ECHANTILLONAGE ET ESTIMATION

0/. Introduction
I/. Estimation
II/.Espérance mathématique d'une moyenne
III/.Espérance mathématique d'une proportion
IV/. Estimation ponctuelle d'un paramètre
V/. Distribution d'échantillonnage
VI/. Distribution de probabilité de la variable aléatoire
VII/. Estimation par intervalle de confiance de µ
VIII/. Intervalle de confiance pour une proportion
IX/. Taille de l'échantillon garantissant la précision de l'intervalle de confiance

0/. Introduction

Statistique inférentielle

Inférentielle du nom inférence synonyme de ***déduction*** c'est-à-dire que à partir des données d'un échantillon on pourra ***déduire*** les paramètres d'une population

Statistique inférentielle

Elle développe des procédés permettant de généraliser à toute une population des résultats observés sur un échantillon, tout en étant capable de mesurer les chances que ces généralisations s'avèrent exactes.

I/. Estimation

Voici la définition de quelque termes que nous utiliserons dans ce chapitre.

Définition 1: On appelle ***PARAMETRE*** toute mesure calculée à partir de l'ensemble des données de la population.

Définition 2: On appelle ***STATISTIQUE*** toute mesure calculée à partir des données échantillonnales.

Définitions 3: On appelle ***ESTIMATION*** d'un paramètre le procédé par lequel on cherche à déterminer la valeur d'un paramètre d'une population.

Définition 4: On appelle ***ESTIMATEUR*** la statistique utilisée pour effectuer l'estimation; c'est une variable aléatoire.

Définition 5: On appelle ***VALEUR ESTIMEE*** la valeur que prend l'estimateur une fois l'échantillon tiré; c'est une valeur de la variable aléatoire que constitue l'estimateur.

Statistique descriptive :

Résumer les mesures sur un échantillon (moyenne, variance,...) Représenter les mesures (histogramme, distribution)

Statistique inférentielle :

Généraliser les propriétés d'un échantillon à une population en prenant en compte les fluctuations d'échantillonnage

il faut modéliser les observations (par des variables aléatoires) :

on fait appel à la théorie des probabilités

Tests d'hypothèses :

Contrôler la validité d'un modèle

Comparer un échantillon à une référence

Statistique décisionnelle :

Savoir prendre une décision alors que les résultats sont exprimés en termes de probabilité (i.e. de pourcentage de chances, de risques)

II/.Espérance mathématique d'une moyenne

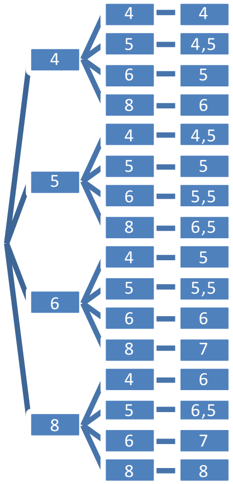
Théorème 1:

La moyenne d'échantillon est un estimateur sans biais de la moyenne de la population à laquelle il appartient, c'est-à-dire

$$E(\bar{X}) = \mu.$$

Exemple 1:

Soit la population {4,5,6,8}. Dans cette population, considérons la variable aléatoire $\bar{X}2$ représentant la moyenne d'un échantillon de taille 2 tiré avec remise.



D’où la distribution de probabilité suivante:

x	4	4,5	5	5,5	6	6,5	7	8
f(x)	1/16	2/16	3/16	2/16	3/16	2/16	2/16	1/16

On a donc

E(X) = 4 (1/16) + 4,5 (2/16) + ……….. +8 (1/16) = 5,75

De plus la moyenne de la population est

μ = (4+5+6+8)/4 = 5,75

Ce qu’il fallait vérifier.

III/.Espérance mathématique d'une proportion

Théorème 2:

La proportion d'individus présentant un caractère donné dans un échantillon est un estimateur sans biais de la proportion des individus dans la population à laquelle appartient l'échantillon, c'est-à-dire

$$E(\overline{P}) = \pi$$

Exemple 2:

Soit la population {4,5,6,8}. Dans cette population, considérons la variable aléatoire $\overline{P}_2$ représentant la proportion de nombres impair dans un échantillon de taille 2 tiré avec remise. L'ensemble des résultats possibles est:

D’où la distribution de probabilité suivante:

P	0	1/2	1
f(P)	9/16	6/16	1/16

On a donc E(P) = 0 (9/16) +(1/2) (6/16) + 1 (1/16) = ¼
De plus, la proportion des nombres impairs dans la population est

$$\pi = \frac{1}{4}$$

Ce qu’il fallait vérifier.

