Echantillonnages et estimations

M.AIT OUDRA M.

M^{lle}.TOUATE S.

PLAN DU COURS

- Introduction générale;
- Chapitre I: Rappel sur les différentes lois de probabilité;
- Chapitre II: Théorie d’échantillonnage;
- Chapitre III: Estimation ponctuelle;
- Chapitre IV: Estimation par intervalle de confiance;
- Chapitre V: Théorie des tests

BIBLIOGRAPHIE INDICATIVE

- «Statistiques pour l'économie et la gestion» Anderson, Sweeney et Williams;
- «Eléments de statistique d'aide à la décision: cours et exercices résolus» par M.ELHAFIDI et D.TOUIJAR;
- «Théorie des sondage: échantillonnage et estimation en populations finies» par Yves Tillé;
- « Méthodes statistiques », P. TASSI;
- « Théorie des sondages »C. GOURIEROUX;
- « Méthodes statistiques de la gestion » J.L. BOURSIN

Les méthodes statistiques sont utilisées dans presque tous les domaines:

- dans le domaine industriel: la fiabilité des matériels, le contrôle de qualité (Etude des caractéristiques de pièces d'une chaine de fabrication), l'analyse des résultats de mesure et leur planification, la prévision, …
- dans le domaine de l'économie et des sciences sociales: les modèles économétriques (Prévoir l'évolution de la vente d'un produit), les sondages, les enquêtes d'opinion, les études quantitatives de marché,…

Après le recueil de données, la **démarche statistique** consiste à traiter et interpréter les informations recueillies.

Elle comporte deux grands aspects: l’aspect descriptif ou exploratoire et l’aspect inférentiel ou décisionnel.

Introduction générale

- Nous avons vu jusqu'à présent des méthodes destinées à observer, à d´écrire et à modéliser un phénomène:

- **Statistique descriptive (S1):** repose sur l'observation des phénomènes concrets (passés). Son but est de résumer, structurer et représenter l'information;

- **Probabilité (S2):** théorie mathématique permettant de modéliser des phénomènes où le hasard intervient et d´écrire des expériences aléatoires.

Introduction générale

- **Statistique inférentielle (S3):** statistique inductive (démarche inductive). son but est d'étendre (inférer=tirer les conséquences), à la population toute entière, les propriétés constatées sur un échantillon.

- Le but de l'inférence statistique est de généraliser les résultats obtenus auprès d'un échantillon **représentatif** pour décrire la population globale.

Introduction générale

- La statistique inférentielle a un aspect décisionnel et le calcul des probabilité y joue un rôle fondamental.

- Etude statistique= étude des caractéristiques d'un ensemble d'objets (population composée d'individus.

- Recensement: les valeurs sont disponibles sur l'ensemble de la population.

Introduction générale

- Sondage: étude d'une partie de la population (un échantillon).

- Le statisticien n'étudie pas le caractère sur l'ensemble de la population mais sur un échantillon extrait de la population pour plusieurs raisons, entre autres:

Introduction générale

– La taille de la population peut être très importante et le coût de l'enquête serait trop important (coût et temps);

– L'accès à tous les individus de la population est matériellement impossible (compléxité, population indéfinie)

Introduction générale

- Un bon échantillon (de qualité) doit constituer une image réduite de l’ensemble de la population (**représentatif)** dont on va étudier un caractère bien défini. Dans le cas contraire on dit que l’échantillon est **biaisé.**

Introduction générale

- Comment choisir un échantillon pour qu'il soit représentatif? (techniques d'échantillonnage)

- Comment les paramètres de la population peuvent- ils être estimés à partir de l'échantillon? (estimation)

Introduction générale

- L’échantillonnage désigne l'opération destinée à sélectionner une fraction d'une population, afin de conduire des analyses.
- Méthodes de prélèvement d’un échantillon:
- Méthode des quotas;
- Échantillonnage aléatoire;
- Échantillonnage au hasard simple;
- Échantillonnage stratifié;
- Échantillonnage par grappe;…

- La problématique de l’inférence statistique consiste, à partir d’un échantillon de données (technique d échantillonnage, chapitre2) provenant d’une population de loi de probabilité inconnue, à déduire des propriétés sur cette population : quelle est sa loi (problème d’estimation, chapitre 3 et4), comment prendre une décision en contrôlant au mieux le risque de se tromper (problème de test chapitre 5).

Introduction générale

- L'échantillonnage permet aux statisticiens de **tirer** des conclusions au sujet d'un tout en y examinant une partie. Il nous permet d'**estimer** des caractéristiques d'une population en observant directement une partie de l'ensemble de la population. Les chercheurs ne s'intéressent pas à l'échantillon lui-même, mais à ce qu'il est possible d'apprendre à partir de l'enquête et à la façon dont on peut appliquer cette information à l'ensemble de la population.

CHI:LOIS USUELLES CONTINUES

- Loi normale très utilisée en statistique inférentielle;

- Importante = une loi approchée par de nombreux phénomènes naturels;

- Dépend de deux paramètres;

- Elle est symétrique.

