STATISTIQUES INFERENTIELLES

Comme nous l'avons déjà vu l'objet des statistiques est d'étudier un ensemble de caractères sur une population, qui est constituée d'un ensemble d'individus.

Pour cela, on peut recueillir l'information sur chaque individu de la population, ce qui est le cas pour un recensement par exemple. Cette exhaustivité n'est pas toujours possible ni souhaitable, on extrait alors l'information sur un certain nombre d'individus de la population ; c'est ce qu'on appelle un échantillon.

1. Echantillonnage d'une population

Voici quelques raisons d'échantillonner :

1.1. Les façons d'échantillonner

Pour prélever un échantillon, il existe plusieurs méthodes :

- **Echantillon au hasard**[1]

- **Méthode des quotas**

- **Méthode en cascade**

Lorsque le tirage de l'échantillon se fait sans remise, on dit que l'échantillon est _ _ _ _ _ _ _ _ _, au contraire, lorsque le tirage de l'échantillon se fait avec remise, on dit que l'échantillon est _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _.

[1] C'est la technique la plus simple à mettre en œuvre. Elle est assimilée au tirage de boules, au hasard, dans une urne. On suppose dans la suite du cours les échantillons comme pris au hasard.

Bien sûr les résultats obtenus sur l'échantillon ne sont pas identiques à ceux qui auraient pu l'être sur la population. Le but du cours est de fournir des outils permettant de déduire les paramètres de la population, connaissant ceux de l'échantillon.

Il existe cependant deux sources de distorsions importantes :

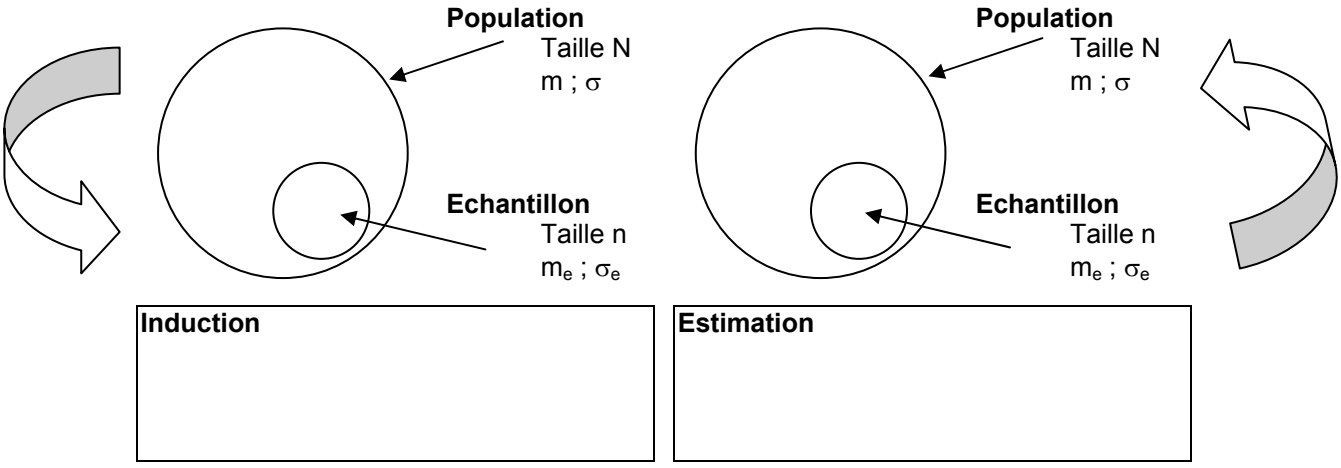
- **Le biais sur les mesures**

- **Le biais de recrutement**

Exemple : Il semble peu opportun de réaliser une enquête dans la rue un jour de semaine, entre 14 h et 17 h, sur l'activité professionnelle des personnes.