IV/. Estimation ponctuelle d’un paramètre

Une estimation ponctuelle d’un paramètre consiste à évaluer la valeur de ce paramètre à partir d’une valeur unique déterminée sur un échantillon. Il est à noté que la statistique utilisée doit satisfaire un certain nombre de critère à fin de bien représenter le paramètre à estimer. On a vu celui de l’estimateur sans biais, le reste de ces critères sera exposer ultérieurement.

V/. Distribution d'échantillonnage

Une estimation ponctuelle une opération facile à réaliser. cependant les résultats trouvés qui sont censés données une image fidèle de la population changent de valeurs à chaque fois qu'on change d'échantillon et par conséquent les chances qu'ont ces estimateurs d'être exactes sont très faibles. Il convient donc de pouvoir estimer un paramètre tout en étant capable d'évaluer les chances qu'a cette estimation de se réaliser. Pour cela, nous effectuons ce qu'on appelle une estimation par intervalle de confiance d'un paramètre de la population. Le problème consiste donc à trouver les bornes d'un intervalle tout en fixant à l'avance la probabilité pour que le paramètre tombe dans cet intervalle.
Avant d'atteindre cet objectif, il est nécessaire d'étudier la distribution d'échantillonnage de la statistique (variable aléatoire) que l'on utilise comme estimateur. Les résultats sont donnés par les deux théorème suivants:

Théorème 3:

La variable aléatoire $\overline{X}$, représentant la moyenne d'échantillon de taille n, possède les caractéristiques

Si l'une des deux conditions suivantes est vérifiée:

$$E(\overline{X}) = \mu_{\overline{X}} = \mu \qquad et \qquad V(\overline{X}) = \sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

a) Tirage avec remise dans une population finie de taille N;

b) Tirage avec ou sans remise dans une population très grande ou infinie.

N.B: une population est grande dés que n ≤0,05 N; c'est-à-dire que la taille de l'échantillon est inférieur à 5% de la taille de la population.

Théorème 4:

La variable aléatoire $\overline{X}$, représentant la moyenne d'échantillon de taille n tiré sans remise d'une population finie de taille N, possède les caractéristiques

$$E(\overline{X}) = \mu_{\overline{X}} = \mu \qquad et \qquad V(\overline{X}) = \sigma^2_{\overline{X}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right).$$

Remarque: On appelle aussi erreur type l'écart type de la variable aléatoire X et on le note $\sigma_{\overline{X}}$

Exemple 3:

A partir de la population {4,5,6,8} dans laquelle on a effectué le tirage de tous les échantillons de taille 2 pris avec remise, on a la distribution de probabilité suivante pour la variable aléatoire X

x	4	4,5	5	5,5	6	6,5	7	8
f(x)	1/16	2/16	3/16	2/16	3/16	2/16	2/16	1/16

On sait que

$$V(\overline{X}) = E\left(\overline{X}^2\right) - \left(E(\overline{X})\right)^2$$

et on trouve $V(\overline{X}) = 1{,}09375 \quad et \quad \frac{\sigma^2}{n} = \frac{2{,}1875}{2} = 1{,}09375$ Classeur1.xlsx

Ce qu'il fallait vérifier.

Exemple 4:

A partir de la population {4,5,6,8} procédons au tirage de tous les échantillons de taille 2 pris sans remise, on a la distribution de probabilité suivante pour la variable aléatoire X

X	4,5	5	5,5	6	6,5	7
f(x)	2/12	2/12	2/12	2/12	2/12	2/12

On trouve Classeur1.xlsx

$$V(X)=0,7291667 \quad et \quad \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)=1,09375\times\frac{2}{3}=0,7291667$$

Ce qu'il fallait vérifier.

VI/. Distribution de probabilité de la variable aléatoire

Théorème central limite:

1) Quand la population est distribuée normalement, la distribution de $\bar{X}_n$ sera normale quelque soit n.
2) Quand la population n'est pas distribuée normalement et que n ≥ 30 alors la distribution de $\bar{X}_n$ sera normale.

$$\bar{X}_n \longrightarrow N(\mu_{\bar{X}} ; \sigma^2_{\bar{X}})$$

Exemple :

Le poids de 60 étudiants est distribué normalement avec μ = 64 kg et $\sigma^2 = 20\,kg^2$.

