République algérienne démocratique et populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Université Abou-bekr Belkaid –Tlemcen

Faculté des sciences humaines et sociales

Département des sciences sociales

COURS 4 et 5

Destiné aux étudiants en 2ème année (Semestre4)

Filière : Démographie

LES TECHNIQUES DE SONDAGE

Responsable du Module : Mme. MORTAD Nedjlaà (M.C.A)

Année universitaire : 2019 - 2020

Chapitre II : L'estimation

Cours 4 : L'estimation

Ce cours s'intéresse à répondre à la problématique suivante: comment, à partir d'informations (couple moyenne/écart type ou proportion) calculées sur un échantillon, retrouver ou plutôt estimer celles d'une population entière ?

1- Estimation : principe

L'estimation désigne le procédé par lequel on détermine les valeurs inconnues des paramètres de la population à partir des données de l'échantillon. Pour cela, on utilise des distributions théoriques, c'est à dire des variables aléatoires dont on connait les lois de probabilités[1]. Il s'agit d'un phénomène réciproque de l'échantillonnage.

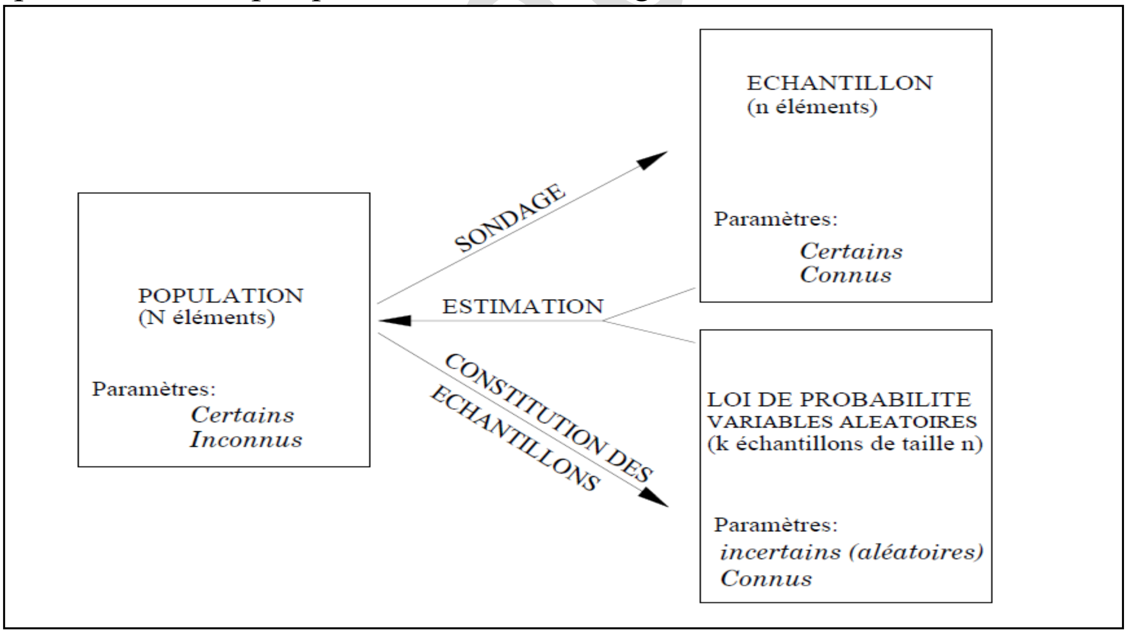

Figure 4 : Principe général de l'estimation[2]

[1] - Théorie statistique de l'échantillonnage ; https://step.ipgp.fr/images/4/49/C5a.pdf
[2] - Idem

Le but donc, est d'estimer à partir d'un échantillon, la ou les valeurs numériques d'un ou de plusieurs paramètres de la population considérée et de déterminer la précision de cette ou de ces estimations. L'estimateur est une nouvelle variable aléatoire construite à partir des données expérimentales de l'échantillon et dont la valeur se rapproche du paramètre que l'on cherche à connaitre.

Nous distinguons deux cas : celui où l'on cherche à estimer une moyenne μ d'une variable aléatoire définie sur une population et celui où l'on cherche à estimer la proportion d'individus p possédant un caractère précis dans la population.