1.2. Induction et estimation

Etant donnée une population de taille N et un échantillon de taille n, deux configurations sont possibles :



Remarque :

- *Lorsque la taille de la population (N) est importante et que le fait de prélever un échantillon exhaustif (de taille n petite par rapport à N) de la population ne modifie pas notablement le caractère étudié, on peut alors considérer l'échantillon comme non exhaustif.*
- *Une population finie (de taille N) avec un échantillonnage non exhaustif peut être considérée comme infinie.*

2. Induction

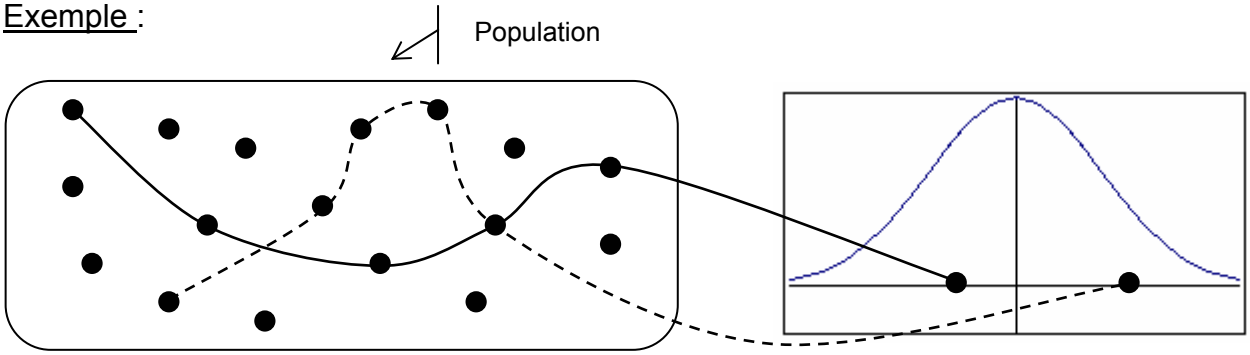
Dans ce paragraphe, on considère que les paramètres de la population sont connus (moyenne, écart-type...).

On considère tous les échantillons de taille n que l'on peut extraire d'une population de taille N, de moyenne m et d'écart-type σ.

Pour chaque échantillon, on peut calculer la moyenne d'une certaine propriété qui varie d'un échantillon à l'autre.

On obtient alors, une variable aléatoire $\overline{X}$, égale à la moyenne des éléments sur tout échantillon de taille n, dont la loi de probabilité est appelée distribution d'échantillonnage des moyennes.

Exemple :



Les divers échantillons (ici de taille 5) possibles donnent différentes valeurs du paramètre. Toutes ces valeurs forment la distribution d'échantillonnage.

Remarque : On peut aussi considérer une variable aléatoire F, égale à la proportion de réussite sur tout échantillon de taille n, dont la loi de probabilité est appelée distribution d'échantillonnage des fréquences.

2.1. Etude d'un exemple

Soient les notes de mathématiques d'un étudiant : 4 ; 5 ; 8 ; 10 ; 12 ; 13.

1. Calculer la moyenne et l'écart-type de la population des notes.
2. Former tous les échantillons exhaustifs possibles de taille 2.
3. Calculer l'espérance et l'écart-type de la distribution d'échantillonnage des moyennes.
4. Recommencer dans le cas d'un échantillon non exhaustif.

Solution :

1. On a m = _ _ _ _ _ _ _ et σ = _ _ _ _ _ _ _ .
2.

4 ; 5	m_e = 4,5	5 ; 8	m_e = 6,5	8 ; 12	m_e = 10
4 ; 8	m_e = 6	5 ; 10	m_e = 7,5	8 ; 13	m_e = 10,5
4 ; 10	m_e = 7	5 ; 12	m_e = 8,5	10 ; 12	m_e = 11
4 ; 12	m_e = 8	5 ; 13	m_e = 9	10 ; 13	m_e = 11,5
4 ; 13	m_e = 8,5	8 ; 10	m_e = 9	12 ; 13	m_e = 12,5

3. La variable aléatoire, $\overline{X}$, égale à la moyenne des éléments sur tout échantillon exhaustif de taille 2 suit la loi de probabilité :

4,5	6	6,5	7	7,5	8	8,5	9	10	10,5	11	11,5	12,5
2/30	2/30	2/30	2/30	2/30	2/30	4/30	4/30	2/30	2/30	2/30	2/30	2/30

Donc $E(\overline{X})$ = _ _ _ _ _ _ _ et $\sigma(\overline{X})$ = _ _ _ _ _ _ _ .

4. De manière identique, la variable aléatoire $\overline{X}$, égale à la moyenne des éléments sur tout échantillon non exhaustif de taille 2 suit la loi de probabilité :

4	4,5	5	6	6,5	7	7,5	8	8,5	9	10	10,5	11	11,5	12	12,5	13
1/36	2/36	1/36	2/36	2/36	2/36	2/36	3/36	4/36	4/36	3/36	2/36	2/36	2/36	1/36	2/36	1/36

Donc $E(\overline{X})$ = _ _ _ _ _ _ _ et $\sigma(\overline{X})$ = _ _ _ _ _ _ _ .