Alors la variable aléatoire d'un échantillon de taille 20 tiré:

a) Avec remise est caractérisée par:

$$\mu_{\overline{X}} = 64 \text{ kg et } \sigma^2_{\overline{X}} = \frac{20}{20} = 1 \; ; \quad \bar{X}_{20} \longrightarrow N(64 ; 1)$$

b) Sans remise est caractérisée par:

$$\mu_{\overline{X}} = 64 \text{ kg et } \sigma^2_{\overline{X}} = \frac{20}{20}\left(\frac{60-20}{60-1}\right) = 0{,}68 \; ; \quad \bar{X}_{20} \longrightarrow N(64 ; 0{,}68)$$

Exercice :

1°) A partir de l'exemple précédant; trouvez la probabilité qu'un échantillon de taille 20 pris avec remise ait une moyenne supérieure à 66 kg.

Réponse:

$$P(\overline{X}_{20} \geq 66) = P\left(\frac{\overline{X}_{20} - 64}{1} \geq \frac{66-64}{1}\right) = P\left(Z \geq \frac{66-64}{1}\right) =$$

$$P(Z \geq 2) = 1 - p(Z \prec 2) = 1 - 0{,}9772 = 0{,}0228$$

table de la loi normal.xlsx

2°) calculez

$$P(63 \le \overline{X}_{20} \le 65)$$

Réponse:

$$P(63 \le \overline{X}_{20} \le 65) =$$

$$P\left(\frac{63-64}{1} \le \frac{\overline{X}_{20}-64}{1} \le \frac{65-64}{1}\right) =$$

$$P(-1 \le Z \le 1) =$$

$$2P(Z \le 1) - 1 = 2(0,8413) - 1 = 0,6826$$

table de la loi normal.xlsx

3°) $Trouvez\ L_1\ et\ L_2\ tel\ que\ P(L_1 \leq \bar{X}_{20} \leq L_2) = 0{,}95$

Réponse:

$$0{,}95 = P(L_1 \leq \bar{X}_{20} \leq L_2) = P\left(\frac{L_1 - \mu}{\sigma} \leq Z \leq \frac{L_2 - \mu}{\sigma}\right) =$$

$$P(-a \leq Z \leq a) = 2P(Z \leq a) - 1$$

On déduit facilement que: $P(Z \leq a) = 0{,}975$

Et d'après la table de la loi normale on trouve que a = 1,96. table de la loi normal.xlsx

$$or\ -a = \frac{L_1 - \mu}{\sigma}\ et\ a = \frac{L_2 - \mu}{\sigma}$$

On trouve facilement que: $L_1 = 62{,}04\ et\ L_2 = 65{,}96.$

VII/. Estimation par intervalle de confiance de μ

Définition 1: On appelle intervalle de confiance un intervalle de la forme, $\left[L_1 ; L_2\right]$ Ayant une certaine probabilité de contenir la valeur d'un paramètre.

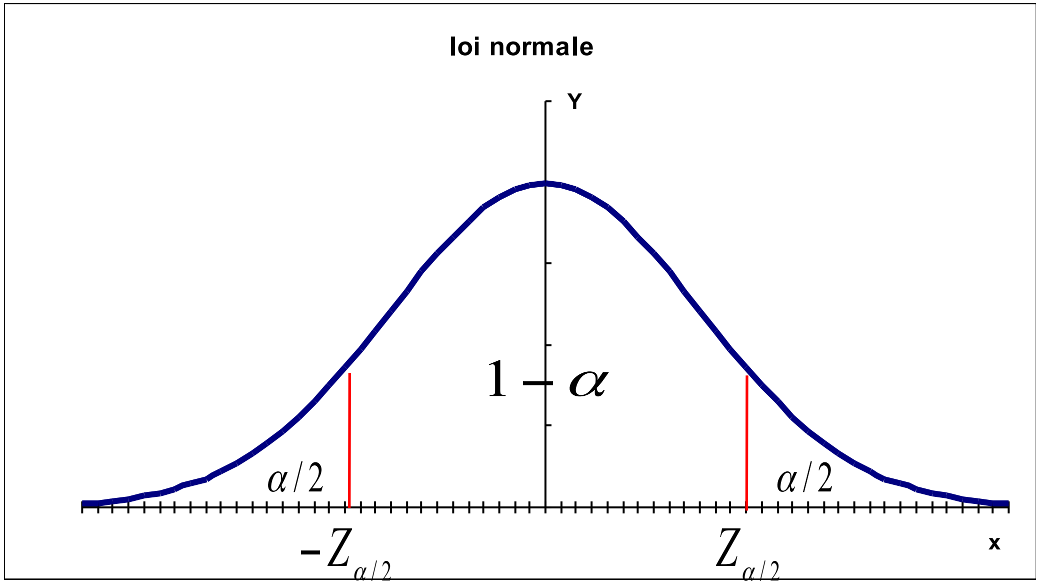
Définition 2: On appelle niveau de confiance, noté $1-\alpha$; la probabilité qu'a l'intervalle de confiance de contenir la valeur du paramètre μ.