ChI: LOIS USUELLES CONTINUES

I. LOI NORMALE

A. Loi normale générale

a. Définition

On dit qu'une v.a.r X suit une loi Normale de paramètres σ et μ si :

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; \forall\, x \in \mathbb{R}$$

ChI: LOIS USUELLES CONTINUES

$X = N(\mu; \sigma)$ $X = N(\mu; \sigma^2)$

b. Espérance et Variance :

$E(X) = \mu$ $V(X) = \sigma^2$

c. Caractéristiques :

- La courbe de la loi Normale possède la forme en « CLOCHE »
- La distribution normale est symétrique par rapport à la droite verticale : X=μ

ChI: LOIS USUELLES CONTINUES

- Points d’inflexion sont situés à une distance σ de cet axe de symétrie

- f atteint son maximum lorsque x= μ

ChI: LOIS USUELLES CONTINUES

$$P(X > \mu) = P(X < \mu) = 50\ \%$$

- **Remarque :** La loi Normale générale n'est pas Tabulée

B. La loi Normale Centrée et Réduite :

a. Variable Centré et Réduite :

- Soit X une v.a :
- $[X - E(X)]$ S'appelle Variable Centrée.
- $U = \dfrac{X - E(X)}{\sigma}$ S'appelle Variable Centrée et Réduite.

ChI: LOIS USUELLES CONTINUES

- Si une V.A suit une loi normale générale, il est difficile de calculer sa fonction de répartition F(x).

- Pour tous les calculs, on se ramène à la fonction de répartition de la loi normale centrée réduite **(une loi TABULEE)**.

- Centrer et réduire une variable, c'est raisonner en nombre d'écart type par rapport à la moyenne.

ChI: LOIS USUELLES CONTINUES

V.A centrée réduite a pour espérance 0 et écart type 1

- E(U)= 0

- V(U)= 1

- La densité de probabilité de la loi normale centrée réduite : $f(U) = \frac{1}{\sqrt{2\pi}} . e^{\frac{-U^2}{2}} \quad \forall x \in \mathbb{R}$

ChI: LOIS USUELLES CONTINUES

b. Théorèmes

si

$$X = N(\mu_X;\sigma_X)$$

Quelle est la loi de Y=a X+b?

ALORS

$$Y = aX + b = N(a.\mu_X + b;\sqrt{a^2.\sigma_X^2}$$

Chapitre I (suite)

$X_1 = N(\mu_1;\sigma_1)$ $\qquad$ $X_2 = N(\mu_2;\sigma_2)$

- Sachant que X1 et X2 sont indépendant, quelle est la loi de X= aX1+bX2 ?
- **<u>Conclusion</u>**

$$X = aX_1 + b\,X_2 = N(a\mu_1 + b\mu_2; \sqrt{a^2.\sigma_1^2 + b^2.\sigma_2^2}$$

Chapitre I (suite)

- **Exercice**

- Soit $X = N(-2\,;5)$. Donner la loi de probabilité de $Y = 3X - 2$
- Soit $X_1 = N(-1;6)$ et $X_2 = N(2;5)$ .Donner la loi de probabilité de $X = -2X_1 + 4X_2$

X1 et X2 sont indépendantes

Chapitre I (suite)

- **<u>C.Théoréme Central Limit (T.C.L)</u>**

Le TCL sera très précieux puisqu'il nous expliquent que si on fait la somme d'un très grand nombre de variables aléatoire de la loi quelquonque, cette somme suit approximativement une loi normale.

Chapitre I (suite)

- Soit $X_1; X_2; \ldots; X_n$ avec (n) variables aléatoires identiquement distribuées et indépendantes. Alors :

$$X = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

$$E(X) \text{ et } V(X) : \text{si } x \mapsto \infty$$

- suit une loi Normale de paramètres

Chapitre I (suite)

D. Calcul des probabilités

a. fonction de répartition de $N(0;1)$

Soit $X = N(\mu;\sigma)$ et $U = \frac{X-\mu}{\sigma} = N(0;1)$

Alors la fonction de répartition (U) est notée $\pi(U)$

Avec $\pi(u) = P(U < u))$

Cette fonction est **Tabulée**

Chapitre I (suite)

$$P(X < a) = P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = P\left(U < \frac{a-\mu}{\sigma}\right)$$

$$P(X \geq a) = 1 - p(X < a) = 1 - \pi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right)$$

Chapitre I (suite)

La lecture de la table

- Elle donne la valeur de $\pi(u)$ pour (u) connue
- Elle donne la valeur de (u) pour $\pi(u)$ connue

EXEMPLE

- Pour les valeurs négatives $\pi(-u) = 1 - \pi(u)$
- Pour les valeurs $\pi(u)$ inférieures à 0,5, on utilise la symétrie de la distribution

Chapitre I (suite)

- Pour certaines valeurs qui ne figurent pas sur la table, on utilise **l'intérpolation linéaire**

Chapitre I (suite)

- ***Exercices d'applications***

- ***Exercice1***
- Soit $X = N(\mu = 2; \sigma^2 = 25)$
- Calculer $P(X < 4) et P(X > 3)$
- La V.A X suit une loi normale d'espérance 550 et d'Ecart type 100.
- Quelle est la probabilité pour que X soit moins de 650, plus de 746, moins de 500, entre 550 et 600

Chapitre I (suite)

Exercice II

- Le temps (en minutes) de fabrication d'un bien par un ouvrier est représenté par une variable normale (X). On considère deux ouvriers A et B travaillent indépendamment l'une de l'autre.
- Pour A $X_A = N(N_A = 60; \sigma_A^2 = 36)$
- Pour B $X_B = N(N_B = 45; \sigma_B^2 = 64$

Chapitre I (suite)

- Si A et B commencent à travailler au même instant, quelle est la probabilité que A termine avant B, la première unité produite ?

