

STATISTIQUE INFÉRENTIELLE

THÉORIE D'ÉCHANTILLONNAGE

Benchikh Tawfik

Faculté de Médecine, UDL, SBA
1^{ère} année Médecine

07 Février 2024



PLAN DE COURS

1 CARACTÉRISTIQUES D'UN ÉCHANTILLON.

- Étude de la statistique $\bar{X}$
- Étude de la statistique S^2 ou variance empirique

2 CAS PARTICULIER DES ÉCHANTILLONS GAUSSIENS



- Hypothèse de la biostatistique: la population >>> l'échantillon.
- Choisir un échantillon représentatif: il doit "ressembler" le plus possible à la population d'où il est extrait. Les individus de l'échantillon doivent avoir les mêmes caractéristiques que ceux de la population.
- L'idéal est donc d'avoir un l'échantillon choisi au hasard (où aléatoire): tous les individus de la population ont la même probabilité d'être choisi (tirage avec remise appelé tirage non exhaustif).

- **Remarque :** Des tirages avec remise forment des événements indépendants (très important). Si la population est très grande ou infinie, on pourra toujours considérer que le tirage est non exhaustif: le fait de remettre ou non l'élément tiré devient sans importance.
- Étude Statistique: étude des caractéristiques (variables statistiques) sur les individus de la population l'échantillon
- Les variables statistiques qui la décrivent (population) peuvent être considérées comme des variables aléatoires.
- La répartition d'une variable statistique X sur la population est décrite par une loi de probabilité

- Les individus dans la population varient $\Rightarrow$ Les échantillons varient aussi. Comment traiter cette variation ?
 - Étude des fluctuations d'échantillonnage d'une proportion observée,
 - Étude des fluctuations d'échantillonnage d'une moyenne observée,
 - Estimations : Intervalles de pari, Intervalles de confiance,
 - Tests d'hypothèses (conformité, homogénéité, ajustement à une loi, ect),

DÉFINITION D'UN ÉCHANTILLON ALÉATOIRE

- On étudie une caractéristique X d'une population de taille finie ou infinie.
- La composition de la population, vis à vis du caractère X , est entièrement définie par la connaissance de la loi de X (f_x ou p_x).
- Soit E l'expérience consistant à choisir au hasard un élément de la population. Avant le tirage, on se propose de prévoir la valeur du caractère X que l'on obtiendra.
- À l'expérience E , est associée une variable aléatoire X dont la loi est f_x (ou p_x) .

DÉFINITION D'UN ÉCHANTILLON ALÉATOIRE

- On réalise n fois la même expérience E , dans des conditions indépendantes, par exemple en remettant dans la population l'élément tire.
- À ces n expériences, on associe n variables aléatoires indépendantes X_i , suivant la même loi que la variable aléatoire X .
- Par définition, l'ensemble $(X_1, X_2, \dots, X_n)$, de n variables aléatoires indépendantes suivant la même loi qu'une variable aléatoire X , appelée variable aléatoire parente, est un échantillon aléatoire, note en abrégé (X_i) , $i \in (1, n)$ ou $\underline{X}$.
- Une réalisation de l'échantillon sera notée $(x_1, x_2, \dots, x_n)$, ce n'est plus une variable aléatoire.

EXEMPLE D'UN ÉCHANTILLON ALÉATOIRE

- Le diamètre X d'un globule rouge est aléatoire, avec $\sigma = 0.617$, mais on ne connaît pas sa moyenne.
- On mesure le diamètre de 200 globules rouges d'un patient. On obtient un diamètre moyen $\bar{x} = 7.07$.
- Si on mesure 200 autres globules, on aura une autre moyenne.
- La moyenne de 200 globules est une variable aléatoire.
- On a une distribution probabiliste pour la moyenne → La distribution de cette moyenne est la distribution d'échantillon

DÉFINITION: DISTRIBUTION D'ÉCHANTILLONNAGE

- **Définition:** La distribution d'échantillonnage est la distribution des moyennes de tous les échantillons possibles de taille n venant de la population.

DÉFINITION D'UNE STATISTIQUE

- Une **statistique** définie sur un échantillon aléatoire X de taille n est une fonction des variables X_i .

$$T(X_1, \dots, X_n)$$

- À un échantillon, on peut associer différentes statistiques. La théorie de l'estimation consiste à définir des statistiques particulières, appelées *estimateurs*.