2.2. Cas d'une moyenne

On considère tous les échantillons non exhaustifs de taille n que l'on peut extraire d'une population de taille N, de moyenne m et d'écart-type σ. On appelle X la variable aléatoire décrivant le caractère étudié sur la population.

On considère la variable aléatoire $\overline{X}$, égale à la moyenne des éléments sur tout échantillon de taille n ; on s'intéresse donc à la distribution d'échantillonnage des moyennes.

Un n-échantillon est donné par le n-uplet $(X_1 ; X_2 ; ... ; X_n)$ où les X_i sont des variables aléatoires indépendantes de même loi que X. On a alors $\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$.

$$E(\overline{X}) = E\left(\frac{\sum X_i}{n}\right) = \frac{1}{n} \cdot E\left(\sum X_i\right) = \frac{1}{n} \cdot \sum E(X_i) = \frac{n \cdot E(X)}{n} = m \cdot$$

et

$$V(\overline{X}) = V\left(\frac{\sum X_i}{n}\right) = \frac{1}{n^2} \cdot V\left(\sum X_i\right) = \frac{1}{n^2} \cdot \sum V(X_i) = \frac{n \cdot V(X)}{n^2} = \frac{V(X)}{n} \cdot$$

- Pour un échantillon non exhaustif de taille n, on a :

$E(\overline{X})$ = _ _ _ _ _ _ _ et $\sigma(\overline{X})$ = _ _ _ _ _ _ _ .

- Pour un échantillon exhaustif[2] de taille n, on a :

$E(\overline{X})$ = _ _ _ _ _ _ _ et $\sigma(\overline{X})$ = _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ .

[2] Le coefficient $\sqrt{\frac{N-n}{N-1}}$ s'appelle le coefficient d'exhaustivité.

Le théorème de la limite centrale nous permet d'affirmer que pour de grandes valeurs de n ($n \geq 30$) la distribution d'échantillonnage de la moyenne est approximativement la loi normale _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ [respectivement _ pour un échantillonnage exhaustif].

Remarque : Si sur la population la loi suivie est normale alors la distribution d'échantillonnage est exactement une loi normale quelle que soit la valeur de n.

Exemple : Reprendre l'exercice de 2.1. et retrouver de façon théorique (à l'aide du cours ci-dessus) les résultats précédents.

2.3. Cas d'une proportion

Soit une population de taille N avec une probabilité de réussite p et une probabilité d'échec (1 – p).

Considérons tous les échantillons non exhaustifs de taille n extraits de cette population et pour chaque échantillon, on détermine la proportion de succès.

On considère Y la VA égale au nombre de succès sur un échantillon de taille n donné. Y suit la loi binomiale _ _ _ _ _ _ _ _ _ _ _ .

Si $n \geq 30$ on peut considérer que Y suit approximativement la loi normale

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

On a alors $F = \frac{Y}{n}$ qui représente la variable aléatoire égale à la proportion de réussites sur l'échantillon. On obtient alors une variable aléatoire F telle que :

Lorsque n est assez grand ($n \geq 30$) la loi de F est approximativement _ _ _ _ _ _ _ _ _ [respectivement _ dans le cas d'un échantillonnage exhaustif] ; donc :

- Pour un échantillon non exhaustif :

E(F) = _ _ _ _ _ _ _ _ _ _ _ et $\sigma(F)$ = _ _ _ _ _ _ _ _ _ _ _

- Pour un échantillon exhaustif :

E(F) = _ _ _ _ _ _ _ _ _ _ _ et $\sigma(F)$ = _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

Exemple : Une enquête écrite est réalisée sur un échantillon de 250 clients ciblés.
Une étude statistique a montré que le taux de défection moyen, dans ce genre d'enquête, est de 15 %.

Les questions peuvent alors être :

- Quelle est la gamme plausible des défections sur l'échantillon ?
- L'échantillon est-il bien ciblé ?

Solution :

On considère Y la VA égale au nombre de défections sur l'échantillon. Y suit la loi binomiale $\mathcal{B}(250 ; 0{,}15)$. Comme $n \geq 30$ on peut considérer que Y suit la loi normale $\mathcal{N}(250\times0{,}15 ; \sqrt{250\times0{,}15\times(1-0{,}15)})$

On a alors $F = \frac{Y}{n}$ qui représente la variable aléatoire égale à la proportion de défections sur l'échantillon.

