STATISTIQUES

Vocabulaire des statistiques:

L'ensemble des méthodes qui permettent de recueillir, de classer, d'étudier et d'exploiter des phénomènes sur un groupe ou une famille d'objets est appelé la **Statistique**.

L'ensemble de référence sur lequel se déroule l'étude est appelé population statistique, tout élément de cet ensemble est un **individu statistique** (ou **unité**), la propriété des éléments de l'ensemble que l'on étudie est appelée : **caractère statistique** (ou **variable statistique**).

La collecte des résultats est appelée **relevé statistique**:

Si la population est trop importante on étudie seulement un sous-ensemble de la population totale. Dans ce cas là on dit que l'on procède à un **sondage**, le groupe étudié est alors appelé un **échantillon**.

Le caractère statistique peut à correspondre à des séries de mesure, on dit dans ce cas là qu'il est **quantitatif**, sinon on dit qu'il est **qualitatif**. Le caractère est **discret** lorsque les valeurs qu'il prend sont en nombre fini et peu important. S'il peut prendre toutes les valeurs d'un certain intervalle on dira qu'il est **continu**.

POPULATION		Parc automobile de la ville.	Le lycée.	Le lycée.	La cour du lycée.
INDIVIDU		Un véhicule.	Un élève.	Une classe.	Un arbre.
ECHANTILLON		Ensemble des véhicules garés au lycée.	Ens. des élèves de seconde.	Les classes du second cycle.	Arbres de plus de 2 mètres.
CARACTERE	QUANTITATIF	Age	Age Poids Taille	Nbre. d'élèves Nbre. de filles Nombre de redoublants	Taille
	QUALITATIF	Marque Couleur	Profession des parents Ethnie Loisir préféré	Matière principale	Essence
CLASSE		Les Peugeot Les véhicules de 10 ans	La classe 1 m 50 La classe des planteurs La classe sport	5 à 10 filles	Les manguiers

Série statistique simple

Pour une exploitation aisée, les résultats sont présentés dans des tableaux numériques et dans des graphiques. A cet effet on partage l'ensemble des valeurs prises par le caractère en plusieurs classes appelées **modalités.** Dans un caractère discret chaque classe peut-être l'une des valeurs prise par le caractère. Par contre dans le cas d'un caractère continu ou d'un caractère discret prenant beaucoup de valeurs on partage l'ensemble des valeurs possibles en intervalles (si possible de même amplitude). Le nombre d'individus appartenant à une classe donnée est appelé **effectif** de la classe. Lorsque le caractère est quantitatif, on présente les classes par valeurs croissantes du caractère. La somme des effectifs des classes dont le caractère est inférieur ou égal à un caractère k d'une classe C donnée est appelée **effectif cumulé** de la classe C.

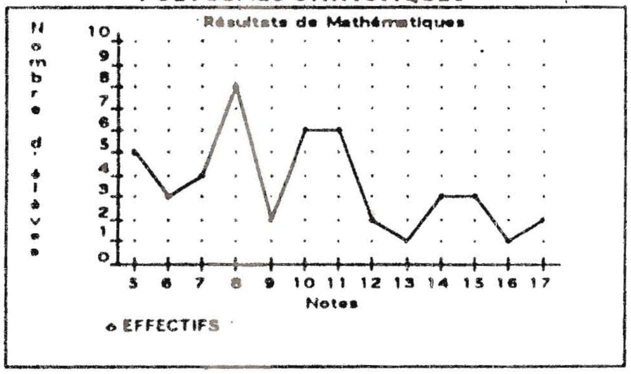
Exemple: Les moyennes trimestrielles en mathématiques des élèves d'une classe sont:

8	15	5	8	14	7	7	11	10	13	10	9	14	8	14	
15	10	8	7	17	12	12	5	11	10	16	10	8	5	7	
15	8	11	6	8	8	6	9	5	10	5	11	11	6	11	17