Définition 3: On appelle risque d'erreur ou niveau de signification noté α la probabilité qu'a l'intervalle de confiance de ne pas contenir la valeur du paramètre μ.

Remarque 1: Dans le contexte où la moyenne μ d'une population est connue, le problème qui retient ici notre attention est de trouver deux valeurs $L_1 \ et \ L_2$ telles que $\mu \in [L_1 ; L_2]$ avec un niveau de confiance $1-\alpha$; c'est-à-dire un intervalle tel que:

$$p(L_1 \leq \mu \leq L_2) = 1-\alpha$$

Représentons par $Z_{\alpha/2}$ la valeur standardisée de L_2 et par $-Z_{\alpha/2}$ celle de L_1 . On a



$$et \ p(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) = 1-\alpha = p(-Z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \leq Z_{\alpha/2}) =$$

$$p(\bar{X} - Z_{\alpha/2}\,\sigma_{\bar{X}} \leq \mu \leq \bar{X} + Z_{\alpha/2}\,\sigma_{\bar{X}})$$

Conclusion:

On a $\mu \in [L_1 ; L_2]$ avec un niveau de confiance $1-\alpha$ fixé à l'avance, où α représente le risque d'erreur. Pour un échantillon donné, la variable aléatoire $\overline{X}$ prend la valeur $\mu_{\bar{x}}$ et on a l'intervalle de confiance

$$\left[\mu_{\bar{X}} - Z_{\alpha/2}\sigma_{\bar{X}} \,;\, \mu_{\bar{X}} + Z_{\alpha/2}\sigma_{\bar{X}}\right]$$

Remarque 2:

Si σ de la population est inconnu, on utilise une valeur estimée ponctuelle de celui-ci à partir de l'échantillon. Pour montrer que l'on a procédé ainsi, on utilise dans ce cas le symbole $S_{\bar{x}}$ pour représenter l'erreur type $\sigma_{\bar{x}}$ (Application voir exercice 2).

Dans la suite on convient d'utiliser:

$$Z_{\alpha/2} = 2{,}58 \quad si \quad \alpha = 1\%$$

$$Z_{\alpha/2} = 1{,}96 \quad si \quad \alpha = 5\%$$

$$Z_{\alpha/2} = 1{,}65 \quad si \quad \alpha = 10\%$$

Exemple 1:

Une machine est réglé pour verser un certain mélange dans une boîte avec σ = 3,2 g. parmi l'ensemble de la production, on prélève au hasard et avec remise 30 boîtes pour chacune desquelles on a noté le poids. Sachant que le poids moyen obtenu à partir de l'échantillon est de 165 g, construire un intervalle de confiance à 95% pour le poids moyen des boîtes remplies par cette machine.

Réponse:

Puisque n≥30 la variable aléatoire suit une loi normale dont les paramètre sont:

$$\begin{cases} \mu_{\bar{X}} = 165g \; car\; la\; moyenne\; est\; un\; estimateur\; sans\; biais \\ \sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{3,2}{\sqrt{30}} = 0,584 \; car\; le\; prélèvement\; est\; fait\; \underline{avec remise} \\ 1-\alpha = 0,95 \; d'où\; \alpha = 5\% \; soit\; Z_{\alpha/2} = 1,96 \end{cases}$$

$$d'où\; L_1 = 165 - 1,96 \times 0,584 = 163,85$$

$$et \qquad L_2 = 165 + 1,96 \times 0,584 = 166,15$$

$$on\; a\; ainsi\; P(\mu \in [163,85;166,15]) = 0,95$$

Il y a 95% des chances pour que le poids moyen de toutes les boîtes remplies par cette machine soit compris entre 163,85 et 166,15g.

Exemple 2: c'est une application de la remarque 2

La moyenne et l'écart type du résultat cumulatif d'un échantillon de taille 36 étudiants d'une université sont 2,6 et 0,3 respectivement. Trouvez un intervalle de confiance à 99% pour la moyenne des résultats cumulatifs de tous les étudiants de cette université.

