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 1 et 2

Destiné aux étudiants Master 1

Filière : Démographie

LES TECHNIQUES DE SONDAGE

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

Année universitaire : 2020 - 2021

Chapitre I : L'échantillonnage

Cours 1 : L'échantillonnage

1- Introduction

L'étude de certaines caractéristiques d'une population, quand on ne dispose pas encore de données est souvent très onéreuse, elle nécessite d'examiner, d'observer des éléments de cette population. La manière de recueillir ces données fait l'objet d'une théorie mathématique appelée : théorie des sondages ou encore théorie de l'échantillonnage (en anglais : sampling theory)[1].

Les sondages s'effectuent en tirant des échantillons de personnes, d'entreprises ou autres (nommés unités) que l'on enquête afin d'obtenir l'information souhaitée[2].

Lorsqu'on désire collecter cette information, deux possibilités se présentent : la première consiste à observer ou à enquêter tous les éléments de la population, c'est une enquête complète ou enquête exhaustive ou recensement. La seconde vise à interroger une partie de la population, et il s'agit d'une enquête partielle ou sondage. Les éléments de la population qui sont réellement étudiés forment l'échantillon et l'acte de choisir ces éléments est appelé : échantillonnage.

[1] - KHERRI abdenacer , Support pédagogique de cours : échantillonnage, statistique de gestion, 2013/2014, p.2, www.sg-ehec.jimdo.com

[2] - LAVALLEE pierre, le sondage indirect ou la méthode généralisée ou partage des poids, Ed. de l'université de Bruxelles, Ellipses, 2002,p1.

2- Vocabulaire[3][4][5]

- Echantillon : Un sous ensemble de la population considérée, c'est l'ensemble aussi des unités de base sélectionnées et réellement observées au cours d'un sondage.
- Echantillonnage : la sélection d'une partie dans un tout (la sélection d'une partie dans la population), l'échantillon sélectionné doit être représentatif de la population.
- Population : ensemble sue l'on observe et qui sera soumis à une analyse statistique, chaque élément de cet ensemble est un individu ou unité statistique.
- Unité de base : unité d'échantillonnage ou unité de sondage, c'est l'élément pris en considération dans l'enquête.
- Enquête : ensemble des opérations de collecte et de traitement de données relatives à quelques domaines que ce soit.
- Sondage : enquête incomplète, enquête partielle ou enquête par échantillonnage, c'est une enquête au cours de laquelle seulement une partie des unités de base de la population est observée.
- Base de sondage : énumération ou présentation ordonnée de toutes les unités de base constituant la population.
- Fraction ou taux de sondage : proportion des unités de la population qui font partie de l'échantillon. C'est le rapport entre la taille de l'échantillon n, et la taille de la population N.

$$f = n / N x 100$$

[3] - EL MARHOUM adil, Cours d'échantillonnage et estimation : licence fondamentale en sciences économiques et gestion. Université Mohammed V, Rabat, 2014/2015, p3.www.coursdefsjes.com

[4] - Christophe Chesneau. Éléments de théorie des sondages. Master. France, 2016,P8. https://cel.archives-ouvertes.fr/cel-01292370

[5] -_KHERRI abdenacer , Support pédagogique de cours : échantillonnage, statistique de gestion, 2013/2014, p.2, www.sg-ehec.jimdo.com

- Recensement : enquête complète ou enquête exhaustive, c'est une enquête au cours de laquelle toutes les unités de base de la population sont observées.
- Population finie : une population qui consiste en un nombre fini d'éléments.
- Population infinie : une population est infinie s'il n'y a pas de limite au nombre d'éléments qu'il contient.
- Population homogène : une population avec des éléments qui possèdent les mêmes caractéristiques.
- Population non homogène : une population avec des éléments qui ne possèdent pas les mêmes caractéristiques.
- Paramètre : Caractéristique numérique d'une population telle que la moyenne de la population « μ », l'écart type de la population « σ » et la proportion de la population « P ».
- Distribution d'échantillonnage : Distribution de probabilité composée de toutes les valeurs possibles d'une statistique d'échantillon.
- Tirage exhaustif : tirage sans remise.
- Tirage non exhaustif : tirage avec remise.
- Méthodes d'échantillonnage : ensemble des méthodes permettant de réaliser un sondage (de prélever un échantillon de données) au sein d'une population, de manière à reproduire un échantillon aussi représentatif que possible de cette population.

