

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 Master 1
Filière : Démographie

LES TECHNIQUES DE SONDAGE

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

Année universitaire : 2020 - 2021

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 connaît les lois de probabilités¹. Il s'agit d'un phénomène réciproque de l'échantillonnage.

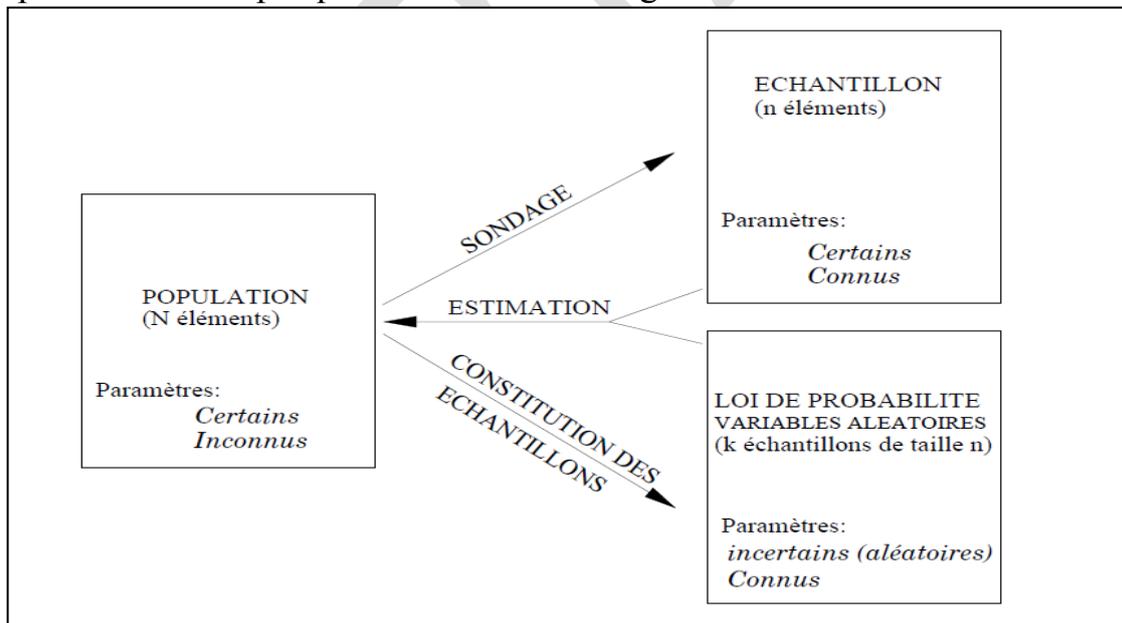


Figure 4 : Principe général de l'estimation²

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 connaître.

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 confiance³.

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

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 des résultats de l'échantillon, ce nombre est appelé une estimation ponctuelle du paramètre⁴.

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 .