Réponse:

a) Puisque (n = 36) ≥30 la variable aléatoire suit une loi normale dont les paramètres sont:

$b)\ \mu_{\bar{\bar{X}}} = 2{,}6$ car la moyenne est un estimateur sans biais

c) Puisque l'écart type de la population σ est inconnu, on utilise donc une valeur estimée ponctuelle; ici S = 0,3 d'où $S_{\bar{X}} = \frac{S}{\sqrt{n}} = \frac{0{,}3}{\sqrt{36}} = 0{,}05$

d) 1-α = 0,99 donc α= 1% soit $Z_{\alpha/2} = 2{,}58$ d'où:

$$L_1 = 2{,}6 - 2{,}58 \times 0{,}05 = 2{,}47$$

$$L_2 = 2{,}6 + 2{,}58 \times 0{,}05 = 2{,}73$$

Conclusion: l'intervalle [2,47;2,73] possède 99% des chances de contenir la moyenne μ du résultat cumulatif des étudiants de cette université.

Exemple 3:

Dans le but d'estimer le résultat moyen des 150 étudiants qui suivent un cours de statistique, on choisi au hasard et sans remise 30 individus parmi ceux-ci. A partir de cet échantillon, on obtient une moyenne de 78,6 avec un écart type de 15,6. si le risque d'erreur associé à cette estimation est de 10% trouvez l'intervalle de confiance.

Réponse:

a) Puisque (n = 30) ≥30 la variable aléatoire suit une loi normale dont les paramètres sont:

$b)\ \mu_{\bar{\bar{X}}} = 78{,}6$ car la moyenne est un estimateur sans biais

c) Puisque l'écart type de la population σ est inconnu, on utilise donc une valeur estimée ponctuelle; ici S = 15,6. d'autre part l'échantillon est choisi sans remise d'où

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{15{,}6}{\sqrt{30}}\sqrt{\frac{150-30}{150-1}} = 2{,}556$$

d) α=10% donc $Z_{\alpha/2} = 1{,}65$ d'où:

$$L_1 = 78{,}6 - 1{,}65 \times 2{,}556 = 74{,}38$$

$$L_2 = 78{,}6 + 1{,}65 \times 2{,}556 = 82{,}82$$

Conclusion: On a µ€[74,38;82,82] avec un niveau de confiance de 90% c'est-à-dire que l'intervalle [74,38;82,82] possède 90% des chances de contenir la valeur du résultat moyen de l'ensemble des individus suivant le cours de statistique.

Théorème :

Lorsque n < 30 et que la population est distribuée normalement, l'intervalle de confiance pour µ si

a) σ est connu est $\left[\mu_{\bar{X}} - Z_{\alpha/2}\,\sigma_{\bar{X}} ; \mu_{\bar{X}} + Z_{\alpha/2}\,\sigma_{\bar{X}}\right]$

b) σ est inconnu est $\left[\mu_{\bar{X}} - t_{\alpha/2}\,S_{\bar{X}} ; \mu_{\bar{X}} + t_{\alpha/2}\,S_{\bar{X}}\right]$

Où $t_{\alpha/2}$ est la valeur de la distribution de Student telle que $P(T \leq t_{\alpha/2}) = 1 - \dfrac{\alpha}{2}$ avec un nombre de degré de liberté noté $\nu = n - 1$

Théorème :

Lorsque n < 30 et que la population n'est pas distribuée normalement, l'intervalle de confiance pour μ si

a) σ est connu est $\left[\mu_{\bar{X}} - k\,\sigma_{\bar{X}} ; \mu_{\bar{X}} + k\,\sigma_{\bar{X}}\right]$

b) σ est inconnu est $\left[\mu_{\bar{X}} - k\,S_{\bar{X}} ; \mu_{\bar{X}} + k\,S_{\bar{X}}\right]$

Où k provient de l'inégalité de Bienaymé Tchebychev; c'est une valeur telle que $k^2 = \frac{1}{\alpha}$

résumé

Conditions	Intervalle de confiance pour μ		
Sur l'échantillon	n ≥ 30	n < 30	
Sur la population	Distribution normale ou non	Distribution normale	Distribution inconnue ou non normale
σ connu	$\left[\mu_{\bar{X}} - Z_{\alpha/2}\,\sigma_{\bar{X}} ; \mu_{\bar{X}} + Z_{\alpha/2}\,\sigma_{\bar{X}}\right]$	$\left[\mu_{\bar{X}} - Z_{\alpha/2}\,\sigma_{\bar{X}} ; \mu_{\bar{X}} + Z_{\alpha/2}\,\sigma_{\bar{X}}\right]$	$\left[\mu_{\bar{X}} - k\,\sigma_{\bar{X}} ; \mu_{\bar{X}} + k\,\sigma_{\bar{X}}\right]$
σ inconnu	$\left[\mu_{\bar{X}} - Z_{\alpha/2}\,S_{\bar{X}} ; \mu_{\bar{X}} + Z_{\alpha/2}\,S_{\bar{X}}\right]$	$\left[\mu_{\bar{X}} - t_{\alpha/2}\,S_{\bar{X}} ; \mu_{\bar{X}} + t_{\alpha/2}\,S_{\bar{X}}\right]$	$\left[\mu_{\bar{X}} - k\,S_{\bar{X}} ; \mu_{\bar{X}} + k\,S_{\bar{X}}\right]$

