Glossaire de statistique pour les professeurs de mathématiques de collège

Glossaire à l'usage des professeurs (ou futurs professeurs) de mathématiques ; c'est sur des exemples que les notions seront introduites auprès des élèves.

Nous définissons tout d'abord le vocabulaire statistique concernant les *données observées sur une population finie* (une classe d'élèves par exemple), puis le vocabulaire supplémentaire concernant les données observées sur un *échantillon* extrait d'une population finie, enfin, nous reprenons l'interprétation des mêmes termes dans le contexte de *données expérimentales* (durée d'un même trajet chaque jour ouvrable de huit semaines consécutives, même moyen de locomotion, départ à la même heure, …).

Les données observées sur une population finie

étude statistique	C'est dans le but de rechercher de l'information sur une question donnée que l'on peut entreprendre une *étude statistique* ; il s'agit alors de définir la *population* et les différents *caractères* (appelés aussi *variables*) définis sur cette population qui pourront apporter l'information cherchée.
population (statistique)	ensemble fini et homogène auprès duquel on recherche l'information, noté $E = \{e_1, ..., e_n\}$ Ce peut être une population humaine ou une population d'entreprises ou de pays, … Cet ensemble n'est pas ordonné. Lorsque l'ensemble est ordonné, les douze mois d'une année donnée par exemple, les données observées sont appelées *séries chronologiques* ; le questionnement, et donc le traitement statistique, diffèrent en général de ce qui est présenté ici.
individus (statistiques)	ou *unités statistiques* : éléments de la *population*
*caractère (*ou *variable)*	application définie sur la *population* ; si on note *X* le caractère, il peut être identifié au *n*-uple $\left(X(e_1), ..., X(e_n)\right)$ des *observations* de *X* sur *E* Dans le cadre d'une enquête par questionnaire, pour une question autorisant plusieurs réponses (sports pratiqués par exemple), on construit autant de *caractères dichotomiques* (oui, non) que de réponses possibles
série statistique (à 1 *variable) de taille n*	*n*-uple de valeurs pouvant être considéré comme le *n*-uple des *observations* d'un caractère sur une population de taille *n*
observations	images par un caractère *X* des individus statistiques, termes de la série statistique
*caractère qualitatif (*ou *variable catégorielle)*	*caractère* à valeurs dans un ensemble fini dont les éléments sont appelés *modalités*
*caractère quantitatif (*ou *variable réelle)*	*caractère* à valeurs dans l'ensemble des nombres réels