On a F qui suit la loi $\mathcal{N}(0{,}15 ; 0{,}02)$.

Si on veut avoir une plage de valeurs où doit se trouver la proportion f_e de l'échantillon, il faut choisir un risque, par exemple 4,5 %, et alors :

$P(0{,}15 - a \leq F \leq 0{,}15 + a) = 0{,}955$ ssi a 0,04.

Donc, au risque de 4,5 %, f_e doit se trouver dans l'intervalle [11 % ; 19 %]. Cet intervalle est appelé intervalle de confiance à 95,5 % (ou au seuil de risque 4,5 %).

Après l'enquête, il est alors possible de valider, ou non, l'échantillon.

3. Estimation

L'objet de l'estimation est d'obtenir les paramètres d'une population à partir d'observations établies sur un échantillon de cette population.

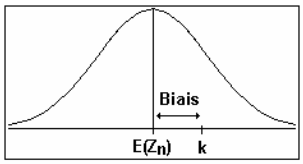
3.1. Estimateur – Estimation ponctuelle

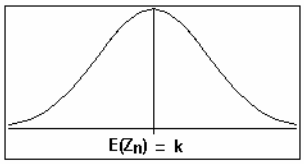
On appelle estimateur d'un paramètre k, une variable aléatoire dont le but est d'estimer au mieux la valeur du paramètre k.

On dit que l'estimateur Z_n est un estimateur sans biais du paramètre k si $E(Z_n) = k$.

Remarque : Si $E(Z_n) \to k$ quand $n \to \infty$, l'estimateur est asymptotiquement sans biais.

Exemple :





Si de plus, $V(Z_n) \to 0$ quand $n \to \infty$, l'estimateur Z_n converge (en probabilité) vers k. L'estimateur Z_n est alors un estimateur absolument correct[3] du paramètre k.

a. Estimation ponctuelle d'une moyenne

On note m la moyenne de la population mère (paramètre inconnu) et X la variable aléatoire décrivant le caractère étudié.

On prélève au hasard un échantillon non exhaustif de taille n.

Un estimateur naturel de m est la variable aléatoire $\overline{X}$ telle que :

$\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ où les X_i sont des variables aléatoires indépendantes de même loi que X telles que (X_1 ; X_2 ; ... ; X_n) forme un n-échantillon.

L'estimateur $\overline{X}$ est sans biais :

$$E(\overline{X}) = E\left(\frac{\sum X_i}{n}\right) = \frac{1}{n} \cdot E\left(\sum X_i\right) = \frac{1}{n} \cdot \sum E(X_i) = \frac{n \cdot E(X)}{n} = m \cdot$$

La réalisation de $\overline{X}$ sur un échantillon donné est m_e la moyenne sur l'échantillon.

Une estimation ponctuelle de m est alors _ _ _ _ _ _ _ _ _ _ _ .

Remarque :

$V(\overline{X}) = V\left(\dfrac{\sum X_i}{n}\right) = \dfrac{1}{n^2} \cdot V\left(\sum X_i\right) = \dfrac{1}{n^2} \cdot \sum V(X_i) = \dfrac{n \cdot V(X)}{n^2} = \dfrac{V(X)}{n} \to 0$ *donc l'estimateur est absolument correct.*

Exemple : On s'intéresse à la durée d'attente des personnes à un centre de renseignements téléphoniques avant que la communication ne soit amorcée. On a prélevé, au hasard, la durée d'attente (en secondes) de 100 contacts :

Durée (s)	[7,5 ; 11,5[	[11,5 ; 15,5[	[15,5 ; 19,5[	[19,5 ; 23,5[	[23,5 ; 27,5[
Effectif	12	25	36	18	9

La moyenne sur l'échantillon est m_e = _ _ _ _ _ _ _ _ _ _ _ _ _ donc on peut dire que l'estimation ponctuelle de la durée d'attente moyenne sur la population des appels est de _ _ _ _ _ _ _ _ _ _ _ s.

b. Estimation ponctuelle d'une proportion

On sait que la population mère contient une proportion p (inconnue) d'individus ayant une propriété donnée.

[3] L'estimateur est dit seulement correct quand il est asymptotiquement sans biais.

On prélève, au hasard, un échantillon non exhaustif de taille n et on note Y la variable aléatoire égale au nombre d'individus ayant la propriété dans l'échantillon.

Y suit la loi binomiale $\mathcal{B}(n ; p)$ qui peut être approchée pour $n \geq 30$ par $\mathcal{N}(np ; \sqrt{np(1-p)})$.

