Les premières études statistiques étaient des recensements démographiques : on en a conservé le vocabulaire.

Population : C'est l'ensemble sur lequel porte l'étude statistique.

Individu : C'est un élément de la population.

Caractère : C'est l'aspect que l'on observe sur les individus. Un caractère permet de déterminer une partition de la population selon ses diverses valeurs (par exemple le genre est un caractère à deux modalités : masculin ou féminin).

Lorsque les différentes valeurs d'un caractère sont des nombres, le caractère est *quantitatif.* Dans le cas contraire, le caractère est *qualitatif.*

I EFFECTIFS - FRÉQUENCES

1 EFFECTIFS ET FRÉQUENCES

L'effectif d'une valeur du caractère étudié est le nombre d'individus de la population ayant cette valeur.
La fréquence d'une valeur est le quotient de l'effectif de cette valeur par l'effectif total de la population.
(la fréquence peut être exprimée en pourcentage)

$$\text{féquence} = \frac{\text{effectif de la valeur}}{\text{effectif total}}$$

2 EFFECTIFS CUMULÉS ET FRÉQUENCES CUMULÉES

On étudie un caractère quantitatif dans une population.

— Les différentes valeurs x_i du caractère quantitatif constituent une série statistique notée (x_i).

— On note n_i l'effectif de la valeur x_i.

L'effectif cumulé croissant de la valeur x_i est la somme des effectifs de toutes les valeurs inférieures ou égales à x_i.
La fréquence cumulée croissante de la valeur x_i est la somme des fréquences de toutes les valeurs inférieures ou égales à x_i.

EXEMPLE

Dans un service de maintenance, on a répertorié le nombre d'interventions par jour sur un mois. On a obtenu la distribution suivante :

Nombre d'interventions x_i	3	5	6	7	8	9
Nombre de jours n_i	2	4	9	6	3	1

Le nombre total de journées d'intervention est $2+4+9+6+3+1=25$. Les fréquences des différentes valeurs du nombre d'intervention sont :

Nombre d'interventions x_i	3	5	6	7	8	9
Nombre de jours n_i	2	4	9	6	3	1
Fréquence $f_i = \dfrac{n_i}{25}$	0,08	0,16	0,36	0,24	0,12	0,04

Le tableau suivant donne les effectifs cumulés croissants ainsi que les fréquences cumulées croissantes :

Nombre d'interventions x_i	3	5	6	7	8	9
Nombre de jours n_i	2	4	9	6	3	1
Effectif cumulé	2	6	15	21	24	25
Fréquence cumulée	0,08	0,24	0,6	0,84	0,96	1

II CARACTÉRISTIQUES D'UNE SÉRIE STATISTIQUE

1 CARACTÉRISTIQUES DE POSITION

LA MOYENNE

On considère la série statistique donnée par le tableau ci-contre.
On note $N = n_1 + n_2 + \cdots + n_p$ l'effectif total

Valeur x_i	x_1	x_2	$\cdots$	x_p
Effectif n_i	n_1	n_2	$\cdots$	n_p

La moyenne d'une série statistique est le quotient noté $\overline{x}$ de la somme de toutes les valeurs de cette série par l'effectif total.

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{N}$$

REMARQUE

Soit $f_i = \dfrac{n_i}{N}$ la fréquence de la valeur x_i alors, la moyenne $\overline{x} = f_1x_1 + f_2x_2 + \cdots + f_px_p$.

EXEMPLE

Avec la série statistique précédente :

Nombre d'interventions x_i	3	5	6	7	8	9
Nombre de jours n_i	2	4	9	6	3	1
Fréquence f_i	0,08	0,16	0,36	0,24	0,12	0,04

Le nombre moyen d'interventions par jour est :

$$\overline{x} = \frac{2\times3+4\times5+9\times6+6\times7+3\times8+1\times9}{25} = 6{,}2$$

ou en utilisant les fréquences :

$$\overline{x} = 0{,}08\times3+0{,}16\times5+0{,}36\times6+0{,}24\times7+0{,}12\times8+0{,}04\times9 = 6{,}2$$

LA MÉDIANE

La médiane d'une série statistique est une valeur telle qu'il y ait autant d'observations ayant une valeur supérieure à la médiane que d'observations ayant une valeur inférieure à la médiane.

La médiane d'une série statistique de N valeurs rangées par ordre croissant est le nombre M_e défini par :

- si l'effectif N est impair, la médiane M_e est la valeur centrale du caractère c'est à dire la valeur de rang $\dfrac{N+1}{2}$ de la série ordonnée.
- si l'effectif N est pair, la médiane M_e est la demi-somme des deux valeurs centrales du caractère c'est à dire la moyenne des valeurs de rangs $\dfrac{N}{2}$ et $\dfrac{N}{2}+1$ de la série ordonnée.