DÉFINITION D'UNE STATISTIQUE: EXEMPLE

- Soit $\bar{X}$ la statistique:

$$\bar{X} = T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i,$$

c'est-à-dire la fonction moyenne arithmétique des n observations d'un échantillon.

- La statistique $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$
- La statistique S^{*2} définie par $S^{*2} = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

LA STATISTIQUE $\bar{X}$

- La statistique $\bar{X}$ est la fonction des variables X_i , définie par:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i .$$

- Un calcul rapide donne les résultats suivants:

$$\mathbf{E}(\bar{X}) = m \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} .$$



- **Comportement asymptotique:**

- **Loi faible des grands nombres:** $\bar{X}$ converge en probabilité vers m quand n tend vers l'infini.
- Appliquons le théorème central limite à la variable aléatoire Y_n :

$$Y_n = \frac{\sum_{i=1}^n X_i - nm}{\sigma\sqrt{n}} = \frac{\bar{X} - m}{\sigma\sqrt{n}}.$$

La variable Y_n converge en loi vers une variable suivant la loi normale $\mathcal{N}(0, 1)$ quand n tend vers l'infini.

APPLICATION: LOI D'UN POURCENTAGE

- Soit K la v.a. représentant le nombre de succès au cours d'une suite de n épreuves indépendantes, la probabilité de succès au cours d'une épreuve étant égale à p .
- La loi de la variable aléatoire K est la loi binomiale $\mathcal{B}(n, p)$.
- Posons $F = \frac{K}{n}$, fréquence empirique du nombre de succès:

$$\mathbb{E}(F) = p \quad \text{Var}(F) = \frac{pq}{n} \quad q = 1 - p.$$

- En appliquant le théorème central limite pour n grand, on démontre que la variable aléatoire F converge en loi vers une variable aléatoire suivant une loi normale:

$$\mathcal{N}(p; \sqrt{\frac{pq}{n}}).$$

- La variance empirique S^2 d'un échantillon de taille n est définie par:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 .$$

PROPRIÉTÉS CARACTÉRISTIQUES DE LA STATISTIQUE S^2 :

- **Espérance mathématique de S^2 :**

$$\mathbb{E}(S^2) = \frac{n-1}{n} \sigma^2 = \left(1 - \frac{1}{n}\right) \sigma^2.$$

- Si la taille n de l'échantillon est grande, l'espérance de S^2 a pour valeur limite σ^2 .

- **Variance de S^2 :**

$$\text{Var}(S^2) = \frac{n-1}{n^3} \left[(n-1)\mu_4 - (n-3)\sigma^4 \right],$$

μ_4 étant le moment centré d'ordre 4 de la variable X .

- Si la taille n de l'échantillon est grande, la variance de S^2 a pour valeur limite:

$$\frac{\mu_4 - \sigma^4}{n}.$$

ÉCHANTILLONS GAUSSIENS

- Si les X_i suivent une loi Normale $\mathcal{N}(m, \sigma)$, on a:
 - La variable $\bar{X}$, la somme de n variables aléatoires indépendantes gaussiennes, est une variable gaussienne. Donc, quelle que soit la valeur de n :

$$\bar{X} \rightsquigarrow \mathcal{N}\left(m, \frac{\sigma^2}{n}\right).$$

- De la décomposition de la statistique S^2 , en déduit le résultat suivant:

$$\frac{nS^2}{\sigma^2} \rightsquigarrow \chi^2(n-1)$$

- Si on connaît la valeur de m , alors utilise la statistique T

$$T = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2.$$

et on a

$$\frac{nT^2}{\sigma^2} \rightsquigarrow \chi^2(n)$$



- Si X est de loi $\mathcal{N}(0, 1)$, alors X^2 est de loi χ_1^2 .
- Si $X_1, \dots, X_n$ sont indépendantes et de même loi $\mathcal{N}(m, \sigma)$, alors:
 - $\sum_{i=1}^n X_i$ est de loi $\mathcal{N}(nm, \sqrt{n}\sigma)$.
 - $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est de loi $\mathcal{N}(m, \frac{\sigma}{\sqrt{n}})$.
 - $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - m)^2$ est de loi χ_n^2 .
 - $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ est de loi χ_{n-1}^2 .
- Les lois de Student et de Fisher-Snedecor sont toujours utilisées par l'intermédiaire de tables ou à l'aide d'un logiciel de statistique.