Rappels et précisions préalables

Fiche résumé - Vocabulaire utilisé en statistique -

Désignation	Courte définition	Exemple
Population	Ensemble des individus étudiés, répondant à une propriété donnée (modalité d'une variable qualitative). La population étudiée est plus ou moins précise ; elle peut-être caractérisée par les paramètres exacts d'une variable aléatoire continue considérée dans l'étude. (*la taille de la population n'est en général pas connue avec précision, c'est pourquoi on ne la considère pas dans l'inférence statistique*)	Population des personnes atteintes par une maladie M (caractérisée, par exemple, par la moyenne exacte et l'écart-type exact de l'âge de ces personnes)
Individu	Un élément de la population étudiée, (vient du latin *individuum*, signifiant ce qui est indivisible) ; un des objets étudiés. (*en biologie, il n'existe pas deux individus totalement identiques*)	un patient, un écologue, une micropipette ; un yaourt, un ours des Pyrénées ; un grain de maïs ; une parcelle cultivée ; un tube à essai ; une personne atteinte par une maladie M ; un logiciel ; un profil Facebook
Unité statistique	Autre façon de désigner un individu	un patient ; un écologue ; une micropipette ; un yaourt ; un ours des Pyrénées ; un grain de maïs ; une parcelle cultivée ; un tube à essai ; une personne atteinte par une maladie M
Echantillon	Ensemble d'individus extraits d'une population. Il s'agit donc d'un sous ensemble de la population (de bien plus petite talle que la population, ce qui en rend l'étude réalisable) Un échantillon est dit représentatif de la population dont il est issu s'il a été tiré au hasard (indispensable en statistique inférentielle) (*l'analyse d'un Big Data peut conduire à étudier l'ensemble complet des données d'une population*)	28 poissons pêchés dans un étang normand ; 65 parcelles de culture d'une nouvelle variété céréalière ; N yaourts issus de la production d'une journée de la chaîne agroalimentaire de l'usine A; 42 micropipettes utilisées dans un laboratoire de recherche, 15 échantillons de salive de personnes touchées par un virus V
Variable	C'est le caractère étudié dans la population et qui varie d'un individu à l'autre. Variable aléatoire [VA] : application définie sur l'ensemble de définition. Les variables manipulées en statistiques sont qualitatives (nominales ou ordinales) ou quantitatives (discrètes ou bien continues). *(C'est le support de l'information, par exemple de l'information biologique étudiée. Dans le logiciel R, vous manipulez un vecteur)*	Concentration du micropolluant dans l'eau du robinet (en ng/ml); nombre de personnes ayant résisté au virus V dans un échantillon de N personnes atteintes par le virus ; Etat du patient (guéri/malade) ; stades de développement d'une maladie grave (non développée, premier stade, stade avancé, stade sévère).
Observation	Valeurs prises par la variable pour un individu donné (exemple de notation : $X=x_i$) On s'intéresse par exemple à la probabilité $P(X>x_i)$ On s'intéresse ainsi aux couples (i, x_i)	La mesure de la taille d'une des abeilles suivie à l'aide d'un nanoémetteur, a donné x_i=1,97 cm
Modalité	Les modalités sont les valeurs prises par la variable associée au caractère étudié. *(La variable manipulée dans un logiciel regroupe l'ensemble des valeurs prises par le caractère étudié. Il s'agit d'un vecteur)*	X : Nombre d'abeilles qui ne sont pas rentrées à leur ruche sur un échantillon de 8 abeilles suivies à l'aide d'un nanoémetteur; X peut prendre les modalités : 0,1,2,3,4,6,7,8 Etat du patient : «guéri(e)», «stationnaire», «moins malade», «encore plus malade »

Calcul des paramètres statistiques et de leurs estimations

▶ paramètres de position, paramètres de dispersion, facteurs de forme

Statistique descriptive **Décrire** *ab minima* **/ Résumer** **la distribution**

Paramètres

- **Position :** moyenne, mode, médiane, min, max…
  (*ordre de grandeur*); tendance centrale
- **Dispersion :** écart-type, variance, quantiles, étendue, …
  (*répartition autour de l'ordre de grandeur*)
- **Forme :** symétrie, aplatissement,…
  (tendance)

Tableaux synthétiques rassemblant **paramètres** caractéristiques et leurs **Intervalles de Confiance**

Graphes Histogrammes, boîtes à moustache, lissage normal, …

Test normalité de la distribution (Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk, Chi-deux,...)