EXEMPLE

Dans la série précédente, l'effectif total $N = 25$ donc la médiane est la valeur du caractère de rang 13 soit $Me = 6$.

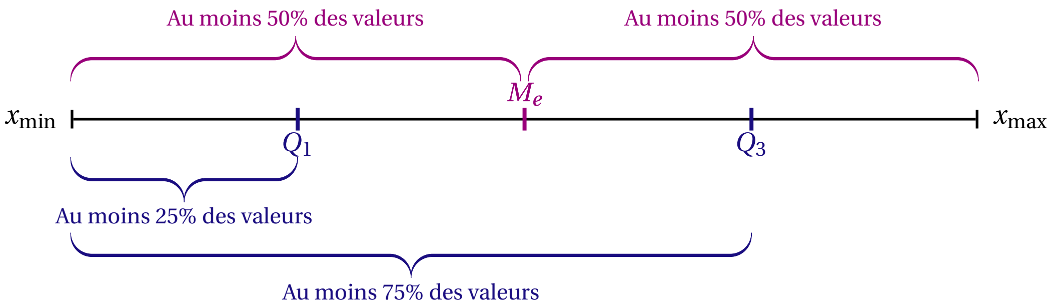
LES QUANTILES

1. LES QUARTILES

Les quartiles au nombre de trois Q_1, Q_2 et Q_3 partagent l'ensemble étudié de N éléments préalablement classés par valeurs croissantes, en quatre sous ensembles.

— Le premier quartile noté Q_1 est la plus petite valeur de la série statistique telle qu'au moins 25 % des valeurs de la série sont inférieures ou égales à Q_1.

— Le troisième quartile noté Q_3 est la plus petite valeur de la série statistique telle qu'au moins 75 % des valeurs de la série sont inférieures ou égales à Q_3.



2. LES DÉCILES

Les déciles au nombre de neuf $D_1, D_2, \dots, D_9$ partagent l'ensemble étudié de N éléments préalablement classés par valeurs croissantes, en dix sous ensembles.

— Le premier décile noté D_1 est la plus petite valeur de la série statistique telle qu'au moins 10 % des valeurs de la série sont inférieures ou égales à D_1.

— Le neuvième décile noté D_9 est la plus petite valeur de la série statistique telle qu'au moins 90 % des valeurs de la série sont inférieures ou égales à D_9.

REMARQUE

Le deuxième quartile Q_2 et le cinquième décile D_5 sont égaux à la médiane.

VOCABULAIRE

La moyenne, la médiane, les quantiles sont appelés caractéristiques de position d'une série statistique.

2 CARACTÉRISTIQUES DE DISPERSION

— L'étendue est la différence entre la plus grande et la plus petite valeur d'une série statistique.

— L'écart interquartile est égal à la différence entre le troisième et le premier quartiles.

— L'écart interdécile est égal à la différence entre le neuvième et le premier déciles.

3 BOÎTES À MOUSTACHES

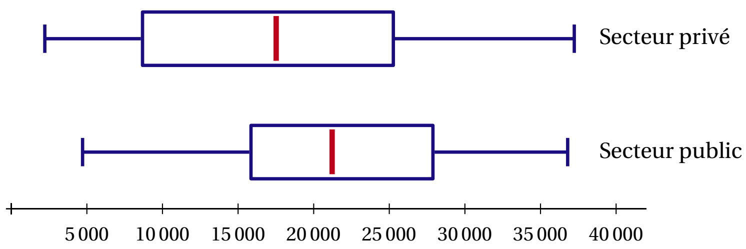
La représentation graphique de la dispersion d'une série statistique se fait à l'aide de diagramme en boîte appelés aussi « boîte à moustaches » ou « box-plot ».
Pour une catégorie donnée, on construit, en face d'un axe permettant de repérer les quantiles de la variable étudiée, un rectangle dont la longueur est égale à l'écart interquartile $Q_3 - Q_1$, la médiane est représentée par un trait. On ajoute alors des segments aux extrémités menant jusqu'aux valeurs extrêmes, ou jusqu'aux premier et neuvième déciles.

EXEMPLE

Le tableau suivant donne la distribution du revenu salarial par secteur d'activité en France en 2014.