caractère quantitatif discret	*caractère quantitatif* dont l'ensemble des valeurs possibles est discret
caractère quantitatif continu	*caractère quantitatif* dont l'ensemble des valeurs possibles est un intervalle ; les valeurs sont alors bien souvent regroupées en *classes*
$\left[X \in I\right]$	notation en statistique et probabilités de $X^{-1}\left(I\right)$, image réciproque par l'application *X* d'un sous-ensemble *I* de l'ensemble d'arrivée : $\left[X \in I\right] = \left\{e \in E \,;\, X\left(e\right) \in I\right\}$; généralisation $\left[X = x\right], \left[X \le x\right], \ldots$
codage	numérotation (de 1 à *r* des *r* modalités d'un *caractère qualitatif*) Un caractère qualitatif codé apparaît comme un caractère quantitatif discret mais faire des calculs sur les codes n'a aucun sens. Un caractère quantitatif peut être considéré comme caractère qualitatif tant que l'on ne fait aucun calcul sur les valeurs prises par le caractère. Lorsque le caractère a deux modalités (caractère dichotomique), on préfère souvent les coder 0 et 1.
caractère dichotomique	*caractère qualitatif* à deux modalités ; lorsque ces modalités sont codées 0 et 1, alors le caractère qualitatif codé est un caractère quantitatif appelé *indicatrice*
indicatrice	caractère quantitatif ne prenant que deux valeurs 0 et 1 Il s'agit donc de la fonction caractéristique du sous-ensemble *A* de la population prenant la valeur 1 (on parle alors de l'indicatrice de *A*). La moyenne de l'indicatrice de *A* est la proportion *p* de *A* dans la population et sa variance est égale à $p(1-p)$.
classes	intervalles de l'ensemble des réels ; deux à deux disjoints, la réunion est un intervalle recouvrant l'ensemble des valeurs d'un *caractère quantitatif continu*. Notation : $[x_{i-1}, x_i[\,;\, i \in \{1,...r\}$ avec $x_0 < x_1 < ... < x_r$ et $X(E) \subset \left[x_0, x_r\right]$. on suppose que, pour chaque classe, les observations de la classe sont uniformément réparties dans la classe ; le regroupement des valeurs en classes fait perdre de l'information pour davantage de lisibilité (histogramme)
catégorie	sous-ensemble d'une *population* de tous les *individus* prenant une même *modalité* d'un *caractère qualitatif* L'ensemble des catégories associées à un caractère qualitatif est la partition de la population engendrée par le caractère qualitatif. On définira de même la partition engendrée par un caractère quantitatif discret ou un caractère quantitatif continu dont les valeurs sont regroupées en classes.
effectif	nombre d'éléments d'un sous-ensemble de la population (effectif ou "fréquence" ou "fréquence absolue" dans quelques ouvrages français "frequence" en anglais, "frecuencia absoluta" en espagnol)
fréquence (relative)	rapport de l'*effectif* d'un sous-ensemble de la population sur l'*effectif* de la *population* ("relative frequence" en anglais, "frecuencia relativa" en espagnol)
distribution d'effectifs (d'un caractère)	ensemble $\left\{\left(x_i, n_i\right), i \in [\![1, r]\!]\right\}$ où $\left\{x_i, i \in [\![1, r]\!]\right\}$ est l'ensemble des *r* valeurs distinctes du caractère *X* et n_i l'effectif du sous-ensemble de la population prenant la valeur x_i ; $n = \sum_{i=1}^{r} n_i$ Pour un caractère quantitatif continu dont les valeurs sont regroupées en classes, on remplace les valeurs par les classes de valeurs. La distribution d'effectifs est souvent présentée dans un tableau à 2 lignes ou 2 colonnes dont la première est l'ensemble des valeurs (ou des classes) (rangées usuellement dans l'ordre croissant lorsque l'ensemble des valeurs est ordonné) et la deuxième est l'ensemble des effectifs correspondants.

distribution de fréquences (d'un caractère)	ensemble $\{(x_i, f_i)\,;\, i \in [\![1, r]\!]\}$ où $\{x_i\,;\, i \in [\![1, r]\!]\}$ est l'ensemble des *r* valeurs distinctes du caractère *X* et f_i la fréquence du sous-ensemble de la population prenant la valeur x_i ; $f_i = n_i / n$ avec $n = \sum_{i=1}^{r} n_i$ Pour un caractère quantitatif continu dont les valeurs sont regroupées en classes, on remplace les valeurs par les classes. La distribution de fréquences est souvent présentée dans un tableau 2 lignes ou 2 colonnes dont la première est l'ensemble des valeurs (rangées usuellement dans l'ordre croissant lorsque l'ensemble des valeurs est ordonné) et la deuxième est l'ensemble des fréquences correspondantes.
effectifs cumulés	$N_0 = 0$, et pour $i = 1, ..., r$, $N_i = \sum_{j=1}^{i} n_j$ pour un caractère quantitatif (ou un caractère qualitatif dont les modalités sont ordonnées)
fréquences cumulées	$F_0 = 0$, et pour $i = 1, ..., r$, $F_i = \sum_{j=1}^{i} f_j$; on a donc : $F_i = N_i / n$ pour un caractère quantitatif (ou un caractère qualitatif dont les modalités sont ordonnées)
fonction de répartition (f.d.r.)	d'un caractère quantitatif, définie sur $\mathbb{R}$ par : $F(x) = Freq([X \le x])$ Cas discret : $F(x) = \sum_{i=1}^{r} F_i \mathbf{1}_{[x_i, x_{i+1}[}(x)$, avec $x_{r+1} = +\infty$ *F* est une fonction en escalier, croissante, à valeurs dans [0, 1], nulle sur $]-\infty, x_1[$, égale à 1 sur $[x_r, +\infty[$, continue sauf aux points d'abscisse x_i où elle admet une discontinuité de saut f_i et où elle est continue à droite. Cas continu : $F(x) = \sum_{i=1}^{r} \left[\frac{x - x_{i-1}}{x_i - x_{i-1}} (F_i - F_{i-1}) + F_{i-1} \right] \mathbf{1}_{[x_{i-1}, x_i[}(x) + \mathbf{1}_{[x_r, \infty[}(x)$ *F* est une fonction continue, affine par morceaux, croissante, à valeurs dans [0, 1], nulle sur $]-\infty, x_0]$, égale à 1 sur $[x_r, +\infty[$. Si, pour tout *i* de $\{1, ..., r\}$ $f_i > 0$, la restriction de *F* à $[x_0, x_r]$ est une bijection à valeurs dans $[0,1]$. On utilise ici à nouveau l'hypothèse d'uniforme répartition des observations dans les classes.
diagramme en secteurs	type de graphique permettant de représenter la *distribution d'effectifs* (ou *de fréquences*) d'un *caractère qualitatif* ; il s'agit d'un disque (parfois d'un demi-disque) composé de secteurs angulaires (représentant les modalités) dont les mesures d'angle sont proportionnelles aux effectifs (ou aux fréquences) des modalités Il peut être également utilisé pour représenter, par exemple, la répartition du budget de la commune selon les différents postes budgétaires. Dans ce cas, il ne s'agit pas d'un diagramme d'effectifs ou de fréquences d'un caractère qualitatif. Les euros ne sont pas des effectifs, les différents postes budgétaires ne sont pas les modalités d'un caractère qualitatif.
diagramme en barres	type de graphique permettant de représenter la *distribution d'effectifs* (ou *de fréquences*) d'un *caractère qualitatif* ; on place les modalités de la variable sur un axe horizontal (à égales distances les unes des autres) et on élève au dessus de ces modalités des barres de hauteurs proportionnelles aux effectifs (ou aux fréquences) des modalités L'ordre des modalités sur l'axe horizontal peut parfois donner une information trompeuse. (diagramme parfois appelé en "tuyaux d'orgue", les barres ou les tuyaux d'orgue peuvent être présentés horizontalement et non verticalement)
diagramme en bâtons	type de graphique permettant de représenter la *distribution d'effectifs* (ou *de fréquences*) d'un *caractère quantitatif discret* ; on place les *r* valeurs distinctes de la variable sur un axe horizontal représentant la droite des nombres réels et on élève au dessus de ces valeurs des bâtons de hauteurs