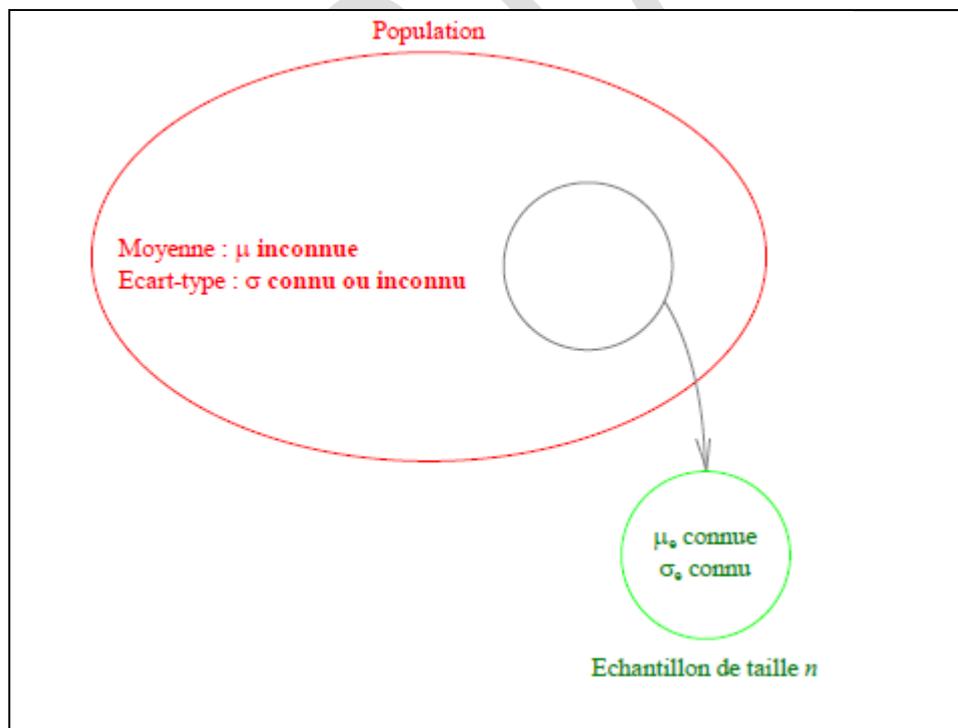


Figure 5 : Estimation ponctuelle d'une moyenne⁵

⁴-BAILLARGEON gerald , Probabilités, statistique et techniques de régression, ed SMG, 1989, P230.

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

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

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

Exemple 1 :

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 $E(\bar{X}) = 4,57\text{cm}$. Estimer la longueur moyenne des vis de la production journalière.

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

⁵ COSTANTINI gille, Echantillonnage-estimation, statistiques inférentielles, BTS 2^{ème} année, p 5 <https://docplayer.fr/13457270-Partie-a-echantillonnage.html>

⁶ - COSTANTINI gille, Echantillonnage-estimation, statistiques inférentielles, BTS 2^{ème} 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 : $\mu_e = 176\text{cm}$ et $\sigma_e = 6\text{cm}$. Estimer les paramètres de la population ?

Solution :

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

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

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

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

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'échantillonnage⁷.

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 recherché⁸.

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'erreur⁹.

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 gerald , Probabilités, statistique et techniques de régression, ed SMG, 1989,, p232.

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

⁹ - EI 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.

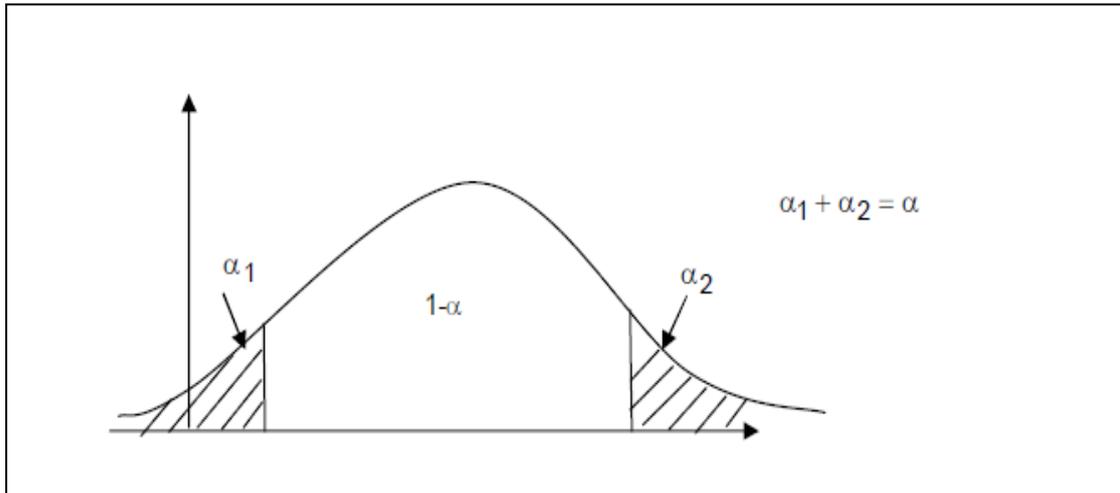


Figure 8 : Principe de l'estimation par intervalle de confiance : proposer un encadrement d'un paramètre inconnu d'une population dont la loi, elle, est connue¹⁰

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 ($1-\alpha$) cas sur 100, l'intervalle recouvre la vraie valeur du paramètre¹¹.

- Niveau de confiance : la probabilité ($1-\alpha$) 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 $\alpha=5\%$, 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 $X_1, X_2, X_3, \dots, X_n$ 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 suffisamment 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 :

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} \sim N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

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

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

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 :

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

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

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

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

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

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

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 utilisés¹³

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 \bar{X} , suit d'après le théorème central limite, approximativement une loi normale.

Exemple 1¹⁴:

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 $\bar{X} = 219$

L'écart type de la population normale : $\sigma = \sqrt{9} = 3 \text{ kg/cm}^2$

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.

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

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

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

$$I_{95\%} = [217,14 ; 220,86]$$

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.