En outre, il existe deux formes d'estimations : l'estimation ponctuelle et l'estimation par intervalle de confiance[3].

[3] - El MARHOUM adil, Echantillonnage et estimation, méthodes quantitatives, université Mohamed V- AGDAL , 2013 p80.

Cours 5
Estimation d'une moyenne
Estimation ponctuelle d'une moyenne

1-Estimation d'une moyenne

1.1-Estimation ponctuelle d'une moyenne

Lorsqu'une caractéristique d'une population (un paramètre) est estimée par un seul nombre, déduit les résultats de l'échantillon, ce nombre est appelé une estimation ponctuelle du paramètre[4].

On considère une variable X sur une population de moyenne (ou espérance) μ inconnue et d'écart type σ inconnu (ou connu). Admettons qu'un échantillon de taille n est prélevé (tirage avec remise ou assimilé) sur lequel on a calculé la moyenne μ_e et l'écart type σ_e.

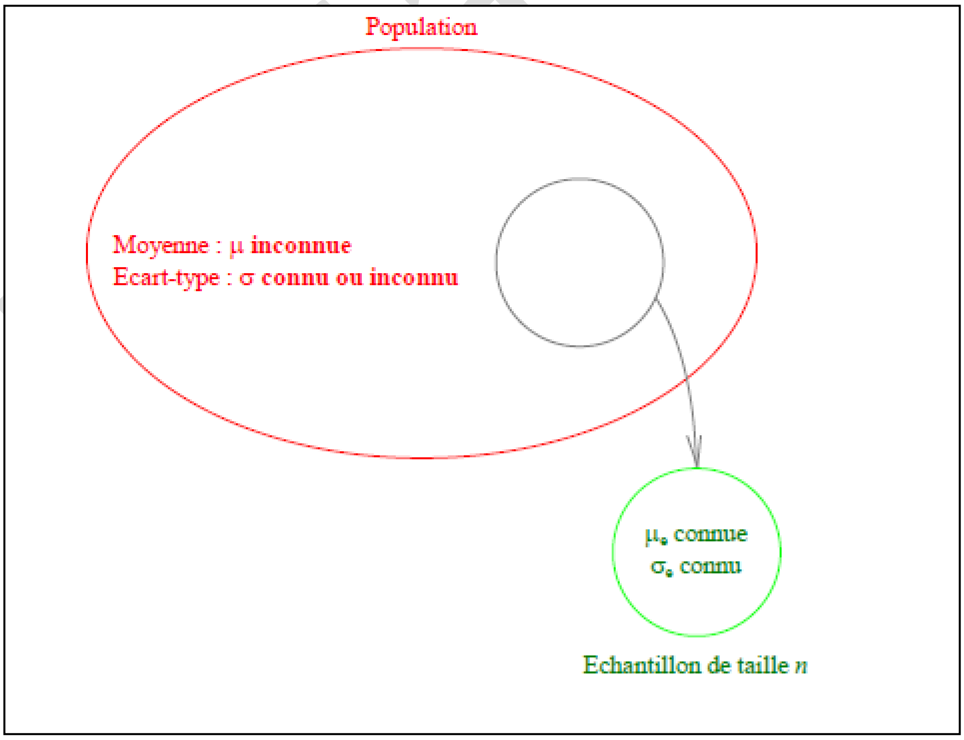

Figure 5 : Estimation ponctuelle d'une moyenne[5]

[4] -BAILLARGEON gérald , Probabilités, statistique et techniques de régression, ed SMG, 1989, P230.

