RESUMÉ DES COURS

→ statistiques descriptives :

- moyenne : $\frac{1}{n}\sum_{i=1}^{n} x_i$ est sensible aux valeurs extrêmes ou aberrantes mais offre des propriétés algébriques intéressantes : connaître les moyennes de 2 sous-populations permet de calculer la moyenne à l'intérieur de la population totale.

- médiane : c'est la valeur-seuil telle que, si on ordonne la population selon 1 critère donné, elle la sépare en 2 sous-population de taille égale.
elle n'est pas sensible aux valeurs extrêmes ou aberrantes.

- tableau croisé :

On étudie les lois de X et Y ainsi que la relation entre les 2 variables
Si on suppose que X ne prend que 2 modalités : i et j
et que Y ... : k et l

on a :

X \ Y	k	l	
i	n_{ik}	n_{il}	$n_{i.}$
j	n_{jk}	n_{jl}	$n_{j.}$
	$n_{.k}$	$n_{.l}$	n

les fréquences peuvent être interprétées comme des probabilités

$F_{ik} = \frac{n_{ik}}{n} = \mathbb{P}(X=i, Y=k)$ $F_{jk} = \frac{n_{jk}}{n} = \mathbb{P}(X=j, Y=k)$

$F_{il} = \frac{n_{il}}{n} = \mathbb{P}(X=i, Y=l)$ $F_{jl} = \frac{n_{jl}}{n} = \mathbb{P}(X=j, Y=l)$

avec : $F_{.k} = \frac{n_{.k}}{n} = \mathbb{P}(Y=k)$

$F_{.l} = \frac{n_{.l}}{n} = \mathbb{P}(Y=l)$

$F_{i.} = \frac{n_{i.}}{n} = \mathbb{P}(X=i)$

$F_{j.} = \frac{n_{j.}}{n} = \mathbb{P}(X=j)$

et : $F_{i|k} = \frac{n_{ik}}{n_{.k}} = \mathbb{P}(X=i | Y=k)$

$F_{i|l} = \frac{n_{il}}{n_{.l}} = \mathbb{P}(X=i | Y=l)$

$F_{j|k} = \frac{n_{jk}}{n_{.k}} = \mathbb{P}(X=j | Y=k)$

$F_{j|l} = \frac{n_{jl}}{n_{.l}} = \mathbb{P}(X=j | Y=l)$

- variance : $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ est 1 indicateur de dispersion.

Plus elle est élevée et plus les valeurs sont dispersées (moins elles sont homogènes).

- écart-type = racine carrée de la variance

- covariance entre X et Y :

$cov(X;Y) = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$

Si X et Y sont indépendantes alors : $cov(X;Y) = 0$

Si $cov(X;Y) \neq 0$ alors X et Y ne sont pas indépendantes

Si $cov(X;Y) = 0$ alors on ne peut pas conclure.

- coefficient de corrélation linéaire :

$$corr(X;Y) = \frac{cov(X;Y)}{\sigma_X \sigma_Y} \in [-1;1]$$

on cherche 1 relation de type $Y = aX + b$

I) si $|corr(X;Y)| \simeq 1$ alors la corrélation linéaire entre les 2 variables est forte

I-1) si $corr(X;Y) \simeq 1$ alors Y est 1 fonction affine croissante de X $(a > 0)$

I-2) si $corr(X;Y) \simeq -1$ alors Y est 1 fonction affine décroissante de X $(a < 0)$

II) si $corr(X;Y) = 0$ alors X et Y ne sont pas corrélées linéairement

mais on peut avoir 1 autre type de liaison ($Y = X^2$, par exemple)

* $a = \frac{cov(X;Y)}{V[X]}$

$b = \bar{Y} - a\bar{X}$

* la droite $Y = aX + b$ est appelée droite de régression linéaire

---

→ les espaces probabilisés : $(\Omega, \mathcal{A}, \mathbb{P})$

* $\mathcal{A}$ est 1 tribu : $\mathcal{A} \subset \mathcal{P}(\Omega)$

$\Omega \in \mathcal{A}$

$\forall A \in \mathcal{A}, A^c \in \mathcal{A}$

$\forall (A_i)_{i=1..n} \in \mathcal{A}, \bigcup_{i=1}^{n} A_i \in \mathcal{A}$

* $\mathbb{P}$ est 1 mesure de probabilité sur $\mathcal{A}$:

$\mathbb{P} : \mathcal{A} \longrightarrow [0;1]$

$A \longmapsto \mathbb{P}(A) = \mathbb{P}(\{\omega \in \Omega \mid \omega \in A\})$

---

pq : $\mathbb{P}(\emptyset) = 0$

$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

$\mathbb{P}(A \cup B) = \mathbb{P}(B) + \mathbb{P}(A) - \mathbb{P}(A \cap B)$

si $A \subset B$ alors : $\mathbb{P}(A) \leq \mathbb{P}(B)$

[rappel : Soit $(A_n)_{n \geq 1}$ 1 suite d'événements : $\mathbb{P}(\bigcup_{i=1}^{n} A_i)$ est 1 suite croissante et majorée par 1 donc elle converge

$\mathbb{P}(\bigcap_{i=1}^{n} A_i)$ est 1 suite décroissante et minorée par 0 donc elle converge]

- probabilité conditionnelle :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} \text{ avec } \mathbb{P}(B) \neq 0$$

- formule des probabilités totales :

Soit $(B_i)_{i=1,\dots,n}$ 1 partition de Ω : $\begin{cases} \bigcup_{i=1}^{n} B_i = \Omega \\ B_i \cap B_j = \emptyset \text{ si } i \neq j \end{cases}$

alors : $\forall A \in \mathcal{A}$, $\mathbb{P}(A) = \mathbb{P}(A\cap\Omega) = \mathbb{P}(A\cap(\bigcup_{i=1}^{n} B_i)) = \sum_{i=1}^{n} \mathbb{P}(A\cap B_i) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$

[1 exemple : $\forall n \in \mathbb{N}$, $A_n : \Omega \to \{a;b;c\}$

on a : $\mathbb{P}(A_n=a) = \mathbb{P}(A_n=a|A_{n-1}=a)\mathbb{P}(A_{n-1}=a) + \mathbb{P}(A_n=a|A_{n-1}=b).\mathbb{P}(A_{n-1}=b) + \mathbb{P}(A_n=a|A_{n-1}=c)\mathbb{P}(A_{n-1}=c)$

$\mathbb{P}(A_n=b) = \mathbb{P}(A_n=b|A_{n-1}=a)\mathbb{P}(A_{n-1}=a) + \mathbb{P}(A_n=b|A_{n-1}=b)\mathbb{P}(A_{n-1}=b) + \mathbb{P}(A_n=b|A_{n-1}=c)\mathbb{P}(A_{n-1}=c)$

$\mathbb{P}(A_n=c) = \mathbb{P}(A_n=c|A_{n-1}=a)\mathbb{P}(A_{n-1}=a) + \mathbb{P}(A_n=c|A_{n-1}=b)\mathbb{P}(A_{n-1}=b) + \mathbb{P}(A_n=c|A_{n-1}=c)\mathbb{P}(A_{n-1}=c)$

et donc les probabilités de A_n en fonction des A_{n-1}]

- formule de Bayes :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

- A et B sont 2 événements indépendants $\Leftrightarrow \mathbb{P}(A\cap B) = \mathbb{P}(A).\mathbb{P}(B)$

→ les variables aléatoires discrètes :

$X : \Omega \to X(\Omega)$ qui prend 1 nb fini ou infini de valeurs.

- loi de X :

$\forall x \in X(\Omega)$, $\mathbb{P}(X=x)$

- fonction de répartition de X :

$$F : \mathbb{R} \to [0;1]$$
$$x \mapsto \mathbb{P}(X \leq x)$$

- espérance de X : $\mathbb{E}[X] = \sum_{x\in X(\Omega)} x\,\mathbb{P}(X=x)$

- si $Y = g(X)$: $Y : \Omega \to g(X(\Omega))$

$$\mathbb{P}(Y=y) = \mathbb{P}(g(X)=y) = \mathbb{P}(X \in g^{-1}(y))$$

- variance de X : $\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ formule de Huyghens-König

avec : $\mathbb{E}[X^2] = \sum_{x\in X(\Omega)} x^2\,\mathbb{P}(X=x)$

- covariance de X et Y : $\text{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