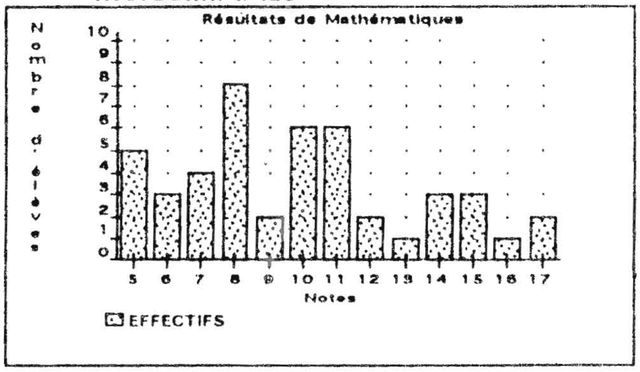
NOTES	5	6	7	8	9	10	11	12	13	14	15	16	17
EFFECTIFS	5	3	4	8	2	6	6	2	1	3	3	1	2
CUMULES	5	8	12	20	22	28	34	36	37	40	43	44	46

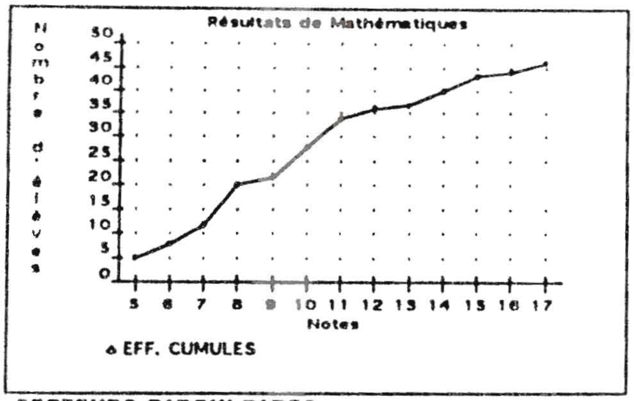
Les résultats sont ensuite représentés dans des diagrammes: **diagramme en bâtons** ou **polygone statistique, histogramme** et **diagramme circulaire.**

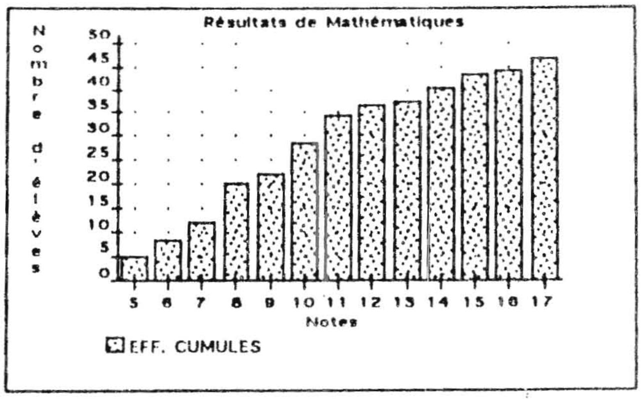
POLYGONES STATISTIQUES



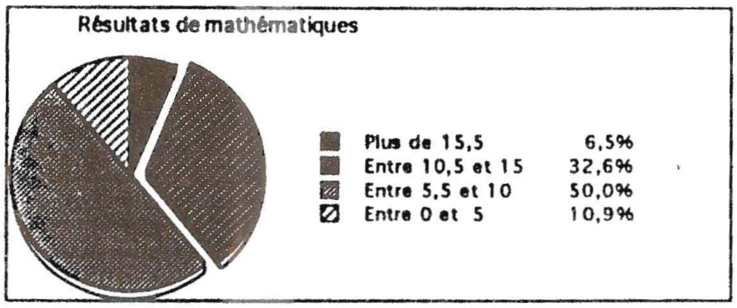
HISTOGRAMMES







SECTEURS CIRCULAIRES



On appelle **fréquence** d'une classe x_i le rapport entre l'effectif n_i de cette classe et l'effectif total N de la population. La **fréquence cumulée** d'une classe sera le rapport de l'effectif cumulé de cette classe avec l'effectif total ou la somme des fréquences des classes dont le caractère est inférieur ou égal à celui de la classe étudiée.