La variable aléatoire $F = \frac{Y}{n}$ suit la loi $\mathcal{N}(p ; \sqrt{\frac{p(1-p)}{n}})$ donc F est un estimateur sans biais de p (en effet E(F) = p).

La réalisation de F sur un échantillon donné est f_e la proportion sur l'échantillon.

Une estimation ponctuelle de p est _ _ _ _ _ _ _ _ _ _ _ .

Remarque :

$V(F) = \sqrt{\frac{p(1-p)}{n}} \rightarrow 0$ donc l'estimateur est absolument correct.

Exemple :

Dans un centre de renseignements téléphoniques, une étude statistique a été réalisée pour déterminer le pourcentage de communications n'aboutissant pas pour des raisons techniques.

Sur un échantillon de 2 000 communications, 60 n'ont pas abouti. Estimer le pourcentage de communications qui n'ont pas abouti.

On a f_e = _ _ _ _ _ _ _ _ _ _ _ .

Donc l'estimation ponctuelle sur la population est de _ _ _ _ _ _ _ _ _ _ _ %.

c. Estimation ponctuelle d'une variance – d'un écart-type

La population mère admet une moyenne m et un écart-type σ inconnus.

Un estimateur logique de la variance σ^2, est : $S_n^2 = \frac{1}{n} \cdot \sum (X_i - \overline{X})^2$.

Cherchons à calculer $E(S_n^2)$:

$$E(S_n^2) = E\left[\frac{1}{n} \cdot \sum X_i^2 - \frac{2}{n} \cdot \sum X_i \overline{X} + \frac{1}{n} \cdot \sum \overline{X}^2\right] = E\left[\frac{1}{n} \cdot \sum X_i^2 - 2 \cdot \overline{X} \cdot (\frac{1}{n} \sum X_i) + \overline{X}^2\right]$$

$$E(S_n^2) = E\left[\frac{1}{n} \cdot \sum X_i^2 - 2 \cdot \overline{X}^2 + \overline{X}^2\right] = E\left[\frac{1}{n} \cdot \sum X_i^2 - \overline{X}^2\right] = \frac{1}{n} \cdot \sum E[X_i^2] - E\left[\overline{X}^2\right]$$

Or tous les X_i ont même loi que X donc pour un échantillonnage non exhaustif :

$E[X_i^2] = V(X_i) + E(X_i)^2 = \sigma^2 + m^2$

$E[\bar{X}^2] = V(\bar{X}) + E(\bar{X})^2 = \frac{\sigma^2}{n} + m^2$.

Donc $E(S_n^2) = \sigma^2 + m^2 - (\frac{\sigma^2}{n} + m^2) = \frac{n-1}{n}\sigma^2$.

Comme $E(S_n^2) = \frac{n-1}{n}\sigma^2$ il y a un biais, donc un estimateur sans biais de σ^2 est $\frac{n}{n-1}S_n^2$ on le note S'^2_n.

Une estimation ponctuelle de σ est s = _ _ _ _ _ _ _ _ _ _ _ (σ_e = écart-type de l'échantillon).

Remarque : Pour un échantillonnage exhaustif il faut apporter la correction suivante :

$E[X_i^2] = V(X_i) + E(X_i)^2 = \sigma^2 + m^2$

$E[\bar{X}^2] = V(\bar{X}) + E(\bar{X})^2 = \frac{\sigma^2}{n} \times \frac{N-n}{N-1} + m^2$.

Donc $E(S_n^2) = \sigma^2 + m^2 - (\frac{\sigma^2}{n} \times \frac{N-n}{N-1} + m^2) = \frac{n-1}{n} \times \frac{N}{N-1}\sigma^2$.

Un estimateur sans biais de σ^2 pour un échantillonnage exhaustif est $\frac{n}{n-1} \times \frac{N-1}{N} S_n^2$.

Définition : s est appelée déviation standard et est souvent notée σ_{n-1} sur les calculatrices.

Exemple : Reprenons l'exemple du 3.1.a (délai avant qu'une communication téléphonique soit amorcée), σ_e = 4,48 donc une estimation ponctuelle sur la population de l'écart-type est s = _ _ _ _ _ _ _ _ _ _ _ .

Remarque : Une estimation ponctuelle de l'écart-type de la loi d'échantillonnage d'une proportion, c'est à dire $\sqrt{\frac{p(1-p)}{n}}$ est $\sqrt{\frac{f_e(1-f_e)}{n-1}}$ dans le cas non exhaustif et $\sqrt{\frac{f_e(1-f_e)}{n-1}} \times \sqrt{\frac{N-n}{N}}$ dans le cas exhaustif.