	proportionnelles aux effectifs (ou aux fréquences) des valeurs
histogramme	type de graphique permettant de représenter la *distribution d'effectifs* (ou *de fréquences*) d'un *caractère quantitatif continu* dont les valeurs sont regroupées en classes ; on place les classes sur un axe horizontal représentant la droite des nombres réels et on élève au dessus de ces classes des rectangles de mesures d'aires proportionnelles aux effectifs (ou aux fréquences) ; les rectangles sont en cohérence avec l'hypothèse faite selon laquelle les observations d'une même classe sont uniformément réparties dans la classe
graphe des effectifs cumulés	graphe de la fonction définie sur $\mathbb{R}$ par: $N(x) = card([X \leq x]) = n\,F(x)$ où F est la *fonction de répartition* de X
graphe des fréquences cumulées	graphe de la *fonction de répartition* de X
indice de position ou *de tendance centrale*	résumé numérique d'une distribution d'effectifs ou de fréquences d'un caractère quantitatif correspondant à une valeur située au "milieu" (selon un critère à préciser) de la distribution, par exemple, le mode, la moyenne, la médiane, … Il s'agit d'un résumé numérique d'une distribution qui entraîne une grande perte d'information ; aussi, il faut éviter de l'appeler "caractéristique" qui a un autre sens en mathématiques, d'autant plus que ce mot est parfois utilisé à la place de "caractère". Dans le cas où les données sont observées sur un *échantillon*, on appelle "paramètres" les indices concernant la population (qu'il s'agit d'estimer à partir des indices similaires calculés sur l'échantillon).
indice de dispersion	résumé numérique d'une distribution d'effectifs ou de fréquences d'un caractère quantitatif mesurant le plus ou moins grand "étalement" de la distribution, par exemple, l'étendue, l'écart-type, l'écart interquartile, …
mode, *classe modale*	(*indice de position d'un caractère quantitatif*) le *mode* est la valeur du caractère correspondant à l'effectif (ou la fréquence) maximal ; dans le cas d'un caractère quantitatif continu dont les données sont regroupées en classes, la *classe modale* est celle dont l'effectif (ou la fréquence) par unité du caractère est maximal un caractère peut avoir plusieurs modes ou classes modales
étendue	(*indice de dispersion d'un caractère quantitatif)* l'étendue est la différence $x_r - x_1$ entre la valeur maximale du caractère x_r (appelée *max*) et la valeur minimale x_1 (appelée *min*) pour un caractère quantitatif continu dont les valeurs sont regroupées en *classes* (cf. notation introduite), l'étendue est la différence $x_r - x_0$
moyenne (arithmétique)	(*indice de position d'un caractère quantitatif*) la *moyenne* est la somme des n valeurs du caractère divisée par n ; elle correspond à la valeur commune qu'auraient les n individus de la population s'ils se partageaient de façon égale la somme des valeurs positives ou négatives du caractère ; la moyenne a la dimension du caractère X et elle est notée $\overline{X}$ ou $\overline{x}$ (la notation μ est utilisée pour la moyenne de la population dont serait extrait un échantillon) $\overline{x} = \frac{1}{n}\sum_{k=1}^{n} X(e_k) = \frac{1}{n}\sum_{i=1}^{r} n_i x_i = \sum_{i=1}^{r} f_i x_i$ avec $n = \sum_{i=1}^{r} n_i$ et $f_i = \frac{n_i}{n}$