Diapositive 35

Diapositive 36

Population finie		Population grande (n<0,05N) ou infinie
Avec remise	Sans remise	Avec ou sans remise
$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

Exemple 1:

A partir d'un échantillon aléatoire de taille 50 pris sans remise dans l'ensemble des joueurs d'un jeu d'enfant de la région, on a obtenu un âge moyen de 10,5 années et un écart type de 0,8 année. Construire un intervalle de confiance à 90% pour l'âge moyen μ de la population des joueurs.

Réponse:

a) Puisque (n = 50) ≥30 la variable aléatoire suit une loi normale dont les paramètres sont:

$b)\ \mu_{\bar{\bar{X}}} = 10{,}5$ car la moyenne est un estimateur sans biais

c) Puisque l'écart type de la population σ est inconnu, on utilise donc une valeur estimée ponctuelle; ici S = 0,8.

d) d'autre part la population est grande, car Région ⟹ N>1000 donc n < 0,05 N

ainsi $S_{\bar{X}} = \frac{S}{\sqrt{n}} = \frac{0{,}8}{\sqrt{50}} = 0{,}113\ldots.$

e) α=10% donc $Z_{\alpha/2} = 1{,}65$ d'où:

$$L_1 = 10{,}5 - 1{,}65 \times 0{,}113\ldots = 10{,}31$$

$$L_2 = 10{,}5 + 1{,}65 \times 0{,}113\ldots = 10{,}69$$

<u>Conclusion:</u> On a μ€[10,31;10,69] avec un niveau de confiance de 90% c'est-à-dire que l'intervalle [10,31;10,69] possède 90% des chances de contenir l'âge moyen de la population.

Remarque:

Soit une population de taille N sur laquelle est observée une caractéristique dont on connaît la moyenne µ et la variance σ^2. On a:

$$\sigma^2 = \frac{\sum_{i=1}^{i=N} n_i (x_i - \mu)^2}{N}$$

Si on prélève n individus dans cette population leur moyenne sera:

$$\bar{X} = \frac{\sum_{j=1}^{j=n} n_j x_j}{n}$$

appelée moyenne d'échantillon, et leur variance:

$$\boxed{S^2 = \frac{\sum_{j=1}^{j=n} n_j (x_j - \bar{X})^2}{n-1} = \frac{n\sum_{j=1}^{j=n} n_j x_j^2 - \left(\sum_{j=1}^{j=n} n_j x_j\right)^2}{n(n-1)}}$$

appelée variance d'échantillon.

Exemple 2:

Dans une association regroupant 50 membres, on choisit au hasard et sans remise 15 individus pour lesquels on obtient les poids en kg:

75	68	65	72	72	69	63	65	70	60	68	73	78	67	71

Si X est la variable aléatoire représentant le poids des individus à partir de cet échantillon, déterminez l'intervalle de confiance à 95% pour μ

a) Si l'on suppose que la distribution du poids de l'ensemble des individus de cette association obéit à une loi normale.

b) Si l'on ne peut pas supposer que la distribution du poids de l'ensemble des individus de l'association obéit à une loi normale.

Réponse:

Poids en Kg (Xi)	75	68	65	72	72	69	63	65	70	60	68	73	78	67	71	**1036**
Xi*Xi	5625	4624	4225	5184	5184	4761	3969	4225	4900	3600	4624	5329	6084	4489	5041	**71864**

$$\sum_{i=1}^{i=15} x_i = 1036$$

$$\sum_{i=1}^{i=15} x_i^2 = 71864$$

D'où

$$\bar{x} = \frac{\sum_{i=1}^{i=15} n_i x_i}{n} = \frac{1036}{15} = 69{,}07\,Kg$$

$$S^2 = \frac{n\sum_{j=1}^{j=n} n_j x_j^2 - \left(\sum_{j=1}^{j=n} n_j x_j\right)^2}{n(n-1)} = \frac{15*71864-(1036)^2}{15*14} = \frac{4664}{210} = 22{,}21\,Kg^2$$

$$Soit\ S = \sqrt{S^2} = \sqrt{22{,}21} = 4{,}71\,Kg$$

a) La population suit une loi normale.