Une estimation ponctuelle $\hat{\mu}$ de la moyenne μ de la population est :

$$\hat{\mu} = \mu_e \quad \text{ou bien} \quad \hat{\mu} = E(\overline{X})$$

Une estimation ponctuelle $\hat{\sigma}$ de l'écart type σ_e de la population est :

$$\hat{\sigma} = \sqrt{\frac{n}{n-1}}\,\sigma_e$$

Le coefficient $\sqrt{\frac{n}{n-1}}$ est appelé : correction de biais. Plus l'échantillon est grand ($n \geq 30$), plus ce coefficient s'approche de 1, dans ce cas, on peut estimer $\hat{\sigma} \simeq \sigma_e$[6]

Exemple1 :

Une usine produit des vis cruciformes. On souhaite estimer la moyenne des longueurs des vis dans la production de la journée qui s'élève à 10000 pièces. On prélève un échantillon de 150 vis, et on obtient une moyenne $E(\bar{X}) = 4{,}57$cm. Estimer la longueur moyenne des vis de la production journalière.

Solution : $\hat{\mu} = 4{,}57$cm

---

[5] COSTANTINI gille, Echantillonnage-estimation, statistiques inférentielles, BTS 2ème année, p 5 https://docplayer.fr/13457270-Partie-a-echantillonnage.html

[6] - COSTANTINI gille, Echantillonnage-estimation, statistiques inférentielles, BTS 2ème année, p6 https://docplayer.fr/13457270-Partie-a-echantillonnage.html

Exemple 2 :

Une université comporte 1500 étudiants. On mesure la taille de 20 d'entre eux. La moyenne μ_e et l'écart type σ_e calculés à partir de cet échantillon sont : μ_e = 176cm et σ_e= 6cm. Estimer les paramètres de la population ?

Solution :

$\hat{\mu} = \mu_e = 176cm$

$\hat{\sigma} = \sqrt{\frac{n}{n-1}}\ \sigma_e$

$\hat{\sigma} = 6\ \sqrt{\frac{20}{19}}$

$\hat{\sigma} = 6{,}16cm$

Cours 6

Estimation par intervalle de confiance d'une moyenne

Ecart type σ de la population connu (ou variance connue)

1-Estimation par intervalle de confiance d'une moyenne

1.1-Principe

Les estimations ponctuelles, bien qu'utiles, ne fournissent aucune information concernant la précision des estimations, elles ne tiennent pas compte de l'erreur possible dans l'estimation, erreur attribuable aux fluctuations d'échantillonnage[7].

L'estimation par intervalle de confiance consiste à déterminer autour de la valeur estimée un intervalle dont on a de fortes chances de croire qu'il contient la vraie valeur du paramètre recherché[8].

Si on s'intéresse à un paramètre θ, dont on possède un estimateur T, l'estimation par intervalle de confiance consiste à déterminer de part et d'autre de T les bornes T1 et T2 d'un intervalle qui a une forte probabilité de contenir θ. Cette probabilité est appelée niveau de confiance et désignée par $(1-\alpha)$. α est alors un risque d'erreur[9].

Les limites T1 et T2 sont telles que :

$$P(T1 \leq \theta \leq T2) = 1-\alpha$$

L'intervalle $[T1, T2]$ est appelé intervalle de confiance.

T1 : limite inférieure de l'intervalle de confiance.

T2 : limite supérieure de l'intervalle de confiance.

---

[7]- BAILLARGEON gérald , Probabilités, statistique et techniques de régression, ed SMG, 1989,, p232.

[8] - El MARHOUM adil, Echantillonnage et estimation, méthodes quantitatives, université Mohamed V-AGDAL , 2013 p80.

[9] - El MARHOUM adil, Echantillonnage et estimation, méthodes quantitatives, université Mohamed V-AGDAL , 2013 p88.

1- α : la probabilité associée à l'intervalle d'encadrer la vraie valeur du paramètre.