3- Echantillonnage

L'échantillonnage est le procédé utilisé pour choisir un échantillon et qui est la base de l'enquête par sondage[6]. C'est la phase qui consiste à sélectionner les individus que l'on souhaite interroger au sein d'une population de base[7] (Figure 1).

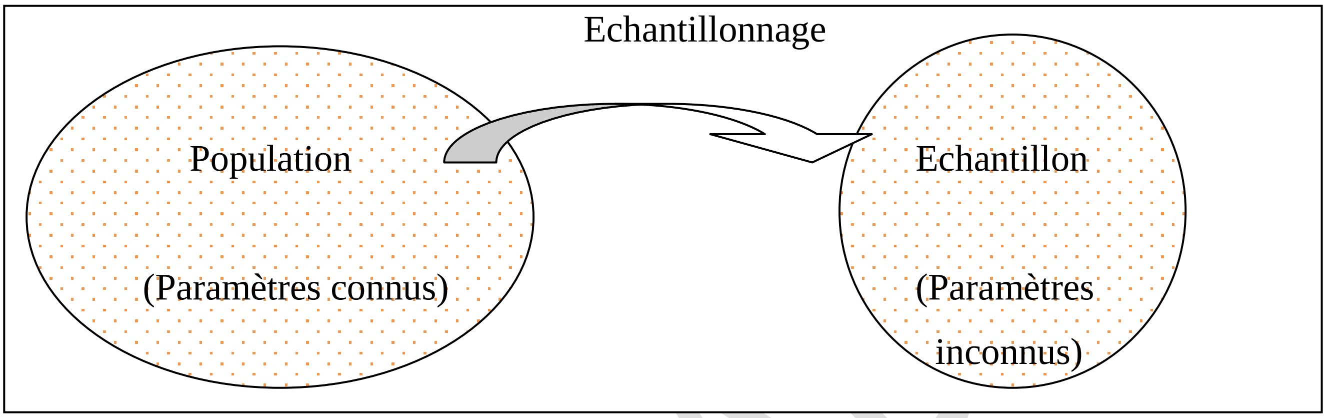

Figure 1 : L'échantillonnage : Passage de la population vers l'échantillon

Pour chaque échantillon, on peut calculer une statistique (moyenne, écart type, variance, etc…) qui variera avec l'échantillon. Pour tous les échantillons, on obtient alors une distribution de la statistique que l'on nomme la distribution d'échantillonnage[8]. Pour la conformité des résultats obtenus, il est important que les échantillons soient représentatifs de la population étudiée.

---

6 - http://www.larousse.fr

7 - http://www.defnitions-marketing.com

8 - KHERRI abdenacer , Support pédagogique de cours : échantillonnage, statistique de gestion, 2013/2014, p.3, www.sg-ehec.jimdo.com

4-Théorie de l’échantillonnage

La théorie des sondages ou théorie de l’échantillonnage est un ensemble d’outils statistiques permettant l’étude d’une population statistique à partir de l’examen d’un échantillon tiré de celle-ci[9]. Elle consiste à déterminer des propriétés sur des échantillons tirés au hasard parmi une population dont on connait les propriétés.

Le tirage d'éléments dans une population peut-être fait de façon exhaustive (c'est-à- dire sans remise) ou de façon non-exhaustive (avec remise). Dans ce dernier cas, les tirages sont indépendants.

Une population finie dans laquelle on procède à un échantillonnage avec remise peut être théoriquement considérée comme infinie. Dans la pratique, il en va de même pour des populations finies mais de grandes tailles.