variance et écart-type	(*indices de dispersion d'un caractère quantitatif*) la *variance* est la moyenne des carrés des écarts à la moyenne, soit : $\mathrm{var}(X)=\frac{1}{n}\sum_{k=1}^{n}\left(X(e_k)-\overline{x}\right)^2=\frac{1}{n}\sum_{i=1}^{r}n_i\left(x_i-\overline{x}\right)^2=\sum_{i=1}^{r}f_i\left(x_i-\overline{x}\right)^2$ la variance est de dimension le carré de la dimension de X ; on vérifie aisément que la variance est égale à la moyenne des carrés moins le carré de la moyenne (2ème formule de la variance utilisée dans les calculs), soit : $\mathrm{var}(X)=\frac{1}{n}\sum_{k=1}^{n}\left(X(e_k)\right)^2-\overline{x}^2=\frac{1}{n}\sum_{i=1}^{r}n_i x_i^2-\overline{x}^2=\sum_{i=1}^{r}f_i x_i^2-\overline{x}^2$ l'*écart-type* est la racine carrée de la variance (de dimension celle de X) ; il est noté s (la notation σ est utilisée pour l'écart-type de la population dont serait extrait un échantillon).
médiane	(*indice de position d'un caractère quantitatif*) intuitivement, la médiane partage la population en deux parties de même effectif ; plus précisément, un réel m est médiane si au moins la moitié de la population prend des valeurs inférieures ou égales à m et au moins la moitié de la population prend des valeurs supérieures ou égales à m, i.e. (définition d'un *quantile d'ordre 0.5*) : $Freq([X\le m])\ge 0.5$ et $Freq([X\ge m])\ge 0.5$; lorsque l'on range les valeurs des n individus de la plus petite à la plus grande, si n est impair, l'unique médiane est la $[(n+1)/2]^{\text{ème}}$ valeur, si n est pair, n'importe quelle valeur comprise au sens large entre la $n^{\text{ème}}$ et la $(n+1)^{\text{ème}}$ est médiane ; par convention, pour avoir unicité, on prend la demi-somme de ces deux valeurs
quantile d'ordre p ($p\in]0,1[$)	q est quantile d'ordre p si la proportion d'observations inférieures ou égales à q est supérieure ou égale à p et si la proportion d'observations supérieures ou égales à q est supérieure ou égale à $1-p$, i.e. si $Freq([X\le q])\ge p$ et $Freq([X\ge q])\ge 1-p$; l'ensemble des quantiles d'ordre p est un intervalle fermé non vide ; pour avoir unicité, différentes conventions sont utilisées
fonction quantile	$\forall p\in]0,1[,\ q(p)=\inf\{x\in\mathbb{R}\ ;\ F(x)\ge p\}$ $q(p)$ est la borne inférieur de l'intervalle des quantiles d'ordre p ; $q(p)$ appartient à $X(E)$, c'est la plus petite valeur q de la série des observations ordonnées en croissant telle que la proportion d'observations inférieures ou égales à q soit au moins égale à p
médiane, quartiles, déciles	On appelle *médiane* un quantile d'ordre 0.5. On appelle *1er quartile* (resp. *2ème quartile*, resp. *3ème quartile*) un quantile d'ordre 0.25 (resp. 0.50, resp. 0.75). On appelle *1er décile* (resp. *2ème décile*, …, resp. *9ème décile*) un quantile d'ordre 0.10 (resp. 0.20, …, resp. 0.90). m est médiane $\Leftrightarrow$ m est 2ème quartile $\Leftrightarrow$ m est 5ème décile dans le secondaire, la définition donnée pour le 1er (resp. 3ème) quartile est $q(0.25)$ (resp. $q(0.75)$, la définition donnée pour le 1er (resp. 9ème) décile est $q(0.1)$ (resp. $q(0.9)$).
intervalle interquartile, intervalle interdécile	On appelle *intervalle interquartile* (resp. *intervalle interdécile*) l'intervalle dont les extrémités sont le 1er et le 3ème quartiles (resp. le 1er et le 9ème déciles) après avoir éventuellement utilisé une convention pour l'unicité.