- $\mathbb{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\,\text{cov}(X,Y)$

- vecteur aléatoire discret :

$$X : \Omega \to X_1(\Omega) \times X_2(\Omega) \times X_3(\Omega)$$
$$\omega \mapsto \begin{pmatrix} X_1(\omega) \\ X_2(\omega) \\ X_3(\omega) \end{pmatrix} \quad \text{avec } (X_i)_{i=1,2,3} \text{ variables aléatoires discrètes}$$

- loi de X :

$\forall (x_1 \in X_1(\Omega)), \forall x_2 \in X_2(\Omega), \forall x_3 \in X_3(\Omega), \mathbb{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3)$

- loi marginale de X_1 :

soit $x_1 \in X_1(\Omega)$, $\mathbb{P}(X_1 = x_1) = \sum_{\substack{x_2 \in X_2(\Omega) \\ x_3 \in X_3(\Omega)}} \mathbb{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3)$

- loi conditionnelle de X sachant que $Y = y$:

soit $x \in X(\Omega)$, $\mathbb{P}(X = x \mid Y = y) = \dfrac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$

- indépendance de 2 vad :

$\forall x \in X(\Omega), \forall y \in Y(\Omega), \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) . \mathbb{P}(Y = y)$

- indépendance de n vad :

$\forall x_1 \in X_1(\Omega), \ldots, \forall x_n \in X_n(\Omega), \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$

- si X et Y sont ind alors : $\mathbb{E}[XY] = \mathbb{E}[X] . \mathbb{E}[Y]$

$cov(X;Y) = 0$

→ lois discrètes usuelles :

- loi de Bernoulli de paramètre p : $X : \Omega \to \{0;1\}$

$\mathbb{P}(X = 1) = p$

- loi binomiale de paramètres n et p :

C'est le nombre de succès obtenus si on répète indépendamment n épreuves de Bernoulli

$X : \Omega \to [\![0;n]\!]$

$$\mathbb{P}(X = k) = C_n^k \, p^k (1-p)^{n-k} = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}, \forall k \in [\![0;n]\!]$$

[en exercice : notez : $X = \sum_{i=1}^{n} X_i$ avec X_i qui suit 1 loi de Bernoulli de paramètre p]

- loi géométrique de paramètre $p \in]0;1[$

On considère 1 expérience qui a la probabilité p de réussir, c'est le nombre de tentatives nécessaires pour obtenir 1 succès.

$X : \Omega \to \mathbb{N}^*$

$\forall n \geq 1, \mathbb{P}(X = n) = (1-p)^{n-1} p$

- loi de Poisson de paramètre $\lambda > 0$: $X : \Omega \to \mathbb{N}$

$$\forall k \in \mathbb{N}, \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

→ variables aléatoires à densité :

- f est 1 densité si :

(1) $f : \mathbb{R} \to \mathbb{R}^+$

(2) f est continue sur $\mathbb{R}$ sauf en 1 nb fini de pts

(3) $\int_{-\infty}^{+\infty} f(t)\,dt = 1$

- $\forall (a;b) \in \mathbb{R}^2$, $\mathbb{P}(X \in [a;b]) = \int_a^b f(t)\,dt$

- fonction de répartition : $F_X : \mathbb{R} \to [0;1]$
$x \mapsto \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$

- propriété : $F_X'(x) = f_X(x)$

- $\mathbb{E}[X] = \int_{-\infty}^{+\infty} t\, f(t)\,dt$

- $\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

avec : $\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} t^2 f(t)\,dt$

- densités usuelles :

- loi uniforme : $f(x) = \frac{1}{b-a} \mathbb{1}_{[a;b]}(x)$

- loi exponentielle de paramètre λ : $f(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0;+\infty[}(x)$

- loi normale de paramètre μ et σ^2 : $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

- lois de $aX+b$, X^2, e^X ?