NOTES	5	6	7	8	9	10	11	12	13	14	15	16	17
FREQUENCE	0,11	0,07	0,09	0,17	0,04	0,13	0,13	0,04	0,02	0,07	0,07	0,02	0,04
FRE. CUMUL.	0,11	0,18	0,27	0,44	0,48	0,61	0,74	0,78	0,80	0,87	0,94	0,96	1,00

Remarque : la somme de toutes les fréquences est égale à 1. Chaque fréquence est un réel positif inférieur ou égal à 1.

Paramètres de position:

Dans les définitions suivantes les séries statistiques sont de caractère quantitatif.

On appelle **mode** d'une série statistique, toute valeur du caractère d'effectif (et donc de fréquence) maximum. Lorsque la série est exprimée en classe, on appelle **classe modale** toute classe d'effectif maximum.

On appelle **valeur médiane** d'une série statistique toute valeur m du caractère vérifiant les deux propriétés suivantes: -

- la moitié au moins des individus de la population ont une modalité inférieure ou égale à m.
- la moitié au moins des individus de la population ont une modalité supérieure oû égale à m.

Remarque : les deux paramétres : mode et valeur médiane sont peu utilisés.

Etant donnée une série statistique (x_i , n_i) on appelle **valeur moyenne** , ou **moyenne**, le réel

$$m = \frac{n_1 x_1 + n_2 x_2 + \dots\dots + n_p x_p}{n_1 + n_2 + \dots\dots + n_p}.$$

ou encore $m = f_1 x_1 + f_2 x_2 + \dots + f_n x_n$

Remarque: lorsque le caractère est continu et la série est exprimée en classe, on donnera à x_i une valeur moyenne de l'intervalle correspondant; dans ce cas là le résultat est une estimation de la moyenne, la moyenne exacte pouvant être difficile à calculer.

Exemple: La classe modale de notre exemple est 8 car 8 a le plus grand effectif

La classe médiane est entre 9 et 10 car l'effectif cumulé en 9 est 22 (inférieur à la moitié de l'effectif total) alors que la classe de 10 a un effectif cumulé de 28 (supérieur à la moitié de l'effectif total)

La moyenne exacte est de 9,84

Paramètres de dispersion

Pour toute série statistique (x_i , n_i)de caractère quantitatif et de valeur moyenne m, on appelle **écart moyen** le nombre (peu utilisé) :

$$e = \frac{n_1|x_1 - m| + n_2|x_2 - m| + \dots\dots + n_p|x_p - m|}{n_1 + n_2 + \dots\dots + n_p}.$$

On appelle **écart type moyen** le nombre σ positif dont le carré est défini par:

$$\sigma^2 = \frac{n_1(x_1 - m)^2 + n_2(x_2 - m)^2 + \dots\dots + n_p(x_p - m)^2}{n_1 + n_2 + \dots\dots + n_p}.$$

Remarque: σ^2 est appelé **variance** de la série statistique. On peut aussi calculer σ^2 avec:

$$\sigma^2 = \frac{n_1 x_1^2 + n_2 x_2^2 + \dots\dots + n_p x_p^2}{n_1 + n_2 + \dots\dots + n_p} - m^2$$

Exercice I: Influence du regroupement en classe pour le calcul de la moyenne :
les tailles en cm de 30 élèves sont donnés ci-dessous:
156, 157, 158, 158, 158, 160, 161, 164, 165, 165, 165, 167, 168, 168, 169, 169, 170, 170, 173, 173, 173, 173, 175, 176, 177, 180, 180, 183, 185, 188.

