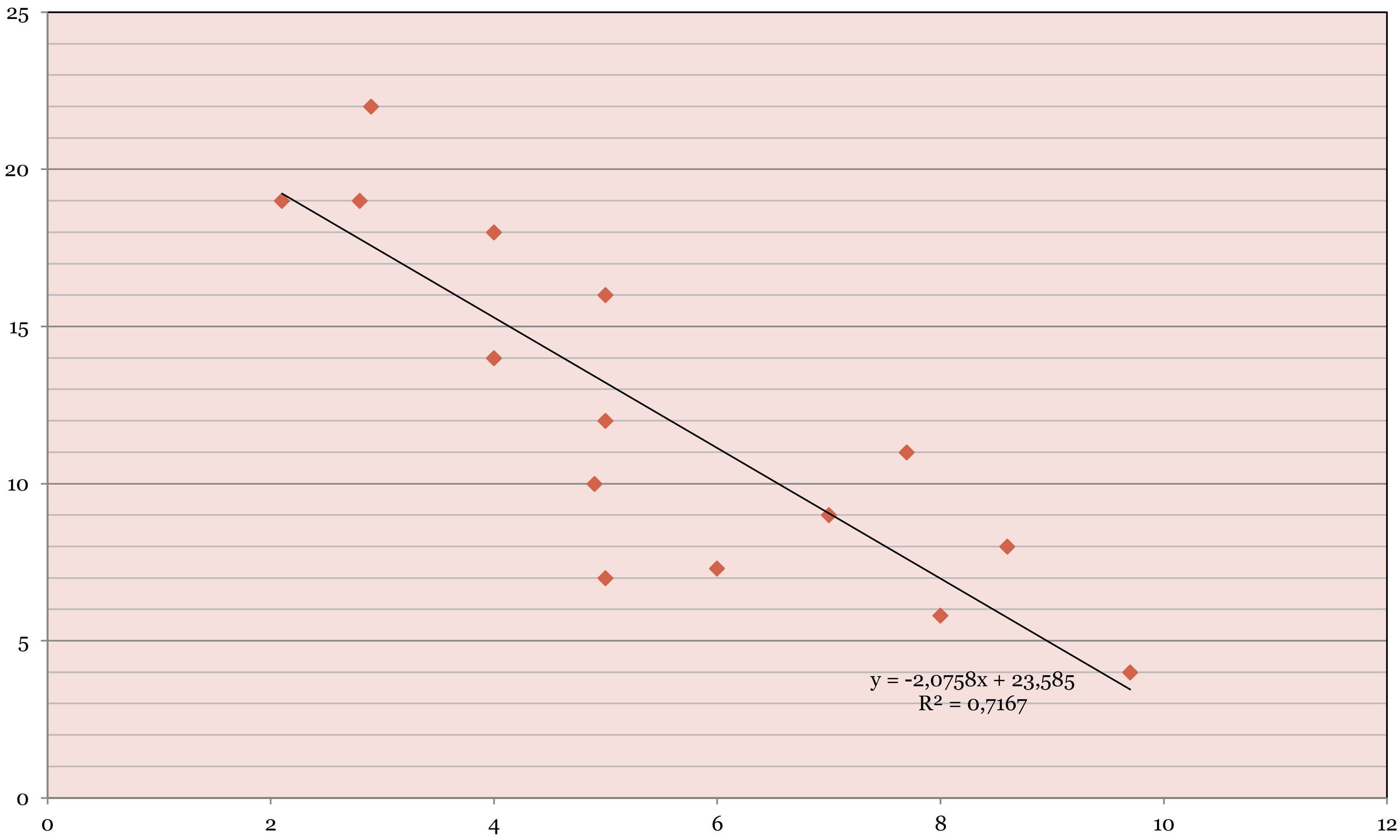
NUAGE 1



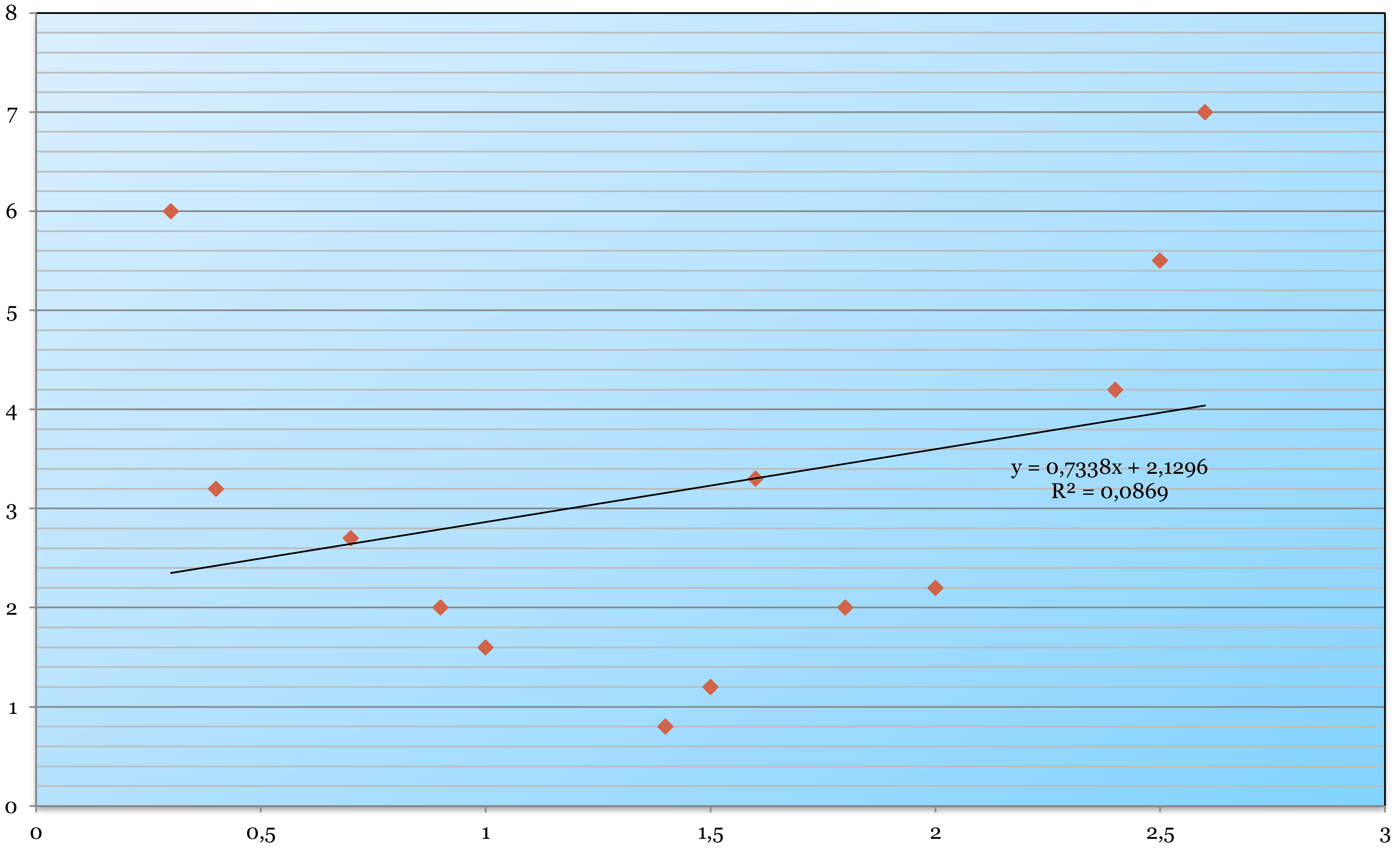
NUAGE 2



NUAGE 3



NUAGE 4



REFERENCES

B.PY, « La statistique sans formule mathématique », Pearson,2010.

H.Carnec, R.Seroux, J.M.Dagoury, M.Thomas, « Itinéraires en statistiques et probabilités »,ellipses,2011.

G.Laget, « Probabilités et Statistiques », Département de Mesures Physiques, IUT1 de Grenoble, http://maths.tetras.org/

M.Mountassir, « Probabilités et Statistiques »,Afrique Orient,2014.

P.Pacé, « Cours de Statistique », Licet,1986