* $Y = X^2$

fonction de répartition de Y ?

soit $y \in \mathbb{R}^+$, $\mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$

[Dans ce cas, toujours passer par la fonction de répartition]

- loi du χ^2 à k degrés de liberté :

$X \sim \chi^2(k) \iff X = \sum_{i=1}^{k} X_i^2$ avec $X_i \sim \mathcal{N}(0;1)$
et $(X_i)_{i=1,\dots,k}$ mutuellement indépendantes

- $\mathbb{E}[X] = \sum_{i=1}^{k} \mathbb{E}[X_i^2] = k\, \mathbb{E}[X_1^2]$ car les $(X_i)_i$ sont de même loi

$= k\left[\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2\right] = k\, \mathbb{V}[X_1] = k$

- $\mathbb{V}[X] = 2k$

<u>démo</u>: $\mathbb{V}\left[\sum_{i=1}^{k} X_i^2\right] = \sum_{i=1}^{k} \mathbb{V}[X_i^2]$ car les $(X_i)_i$ sont mutuellement indépendantes

$= k\,\mathbb{V}[X_1^2]$ car les $(X_i)_i$ sont de même loi

$= k\left[\mathbb{E}[X_1^4] - \mathbb{E}[X^2]^2\right]$

$= k\left[\mathbb{E}[X_1^4] - \mathbb{V}[X_1]\right] = k\left[\mathbb{E}[X_1^4] - 1\right]$

il reste à montrer que : $\mathbb{E}[X_1^4] = 3$?

$$\mathbb{E}[X_1^4] = \int_{-\infty}^{+\infty} x^4 \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,dx = \int_{-\infty}^{+\infty} x^3 \quad x\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,dx$$

On intègre par parties en posant :

$u' = x e^{-\frac{x^2}{2}}/\sqrt{2\pi} \Rightarrow u = -\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$

$v = x^3 \Rightarrow v' = 3x^2$

$$\mathbb{E}[X_1^4] = \left[-x^3\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\right]_{-\infty}^{+\infty} + 3\int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}}\,dx$$

$$= 3 \cdot \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}}\,dx$$

$$= 3 \cdot \mathbb{E}[X_1^2]$$

$$= 3 \cdot 1 = 3$$

donc : $\underline{\mathbb{V}\left[\sum_{i=1}^{k} X_i^2\right]} = k[3-1] = \underline{2k}$

→ <u>intervalle de confiance à 95 % d'1 moyenne :</u>

- on calcule la moyenne sur 1 échantillon $(X_1, .., X_n)$ de la population

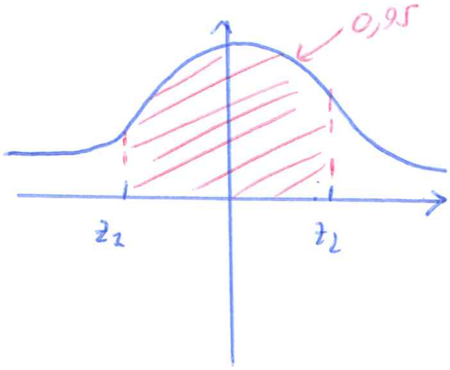
$\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}$

- On suppose que la variable aléatoire suit 1 loi $\mathcal{N}(\mu; \sigma^2)$ [la variance σ^2 est connue]

par conséquent : $\bar{x} \sim \mathcal{N}\left(\mu; \frac{\sigma^2}{n}\right)$

donc : $\underline{\sqrt{n}\left(\frac{\bar{x}-\mu}{\sigma}\right) \sim \mathcal{N}(0;1)}$

- On cherche : z_1 et z_2 tels que : $\mathbb{P}\left(z_1 \leq \sqrt{n}\,\frac{(\bar{x}-\mu)}{\sigma} \leq z_2\right) = 0{,}95$



vrai pour : $z_1 = -1{,}96$ et $z_2 = 1{,}96$

donc : $-1{,}96 \leq \sqrt{n}\,\frac{\bar{x}-\mu}{\sigma} \leq 1{,}96 \iff \mu \in \left[\bar{x} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} \,;\, \bar{x} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}\right]$

→ intervalle de confiance à 95 % d'1 variance :

- on calcule la variance empirique sur 1 échantillon $(X_1, \dots, X_n)$ de la population

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{x})^2$$

1 $\displaystyle\sum_{i=1}^{n} \frac{(X_i-\bar{x})^2}{\sigma^2} \sim \chi^2(n-1)$