écart interquartile, écart interdécile	On appelle *écart interquartile* (resp. *écart interdécile*) la longueur de l'intervalle interquartile (resp. interdécile).
diagramme en boîte (ou en boîte et moustaches)	(en anglais, "box plot" ou "box and whiskers plot") Diagramme représentant une distribution de fréquences construit sur les cinq indices, *min*, q_1, *m*, q_2, *max*, placés sur un axe représentant la droite des réels : la boîte rectangulaire, de largeur arbitraire, est limitée en longueur par le premier et troisième quartile ; à l'intérieur de la boîte est indiquée par un trait la médiane et, de part et d'autre de la boîte, des segments représentent les valeurs extérieures à *l'intervalle interquartile*, les extrémités de ces segments indiquent les valeurs extrêmes (min et max) de la variable. Les extrémités des segments peuvent aussi correspondre aux premier et neuvième déciles ; des croix indiquent alors les observations extérieures à *l'intervalle interdécile*. L'avantage de ce type de diagramme est sa construction rapide et la possibilité de faire une comparaison visuelle de plusieurs distributions.

Les données d'enquête observées sur un *échantillon*

sondage	dans le langage courant, il s'agit d'une enquête d'opinion ("poll" en anglais, "sondeo" en espagnol), en statistique, il s'agit d'une enquête auprès d'une partie de la *population* appelée *échantillon* ("sampling" en anglais, "muestreo" en espagnol)
recensement	s'oppose à *sondage* ; l'enquête est réalisée auprès de la *population* tout entière
base de sondage	liste numérotée de tous les individus de la population, nécessaire pour sélectionner un échantillon aléatoire, notée $E = \{e_1, ..., e_N\}$
échantillon	sous-ensemble de la *population* ; ce peut être un *échantillon aléatoire* (ou *échantillon probabiliste*) ou un *échantillon empirique* On ne parle d'échantillon que lorsque l'on souhaite inférer à la population dont il est issu les indices obtenus sur l'échantillon ; une classe d'élèves est alors une population et non un échantillon.
taille	nombre d'éléments de l'*échantillon* (noté usuellement *n*) ou de la *population* (noté usuellement *N*) ou de sous-populations ou sous-échantillons …
taux de sondage	rapport de la *taille* de l'*échantillon* sur la *taille* de la *population*
échantillon aléatoire (ou *probabiliste*)	échantillon sélectionné selon une *procédure aléatoire* (et en utilisant un *générateur de nombres pseudo-aléatoires* ou une *table de nombres au hasard*)
échantillon empirique	échantillon non aléatoire ; le plus connu est l'échantillon empirique obtenu par la *méthode des quotas*
échantillon avec remise	les *n* individus de l'échantillon sont tirés un à un en remettant dans la population de taille *N* l'individu sélectionné après chaque tirage ; un même individu de la population peut apparaître plusieurs fois dans l'échantillon ; un tel échantillon est représenté par un élément de l'ensemble E^n de cardinal N^n