a) Déterminer la taille moyenne M des trente élèves.
b) Construire le diagramme en bâtons représentant la série statistique.
c) Opérer un regroupement par classes d'amplitude 5 centimètres en prenant pour classes : [155, 160[, [160, 165[... calculer la moyenne M' de la distribution obtenue en prenant pour valeur de chaque classe sa valeur centrale (157,5 ; 162,5 ; ...)
d) Construire l'histogramme montrant la répartition obtenue à la question c).
e) reprendre la question c) en choisissant des classes d'amplitude 10 centimètres ouvertes à gauche :]155, 160];]160, 165]; ... Soit M" la moyenne obtenue.
f) Construire l'histogramme montrant la répartition obtenue à la question e).
g) Comparer M, M' et M".
h) En reprenant les données de l'exercice, présenter les effectifs, effectifs cumulés, fréquences, fréquences cumulées dans un tableau et tracer les diagrammes correspondants. Déterminer le mode, la médiane et la moyenne ainsi que l'écart moyen et l'écart type.

Exercice II:
Dans votre classe étudier le caractère: " année de naissance, donner un effectif et la fréquence de chaque année. Représenter les fréquences dans un diagramme circulaire.

Activités I:

PAYS	Bénin	Cameroun	Côte d'Ivoire	Guinée Conakry	Guinée Bissau	Guinée Equatoriale	Libéria	Togo
X	827	1530	1300	358	72	57	400	403
Y	407	430	480	108	20	15	110	245

Le tableau ci-dessus donne pour huit pays d'Afrique, les productions X de manioc et Y de maïs , exprimées en milliers de tonnes, en 1990. Dans un repère orthonormal, où 1cm représente 10^5 tonnes, marquer pour chaque pays e, le point de coordonnées (X(e), Y(e)). De l'examen des points obtenus peut-on en conclure à l'existence d'une relation entre X et Y.

Activités II

Mathématiques :	3	13	6	11	2	10	7	10	13	10	9	12	0	10	17
Langue Vivante :	8	10	6	7	6	6	12	20	6	14	11	14	7	15	12

Le tableau ci-dessus donne les notes X de mathématiques et Y de langue vivante, sur 20, obtenues par quinze candidats au baccalauréat D. Dans un repère orthonormal, marquer pour chaque élève, le point de coordonnées (x, y)où x et y sont respectivement les notes de mathématiques et de langue vivante. De l'examen des points obtenus peut-on en conclure à l'existence d'une relation entre X et Y.

Activités III.

Y \ X	0	1	2	3	4	5
1	6	4	1	0	0	0
2	3	11	10	5	1	0
3	1	3	16	13	4	1
4	0	1	3	5	8	4

Ce tableau représente, pour un échantillon de 100 familles, le nombre X d'enfants et le nombre Y de chambres occupées. Chaque case correspond donc au nombre de familles possedant x enfants et occupant y chambres. Dans un repère orthonormal représenter le nuage de points (l'ensemble des couples (X, Y)) correspondant à la série étudiée. La grosseur des points représentant un couple (X, Y) sera proportionelle au nombre de familles possédant simultanément ces deux caractères

Disposition pratique permettant d'étudier avec le plus de sécurité une série statistique simple

Classes	X_i	n_i	n_iX_i	$n_iX_i^2$
[4,5 ; 5,5[	5	5	25	125
[5,5 ; 6,5[	6	3	18	108
[6,5 ; 7,5[	7	4	28	196
[7,5 ; 8,5[	8	8	64	512
[8,5 ; 9,5[	9	2	18	162
[9,5 ; 10,5[	10	6	60	600
[10,5 ; 11,5[	11	6	66	726
[11,5 ; 12,5[	12	2	24	288
[12,5 ; 13,5[	13	1	13	169
[13,5 ; 14,5[	14	3	42	588
[14,5 ; 15,5[	15	3	45	675
[15,5 ; 16,5[	16	1	16	256
[16,5 ; 17,5[	17	2	34	578
	Totaux	46	453	4983
		Moyenne	9,85	
		Variance	11,35	
		Ecart type	3,37	

Effectif Total N = 46

453/46

$4983/46 - (453/46)^2$

Racine carrée de la variance prise de préférence sur la valeur approchée avec le maximum de décimales donnée par votre calculatrice

