Théorie d'échantillonnage & Estimations & Tests Statistiques

Notions de base:

-La population est l'ensemble des individus sur lesquels porte l'étude statistique.
-Un échantillon est un sous-ensemble ou une partie d'une population.
-L'individu ou unité statistique est un élément de la population.
-Un caractère statistique (ou variable statistique) c'est ce qui est observé ou mesuré sur les individus d'une population statistique. On note le caractère statistique par X.

Lorsque le caractère statistique est un nombre (taille, note, nombre d'enfants. . .) on parle de caractère quantitatif (continu ou discret).On peut dire aussi un caractère mesurable.

Quand le caractère n'est pas chiffré (langue parlée, secteur d'activité, couleur...) on parle de caractère qualitatif (soit nominal, soit ordinal). On peut dire aussi un caractère non-mesurable.

-Une loi de probabilité décrit de manière théorique le caractère aléatoire d'une expérience dont le résultat dépend du hasard.

Les symboles utilisés

Population	**Échantillon**	**Distributions d'échantillonnage**
$taille\ N$ $moyenne\ \mu$ $variance\ \sigma_x^2$ $écart-type\ \sigma_x$ $la\ proportion\ P$	$taille\ n$ $moyenne\ \overline{X}$ S^2 S f	$E(\overline{X})$ $écart-type\ \sigma_{\overline{x}}$ $moyenne\ E(f)$ $écart-type\ \sigma_f$ $moyenne\ E(S^2)$ $écart-type\ \sigma_{S^2}$

I) Théorie d'échantillonnage:

Théorie de l'échantillonnage = Étude des liaisons existant entre une population et les échantillons de cette population.

De la théorie de l'échantillonnage, nous obtenons 3 Distributions d'échantillonnage:
-Distributions d'échantillonnage de la moyenne.
-Distributions d'échantillonnage de la variance.
-Distributions d'échantillonnage de la proportion.

Chaque distribution d'échantillonnage a 3 propriétés:
1) Une loi de probabilité.

2) Une moyenne.
3) Un écart-type.

1) Distributions d'échantillonnage de la moyenne:

On distingue 3 cas:

1) 1 ^{er} cas : population connue, σ_x connu :
- La loi : loi Normale .$\overline{X} \frown N(\mu_{\overline{x}}\,;\sigma_{\overline{x}})$
- La moyenne : $E(\overline{X}) = \mu_x$
- Lécart-type : $\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$

2) 2 ^{ème} cas : population inconnue, σ_x inconnu et $n \geq 30$:
- La loi : loi Normale .$\overline{X} \frown N(\mu_{\overline{x}}\,;\sigma_{\overline{x}})$
- La moyenne : $E(\overline{X}) = \mu_x$
- Lécart-type : $\sigma_{\overline{X}} = \frac{S}{\sqrt{n}}$

3) 3 ^{ème} cas : population inconnue, σ_x inconnu et $n < 30$:
- La loi : loi de Student .$\overline{X} \frown t(n-1)$
- La moyenne : $E(\overline{X}) = \mu_x$
- Lécart-type : $\sigma_{\overline{X}} = \frac{S}{\sqrt{n}}$

2) Distributions d'échantillonnage de la Proportion:

- La loi : loi Normale .$f \frown N(E(f)\,;\sigma_f)$
- La moyenne : $E(f) = P$
- Lécart-type : $\sigma_f = \sqrt{\frac{P(1-P)}{n}}$

3) Distributions d'échantillonnage de la variance:

- La loi : loi du khi-deux (χ^2).
- La moyenne : $E(S^2) = \frac{n-1}{n}\sigma_x^2$
- Lécart-type : $\sigma_{S^2} = \sqrt{\frac{n-3}{n(n-1)}}\ \ \sigma_x^2$

II) L'estimation

En statistique, un estimateur est une fonction permettant d'estimer un moment d'une loi de probabilité (comme son espérance ou sa variance).

On distingue plusieurs méthodes d'estimation, mais nous allons concentrer sur les plus utilisées : l'estimation ponctuelle et l'estimation par intervalle de confiance.

L'estimation ponctuelle

L'estimation d'un paramètre quelconque θ est ponctuelle si l'on associe une seule valeur à l'estimateur $\widehat{\theta}$ à partir des données observables sur un échantillon.