Exercice III

- Calculer $\pi(1.846)$ $\boldsymbol{\pi}(\mathbf{2.148})$ $\pi(U^*) = 0.0606$ $\pi(U^*) = 0.1562$
- Déterminer la valeur de U

- A la fin de cette partie, l'étudiant doit être capable de:
- Expliquer les caractéristiques d'une loi NORMALE;
- Définir et expliquer la valeur centrée réduite correspondant a n'importe quelle observation d'une variable normale (toute variable normale peut se convertir en une variable normale centrée réduite);
- Déterminer, a l'aide de la distribution normale centrée réduite, la probabilité d'une variable normale se trouvant dans un intervalle donne.

Chapitre I (suite)

II. Loi de khi-deux (Loi de Karl Pearçon)

a. Théorème

Soit $U_1, U_2, \ldots, U_v$ une suite de (*v*) **variables Normales Centrées et réduites indépendantes**.

X suit une loi de chi deux $\boldsymbol{\chi^2}$ à (*v*) *dégrée de liberté.*

$$X = U_1^2 + U_2^2 + \cdots + U_v^2 = \sum_{i=1}^{i=v} U_i^2$$

Chapitre I (suite)

(*v* représente le paramètre)

$X = \chi^2_{(v)}$

b.Caractéristiques de la distribution de $\chi^2_{(v)}$

$$D_{\chi^2} = [0;+\infty[= \mathbb{R}^+$$

- La distribution $\chi^2_{(v)}$ est dissymétrique pour les petites valeurs de **(*v*)**
- La distribution $\chi^2_{(v)}$ Commence à devenir symétrique à partir ***v=30***

Chapitre I (suite)

c. L'espérance et la variance

$X = \chi^2_{(v)}$

- $E(X) = v$
- $V(X) = 2v$

Exemple

$X = \chi^2_{(10)}$ $E(\chi^2_{(10)}) = 10$ $V(\chi^2_{(10)}) = 2 \times 10$

$X = \chi^2_{(5)}$ $E(\chi^2_{(5)}) = 5$ $V(\chi^2_{(5)}) = 2 \times 5$

Chapitre I (suite)

d. Comportement asymétrique (Approximation)

- **Approximation de Fisher:**

Si *v≥30 alors:* $\sqrt{2\chi^2} \quad \sqrt{2\overline{v}-1} \sim N(0;1)$

- **Approximation générale:**

Si *v≥101 alors:* $\dfrac{\chi^2 - v}{\sqrt{2v}} \sim N(0;1)$

Chapitre I (suite)

e. Lecture de table de $\boldsymbol{x^2_{(v)}}$

- La table $\boldsymbol{x^2_{(v)}}$ donne les valeurs de la V.A $\boldsymbol{x^2_{(v)}}$ ayant la probabilité α d'être dépassée.
- Elle donne:
  - La probabilité **α** si $\boldsymbol{x^2_{(v)}}$ et (*v)*sont connus
  - La valeur de $\boldsymbol{x^2_{(v)}}$ si **α** et (*v)* sont connus.

Chapitre I (suite)

Exercices d'application

Exercice I

Déterminera la valeur de (k) dans chacun des cas suivants :

- $P(\boldsymbol{x^2}_{(17)} > k) = 0,25$
- $P(\boldsymbol{x^2}_{(24)} > k) = 0,01$
- $P(\boldsymbol{x^2}_{(29)} < k) = 0,90$

Chapitre I (suite)

Exercice 2 :

1- Calculer le quantile d'ordre 10% pour la V.A $\boldsymbol{x^2}_{\mathbf{(10)}}$.

2- Déterminer la médiane de $\boldsymbol{x^2}_{\mathbf{(15)}}$

Exercice 3 :

Déterminer les valeurs k_1 et k_2 dans chacune d des eux cas suivants :

- $P(k_1 < \boldsymbol{x^2}_{\mathbf{(9)}} < k_2)=0,9$
- $P(k_1 < \boldsymbol{x^2}_{\mathbf{(50)}} < k_2)=0,98$

Chapitre I (suite)

Rappel : le quantile d'ordre (α) de la variable X est noté (q_{α}) *avec* : $P(X< q_{\alpha}) = \alpha$

Rappe2: La médiane est la modalité qui partage la série statistique en deux parties égales.

Autrement dit :

La médiane est le quantile d'ordre $\alpha=0,50$

Chapitre I (suite)

Exercice 4

En utilisant l'approximation de Fisher. Calculer la valeur de k_1 telle que :

$P(\boldsymbol{x^2}_{(60)} < k_1) = 0{,}8413$

Exercice 5 :

Soit $X = \boldsymbol{x^2}_{(200)}$ Calculer la valeur de k telle que :

$P(\boldsymbol{X} > k) = 0{,}0708$

Chapitre I (suite)

III. Loi de Student (William Sealy Gosset)

a. ***Définition***

Soit U=N(0,1) et X= $\boldsymbol{x}^{2}_{(v)}$

Avec U et X sont indépendantes

Alors: $$T = \frac{N(0,1)}{\sqrt{\frac{x^2_{(v)}}{v}}} = \frac{U}{\sqrt{\frac{X}{v}}}$$

Suit une loi student à (*v*) degré de liberté.

- ***Notation :*** T = stud $_{(v)}$

Chapitre I (suite)

b. Espérance mathématique et variance

Soit T = stud $_{(v)}$

- E(T)= 0 (existe si v>1)
- V(T)= $\frac{v}{v-2}$ (existe si v>2)

c. Comportement asymétrique

Soit T = stud $_{(v)}$

Si (v) *t*end vers l'infini *alors* T tend vers N(0,1)

Chapitre I (suite)

- ***Remarque***

La loi de Student est généralement utilisée pour les petits échantillons lorsque la variance de la population est inconnue.