FORMULAIRE

$\overline{X} = \dfrac{x_1 + x_2 + \ldots + x_n}{n} = \dfrac{\sum X}{n}$ $\overline{Y} = \dfrac{y_1 + y_2 + \ldots + y_n}{n} = \dfrac{\sum Y}{n}$ $V(X) = \dfrac{x_1^2 + x_2^2 + \ldots + x_n^2}{n} - \overline{X}^2 = \dfrac{\sum X^2}{n} - \overline{X}^2$ $V(Y) = \dfrac{y_1^2 + y_2^2 + \ldots + y_n^2}{n} - \overline{Y}^2 = \dfrac{\sum Y}{n} - \overline{Y}^2$ $Cov(X,Y) = \dfrac{x_1 y_1 + x_2 y_2 + \ldots + x_n y_n}{n} - \overline{X}\,\overline{Y} = \dfrac{\sum XY}{n} - \overline{X}\,\overline{Y}$	$a = \dfrac{Cov(X,Y)}{V(X)}$ $b = \overline{Y} - a\overline{X}$ $\alpha = \dfrac{Cov(X,Y)}{V(Y)}$ $\beta = \overline{X} - \alpha\overline{Y}$ $r = \dfrac{Cov(X,Y)}{\sqrt{V(X)\,V(Y)}}$ ou $\lvert r \rvert = \sqrt{a\alpha}$

Equation de la droite de régression de Y en X, (D) : $y = a\,x + b$
Equation de la droite de régression de X en Y, (Δ) : $x = \alpha\,y + \beta$.

i	X_i	Y_i	X_i^2	Y_i^2	X_iY_i
1	48	5	2304	25	240
2	43	0	1849	0	0
3	86	18	7396	324	1548
4	72	5	5184	25	360
5	48	2	2304	4	96
6	58	7	3364	49	406
7	144	24	20736	576	3456
8	62	10	3844	100	620
9	96	12	9216	144	1152
10	58	10	3364	100	580
11	72	15	5184	225	1080
12	77	14	5929	196	1078
13	43	5	1849	25	215
14	58	8	3364	64	464
15	120	20	14400	400	2400
Totaux	1085	155	90287	2257	13695
	$\sum X$	$\sum Y$	$\sum X^2$	$\sum Y^2$	$\sum XY$

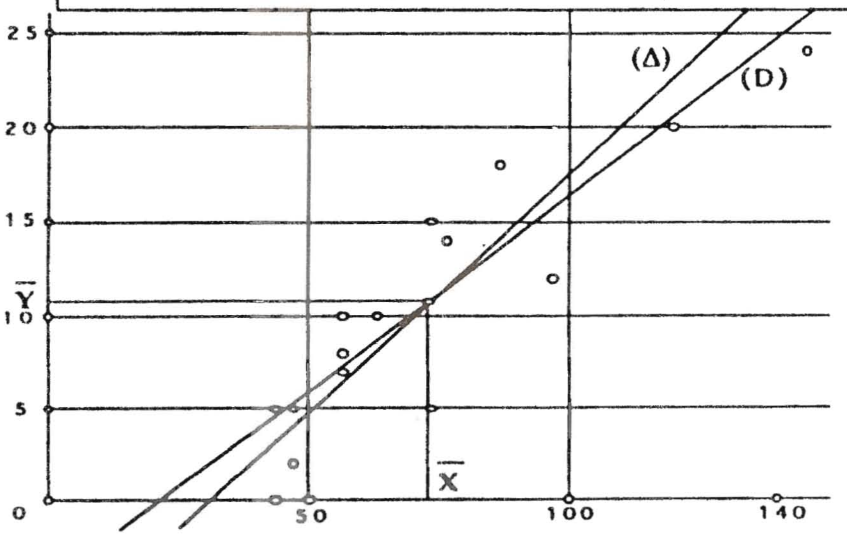
Le tableau ci-contre donne, pour 15 salariés interrogés au cours d'une enquête, le revenu X mensuel et l'épargne Y correspondante ; X et Y sont exprimés en milliers de Francs. Votre mission, si vous l'acceptez, est de trouver une relation entre X et Y, si elle existe.