⇨ **Comparer** **échantillon (s) - population (s)**

Informations (fichier) à décrire du mieux possible pour permettre leur analyse

Rappel : somme de n valeurs

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

i : indice de l'individu, de l'observation ou de la mesure

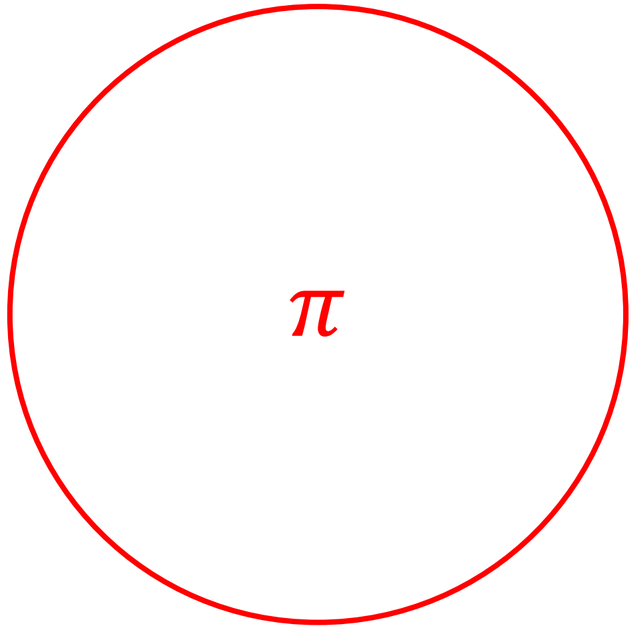
x_i : valeur observée ou mesurée (observation/mesure) pour l'individu i

n : nombre total d'individus, d'observations ou de mesures

Calcul des paramètres statistiques et de leurs estimations

▶ deux mots sur les codes couleurs adoptés

Y variable d'étude qualitative
W variable sous-jacente



Population : caractérisée par

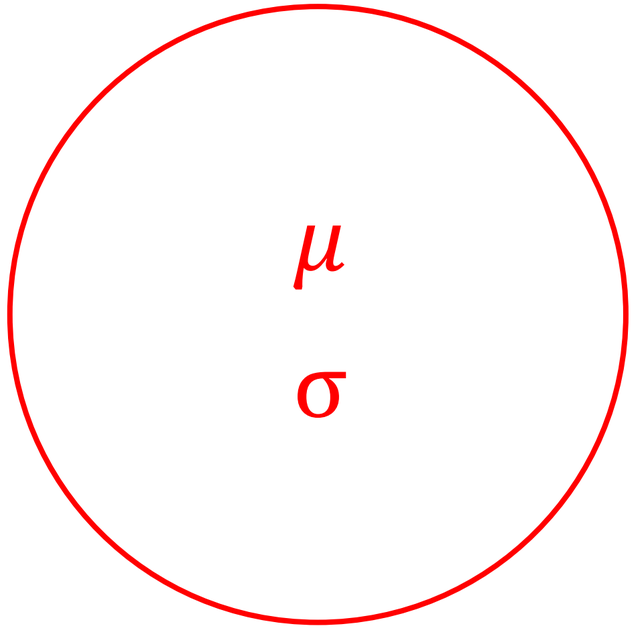
le paramètre exact : π

(fréquence/proportion d'un caractère dans la population)

Calcul des paramètres statistiques et de leurs estimations

▶ deux mots sur les codes couleurs adoptés

X : **variable aléatoire** d'étude, suit généralement une **loi de probabilité**



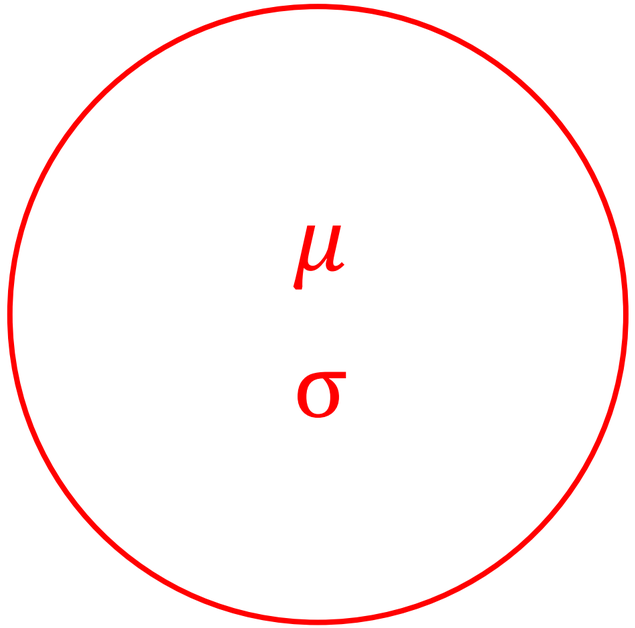
Population : caractérisée par

2 paramètres exacts : μ et σ

Calcul des paramètres statistiques et de leurs estimations

▶ deux mots sur les codes couleurs adoptés

X : variable aléatoire d'étude



Population : caractérisée par

2 paramètres exacts : μ et σ

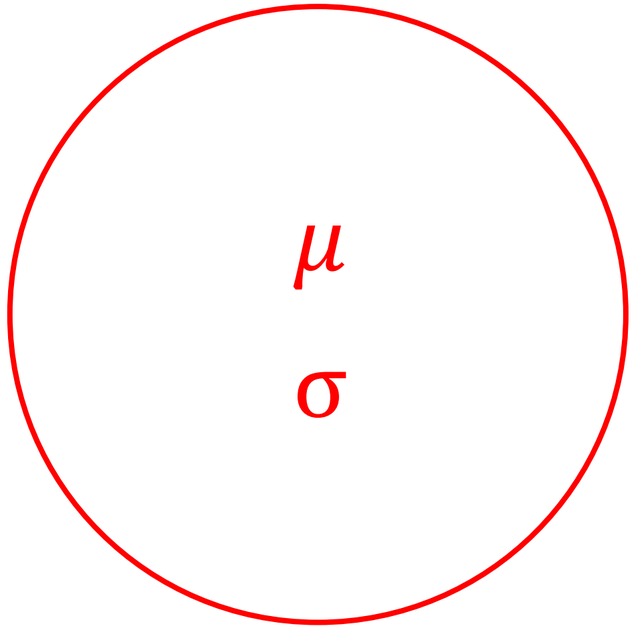
- **rouge** pour ce qui se rapporte **à la population**