3.2. Estimation par intervalle de confiance

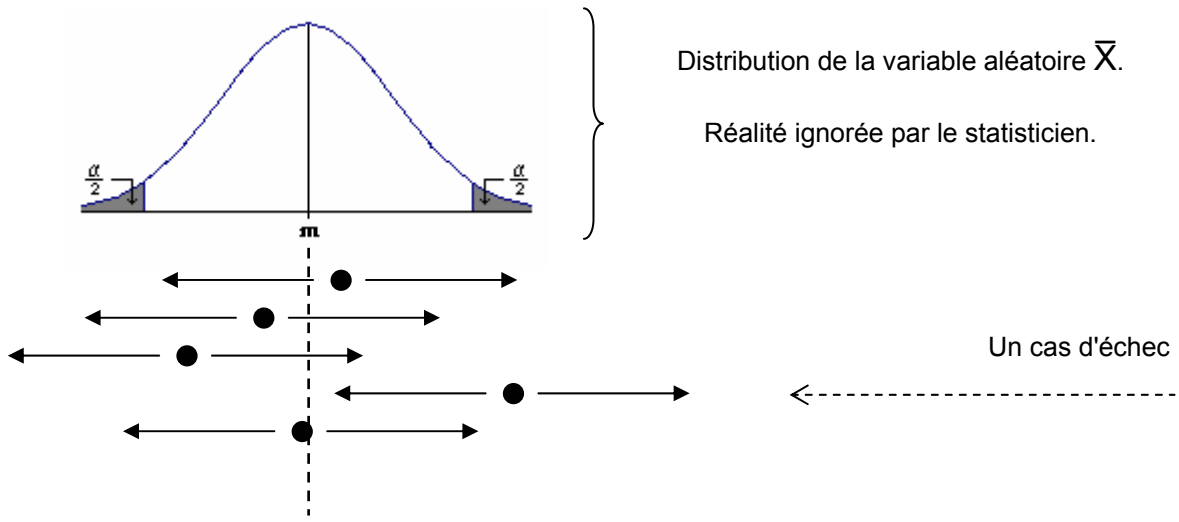
L'estimation ponctuelle dépend, bien sûr, de l'échantillon choisi, ce qui sera toujours le cas mais ne donne aucune information sur la pertinence du résultat.

L'estimation par intervalle de confiance va nous permettre de donner une gamme de valeurs susceptibles d'être prises par le paramètre. On lui affecte de plus un coefficient de crédibilité appelé coefficient de confiance.

Exemple : Calculer un intervalle de confiance à 95 % de la moyenne (on dit aussi au seuil de risque de 5 %).

Ceci veut dire que l'intervalle cherché doit vérifier la propriété suivante :

Exemple :



Le paramètre m appartient à 95 % des intervalles.

Principe général :

Soit Z_n un estimateur du paramètre k :

- Intervalle de confiance bilatéral (symétrique en probabilité) au seuil de risque α :
  On cherche les nombre l_α et l'_α tels que : $P(Z_n - l_\alpha \leq k \leq Z_n + l'_\alpha) = 1 - \alpha$

- Intervalle de confiance unilatéral au seuil de risque α :
  On cherche le nombre l_α tel que : $P(k \geq Z_n + l_\alpha) = 1 - \alpha$ ou $P(k \leq Z_n - l_\alpha) = 1 - \alpha$

Pour un échantillon donné, Z_n admet une réalisation (prend une valeur), on obtient alors un intervalle de confiance de k, au seuil de risque α, associé à l'échantillon choisi.

a. Intervalle de confiance d'une moyenne

On note X la variable aléatoire mesurant le caractère étudié sur la population mère et on considère un échantillonnage non exhaustif.

Cas 1 : X suit la loi $\mathcal{N}(m ; \sigma)$ et σ est connu

On sait qu'un estimateur de m est $\overline{X}$ qui suit la loi $\mathcal{N}(m ; \frac{\sigma}{\sqrt{n}})$.