l'intervalle de confiance s'obtient à partir du théorème Diapositive 30

$$\left[\mu_{\bar{X}} - t_{\alpha/2} S_{\bar{X}} ; \mu_{\bar{X}} + t_{\alpha/2} S_{\bar{X}}\right] \qquad avec \qquad \nu = n-1 = 14 \text{ degrés de liberté}$$

$\mu_{\bar{X}} = 69{,}07 \quad t_{\alpha/2} = t_{2{,}5\%} = 2{,}145$ d'après la table de Student Table student.pdf

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \left(\frac{4{,}71}{\sqrt{15}}\right)\sqrt{\frac{50-15}{50-1}} = 1{,}028...$$

$d'où$

$$L_1 = 69{,}07 - (2{,}145)(1{,}028...) = 66{,}86$$

et

$$L_2 = 69{,}07 + (2{,}145)(1{,}028...) = 71{,}28.$$ On a ainsi

$$\mu \in [66{,}86 ; 71{,}28]$$

avec un niveau de confiance de 95%, où μ est le poids moyen des individus de cette association.

b) La population ne suit pas une loi normale.

l'intervalle de confiance s'obtient à partir du théorème Diapositive 30

$$\left[\mu_{\bar{X}} - kS_{\bar{\bar{X}}} ; \mu_{\bar{X}} + kS_{\bar{\bar{X}}}\right]$$ où

$\mu_{\bar{X}} = 69{,}07$ et $S_{\bar{X}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \left(\frac{4{,}71}{\sqrt{15}}\right)\sqrt{\frac{50-15}{50-1}} = 1{,}028\ldots$

$$et$$

$$k^2 = \frac{1}{\alpha} = \frac{1}{0{,}05} = 20$$

$$donc$$

$$k = \sqrt{20} = 4{,}47 \; d'où$$

$$L_1 = 69{,}07 - (4{,}47)(1{,}028..) = 61{,}80 \; et$$

$$L_2 = 69{,}07 + (4{,}47)(1{,}028..) = 76{,}34$$ On a ainsi

$$\mu \in [61{,}80 ; 76{,}34]$$

avec un niveau de confiance de 95%.

VIII/. Intervalle de confiance pour une proportion

Il arrive très souvent que l'on soit intéressé à estimer, pour une population, la proportion des individus ayant une modalité donnée. On peut construire alors un intervalle de confiance pour cette proportion, π .

Désignons par $\bar{p}$ la valeur estimée ponctuelle de π, c'est-à-dire le pourcentage observé dans l'échantillon et $S_{\bar{P}}$ la valeur estimée ponctuelle de l'erreur type de la variable aléatoire $\bar{P}$.

Théorème :

Si n ≥ 30, n $\bar{p}$ ≥ 5 et n(1- $\bar{p}$) ≥ 5, alors l'intervalle de confiance de π est

$$\left[\bar{p} - z_{\alpha/2} S_{\bar{P}} \,;\, \bar{p} + z_{\alpha/2} S_{\bar{P}}\right]$$ où

$$S_{\bar{P}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$ Lorsque n < 0,05 N peu importe le type de tirage et la taille de la population

$$S_{\bar{P}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \sqrt{\frac{N-n}{N-1}}$$ Lorsque n ≥ 0,05 N lors d'un tirage sans remise d'une population finie.

Exemple

Sachant que 90% des 150 plants de tomates produits par un agriculteur sont encore vivants lorsque vient le temps de les vendre aux clients; déterminez l'intervalle de confiance à 95% sur la proportion du nombre de plants de tomates vivants de l'ensemble de la culture de cet agriculteur.

Réponse:

L'intervalle de confiance est: $\left[\bar{p} - z_{\alpha/2} S_{\bar{P}} ; \bar{p} + z_{\alpha/2} S_{\bar{P}}\right]$ où

$$\bar{p} = 0{,}9; z_{\alpha/2} = z_{2{,}5\%} = 1{,}96\, et$$

$$S_{\bar{P}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{(0{,}9)(0{,}1)}{150}} = 0{,}024...$$

Ici, le facteur de correction n'est pas appliqué du moment qu'on ne connaît pas la taille de la population. Cependant, on peut facilement supposer qu'elle est suffisamment grande pour que l'ont n'est pas à tenir compte de ce facteur. On a:

$$L_1 = 0{,}9 - (1{,}96)(0{,}024..) = 0{,}85$$

$$L_2 = 0{,}9 + (1{,}96)(0{,}024..) = 0{,}95$$

$d'où\, \pi \in [0{,}85 ; 0{,}95]$ c'est-à-dire que le pourcentage du nombre de plants vivants de cet agriculteur se situe entre 85% et 95% avec un niveau de confiance de 95%.