Cours 5

Estimation d'une proportion

Estimation ponctuelle d'une proportion

1- Estimation d'une proportion

1.2- Estimation ponctuelle d'une proportion

On considère un caractère (ou attribut) A sur une population dont la proportion p est inconnue. On suppose que l'on a prélevé un échantillon de taille n (tirage avec remise) sur lequel on a calculé la proportion p_e d'individus ayant le caractère A (Figure 9).

F est la variable aléatoire correspondant à la proportion du caractère A dans un échantillon de taille n pris au hasard. Alors F suit approximativement une loi normale :

$$F \sim N(p; \sigma_p) \text{ où } \sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

Une estimation ponctuelle \hat{p} de la proportion p de l'attribut A dans la population est : $\hat{p} = p_e$

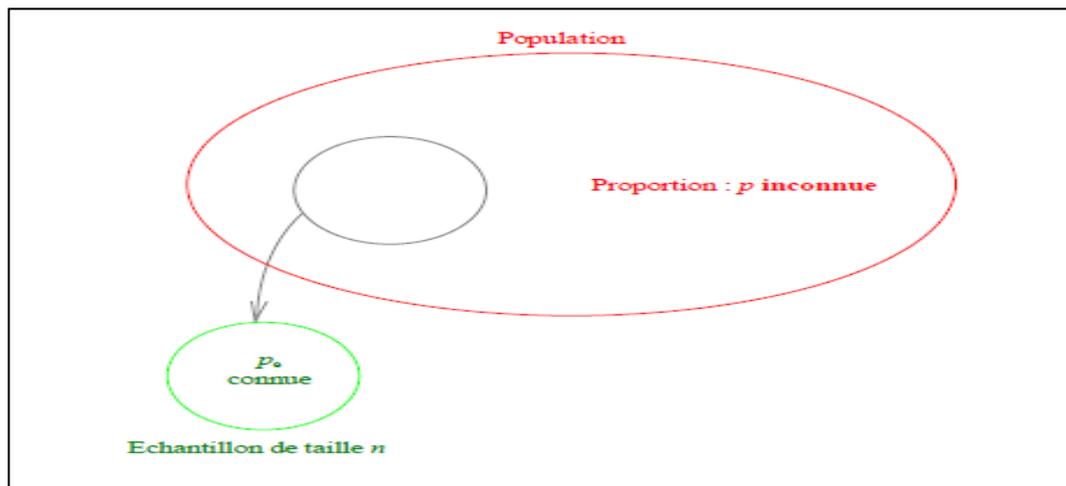


Figure 9 : Estimation ponctuelle de la proportion¹⁵

Une estimation ponctuelle $\widehat{\sigma}_p$ de l'écart type σ_p est selon le cas :

$$\widehat{\sigma}_p = \sqrt{\frac{p_e(1-p_e)}{n}} \sqrt{\frac{n}{n-1}} \quad \text{si } n \leq 30$$

$$\widehat{\sigma}_p = \sqrt{\frac{p_e(1-p_e)}{n}} \quad \text{si } n > 30$$

Exemple 1 :

A quelques jours d'une élection, un candidat fait effectuer un sondage. Sur les 150 personnes interrogées, 45 se disent prêts à voter pour lui aux prochaines élections.

1-Estimer ponctuellement la proportion d'individus prêts à voter pour ce candidat.

2-Estimer ponctuellement l'écart type de la proportion.

¹⁵ - -COSTANTINI gille, Echantillonnage-estimation, statistiques inférentielles, BTS 2^{ème} année , p5
<https://docplayer.fr/13457270-Partie-a-echantillonnage.html>

Solution :

$$p_e = \frac{45}{150} = 0,3$$

Donc : $\hat{p} = p_e = 0,3$

$$\hat{\sigma}_p = \sqrt{\frac{p_e(1-p_e)}{n}} = \sqrt{\frac{0,3 \cdot 0,7}{150}} = 0,037$$

MORTAD.N

Estimation d'une proportion

Estimation d'une proportion par intervalle de confiance

1- Estimation d'une proportion par intervalle de confiance

La proportion d'éléments (individus) possédant un caractère qualitatif dans la population, ayant un niveau de confiance $100(1-\alpha)$ % de contenir la vraie valeur de p est en autant que $n\hat{p} \geq 5$ et $n(1-\hat{p}) \geq 5$, est donnée par l'intervalle de confiance suivant :

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Où \hat{p} représente la valeur observée de p (estimateur ponctuel de p) sur l'échantillon de taille n et $z_{\alpha/2}$ est la valeur de la variable normale centrée réduite.