Les couples (Xi ; Yi) ont un effectif égal à 1, l'effectif total est donc 15

X = 72,33 a = 0,21
Y = 10,33 b = - 4,88
V(X) = 787,02 a = 3,79
V(Y) = 43,69 b = 33,18
Cov(X,Y) = 165,56 r = 0,89
(D) : y = 0,21 x - 4,88
(Δ) : x = 3,79 y + 33,18

Les résultats ci-dessus sont arrondis au centième. Dans vos calculs successifs, utilisez une meilleure précision (celle de vos calculatrices).
Exemple : V(X) = 90287/15 - (72,33)2 donne 787,50 soit une erreur de 0,48

Représentation graphique du nuage de points



Vérifiez que le point de coordonnées ($\overline{X}$, $\overline{Y}$) appartient aux deux droites (D) et (Δ).

TRAVAUX DIRIGES 1°S1

On a effectué une recherche sur la vitesse de propagation de l'influx nerveux dans une fibre nerveuse en fonction du diamètre de la fibre.

On désigne par X le diamètre en microns de la fibre nerveuse et par Y la vitesse en mètres par seconde de l'influx nerveux dans la fibre de diamètre X.

Les résultats sont donnés dans le tableau ci-dessous:

X	2,1	3,2	3,6	3,9	5,1	5,4	6,3	6,7
Y	15	14	18	22	26	38	35	44

X	8	8,8	9,6	11	12,3	13,4	14,2	14,4
Y	50	54	60	72	80	82	84	90

1°) Représenter le nuage de points associé à cette série statistique double. On prendra 1 cm en abscisse pour 1 micron et 1 cm en ordonnée pour 10 m/s.

2°) Calculer les moyennes de chacune des séries de caractère X et Y.

3°) Calculer les variances V(X) et V(Y) de chacune des séries de caractère X et Y.

4°) Trouver les équations des droites de régression de Y en fonction de X et de X en fonction de Y.

5°) Construire ces droites.

6°) Utiliser ces droites pour conjecturer

a) le diamètre d'une fibre nerveuse à travers laquelle, dans les mêmes conditions que l'expérience précédente, l'influx nerveux se propagerait à la vitesse de 100 m/s.

b) la vitesse de l'influx nerveux à travers une fibre nerveuse de diamètre 18 microns.

TRAVAUX DIRIGES 1°S1

On a effectué une recherche sur la vitesse de propagation de l'influx nerveux dans une fibre nerveuse en fonction du diamètre de la fibre.

On désigne par X le diamètre en microns de la fibre nerveuse et par Y la vitesse en mètres par seconde de l'influx nerveux dans la fibre de diamètre X.

Les résultats sont donnés dans le tableau ci-dessous:

X	2,1	3,2	3,6	3,9	5,1	5,4	6,3	6,7
Y	15	14	18	22	26	38	35	44

X	8	8,8	9,6	11	12,3	13,4	14,2	14,4
Y	50	54	60	72	80	82	84	90

1°) Représenter le nuage de points associé à cette série statistique double. On prendra 1 cm en abscisse pour 1 micron et 1 cm en ordonnée pour 10 m/s.

2°) Calculer les moyennes de chacune des séries de caractère X et Y.

3°) Calculer les variances V(X) et V(Y) de chacune des séries de caractère X et Y.

4°) Trouver les équations des droites de régression de Y en fonction de X et de X en fonction de Y.

5°) Construire ces droites.

6°) Utiliser ces droites pour conjecturer

a) le diamètre d'une fibre nerveuse à travers laquelle, dans les mêmes conditions que l'expérience précédente, l'influx nerveux se propagerait à la vitesse de 100 m/s.

b) la vitesse de l'influx nerveux à travers une fibre nerveuse de diamètre 18 microns.