Chapitre I (suite)

IV. Loi de Fisher-Snédécor

a. Définition

$X_1 = \boldsymbol{\chi^2_{(v1)}}$ **et** $X_2 = \boldsymbol{\chi^2_{(v2)}}$

X_1 et X_2 deux variables indépendantes

Alors la variable:

$$F = \frac{X_1/v_1}{X_2/v_2} = \frac{X_1}{X_2} \times \frac{v_2}{v_1}$$

Suit une loi de Fisher de paramètres v_1 et v_2 notée

$$\boldsymbol{F} = \boldsymbol{F}(v_1\,;v_2)$$

Chapitre I (suite)

Remarque

La distribution de Fisher, comme celle de *x^2*, n'est pas symétrique.

b. Règle empirique de Fisher

$$F_\alpha(v_1\,;v_2)=\frac{1}{F_{1-\alpha}(v_2\,;v_1)}$$

- ***Avec*** $F_\alpha(v_1\,;v_2)$ ***est telle que:***

Chapitre I (suite)

$$P\left(F_{(v_1\,;v_2)} > f_\alpha(v_1\,;v_2)\right) = \alpha$$

- La table donne les valeurs de F pour avec la règle empirique :

$$F_{0,95}(v_1\,;v_2) = \frac{1}{F_{0,05}(v_2\,;v_1)}$$

$$F_{0,99}(v_1\,;v_2) = \frac{1}{F_{0,01}(v_2\,;v_1)}$$

CHII:ECHANTILLONNAGE

- **I- Définitions**

  a-Population

  – **Population statistique:** ensemble des unités statistiques (personnes ou objets) auxquelles on s'intéresse et sur lesquelles porte une étude. Et on appelle **individu** chaque élément (entité) de cette population pour lequel des données sont collectées.
  – **Variable Aléatoire Parente** (V.A.P): caractère étudié dans la population statistique (poids, CA, salaire, taille, bénéfice,…), noté X , Y, Z,…La loi de la V.A.P s'appelle « **loi de la population** »

CHII (suite)

- **Paramètres de la population:** Ce sont les caractéristiques numériques de la population statistique
  - ❖Moyenne de la population (μ) = E(X);
  - ❖Variance de la population (σ^2) = V(X);
  - ❖Proportion de la population = P

b-Echantillon

- **Définition:** une partie (sous ensemble) représentative de la population statistique observée. La population d'où on tire l'échantillon s'appelle « **population mère** » ou « **population échantillonnée** »

La constitution d’un échantillon permet de collecter des données pour répondre à une question concernant une population.

Exemple 1:

Les membres d’un parti politique marocain sont supposés soutenir un candidat particulier aux élections du parlement, et les responsables du parti voudraient estimer la proportion d’électeurs favorables à leur candidat. Un échantillon de 400 électeurs a été sélectionné et 160 de ces électeurs ont indiqué être favorables au candidat. Une estimation de la proportion d’électeurs favorables au candidat est donc 160 sur 400 soit 0,40.

Exemple 2:

Un fabriquant de pneu a conçu un nouveau type de pneu permettant d'accroitre le kilométrage effectué. Pour estimer le nombre moyen de kilomètres effectués avec les nouveaux pneus, le fabricant a sélectionné un échantillon de 120 nouveaux pneus, dans le but de les tester. D'aprés les résultats du test, la moyenne de l'échantillon est égale a 36 500 kilomètres. Par conséquent, une estimation du kilométrage moyen pour la population des nouveaux pneus est de 36 500 kilomètres.

Il est important de comprendre que les résultats d’un échantillon fournissent seulement des ***estimations (***l’échantillon contient juste une partie de la population. Une certaine ***erreur d’échantillonnage*** est attendu***).***

Avec des méthodes d’échantillonnage adéquates, les résultats de l’échantillons fournissent de « *bonnes*» estimations des paramètres de la population (une moyenne d’échantillon fournit une estimation de la moyenne de la population, une proportion d’échantillon fournit une estimation de la proportion de la population)

- **Sélectionner un échantillon:**

  comment sélectionner un échantillon à partir d'une population finie ?

  comment sélectionner un échantillon à partir d'une population infinie ?

  – ***Échantillonnage à partir d'une population finie:*** les statisticiens recommandent d'utiliser des méthodes d'échantillonnage probabilistes .

    Le type le plus simple d'échantillons probabiliste est celui dans lequel chaque échantillon de taille n a la même probabilité d'etre sélectionné . On parle d'**<u>échantillon aléatoire simple</u>**

CHII (suite)

- **Echantillon aléatoire simple(E.A.S) (population finie):** La construction d’un échantillon aléatoire simple de taille n est réalisée par un tirage au hasard avec remise de n individus dans l’ensemble de la population. Ainsi, tous les individus seront tirés de manière indépendante et auront une chance égale de faire partie de l’échantillon (la même probabilité d’etre sélectionné).

CHII (suite)

- Autrement dit, On dit qu'un échantillon (X_1, X_2, X_3,…, X_n) est aléatoire simple si et seulement si :
  - Quelque soit i= 1, 2, 3,…,n; la loi de X_i est la même que celle de la V.A.P (X), d'où:
    - $E(X_i) = E(X) = \mu$
    - $V(X_i) = V(X) = \sigma^2$
  - Et les X_i sont indépendants

– ***Échantillonnage à partir d'une population infinie:*** parfois, nous souhaitons sélectionner un échantillon à partir d'une population qui est infiniment grande ou dont les éléments sont génères par un processus pour lequel il n'y a pas de limite quant au nombre d'éléments qui peuvent être génères.