5-Nombre d’échantillons possibles

Le nombre d’échantillons de n éléments qui peuvent être isolés d’une population de N éléments est calculé selon les deux cas de tirage distingués :

- Tirage exhaustif (sans remise) : nombre d’échantillons possibles est une combinaison : $\mathbf{C_N^n} = \frac{\boldsymbol{N}!}{\boldsymbol{n}!(\boldsymbol{N}-\boldsymbol{n})!}$
- Tirage non exhaustif (avec remise) : nombre d’échantillon possibles est : $\mathbf{N^n}$

[9] -VAILLANT jean, Initiation à la théorie de l’échantillonnage, octobre 2005, P1, http ://monnano.weebly.com/uploads/1/6/6/3/1663287/theoechtage

Cours 2

Distribution d'échantillonnage

Loi d'échantillonnage des moyennes

1- Distribution d'échantillonnage

D'une façon générale, la distribution d'échantillonnage caractérise les fluctuations d'échantillonnage de toute statistique (moyenne, proportion, variance,…) calculée sur tous les échantillons possibles de même taille[10].

Pour chaque distribution d'échantillonnage, on peut calculer une moyenne, un écart type, une variance etc…On peut parler alors de la moyenne de la distribution d'échantillonnage des moyennes, de la moyenne de la distribution d'échantillonnage des proportions etc..

Si on suppose que ces échantillons sont prélevés au hasard et que le tirage de ces échantillons est effectué avec remise. L'ensemble de ces échantillons de taille n est appelé échantillonnage de taille n. Dans ces conditions on peut étudier :

- La loi d'échantillonnage des moyennes ;
- La loi d'échantillonnage des fréquences[11].

[10] -BAILLARGEON gérald , Probabilités, statistique et techniques de régression, ed SMG, 1989, P220.
[11] - Nous nous limitons à ces deux lois uniquement pour le programme de licence 2ème année.

2- Loi d'échantillonnage des moyennes

On dispose d'une population sur laquelle est définie une variable aléatoire X dont on connait l'espérance (ou la moyenne) μ et l'écart type σ.

On s'intéresse aux échantillons de taille n, on calcule leurs moyennes qui seront nettement différentes les unes des autres.

$\bar{X}$ est la variable aléatoire qui, à chaque échantillon de taille n, associe sa moyenne ($\bar{X}$ s'appelle la distribution des moyennes des échantillons)[12] (Figure 2).