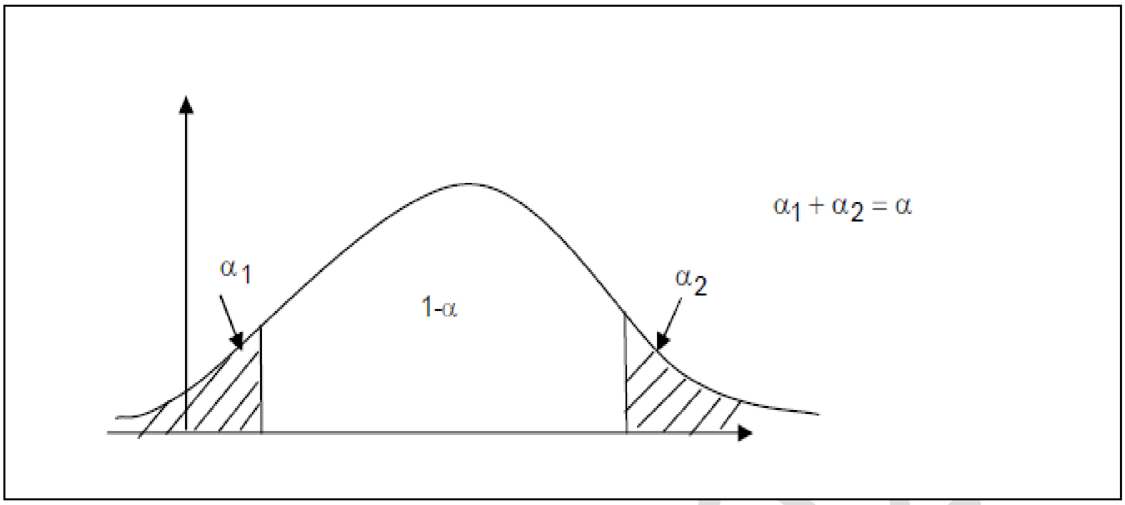

Figure 8 : Principe de l'estimation par intervalle de confiance : proposer un encadrement d'un paramètre inconnu d'une population dont la loi, elle, est connue[10]

Cet intervalle de confiance signifie que, si nous répétons l'expérience un grand nombre de fois (prélever plusieurs fois un échantillon de taille n de la même population) dans 100 cas (1- α) cas sur 100, l'intervalle recouvre la vraie valeur du paramètre[11].

•Niveau de confiance : la probabilité (1-α) souvent exprimée en pourcentage s'appelle le niveau de confiance de l'intervalle. Ce niveau est décidé au préalable et peut être aussi élevé (proche de 100%) que l'on veut. On choisit souvent α= 5%, ce qui donne un niveau de confiance de 95%[12].

•Le seuil de risque : la probabilité α, exprimée en pourcentage, est appelée le niveau de risque. A représente la probabilité de se tromper en affirmant que l'intervalle de confiance contient le paramètre θ. A désigne

[10] - Module 2 : estimation par intervalle de confiance, www.foad-mooc.auf.org/IMG/pdf/M02.pdf
[11] - BAILLARGEON gérald , Probabilités, statistique et techniques de régression, ed SMG, 1989, P233.
[12] - BERTRANDIAS jean paul et françoise, Mathématiques pour les sciences de la nature et de la vie, office des publications universitaires : 01, 1994, p 141.

la probabilité pour que l'intervalle que l'on détermine ne contienne pas la vraie valeur du paramètre.

1.2-Intervalle de confiance d'une moyenne (ou d'une espérance)

- **Ecart type σ de la population connu (ou variance connue)**

La variable aléatoire parente X suit une loi de probabilité de paramètre $\mu = E(X)$ inconnu et de variance σ^2 connue. Soit X_1, X_2, X_3,….X_n un n-échantillon aléatoire simple de X.

Si les échantillons sont non exhaustifs (ou extraits d'une population infinie), et si X suit une loi normale, ou si les échantillons sont de taille suffisament grande ($n \geq 30$), alors la variable aléatoire $\bar{X}_n$, qui à tout échantillon de taille n, associe la moyenne des valeurs de X observées sur cet échantillon suit une loi normale de paramètres :

$$\bar{X} = \frac{x1 + x2 + \cdots + xn}{n} \sim N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

Une démonstration utilisant le calcul des probabilités nous permet d'établir le résultat suivant :

$$P(\bar{X} - Z_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_\alpha \frac{\sigma}{\sqrt{n}}) = \alpha$$

D'où la définition suivante :

Si X est une variable aléatoire de moyenne (ou d'espérance) inconnue μ et d'écart type σ connu, et si $\bar{X}$ est la moyenne des valeurs de X observées sur un échantillon de taille n, l'intervalle de confiance du paramètre μ, au seuil de confiance α, est :

$$I\alpha = \left[\bar{X} - Z_\alpha \frac{\sigma}{\sqrt{n}} ; \bar{X} + Z_\alpha \frac{\sigma}{\sqrt{n}}\right]$$