	D1	Q1	Médiane	Q3	D9
Secteur privé	2 218	8 570	17 520	25 377	37 234
Secteur public	4 716	15 744	21 221	27 996	36 797

Source : INSEE



III TEST STATISTIQUE

On considère une population d'effectif N, très grand, dans laquelle on connaît la proportion p des individus présentant un certain caractère.

1 ÉCHANTILLONNAGE

ÉCHANTILLON

Une urne contient des boules blanches et des boules noires en proportion p.
On tire au hasard une boule dans l'urne, on note sa couleur, puis on la replace dans le sac.
En répétant n fois ce procédé, on constitue un échantillon aléatoire de taille n de la population des boules de l'urne.

> Un échantillon de taille n est constitué de n éléments pris au hasard dans la population dont l'effectif N est suffisamment grand par rapport à l'effectif n de l'échantillon pour considérer que le tirage des éléments de l'échantillon s'effectue avec remise.

FLUCTUATION D'ÉCHANTILLONNAGE

Si on prélève plusieurs échantillons d'effectif n, où $n \geqslant 30$, dans la même population, on constate que la fréquence du caractère observé fluctue autour de la proportion p du caractère. Ce phénomène dû au hasard dans la constitution des échantillons est appelé *fluctuation d'échantillonnage*.

2 INTERVALLE DE FLUCTUATION

On prélève un échantillon dans la population et on note f la fréquence d'apparition du caractère observé dans cet échantillon.

INTERVALLE DE FLUCTUATION AU SEUIL DE 95%

> Un intervalle de fluctuation de la fréquence f au seuil de 95% relatif aux échantillons de taille n est un intervalle I tel que, pour au moins 95% de l'ensemble des échantillons possibles, la fréquence observée appartient à I.

PROPRIÉTÉ

> Pour une proportion p comprise entre 0,2 et 0,8, et des échantillons de taille $n \geqslant 25$, l'intervalle $\left[p-\dfrac{1}{\sqrt{n}}\,;p+\dfrac{1}{\sqrt{n}}\right]$ est un intervalle de fluctuation au seuil de 95% de la fréquence f observée.

PRISE DE DÉCISION

Selon la situation étudiée, l'intervalle de fluctuation au seuil de 95% peut permettre :

— de décider si l'échantillon est représentatif de l'ensemble de la population ;

— d'accepter ou pas l'hypothèse que la proportion d'individus présentant le caractère étudié est égale à p.

EXEMPLE

En première partie de soirée une nouvelle série a attiré près de 5,7 millions de téléspectateurs soit 28% de part d'audience.
Pour estimer le degré de satisfaction de cette série auprès du public, on réalise une enquête auprès de 300 personnes. On constate que 96 personnes sur les 300 personnes interrogées ont regardé cette série.
Ce sondage remet-il en question la part d'audience de 28% de cette série ?

1. La fréquence observée de la part d'audience de la série dans l'échantillon de taille 300 est :

$$f = \frac{96}{300} = 0{,}32$$

2. On a ici $p = 0{,}28$ et $n = 300$. Un intervalle de fluctuation au seuil de 95 % de la part d'audience de la série dans les échantillons de taille 300 est :

$$I = \left[0{,}28 - \frac{1}{\sqrt{300}} ; 0{,}28 + \frac{1}{\sqrt{300}}\right]$$

   Soit avec des valeurs approchées à 10^{-2} près des bornes de l'intervalle, un intervalle de fluctuation au seuil de 95 % de la part d'audience de la série est $I = [0{,}22; 0{,}34]$

3. Comme $f = 0{,}32$ appartient à l'intervalle de fluctuation $I = [0{,}22; 0{,}34]$, on accepte l'hypothèse selon laquelle la part d'audience de la série est de 28%.

EXERCICE 1

Pour les trois séries statistiques ci dessous, la médiane est égale à 10. Compléter le tableau ci-dessous par des données pour chacune des séries sachant que :

— La moyenne de la série 1 est égale à 10.

— La moyenne de la série 2 est la plus petite possible.

— La moyenne de la série 3 est la plus grande possible.

indice i	1	2	3	4	5	6	7	8	9	10
Série 1	5									18
Série 2	5									18
Série 3	5									18

EXERCICE 2

Dans une entreprise, il y a 28 cadres et 92 ouvriers. Le salaire moyen des cadres est de 3 450 € et celui des ouvriers est de 1 320 €.

1. Calculer le salaire moyen de l'ensemble des salariés de cette entreprise.
2. a) Quel est le pourcentage d'augmentation du salaire moyen si on verse une prime de 35 € à chaque salarié ?

   b) On augmente le salaire de chaque cadre de 2 % et celui de chaque ouvrier de 4 %.
   Le salaire moyen dans l'entreprise a-t-il augmenté de 3% ?