Ainsi, il n'est pas possible de développer une liste de tous les éléments de cette population. C'est ce qu'on appelle le cas d'une population infinie. Dans ce cas, les statisticiens recommandent de sélectionner un **échantillon aléatoire**.

- **Échantillon aléatoire (population infinie):** un EA de taille n issue d'une population infinie est un échantillon sélectionné qui satisfait les conditions suivantes:

1. Chaque élément sélectionné est issu de la même population;
2. Chaque élément est sélectionné indépendamment des autres.

Exemples des situations impliquant un échantillonnage à partir d'une population infinie (généralement associés à un processus durable): pièces fabriquées sur une chaine de production, transactions bancaires, clients entrants dans un magasin, appels téléphoniques reçus dans un centre de soutien technique, …

CHII (suite)

C- probabilité d'un échantillon (fonction de vraisemblance)

Soit $(X_{1,}\ X_{2,}\ X_{3,\dots,}\ X_n)$ un échantillon aléatoire simple dont la réalisation est $(x_1, x_2, x_3, \dots, x_n)$, la probabilité de cette échantillon est notée: $L(x,\Theta)$, avec (**cas continu**):

$$L(x;\theta) = \prod_{i=1}^{i=n} f(x_i) = f(x_1) \times f(x_2) \times \dots \times f(x_n)$$

CHII (suite)

- Ou (**cas discret**)

$$L(x;\theta) = \prod_{i=1}^{i=n} P(X_i = x_i) = P(X_1 = x_1) \times P(X_2 = x_2) \times P(X_n = x_n)$$

- Ɵ est le (s) paramètre(s) de la loi de la V.A (X).

CH II (suite)

- **Exemple**

soit X=P(λ), calculons L(x,λ)

- **Exercice1**

Soit X=B(n,p)

Calculer L(x;p)

- **Exercice2**

Soit X= B(p)

Calculer L(x;p)

CHII (suite)

- **Exercice 3**

$X=N(\mu;\sigma^2)$

Calculer $L(x;\mu;\sigma^2)$

CHII (suite)

- **II- variables et distribution d'échantillonnage**

Distribution de probabilité appelé distribution d échantillonnage

Supposons une population statistique de taille N. de combien de manières différentes peut on tirer un échantillon de taille n parmi les N individus?

La réponse dépend des conditions de tirage.

Supposons un T.S.R simultané. Le nombre d'échantillons possible est $C_N^n = k$.

CHII (suite)

a-définition

On appelle variable d'échantillonnage toute fonction de l'E.A.S ($X_{1,}$ $X_{2,}$ $X_{3,\dots,}$ X_n)

Exemple:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{i=n} X_i$$

$$S_e^2 = \frac{1}{n}\sum_{i=1}^{i=n} (X_i - \bar{X})^2 =$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{i=n} (X_i - \bar{X})^2$$

CHII (suite)

<u>b-la moyenne d’échantillonnage</u>

Une distribution d échantillonnage de la moyenne d échantillonnage correspond à la distribution de toutes les valeurs possibles de la moyenne d échantillon

- <u>**Exemple:**</u> on effectue une étude démographique sur la fécondité chez la femme citadine et on considère la variable aléatoire X qui désigne <u>***le nombre d ’enfants***</u> par famille.
- On s’intéresse au nombre moyen d’enfants par famille. Pour cela, on prélève 5 échantillons aléatoires et on observe la réalisation des 9 V.A X 1 , X 2,…, X 9
- Pour chacun des 5 échantillons on a:

CHII (suite)

Echantillon	1	2	3	4	5
X_1	2	1	4	4	0
X_2	**1**	**0**	**3**	**4**	**1**
X_3	**1**	**0**	**3**	**0**	**0**
X_4	**1**	**4**	**0**	**1**	**2**
X_5	**3**	**3**	**1**	**0**	**2**
X_6	**2**	**2**	**2**	**3**	**2**
X_7	**5**	**5**	**5**	**2**	**4**
X_8	**2**	**1**	**2**	**4**	**3**
X_9	**4**	**0**	**2**	**1**	**5**
X	**2,3**	**1,8**	**2,4**	**2,1**	**2,1**

CHII (suite)

- On remarque que le nombre moyen d'enfants par famille prend des valeurs différentes selon l'échantillon considéré;
- La moyenne est donc une variable aléatoire

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{i=n} X_i$$

CHII (suite)

❖ **Propriété de la moyenne d'échantillonnage**

Le tableau précédent nous a permis d'obtenir 5 réalisations de la moyenne d'échantillonnage qu'on note:

On s'attend à ce que ces 5 valeurs soient proches de la moyenne μ .

CHII (suite)

- **Propriété 1**

Soit X1, X2, …, Xn un échantillon aléatoire simple de taille n, relatif à la V.A. parente X. L'espérance de la V.A moyenne d'échantillonnage est égale à la moyenne de la population μ

$$E(\bar{X}) = \mathbf{E(X_i) = E(X) = \mu}$$

CH II (suite)

- **Propriété 2**

Soit X1, X2, …, Xn un échantillon aléatoire simple de taille n, relatif à la V.A. parente X. La variance de la V.A moyenne d'échantillonnage est égale à la variance de X divisée par la taille n de l'échantillon:

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

CH II (suite)

❖ **Loi(Distribution) de la moyenne d'échantillonnage**

La loi de $\bar{X}$ dépend de :

- Loi de la population (Normale ou quelconque)
- Variance de la population (Connue ou inconnue)
- Taille de l'échantillon (n petite ou grande)

CH II (suite)

- **Exemple**

La Taille des étudiants d'une université, Notée X est une Variable Normale de moyenne 170 cm et d'écart type 7 cm .On tire au hasard un échantillon de taille n=49. Calculer la probabilité que la moyenne de l'échantillon soit :

a- inférieure à 172 cm

b-comprise entre 168cm et 172cm ?