- Lettres **Majuscules** ▶ **Variables Aléatoires**
- lettres **grecques** ▶ **paramètres exacts** d'une **population**

Calcul des paramètres statistiques et de leurs estimations

▶ deux mots sur les codes couleurs adoptés

X : variable aléatoire d'étude



Population : caractérisée par

2 paramètres exacts : μ et σ

Espérance : μ

Ecart-type : σ ; **Variance** : σ^2

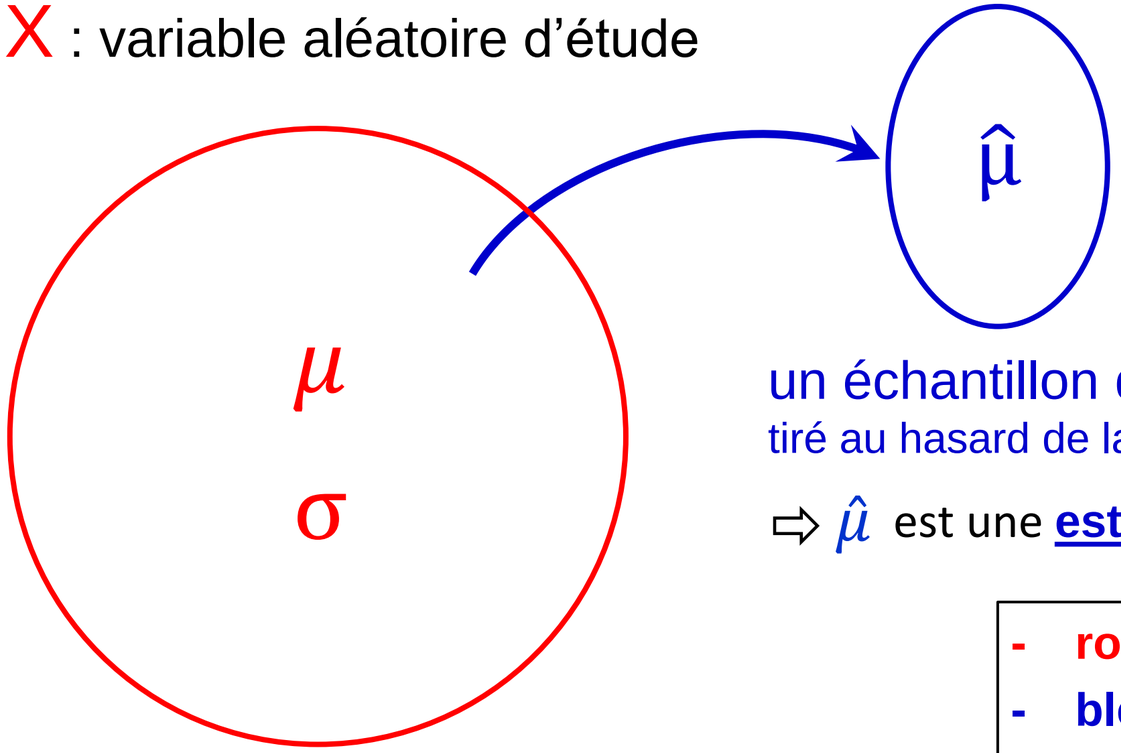
- **rouge** pour ce qui se rapporte **à la population**

- Lettres **Majuscules** ▶ **Variables Aléatoires**
- lettres **grecques** ▶ **paramètres exacts** d'une **population**

Calcul des paramètres statistiques et de leurs estimations

▶ deux mots sur les codes couleurs adoptés

X : variable aléatoire d'étude



⇨ $\hat{\mu}$ est également désignée par $\bar{x}$

$(x_1, x_2 , \dots x_i , \dots x_n)$

un échantillon de taille n
tiré au hasard de la population

⇨ $\hat{\mu}$ est une **estimation** de μ

Population : caractérisée par

2 paramètres exacts : μ et σ

Espérance : μ

Ecart-type : σ ; **Variance** : σ^2

- **rouge** pour ce qui se rapporte **à la population**
- **bleu** pour ce qui se rapporte **à l'échantillon**
- lettres **minuscule** ▶ **observations**, **mesures**, **estimations**
- Lettres **Majuscules** ▶ **Variables Aléatoires**
- lettres **grecques** ▶ **paramètres exacts** d'une **population**
- **chapeaux** ▶ **estimations**