Il faut trouver l_α tel que :

$P(\overline{X} - l_\alpha \le m \le \overline{X} + l_\alpha) = 1 - \alpha$ ssi $P(-l_\alpha \le \overline{X} - m \le l_\alpha) = 1 - \alpha$

ssi $P\left(\frac{-l_\alpha}{\sigma/\sqrt{n}} \le \frac{\overline{X}-m}{\sigma/\sqrt{n}} \le \frac{l_\alpha}{\sigma/\sqrt{n}}\right) = 1-\alpha$ ssi $P\left(T \le \frac{l_\alpha}{\sigma/\sqrt{n}}\right) = \frac{1+(1-\alpha)}{2}$.

En posant $T = \frac{\overline{X}-m}{\sigma/\sqrt{n}}$, la variable aléatoire T suit la loi normale centrée réduite.

Par lecture sur la table de la loi normale centrée réduite on obtient une valeur t_α :

$l_\alpha = \frac{\sigma}{\sqrt{n}} \cdot t_\alpha$.

En remplaçant $\overline{X}$ par sa valeur sur l'échantillon de taille n :

$P(m_e - \frac{\sigma}{\sqrt{n}} \cdot t_\alpha \le m \le m_e + \frac{\sigma}{\sqrt{n}} \cdot t_\alpha) = 1 - \alpha$

On a donc l'intervalle de confiance : $\left[m_e - \frac{\sigma}{\sqrt{n}} t_\alpha ; m_e + \frac{\sigma}{\sqrt{n}} t_\alpha \right]$.

Exemple : Pour l'exemple du 3.1.a, donnons un intervalle de confiance bilatéral à 95 % de la moyenne des temps d'attente sur la population des appels.

On suppose que le temps d'attente sur la population mère suit la loi normale $\mathcal{N}(m ; \sigma)$ où σ est connu et vaut 4.

On sait alors que la variable aléatoire $\overline{X}$ suit la loi normale $\mathcal{N}(m ; \frac{4}{\sqrt{100}} = 0{,}4)$.

En posant $l_\alpha = a$ on cherche a tel que :

$P(-\frac{a}{0{,}4} \le \frac{\overline{X}-m}{0{,}4} \le \frac{a}{0{,}4}) = 0{,}95$ donc $P(T \le \frac{a}{0{,}4}) = 0{,}975$.

Par lecture sur la table de la loi normale centrée réduite : on trouve $\frac{a}{0{,}4} = 1{,}96$ donc :

$P(16{,}98 - 0{,}78 \le m \le 16{,}98 + 0{,}78) = 0{,}95$ donc l'intervalle de confiance à 95 % de l'attente moyenne sur la population est l'intervalle [16,2 ; 17,76].

Cas 2 : X suit $\mathcal{N}(m ; \sigma)$ et σ est inconnu

Dans ce cas, on doit utiliser une estimation de σ. L'estimation naturelle est la déviation standard s, mais l'utilisation de s introduit une source supplémentaire de non-fiabilité. Afin de palier cet inconvénient on doit "élargir" l'intervalle de confiance bilatéral de la moyenne m.

On obtient alors : $\left[m_e - \frac{s}{\sqrt{n}} t_\alpha^* ; m_e + \frac{s}{\sqrt{n}} t_\alpha^* \right]$ où la valeur t_α^* est lue dans la table de la loi de Student à n – 1 degrés de liberté (ddl).

Remarque : En fait la variable aléatoire $\frac{\overline{X} - m}{S'_n / \sqrt{n}}$ *suit la loi de Student à n – 1 degrés de liberté. Cette loi a été inventée pour construire des intervalles de confiance de la moyenne sans pour autant connaître* σ*.*
Dans la pratique, si $n \geq 30$, σ *est estimé par s et on fait comme pour le cas n°1.*

Cas 3 : X suit une loi quelconque

Il faut : $n \geq 30$ et alors :

- Si σ est connu $\overline{X}$ suit approximativement la loi _ _ _ _ _ _ _ _ _ _ _ .
- Si σ est inconnu $\overline{X}$ suit approximativement la loi _ _ _ _ _ _ _ _ _ _ _ .

b. Intervalle de confiance d'une proportion

Ici la proportion p d'individus ayant une propriété donnée est inconnue, l'échantillonnage toujours considéré comme non exhaustif.

Cas 1 : $n \geq 30$

Comme $n \geq 30$, on sait que F suit (approximativement) la loi $\mathcal{N}(p ; \sqrt{\frac{p(1-p)}{n}})$.

Pour un échantillon donné de taille n, l'intervalle de confiance au seuil de risque α est donné par $\left[f_e - t_\alpha \sqrt{\frac{f_e(1-f_e)}{n-1}} ; f_e + t_\alpha \sqrt{\frac{f_e(1-f_e)}{n-1}} \right]$.