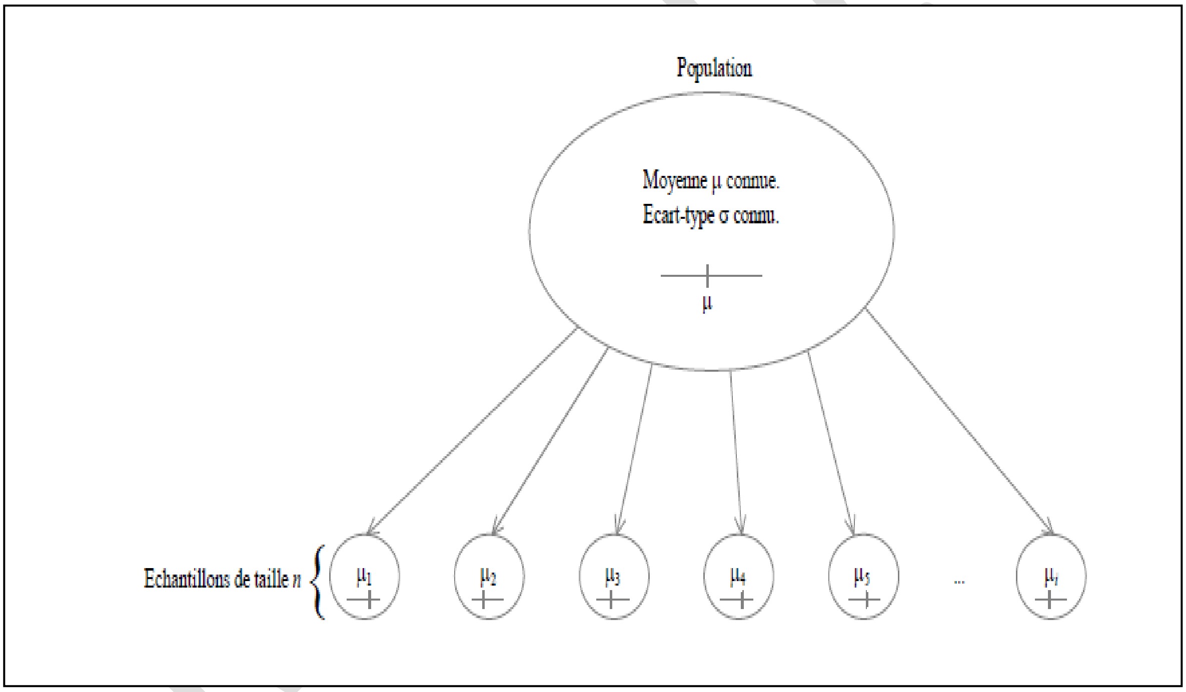

Figure 2 : Distribution d'échantillonnage des moyennes[13]

[12] -COSTANTINI gille, Echantillonnage-estimation, statistiques inférentielles, BTS 2ème année , p1 https://docplayer.fr/13457270-Partie-a-echantillonnage.html

[13] -COSTANTINI gille, Echantillonnage-estimation, statistiques inférentielles, BTS 2ème année , p1 https://docplayer.fr/13457270-Partie-a-echantillonnage.html

Démonstration :

Notons E={x_1 ;x_2 ; x_3 ;…….x_n} un échantillon de n éléments prélevés au hasard dans la population. Pour tout i compris entre 1 et n, notons x_i la variable aléatoire correspondant à la valeur du i-$_{ème}$ élément x_i de l'échantillon. Nous savons, par hypothèse, que :

$$E(X_i) = \mu \text{ et } \sigma(X_i)=\sigma$$

La moyenne $\bar{X}$ des n valeurs de l'échantillon est :

$$\bar{X}= \frac{x1+x2+\cdots+xn}{n}$$

D'après les propriétés de la loi normale, une combinaison linéaire de variables aléatoires qui suivent la loi normale est encore une variable aléatoire qui suit la loi normale. Comme chaque variable aléatoire x_i suit ici la loi normale N(μ, σ), la variable aléatoire moyenne $\bar{X}$ suit donc également une loi normale[14]. Ses paramètres sont :

$$E(\bar{X})= \frac{E(X_1)+E(X_2)+E(X_3)+\cdots+E(X_n)}{n}= \frac{n\mu}{n} = \mu$$

E($\bar{X}$) est la moyenne des moyennes des échantillons tirés.

D'après les propriétés de la variance :

$$V(\bar{X})= \frac{V(X_1)+V(X_2)+V(X_3)+\cdots+V(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\text{D'où : } \sigma(\bar{X})= \frac{\sigma}{\sqrt{n}}$$

[14] - COSTANTINI gille, Echantillonnage-estimation, statistiques inférentielles, BTS 2ème année, p1 https://docplayer.fr/13457270-Partie-a-echantillonnage.html

$\sigma(\bar{X})$ est l’écart type des moyennes des échantillons tirés.

Donc, d’après le *Théorème central limite-*[15] *:*

Théorème

Soit la variable aléatoire X qui suit une loi quelconque sur la population ave : E(X)= μ et σ(X)= σ.

On prélève, au hasard, un échantillon (tirage avec remise ou assimilés) de taille n, avec n ≥ 30, de moyenne $\bar{X}$*. Alors la variable* $\bar{X}$ *suit approximativement une loi normale :*

$$\bar{X} \backsim N\left(\square; \frac{\sigma}{\sqrt{n}}\right)$$

Supposons que tous les échantillons de taille n constitués sans remise à partir d’une population finie de Taille N, on alors[16] :

$$\boxed{E(\bar{X}) = \mu \text{ et } \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

N.B : $\sqrt{\frac{N-n}{N-1}}$ est appelé facteur d’exhaustivité ou facteur de correction[17].