EXERCICE 3

Un concours est organisé dans deux centres d'examens. Dans le premier centre, les garçons ont obtenu 13 de moyenne et les filles 12 de moyenne. Dans le second centre, les garçons ont obtenu 9 de moyenne et les filles 8 de moyenne.
Le président du jury en déduit que les garçons ont eu de meilleurs résultats que les filles Est-ce si sûr ?
Sachant qu'il y avait 58 garçons et 104 filles dans le premier centre, et 87 garçons et 32 filles dans le second centre, calculer la moyenne générale des garçons puis celle des filles. Conclure.

EXERCICE 4

PARTIE A

Le tableau suivant donne la distribution des salaires mensuels nets, en euros, en France en 2015.

Déciles	D1	D2	D3	D4	D5	D6	D7	D8	D9
Hommes	1 262	1 427	1 573	1 728	1 906	2 130	2 451	2 996	3 990
Femmes	1 171	1 288	1 396	1 512	1 650	1 830	2 073	2 432	3 149
Ensemble	1 213	1 357	1 490	1 630	1 797	2 004	2 286	2 752	3 646

Souce : Insee Première (*octobre 2017)*

1. a) Donner le salaire net médian des salariés hommes et des salariées femmes.

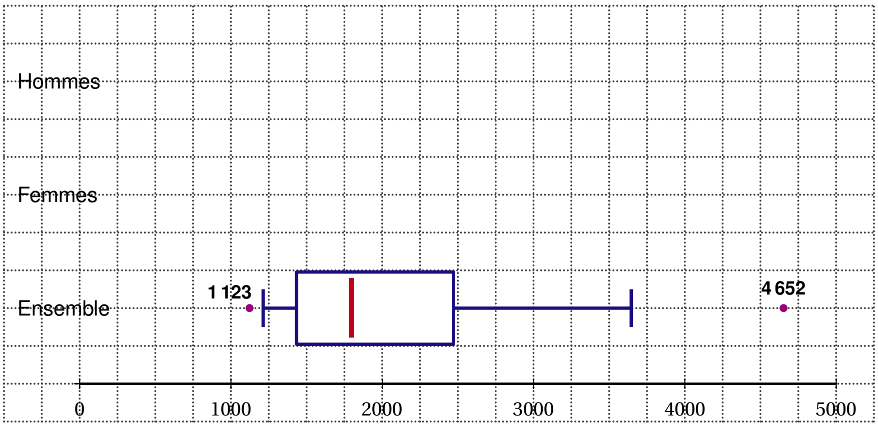
   b) Calculer les variations en pourcentage des déciles des femmes par rapport à ceux des hommes

Déciles	D1	D2	D3	D4	D5	D6	D7	D8	D9
Hommes	1 262	1 427	1 573	1 728	1 906	2 130	2 451	2 996	3 990
Femmes	1 171	1 288	1 396	1 512	1 650	1 830	2 073	2 432	3 149
Écarts en %	−7,2								

   c) Recopier et compléter la phrase :
   « Les écarts de salaire entre femmes par rapport aux hommes . . . le long de l'échelle salariale : de −7,2% pour le 1er décile à . . . pour le 9e décile. »

2. Le montant en euros du premier quartile est de 1 488 euros pour les salariés hommes et de 1 332 euros pour les salariées femmes. Le troisième quartile est de 2 678 euros pour les hommes et de 2 208 euros pour les femmes.

   La distribution des salaires mensuels nets de l'ensemble des salariés est représentée ci-dessous (centiles C5 à C95).



   a) Donner une interprétation du nombre 4 652.

   b) Sur le même graphique, représenter la distribution des salaires nets des hommes et des femmes.

PARTIE B

Le tableau ci-dessous donne le montant en euros du salaire mensuel moyen net selon les catégories socioprofessionnelles en France en 2015.

	Hommes		Femmes	
	Salaires nets (en euros)	Effectifs (en %)	Salaires nets (en euros)	Effectifs (en %)
Cadres	4 451	20,6	3 561	15,6
Professions intermédiaires	2 420	18,9	2 081	20,8
Employés	1 739	16,2	1 591	50,8
Ouvriers	1 765	44,3	1 483	12,8

Souce : Insee Première (*octobre 2017)*

1. a) Calculer le salaire moyen des hommes et celui des femmes.

   b) Recopier et compléter la phrase :

   « En 2015, une salariée gagne, en moyenne . . . % de moins qu'un salarié. »

2. 58,5% des salariés sont des hommes.