CH II (suite)

C- La variance d'échantillonnage et la quasi-variance

❖**Définition:** On rappelle que la variance d'échantillonnage(noté S_e^2)est une V.A. qui associe une valeur numérique à chaque tirage d'échantillon.

$$S_e^2 = \frac{1}{n}\sum_{i=1}^{i=n}(X_i - \bar{X})^2 =$$

CH II (suite)

- Si: $\begin{cases} pop.\,Normale \\ \quad + \\ \mu \; inconnue \end{cases}$

- Alors: $\dfrac{nS_e^2}{\sigma^2} = \chi^2_{(v=n-1)}$

CH II (suite)

- Si **μ** est connue alors:

$$\frac{nS_e^2}{\sigma^2} = \chi^2_{(v=n)}$$

CH II (suite)

- **Propriété:** Soit X1, X2, …, Xn un échantillon aléatoire simple de taille n, relatif à la V.A. parente X de moyenne µ et de variance σ^2. L'espérance de la V.A variance d'échantillonnage est égale à:

$$E(S_e^2) = \frac{(n-1)}{n} \times \sigma^2$$

- La variance de la V.A variance d'échantillonnage est égale à:

$$V(S_e^2) = \frac{\sigma^4 \times 2(n-1)}{n^2}$$

CH II (suite)

- **Remarque**: L'espérance de la variance d'échantillonnage n'est pas une image parfaite de la variance.

- Pour remédier à ce problème, on construit une statistique (VA) qui approchera le mieux σ^2

CH II (suite)

- **Propriété :** Soit X1, X2, …, Xn un échantillon aléatoire simple de taille n, relatif à la V.A. parente X de moyenne μ et de variance σ^2. **L'espérance** de la V.A quasi-variance d'échantillonnage est égale à:

$$E(S^2) = \sigma^2$$

- **La variance** de la V.A quasi-variance d'échantillonnage est égale à:

$$V(S^2) = \frac{2\sigma^4}{(n-1)}$$

CH II (suite)

- Si:

$$\begin{cases} pop.\,Normale \\ + \\ \mu\ inconnue \end{cases}$$

- Alors:

$$\frac{(n-1)S^2}{\sigma^2} = \frac{nS_e^2}{\sigma^2} = \chi^2_{(v=n-1)}$$

- si **μ** est connue :

$$\frac{(n-1)S^2}{\sigma^2} = \frac{nS_e^2}{\sigma^2} = \chi^2_{(v=n)}$$

CH II (suite)

- **Remarque**: la variance et la quasi-variance tendent toutes les deux vers zéro lorsque la taille de l'échantillon tend vers l'infini

CH II (suite)

d- <u>**La proportion (fréquence) aléatoire F**</u>

- **Définition:** Soit X1, X2, …, Xn une suite de variables aléatoires de Bernoulli indépendantes.

$$F = \frac{\sum_{i=1}^{i=n} X_i}{n} \qquad \left(\forall\, X_i = B(p) \, avec \, X_i = \begin{cases} 1 & (succés) \\ 0 & (Echec) \end{cases}\right)$$

- S’appelle proportion (ou fréquence) aléatoire.

$$F = \frac{\sum X_i}{n} = \frac{B(n;p)}{n}$$

CH II (suite)

- **Propriétés de F:**

p = proportion de succès dans la pop. stat)

$$E(F) = P$$

$$V(F) = \frac{p.q}{n}$$

$$\sigma_F = \sqrt{\frac{p.q}{n}}$$

(si le tirage est avec remise)

CH II (suite)

Si le tirage est sans remise

$$V(F) = \frac{p.q}{n} \cdot \frac{N-n}{N-1}$$

$$\sigma_F = \sqrt{\frac{p.q}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

CH II (suite)

- **Loi de probabilité de F**

Lorsque n < 30

Soit $F = \frac{B(n;p)}{n}$

Valeurs possible de F : D_F

Pour $B(n;p) \Rightarrow \{0;1;2;\ldots;n\}$

Donc $\boldsymbol{D_F = \left\{0;\frac{1}{n};\ldots;1\right\}}$

CH II (suite)

- Soit $X = B(n;p) \Rightarrow P(X = k) = C_n^k.p^k.q^{n-k}$

- Si: $F = \dfrac{B(n;p)}{n} = \dfrac{X}{n}$

- Alors: $P(F = l) = p\left(\dfrac{X}{n} = l\right) = P(X = nl)$

- $P(F = l) = C_n^{nl}.p^{nl}.q^{n-nl}$

CH II (suite)

- Si:

$$n \geq 30$$
$$np \geq 5$$
$$nq \geq 5$$

- Alors:

$$\frac{F - P}{\sqrt{\frac{p.q}{n}}} \leadsto N(0;1)$$

CH III: Estimation

- Après avoir prélevé l'échantillon, et étudié les distributions d'échantillonnage, on peut alors généraliser, à la population, les résultats expérimentaux obtenus à partir de l'échantillon

- L'estimation consiste en l'évaluation d'un paramètre de la population à partir de l'observation d'un E.A. La théorie de l'estimation se divise en deux parties:

CH III: Estimation ponctuelle

- ***L'estimation ponctuelle: permet*** d'obtenir une valeur unique calculée à partir d'un E.A., valeur qui sera prise comme estimation du paramètre inconnu.