Si la population est infinie ou que l’échantillonnage est avec remise,

$$\boxed{E(\bar{X}) = \mu \text{ et } \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}}$$

[15] - BOULONNE clément : Echantillonnage : leçon 10, 2012, P 3 https://cboumaths.files.wordpress.com/2012/01/lecon10-2012

[16] --BAILLARGEON gérald , Probabilités, statistique et techniques de régression, ed SMG, 1989, P228.

[17] - Cours en ligne de Said Chermak : Echantillonnage, Estimation, Mai 2012 https://www.youtube.com/watch?v=Xnfu7lkzqK8

Exemple1[18] *:*

Une machine fabrique des rondelles métalliques de diamètre moyen μ= 20mm avec un écart type σ=2mm. On assimile cette population a une population infinie (tirage avec remise), et on prélève un échantillon de taille n= 50. Donner $E(\bar{X})$ et $\sigma(\bar{X})$? si le tirage est présumé exhaustif ; quelles seront de nouveau les valeurs de $E(\bar{X})$ et $\sigma(\bar{X})$?

On donne N=400

<u>Correction</u>

$E(\bar{X})$=μ=20mm.

$\sigma(\bar{X}) = \frac{\sigma_p}{\sqrt{n}} = \frac{2}{\sqrt{50}} = 0.28$

Si le tirage est exhaustif (sans remise), $E(\bar{X})$=μ=20mm.

$\sigma(\bar{X}) = \frac{\sigma_p}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

$\sigma(\bar{X}) = \frac{2}{\sqrt{50}}\sqrt{\frac{400-50}{400-1}} = 0.28\sqrt{\frac{350}{399}}$

$\sigma(\bar{X}) = 0.26$

Exemple2[19] *:*

Soit P une population composée des éléments suivants :P $= \{1,2;3\}$. *On y prélève des échantillons de taille n=2. On effectue des tirages successifs avec remise.*

-Quelles sont tous les échantillons possibles ?

-Déterminer la moyenne $E(\bar{X})$ et l'écart type $\sigma(\bar{X})$.

---

[18] -Idem

[19] - Cours en ligne de Said Chermak : Echantillonnage, Estimation, Mai 2012 https://www.youtube.com/watch?v=Xnfu7lkzqK8

Correction

Le nombre d'échantillons possibles :$N^n = 3^2$=9.

Donc, tous les tirages successifs de taille n=2 sont :

$$\begin{Bmatrix} 11\ 12\ 13 \\ 21\ 22\ 23 \\ 31\ 32\ 33 \end{Bmatrix}$$

A partir de notre population, on peut calculer la moyenne μ :

$\mu = \frac{1+2+3}{3} = 2$

Calculons l'écart type à partir de la variance :

$\sigma^2 = \frac{1}{N}\sum(x_i^2 - \bar{x})^2 \Rightarrow \sigma^2 = \frac{1}{3}\sum(1^2 + 2^2 + 3^2) - 2^2 = 0.66$

$\Rightarrow \sigma = \sqrt{0.66} = 0.81$

Notre population est caractérisée par deux paramètres : μ = 2 et σ = 0.81

Calculons les moyennes pour chaque échantillon puis $E(\bar{X})$:

$\begin{Bmatrix} 1 & 1.5 & 2 \\ 1.5 & 2 & 2.5 \\ 2 & 2.5 & 3 \end{Bmatrix}$ On obtient ici la distribution d'échantillonnage des moyennes

Donc, $E(\bar{X}) = \frac{1+1.5+2+1.5+2+2.5+2+2.5+3}{9} = 2$

Alors, $E(\bar{X}) = \mu = 2$

$\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}} \Rightarrow \sigma(\bar{X}) = \frac{0.81}{\sqrt{2}} = 0.57$

Exercice supplémentaire :

Une population comporte 4 individus dont les masses sont respectivement 4, 8 12 et 15Kg. On tire tous les échantillons possibles de taille 2 à partir de cette population. Vérifier les relations entre la moyenne de la population et la moyenne de la distribution d'échantillonnage des moyennes et entre l'écart type de la population et l'écart type de distribution d'échantillonnage de la moyenne.