1- L'estimateur ponctuel de la moyenne μ_x est : $\overline{X}$

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

2- L'estimateur ponctuel de variance σ_x^2 est : S^2

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

3- L'estimateur ponctuel d'écart-type σ_x est : S

$$S = \sqrt{S^2}$$

4- L'estimateur ponctuel de P est : f

$$f = \frac{Nombre\ des\ cas\ favorables}{Nombre\ des\ cas\ possibles}$$

*Propriétés des estimateurs ponctuels :
1- Sans-biais $\Leftrightarrow$ $E(\widehat{\theta}) = \theta$
2- Efficace $\Leftrightarrow Var(\widehat{\theta}_1) < Var(\widehat{\theta}_2) < ...$

3-Convergent $\Leftrightarrow \lim\limits_{n \longrightarrow +\infty} Var(\widehat{\theta}) = 0$

L'estimation par intervalle de confiance.

En théorie des probabilités et en statistiques, un intervalle de confiance encadre une valeur réelle que l'on cherche à estimer à l'aide de mesures prises par un procédé aléatoire. Un intervalle de confiance doit être associé à un niveau de confiance, en général sous la forme d'un pourcentage, qui minore la probabilité que l'intervalle contienne la valeur à estimer(niveau de confiance (1-α) % *ou* α *est le risque d'erreur ,généralement* $\alpha = 5$ *ou*1 %).

1)L'estimation par intervalle de confiance de la moyenne :

1 ^{er} cas : population connue, σ_x connu :

$$\mu_x \in IC\ [\overline{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma_x}{\sqrt{n}}\ ;\ \overline{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma_x}{\sqrt{n}}]$$

$Z_{1-\frac{\alpha}{2}} = 1.96$ (si $\alpha = 0.05$)
$Z_{1-\frac{\alpha}{2}} = 2.58$ (si $\alpha = 0.01$)

2 ^{ème} cas : population inconnue, σ_x inconnu et $n \geq 30$

$$\mu_x \in IC\ [\overline{X}\ - Z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\ ;\ \overline{X}\ + Z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}]$$

3 ^{ème} cas : population inconnue, σ_x inconnu et $n < 30$:

$$\mu_x \in IC\ [\overline{X}\ - t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\ ;\ \overline{X}\ + t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}]$$

$t_{1-\frac{\alpha}{2}}$ (table de la loi de student 0.05 ou 0.01 two-tails avec n-1)

2)L'estimation par intervalle de confiance de la proportion :

$$\begin{aligned} n * f &\geq 5 \\ n * (1-f) &\geq 5 \end{aligned}$$

$$P \in IC\ [f - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}}\ ;\ f + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}}]$$

3)L'estimation par intervalle de confiance de la variance :

$$\sigma_x^2 \in IC\ [\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}; \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}}]$$

$\chi^2_{1-\frac{\alpha}{2}}$(table de la loi khi-deux $1-\frac{\alpha}{2}$ avec $n-1$)
$\chi^2_{\frac{\alpha}{2}}$ (table de la loi khi-deux $\frac{\alpha}{2}$ avec $n-1$)

III) Tests d'hypothèse:

Principe d'un test d'hypothèse:

Les tests d'hypothèse constituent un autre aspect important de l'inférence statistique. Le principe général d'un test d'hypothèse peut s'énoncer comme suit :

- On étudie une population dont les éléments possèdent un caractère (mesurable ou qualitatif) et dont la valeur du paramètre relative au caractère étudié est inconnue.

- Une hypothèse est formulée sur la valeur du paramètre : cette formulation résulte de considérations théoriques, pratiques ou encore elle est simplement basée sur un pressentiment.

- On veut porter un jugement sur la base des résultats d'un échantillon prélevé de cette population.

Une **hypothèse statistique** est un énoncé (une affirmation) concernant les caractéristiques (valeurs des paramètres, forme de la distribution des observations) d'une population.

Un **test d'hypothèse** (ou test statistique) est une démarche qui a pour but de fournir une règle de décision permettant, sur la base de résultats d'échantillon, de faire un choix entre deux hypothèses statistiques.

Hypothèse nulle ($\mathbf{H_0}$) et hypothèse alternative ($\mathbf{H_1}$)

L'hypothèse selon laquelle on fixe à priori un paramètre de la population à une valeur particulière s'appelle l'hypothèse nulle et est notée H_0. N'importe quelle autre hypothèse qui diffère de l'hypothèse H_0 s'appelle l'hypothèse alternative (ou contre-hypothèse) et est notée H_1 .

C'est l'hypothèse nulle qui est soumise au test et toute la démarche du test s'effectue en considérant cette hypothèse comme vraie.

Les étapes à suivre pour faire un test statistique:

1-Donner les hypothèses :
2-Seuil de signification
3-Conditions d'application
4-La variable appropriée
5-La règle de décision
6-Calculer la valeur de la variable appropriée
7- Conclusion et décision