- ***L'estimation par intervalle: permet de*** déterminer un intervalle qui, avec une grande probabilité fixée a priori, contient la vraie valeur du paramètre inconnu

Estimation ponctuelle

- **Définition de l'estimateur**: Soit X1, X2, …, Xn un E.A.S relatif à la V.A. parente X dont sa loi dépend du paramètre θ. On appelle estimateur du paramètre θ, toute statistique utilisée dans le but d'approcher la valeur inconnue de θ. Un estimateur est donc une **Variable Aléatoire**, c'est une fonction de l'E.A.S.

- Soit $\hat{\theta}$ l'estimateur de θ alors:

$$\hat{\theta} = f(X_1; X_2; \ldots; X_n)$$

Estimation ponctuelle

- **Exemple:**

$$\hat{\theta} = \bar{X} = \frac{\sum X_i}{n} = f(X_1; X_2; \dots; X_n)$$

$$\hat{\theta} = S^2 = \frac{n}{n-1}\sum (X_i - \bar{X})^2 = f(X_1; X_2; \dots; X_n)$$

$$\hat{\theta} = S_e^2 = \frac{1}{n}\sum^{n} (X_i - \bar{X})^2 = f(X_1; X_2; \dots; X_n)$$

$$\hat{\theta} = F = \frac{\sum X_i}{n} = f(X_1; X_2; \dots; X_n)$$

avec $X_i = \begin{cases} 1 \ (succés) \\ 0 \ (Echec) \end{cases}$

Estimation ponctuelle

- **Définition de l'estimation:** C'est la valeur prise par l'estimateur dans l'échantillon réalisé (x_1; x_2; x_3; x_4;…;x_n).
- Une estimation du paramètre θ est une réalisation d'un estimateur de ce paramètre. Une estimation est donc une valeur numérique.

Estimation ponctuelle

II- **Propriétés des estimateurs**

1-Estimateurs sans biais

Définition: $\hat{\theta}$ est un estimateur sans biais du paramètre θ

Si: $E\left(\hat{\theta}\right) = \theta$

$$B\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta$$

Remarque: Si $\hat{\theta}$ est biaisé, le biais (l'erreur) serait alors: $\underline{\underline{\boldsymbol{B}\left(\hat{\boldsymbol{\theta}}\right)}}$

Estimation ponctuelle

- **Interprétation**:

Si $B(\hat{\theta}) = 0 \Rightarrow \hat{\theta}$ est un estimateur sans biais de θ

Si $B(\hat{\theta}) > 0 \Rightarrow \hat{\theta}$ sur estime θ.

Si $B(\hat{\theta}) < 0 \Rightarrow \hat{\theta}$ sous estime θ.

Estimation ponctuelle

- Exercice: montrer que:

$\bar{X}$ est un estimateur sans biais de $\boldsymbol{\mu}$

F est un estimateur sans biais de p

S_e^2 est un estimateur biaisé de σ^2

Calculer le biais et proposer un estimateur sans biais de σ^2 en fonction de S_e^2

Estimation ponctuelle

Définition: On appelle $\widehat{\theta}$ estimateur asymptotiquement sans biais, du paramètre θ

Si $\lim_{n \to \infty} E\left(\widehat{\theta}\right) = \theta$

Exemple

S_e^2 est un estimateur asymptotiquement sans biais de σ^2

Estimation ponctuelle

2-Estimateurs Convergents

- **Définition:** On appelle $\hat{\theta}$ estimateur convergent du paramètre θ si: $\lim_{n \to \infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$
- *Dans la pratique*, pour vérifier la convergence, il suffit de vérifier la **consistance**.
- **Estimateur consistant:** $\hat{\theta}$ un estimateur convergent du paramètre θ si: $\begin{cases} E(\hat{\theta}) = \theta \\ \lim_{n \to \infty} V(\hat{\theta}) = 0 \end{cases}$

Estimation ponctuelle

- Si $\hat{\theta}$ est un estimateur sans biais ou asymptotiquement sans biais de θ , et si sa variance tend vers zéro lorsque n tend vers l'infini, alors $\hat{\theta}$ est un estimateur consistant de θ
- **Théorème:**

Tout estimateur Consistant est convergent

- **Exemple:**

$E(\overline{X}) =$ **μ** et $\lim\limits_{n\to\infty} V(\overline{X}) = \lim\limits_{n\to\infty}\left(\frac{\sigma^2}{n}\right) = 0$ $D'où\ \overline{X}$ est un estimateur convergent car il est consistant

Estimation ponctuelle

3-Estimateurs Efficaces (choix entre estimateurs):

- **Choix entre estimateurs quelconques**: pour choisir entre plusieurs estimateurs quelconque, on utilise le critère de l'**erreur quadratique moyenne** (EQM).

$$EQM\left(\hat{\theta}\right) = E\left(\hat{\theta} - \theta\right)^2$$

$$EQM\left(\hat{\theta}\right) = V\left(\hat{\theta}\right) + B\left(\hat{\theta}\right)^2$$

On choisit l'estimateur à EQM **minimale.**

Estimation ponctuelle

Si $\boldsymbol{EQM}(\widehat{\boldsymbol{\theta}}_1) < EQM(\hat{\theta}_2) \rightarrow \hat{\theta}_1 \; est \; plus \; efficace \; que \; \hat{\theta}_2$