   De quel pourcentage, le salaire net médian de l'ensemble des salariés est-il inférieur au salaire net moyen ?

EXERCICE 5

1. Compléter le tableau ci-dessous qui donne la distribution des salaires mensuels bruts des 100 salariés d'une entreprise.

Salaires en euros	1500	1600	1900	2400	2700	3200	5000
Effectifs	30	25	15	12	8	6	4
Fréquences							

2. a) Donner le montant du salaire mensuel brut médian.

   b) Calculer le pourcentage de la masse salariale totale perçue par les 10% des salariés les mieux rémunérés.

3. Calculer le montant du salaire mensuel brut moyen.

EXERCICE 6

Le tableau suivant donne le montant des salaires annuels exprimés en milliers d'euros d'une petite entreprise.

Salaires	16	18	20	25	30	40
Nombre de salariés	6	9	10	8	5	2

1. Déterminer la médiane, le premier et le troisième quartiles. Interpréter ces résultats et les traduire à l'aide d'un diagramme en boîte.
2. Calculer le montant en euros du salaire moyen annuel de cette entreprise.
3. Soit S la fonction définie sur $\mathbb{R}$ par :

$$S(x) = 6 \times (16-x)^2 + 9 \times (18-x)^2 + 10 \times (20-x)^2 + 8 \times (25-x)^2 + 5 \times (30-x)^2 + 2 \times (40-x)^2$$

   a) Vérifier que $S(x) = 40x^2 - 1776x + 21152$

   b) Déterminer le sens de variations de la fonction S.

EXERCICE 7

Une entreprise de produits alimentaires fabrique et distribue une marque de café dans des sachets de 250 grammes. On suppose que le poids du sachet vide est négligeable.
La machine utilisée pour remplir les sachets est contrôlée selon la procédure suivante.
À chaque heure, un échantillon aléatoire de 30 sachets est prélevé dans la production ; on mesure la masse de chaque sachet et on calcule la masse moyenne $\overline{x}$ de l'échantillon.
Un réglage de la machine est nécessaire si l'un des critères suivants n'est pas vérifié :

— les 30 sachets ont une masse supérieure ou égale à 244 grammes ;

— 50% au moins des sachets ont une masse en grammes comprise dans l'intervalle $[248;252]$;

— 95% au moins des sachets ont une masse en grammes comprise dans l'intervalle $[245;256]$.

Au cours de la production, l'échantillon suivant a été prélevé.

253	249	250	248	244	252	250	250	251	248
251	255	253	252	252	249	253	248	251	253
246	250	245	246	257	251	252	252	250	251

1. Recopier et compléter le tableau suivant :

Masse	244	245	246	247	248	249	250	251	252	253	254	255	256	257
Effectif														

2. Représenter la dispersion de cette série à l'aide d'un diagramme en boîte.
3. Faut-il effectuer un réglage de la machine ?

EXERCICE 8

Chez un fabricant de lames de parquet en chêne rustique on indique :

— longueur moyenne des lames : 45 cm ;

— lames de longueur inférieure à 35 cm : 23 %.

On prélève un échantillon de 60 lames dans le stock, pour vérification. On constate que 18 lames ont une longueur inférieure à 35 cm.

1. Déterminer un intervalle de fluctuation au seuil de 95 % de la fréquence des lames dont la longueur est inférieure à 35 cm dans les échantillons de taille 60.
2. L'affirmation « 23% des lames ont une longueur inférieure à 35 cm » est-elle remise en cause ?

EXERCICE 9

Le directeur commercial affirme que 80 % des consommateurs sont satisfaits de la qualité des produits commercialisés par son entreprise.
On réalise une étude de satisfaction sur un échantillon de 120 personnes. Parmi les personnes interrogées, 90 déclarent être satisfaites des produits.
Déterminer, en justifiant, si l'on doit remettre en question l'affirmation du directeur commercial.

EXERCICE 10

Selon une publication de l'INSEE, 28 % des ménages comprennent une famille avec au moins un enfant mineur.

1. On interroge un échantillon de 100 ménages choisis au hasard, et on constate que dans cet échantillon 35 ménages comprennent une famille avec au moins un enfant mineur.

   Cet échantillon est-il représentatif de la population ?

2. On interroge au hasard 300 ménages qui résident dans le même arrondissement d'une grande agglomération, et on constate également que 105 ménages comprennent une famille avec au moins un enfant mineur.

   Cet échantillon est-il représentatif de la population ?