Remarque :

Cet intervalle de confiance n'est valable que sous les conditions d'application mentionnées précédemment soit : $n\hat{p} \geq 5$ et $n(1-\hat{p}) \geq 5$ ¹⁶.

*Exemple*¹⁷ :

Dans une municipalité, on a effectué un sondage pour connaître l'opinion des contribuables sur un nouveau règlement d'emprunt. D'une liste informatisée de 6000 payeurs de taxes, on a prélevé, par tirage au sort, 150 noms. Sur ces 150, 45 étaient en faveur du nouveau règlement d'emprunt. Déterminer un intervalle de confiance pour p , la proportion vraie de

¹⁶ - BAILLARGEON gérald , Probabilités, statistique et techniques de régression, ed SMG, 1989, P255.

¹⁷ - idem, p255.

contribuables de cette municipalité qui sont en faveur du nouveau règlement d'emprunt, avec les niveaux de confiance de 90%, de 95% et de 99%.

Solution :

L'estimation ponctuelle de p est $\hat{p} = \frac{45}{150} = 0,30$

L'écart type de la proportion d'échantillon est :

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0,30)(0,70)}{150}} = 0,0374$$

Les intervalles de confiances obtenus pour les différents niveaux de confiance sont :

$$Ic_{90\%} = [0,30 - 1,645 \cdot 0,0374; 0,30 + 1,645 \cdot 0,0374]$$

$$Ic_{90\%} = [0,2385; 0,3515]$$

$$Ic_{95\%} = [0,30 - 1,96 \cdot 0,0374; 0,30 + 1,96 \cdot 0,0374]$$

$$Ic_{95\%} = [0,2267; 0,3733]$$

$$Ic_{99\%} = [0,30 - 2,58 \cdot 0,0374; 0,30 + 2,58 \cdot 0,0374]$$

$$Ic_{99\%} = [0,2035; 0,3965]$$

Le calcul de ces intervalles de confiance est valable en autant que : $n\hat{p} \geq 5$ et $n(1-\hat{p}) \geq 5$. Ces conditions d'application sont largement vérifiées : puisque $n\hat{p} = (150)(0,30) = 45$ et $n(1-\hat{p}) = (150)(0,70) = 105$.

Remarques¹⁸

-Dans le calcul de tout intervalle de confiance, plus le niveau de confiance est élevé, plus l'amplitude de l'intervalle est grande.

¹⁸ - BAILLARGEON gérald , Probabilités, statistique et techniques de régression, ed SMG, 1989, p 256.

-Si l'échantillonnage s'effectue sans remise, à partir d'une population finie de taille N, on doit apporter une correction à $\sigma(\hat{p})$, dans ce cas :

$$\sigma(\hat{p}) = \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}$$

Pour considérer que ce facteur de correction est négligeable : $\frac{n}{N} \leq 0,10$.

-Dans le cas d'un tirage sans remise, l'écart type de la statistique concernée est donc plus faible que celui obtenu lors d'un tirage avec remise (meilleure précision).

2- Marge d'erreur associée à l'estimation de p et taille d'échantillon requise

Comme la valeur observée \hat{p} est prédisposée à des fluctuations d'échantillonnage, il y a toujours un écart entre la valeur observée \hat{p} et la valeur réelle p. Cet écart constitue la marge d'erreur dans l'estimation de p.

On l'appelle aussi : la précision du sondage.

Lorsque \hat{p} est utilisée comme estimation de p, alors, pour un niveau de confiance $100(1-\alpha)\%$, la marge d'erreur E sera au plus égale à :

$$z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Par la suite, on peut déterminer la taille minimale de l'échantillon requise pour une marge d'erreur fixée.

Ainsi, pour une marge d'erreur désirée E et un niveau de confiance $100(1-\alpha)\%$; la taille d'échantillon minimale requise est :

$$n = \frac{Z_{\alpha/2}^2 \hat{p}(1-\hat{p})}{E^2}$$

Exemple

Un groupe d'étudiants effectuent un sondage auprès de la population étudiante pour estimer le pourcentage des fumeurs. La population étudiante compte 8000.

Déterminer la taille d'échantillon requise pour assurer une marge d'erreur d'excédant pas 5%, avec un niveau de confiance de 95%, si l'enquête indique que 32% des étudiants fument.

Solution :

Marge d'erreur = 5%

Niveau de confiance : 95% , $Z_{\alpha/2} = 1,96$

$\hat{p} = 0,32$

Alors, $n = \frac{(1,96)^2 (0,32)(0,68)}{(0,05)^2} = 334,37 \approx 335$