échantillon sans remise	les *n* individus de l'échantillon sont tirés un à un sans remettre dans la population de taille *N* les individus déjà sélectionnés ; un tel échantillon est représenté par un arrangement de *n* éléments de *E*, il y a A_N^n tels échantillons. Dans la mesure où l'on ne tient pas compte de l'ordre de tirage des individus de l'échantillon, on peut considérer un échantillon sans remise comme un sous-ensemble de *n* éléments de *E* ; on a $\binom{N}{n}$ tels échantillons.
échantillon aléatoire simple à probabilités égales avec remise	un des N^n échantillons de taille *n* avec remise d'une population de taille *N* (tous les échantillons ayant la même probabilité d'être tiré)
échantillon aléatoire simple à probabilités égales sans remise	un des $\binom{N}{n}$ échantillons de taille *n* sans remise d'une population de taille *N* (tous les échantillons ayant la même probabilité d'être tiré)
échantillon aléatoire stratifié	on dispose dans la *base de sondage* d'une (ou plusieurs) variable catégorielle permettant de définir une partition de la population en sous-populations appelées *strates* ; on tire de façon indépendante, un échantillon aléatoire simple de taille fixée dans chaque strate ; l'échantillon global est la réunion de tous ces échantillons.
échantillonnage aléatoire stratifié proportionnel	échantillon aléatoire stratifié tel que les tailles des échantillons soient proportionnelles aux tailles des strates
méthode des quotas	il s'agit d'une méthode de sélection non aléatoire d'un échantillon tentant d'imiter un échantillon aléatoire stratifié proportionnel ; on connaît la répartition de la population selon deux ou trois variables catégorielles, appelées variables de quotas, l'échantillon doit respecter ces répartitions (appelées quotas)

Au niveau du collège, on travaillera sur des populations et non sur des échantillons.

Au niveau du lycée, on pourra travailler sur des échantillons aléatoires simples à probabilités égales avec remise. Le vocabulaire statistique défini sur la population dans la partie précédente s'applique ici en remplaçant population par échantillon.

Une fois l'échantillon décrit, l'objectif est d'inférer à la population, dont les *paramètres* (proportions, moyenne, écart-type, …) sont inconnus, les informations obtenues (calculées) sur l'échantillon.

> La *loi des grands nombres* permet d'approcher une proportion (ou fréquence) relative à la population (appelée aussi *fréquence théorique*) par une proportion (ou fréquence) relative à l'échantillon (appelée aussi *fréquence empirique* ou observée) et donc une *distribution de fréquences théoriques* (sur la population) par une *distribution de fréquences empiriques* (sur l'échantillon).
>
> Le *théorème central limite* permet d'utiliser l'approximation gaussienne pour estimer par intervalle de confiance les proportions ou moyennes de la population à partir des proportions ou moyennes calculées sur l'échantillon.

On notera que l'observation *x* d'un caractère (qualitatif ou quantitatif) sur *UN individu tiré avec équiprobabilité d'une population finie* est l'observation d'une *variable aléatoire* (catégorielle ou réelle) dont la distribution de probabilité n'est autre que la distribution de fréquences du caractère (le "nombre de cas favorables sur nombre de cas possibles" n'est autre que la fréquence). Nous ne faisons aucune hypothèse sur cette distribution de probabilité, l'aléatoire vient de la procédure de sélection de l'individu (on simule l'équiprobabilité).

Pour un échantillon *aléatoire* de taille *n* tiré avec équiprobabilité et avec remise d'une population finie, la série statistique de taille *n* EST l'observation de *n* variables aléatoires indépendantes et identiquement distribuées (i.i.d.), la distribution de probabilité EST la distribution de fréquences sur la population du caractère considéré (*approche sondage*).

Données expérimentales

étude expérimentale	C'est dans le but de rechercher de l'information sur une question donnée que l'on peut décider de mettre en place un *dispositif expérimental* ; il s'agit alors de définir les conditions expérimentales permettant de réaliser une série d'*expériences* dans les mêmes conditions et de définir la ou les variables qui seront observées pour chacune des expériences et qui pourront apporter l'information cherchée.
série statistique (à 1 variable) de taille n	*n*-uple de valeurs pouvant être considéré comme le *n*-uple des observations d'une variable sur une suite de *n expériences*

Si on note $E = \{e_1, ..., e_n\}$ l'ensemble des expériences, nous sommes dans la situation précédente, observation d'une variable sur un échantillon de taille *n*. Mais de quelle population serait extrait cet échantillon ? On adaptera donc le vocabulaire décrit dans la première partie pour décrire la série numérique sans référence à des individus statistiques ou à une population.

Si l'expérience est reproduite dans les mêmes conditions, *on fera l'hypothèse* que la série statistique de taille *n* est l'observation de *n* variables aléatoires indépendantes et de même distribution de probabilité (i.i.d.) mais on ne connaît en général rien du type de la distribution de probabilité et on pourra être amené à *faire des hypothèses* sur cette distribution de probabilité (*approche modèle*).

C'est par abus de langage que l'on parle encore, dans le cadre de données expérimentales, de *paramètres de la population* ; il s'agit en fait de paramètres de la loi de probabilité du modèle.