

MOTS CLES DE STATISTIQUE

La **statistique** est la discipline qui étudie des phénomènes à travers la collecte de **données**, leur traitement, leur analyse, l'interprétation des résultats et leur **présentation** afin de rendre ces données compréhensibles par tous. (Wikipédia)

Science qui recueille et analyse mathématiquement des données.

Vient de l'ital. statista « homme d'État », la statistique représentant pour lui l'ensemble des connaissances que doit posséder un homme d'État

Population groupe d'individus soumis à une étude. Ensemble des individus ou des éléments de même nature à partir duquel on peut constituer un échantillon lors d'une étude statistique.

Vient du latin populatio, « population, foule »

Individu élément issu de la population.

Echantillon partie d'une population.

XVIII^{es}. « élément représentatif sélectionné comme objet d'étude »

Les variables aléatoires

C'est le type de variables auquel on s'intéresse en statistique et en probabilité (toujours notées en utilisant une grande lettre). Il s'agit d'une variable qui est soumise à une certaine dispersion lors de la répétition d'un processus donné et dont on ne peut pas être certain de la valeur qu'elle prendra. Le résultat de l'expérience (la valeur que prend la variable) est soumis au hasard.

Variable **qualitative** : variable à valeurs non-numériques
(où la moyenne n'a pas de sens). Ses valeurs sont appelées modalités

nominale (ou catégorielle) : absence d'ordre entre les modalités. Par exemple : la couleur des cheveux, la couleur des yeux, porter des lunettes ou ne pas porter de lunettes, etc.

ordinaire : existence d'un ordre Par exemple : assez bien, bien, très bien, excellent.

Variable **quantitative** : variable à valeurs numériques (où la moyenne a un sens)

continue : à valeurs dans un intervalle réel (On ne peut pas compter toutes les valeurs)

discrète : les valeurs sont séparées.

Intervalle de fluctuation à 95 %

Théorique est ici synonyme de vraie valeur, la vraie valeur de la proportion. Celle qui serait obtenue si on pouvait prendre en compte la population dans son ensemble. C'est-à-dire toutes les valeurs possibles de la variable aléatoire.

C'est un intervalle construit autour de la **valeur théorique p** (voir point I/1. du pdf de cours) ou d'une valeur théorique que l'on veut tester (Voir point I/ 3. du pdf de cours) et dont on est sûr qu'il contiendra la réalisation f de la variable aléatoire sur l'échantillon que l'on a extrait de notre population dans 95 % des cas.

Si la valeur que l'on a calculée à partir de notre échantillon n'est pas comprise dans l'intervalle, on peut considérer que la fluctuation est trop importante. Et on sait, si la valeur théorique p des individus qui possèdent le caractère est **connue**, que l'on a 95 % de chances d'avoir raison de considérer que notre échantillon n'est pas représentatif de la population, ou, si on fait un **test sur p**, que l'hypothèse sur la valeur théorique p doit être rejetée.

Car, il serait déraisonnable de considérer qu'un écart aussi grand puisse être dû à un phénomène normal de fluctuation d'échantillon. Les résultats f obtenus à partir d'échantillons tirés au hasard d'une même population vont bien-sûr fluctuer. On parle de fluctuation d'échantillon. Mais au-delà d'un certain niveau, on décide que l'échantillon n'est pas représentatif et donc ne peut pas provenir de la population étudiée ou que la valeur de p est incorrecte.

Mais si la valeur calculée à partir de l'échantillon n'appartient pas à l'intervalle de fluctuation, on ne considère plus qu'il s'agit d'une fluctuation normale d'échantillon.

Donc,

Soit **1** : On considère que l'échantillon n'est pas représentatif de la population dont la proportion p est connue .

Ou **2** : Dans le cas où on testait une valeur pour p , on rejette l'hypothèse sur la valeur de p .

Sans perdre pas de vue que le risque de se tromper est de 5%.

Intervalle de confiance à 95 % sur p

La proportion théorique p est rarement connue dans la pratique.

Il est souvent utile de l'estimer. Pour cela on va construire un intervalle autour de f (la fréquence observée calculée à partir de l'échantillon).

Pour un échantillon suffisamment grand, on sait que la probabilité que p appartienne à cet intervalle est égale à 0,95.

On dit que p est estimé au niveau de confiance de 95 % ou 0,95.

Exercice :

Pour bien fixer les idées, je vous conseillerais de faire un petit schéma illustrant ce que l'on cherche à faire avec un intervalle de fluctuation (p supposé connu et p que l'on veut tester) et puis un schéma illustrant le but d'un intervalle de confiance.