Exemple : Reprendre l'exemple du 3.1.b

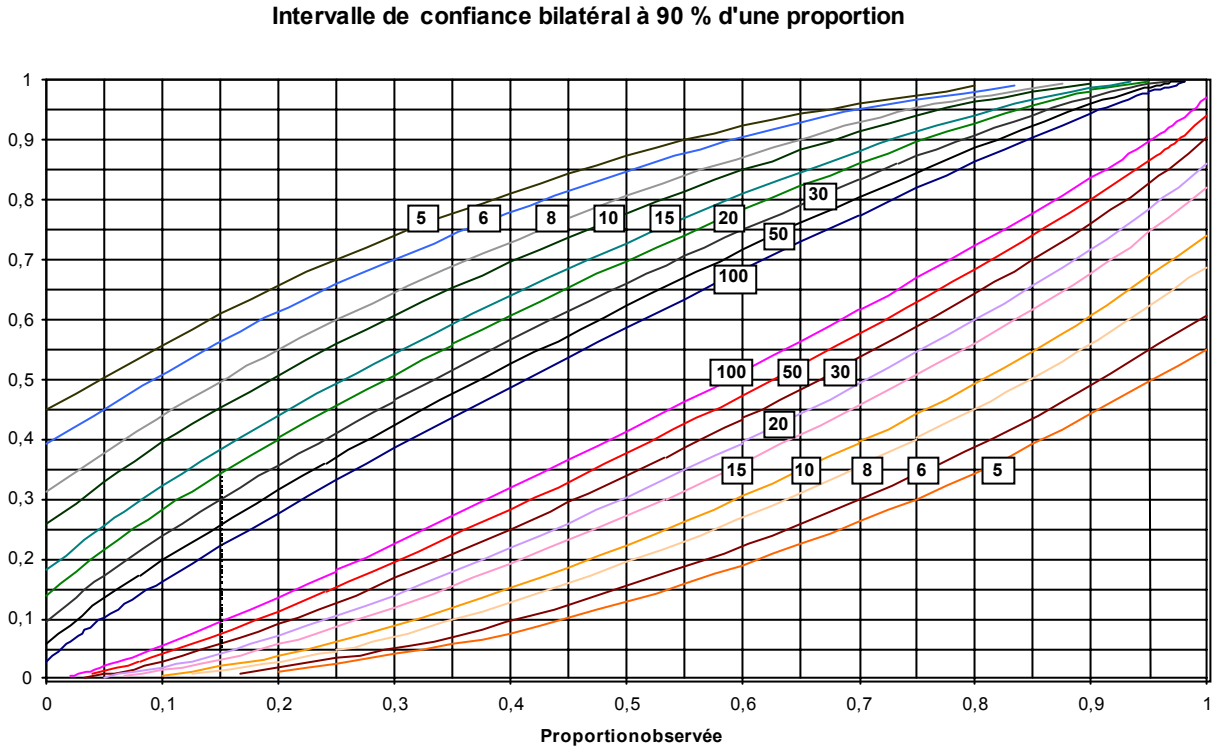
Donner un intervalle de confiance bilatéral de la proportion p d'appels n'aboutissant pas pour des raisons techniques, au seuil de risque 2 %.

Cas 2 : n < 30

Dans ce cas là[4], la seule approche est graphique. On utilise des abaques pour conclure.

Exemple :

Donner l'intervalle de confiance à 90 % de la proportion de l'absentéisme sur l'année dans l'entreprise Alpha (comprenant 500 employés) sachant que sur un échantillon représentatif de 20 employés la proportion est de 15 %.



L'intervalle de confiance de p est donc _ _ _ _ _ _ _ _ _ _ _ _ .

Remarque :

Si l'échantillonnage est exhaustif, il convient d'adapter les résultats ci-dessus :

- *Si σ est connu $\bar{X}$ suit approximativement la loi* $\mathcal{N}(m ; \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}})$.
- *Si σ est inconnu $\bar{X}$ suit approximativement la loi* $\mathcal{N}(m ; \frac{\sigma_e}{\sqrt{n-1}} \times \sqrt{\frac{N-n}{N}})$.
- *F suit approximativement la loi* $\mathcal{N}(p ; \sqrt{\frac{f_e(1-f_e)}{n-1}} \times \sqrt{\frac{N-n}{N}})$.

[4] Les abaques peuvent être utilisés dans tous les cas quelle que soit la valeur de n.