- Les valeurs les plus usuelles de Z_α :

Au seuil de confiance de 95% (ou de risque de 5%) : $\alpha = 0,975$

Et, d'après la table de la loi normale centrée réduite N(0,1) dont la fonction de répartition et notée π), on obtient :

$$\Pi(Z_\alpha) = \frac{\alpha+1}{2} = 0,975 \Leftrightarrow Z_\alpha = 1,96$$

Et dans ce cas :

$$I_{95\%} = \left[\overline{X} - 1,96\,\frac{\sigma}{\sqrt{n}}\ ;\ \overline{X} + 1,96\,\frac{\sigma}{\sqrt{n}}\right]$$

Cet intervalle est appelé intervalle de confiance de la moyenne de la population avec le coefficient de confiance α = 95% ou avec le risque 5%.

De même, au seuil de confiance de 99% (ou de risque de 1%), α=0,99 donne Z_α= 2,575, et l'intervalle de confiance sera :

$$I_{99\%} = \left[\overline{X} - 2,575\,\frac{\sigma}{\sqrt{n}}\ ;\ \overline{X} + 2,575\,\frac{\sigma}{\sqrt{n}}\right]$$

Le seuil de 90% est aussi utilisé ; avec un Z_α = 1,645 et l'intervalle de confiance :

$$I_{90\%} = \left[\overline{X} - 1,645\,\frac{\sigma}{\sqrt{n}}\ ;\ \overline{X} + 1,645\,\frac{\sigma}{\sqrt{n}}\right]$$

seuil de confiance	α	$\alpha/2$	$z_{1-\alpha/2}$
90%	0.1	0.05	1.645
95%	0.05	0.025	1.96
99%	0.01	0.005	2.576

Tableau 1 : Valeurs de Z_α pour les seuils de confiance les plus fréquemment utilisés[13]

[13] - BAILLARGEON gérald , Probabilités, statistique et techniques de régression, ed SMG, 1989, P237.

Remarque :

Pour une population dont le caractère étudié est une variable aléatoire X ne suivant pas nécessairement une loi normale, mais de moyenne m et d'écart type σ, et avec une taille suffisamment grande de l'échantillon considéré (avec lequel on travail). Dans ce cas, la variable aléatoire $\overline{X}$, suit d'après le théorème central limite, approximativement une loi normale.

Exemple 1[14]*:*

Un laboratoire indépendant a vérifié pour le compte de l'office de la protection du consommateur, la résistance à l'éclatement (en kg/cm²) d'un réservoir à essence d'un certain fabricant. Des essais similaires, effectués il y a un an, permettent de considérer que la résistance à l'éclatement est distribuée normalement avec une variance de 9.

Des essais sur un échantillon de 10 réservoirs conduisent à une résistance moyenne à l'éclatement de 219kg/cm².

Estimer par intervalle de confiance la résistance moyenne à l'éclatement de ce type de réservoir avec un niveau de confiance de 95%.

Solution :

La moyenne de l'échantillon $\overline{X} = 219$

L'écart type de la population normale : $\sigma = \sqrt{9} = 3$ kg/cm²

La taille de l'échantillon : n=10

Le niveau de confiance : $1-\alpha = 0{,}95$, $\alpha = 0{,}05$

Sous ces conditions, les limites de l'intervalle sont :

[14] - Idem, p 236.

$$\overline{X} - Z_{\alpha}\ \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\alpha}\ \frac{\sigma}{\sqrt{n}}$$

Avec $Z_{\alpha} = 1{,}96$

$$I_{95\%} = \left[219 - 1{,}96\,\frac{3}{\sqrt{10}}\ ;\ 219 + 1{,}96\,\frac{3}{\sqrt{10}}\right]$$

$$I_{95\%} = [217{,}14\ ;\ 22O{,}86]$$

Donc, on attribue à cet intervalle, un niveau de confiance de 95% de contenir la vraie valeur de la résistance moyenne à l'éclatement pour le réservoir de ce fabricant.