IX/. Taille de l'échantillon garantissant la précision de l'intervalle de confiance

En échantillonnage nous travaillons sur un échantillon avec l'espoir que cet échantillon soit représentatif. Un des critères qui fait qu'un échantillon soit représentatif c'est la taille de cet échantillon; autrement dit à partir de quel taille un échantillon peut-il être considéré comme représentatif?

1°) Pour l'estimation d'une moyenne

Dans l'intervalle de confiance

$$\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

le facteur

$$z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

est appelé erreur possible dans l'estimation de µ. Si on se fixe à l'avance l'erreur E tolérable dans l'estimation de µ, Il suffit de résoudre pour n l'inéquation

$$z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \leq E$$

pour obtenir la taille de l'échantillon nécessaire à l'obtention de la précision qu'on s'est fixée. En élevant au carré chacun des membres de l'inégalité, on obtient

$$z^2_{\alpha/2}\left(\frac{\sigma^2}{n}\right) \leq E^2$$

et ainsi

$$\frac{z^2_{\alpha/2}\ \sigma^2}{E^2} \leq n$$

Théorème:

Si $\bar{x}$ est la valeur estimée de μ, il y a une probabilité de (1-α) que l'erreur d'estimation n'excède pas la valeur E, fixée à l'avance, à condition que la taille de l'échantillon soit

$$n \geq \left(\frac{z_{\alpha/2}\ \sigma}{E} \right)^2$$

Remarque:

Dans le cas où la distribution n'est pas normal alors la valeur minimale de n sera 30 afin de pouvoir appliquer le théorème central limite

Exemple

Sur une population étudiée selon un caractère X, on a relevé une variance égale à 64. on désire construire un intervalle de confiance à 95% pour µ. Celui-ci doit avoir une erreur maximale de 5. quelle devrait être la taille de l'échantillon si

a) La population suit une distribution normale,
b) La distribution de la population est inconnue.

Réponse:

a) La population suit une distribution normale

La taille de l'échantillon vérifie l'inéquation

$$n \geq \left(\frac{z_{\alpha/2} \, \sigma}{E} \right)^2$$

Avec $z_{\alpha/2} = z_{2,5\%} = 1,96$

$$\sigma = \sqrt{64} = 8$$

$$E = 5$$

$$soit$$

$$n \geq \left(\frac{(1,96)(8)}{5} \right)^2 = 9,8$$

Comme la taille n doit être entière, la taille de l'échantillon sera 10.

b) La distribution de la population est inconnue.

Dans ce cas la taille doit nécessairement être au minimum égale à 30.

2°) Pour l'estimation d'une proportion

Théorème:

La taille de l'échantillon est:

$$n \geq \left(\frac{z^2_{\alpha/2} \; \bar{p}\left(1-\bar{p}\right)}{E^2} \right)$$

Où $z_{\alpha/2}$ est la valeur de la variable Z telle que $P(Z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}$, α le risque d'erreur, $\bar{p}$ la valeur estimée ponctuelle de la proportion, et E l'erreur maximale de l'intervalle de confiance.

Remarque:

L'application de la formule nécessite la connaissance de la valeur estimée ponctuelle $\bar{p}$. Dans le cas où cette valeur serais inconnue, on se sert de la valeur maximale que peut prendre l'expression $\bar{p}\left(1-\bar{p}\right)$ à savoir $\frac{1}{2}\left(1-\frac{1}{2}\right)=\frac{1}{4}$. Ainsi la formule devient:

$$n \geq \left(\frac{z_{\alpha/2}}{2E}\right)^2$$

Exemple:

Dans une municipalité, on évalue à 2/3 la proportion des personnes favorables à un projet de réaménagement du centre-ville; déterminez la taille de l'échantillon qu'on doit prélever si l'on désir construire un intervalle de confiance à 90% et ayant une erreur maximale de 0,1 pour $\prod$.

Réponse:

$$on\ a$$

$$n \geq \left(\frac{z^2_{\alpha/2}\ \bar{p}\left(1-\bar{p}\right)}{E^2} \right) \quad où$$

$$z_{\alpha/2} = z_{5\%} = 1{,}65 \quad \bar{p} = \frac{2}{3} \quad et \quad E = 0{,}1$$

$$donc$$

$$n \geq \frac{(1{,}65)^2 \left(\frac{2}{3}\right)\left(\frac{1}{3}\right)}{(0{,}1)^2} = 60{,}5$$

D'où la taille d'échantillon nécessaire est 61.