- **<u>Choix entre estimateurs Sans-biais</u>**: Entre deux estimateurs sans biais, on préfère utiliser celui qui a la variance la plus petite (la variance minimale).
- **<u>Efficacité:</u>**

➢ **Inégalité de CRAMER-RAO:** $\boldsymbol{V}(\widehat{\boldsymbol{\theta}}) \geq \dfrac{\mathbf{1}}{\boldsymbol{I}(\theta)}$

Estimation ponctuelle

- Avec

$$\boldsymbol{I}(\boldsymbol{\theta})\ quantité\ d'information\ de\ fisher$$

$$\mathbf{I}(\theta) = \mathbf{n} \times E\left\{\left(\frac{\partial \log \mathbf{P}(\mathbf{X_i} = \mathbf{x_i})}{\partial \theta}\right)^2\right\}$$

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbf{V}\left\{\frac{\boldsymbol{\partial} \log \mathbf{L}(\mathbf{x}; \boldsymbol{\theta})}{\boldsymbol{\partial\theta}}\right\}$$

Estimation ponctuelle

- **Estimateur efficace:** $\hat{\theta}$ est efficace si:

$$V(\hat{\theta}) = \frac{1}{I(\theta)}$$

III. Méthodes d'estimation ponctuelle

a. La Méthode des moments : Elle consiste à estimer les moments de la population par les moments analogues de l'échantillon.

Estimation ponctuelle

- **<u>Rappel</u>**

$$E\left(X^k\right) = Moment\ non\ centré\ d'ordre k$$

$$E\left(((X - E(X))^k\right) = Moment\ centrée\ d'ordre\ k$$

- **Cas particuliers :**

Si $\boldsymbol{k = 1 \Longrightarrow E\left(X^k\right) = E(X)}$

Si $\boldsymbol{k = 2 \Longrightarrow E[((X - E(X))^2] = V(X) = E\left(X^2\right) - E(X)^2}$

Estimation ponctuelle

- ***b. La Méthode de maximum de vraisemblance***

Pour calculer l'estimation de maximum de vraisemblance (EMV) de θ noté $\hat{\theta}$, on doit suivre 4 étapes : $Calculer\ L(x\,;\theta) = \prod_{k=1}^{n} f(x_i)$

$$Calculer \log L(x\,;\theta) = log\left(\prod_{k=1}^{n} f(x_i)\right) = \sum_{i=1}^{n} log f(x_i)$$

$$Résoudre\ on\ \theta;\ l'équation:\ \frac{\partial\ log L(x\,;\theta)}{\partial\theta} = 0$$

Estimation ponctuelle

$La\ solution\ \hat{\theta}\ peut\ \text{être}\ un\ EMV$

$$Vérifier\ si\ \frac{\partial \log L(x;\theta)}{\partial \theta^2} \leq 0$$

Estimation par intervalle de confiance

- Un estimateur ponctuel est une statistique d'échantillon utilisée pour estimer un paramètre d’une population.
- puisqu’on ne peut s’attendre à ce qu’une estimation ponctuelle soit exactement égale à la valeur du paramètre de la population correspondant, une estimation par intervalle est souvent réalisée en ajoutant une **marge d’erreur** à l’estimation ponctuelle.

Estimation par intervalle de confiance

- Elle permet d'estimer le paramètre θ par un intervalle $[a;b]$
- Elle tient compte de l'erreur d'échantillonnage.
- **Définition:** On dit que $[a;b]$ est un intervalle de confiance au niveau 1- α pour le paramètre θ à partir de l'échantillon réalisé $(x_1, x_2, x_3,\ldots,x_n)$ si on a :

$$P(\theta \in [a;b]) = 1 - \alpha$$

- Les bornes de l'intervalle sont des statistiques relatives à l' E.A.

Estimation par intervalle de confiance

- **1- α** est appelé niveau de confiance de l'intervalle.
- **α** s'appelle « Risque » avec: $\alpha = P[\theta \notin [a;b]]$
- Plus **α** est petite et plus l'intervalle de confiance est grand. Généralement, on considère des intervalles à risques symétriques:

I- Intervalle de Confiance pour une proportion: **$I_C(p)$**

Estimation par intervalle de confiance

Construction de l’ I_C(p): Afin d’estimer une proportion p de « Succès » par exemple; on utilise la fréquence F qui est un estimateur sans biais, convergent et efficace du paramètre p.

Exemple: pour un niveau de confiance 95%; **α** =5%

Estimation par intervalle de confiance

Propriété: Soit X1, X2, …, Xn un échantillon aléatoire simple de taille n, relatif à la V.A. parente X suivant une loi de Bernoulli de paramètre p inconnu**.** Un intervalle de confiance au niveau **1- α** pour **p** peut se présenter comme suit :

$$I_p = \left[f - t_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}} \, ; f + t_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}} \right]$$

Estimation par intervalle de confiance

- Avec $t_{1-\frac{\alpha}{2}}$ est toujours lue sur la table de la loi normale centrée réduite après vérification du T.C.L
- Et $\pi\left(t_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$

Exercice d'application: Parmi un E.A. de 250 électeurs, 108 déclarent vouloir voter pour le président sortant. Tandis que les autres voteront pour l'autre candidat.

- Donner un IC(p) au niveau 90 et 95%
- Combien faudrait-il interroger d'électeurs pour que l'erreur d'estimation ne dépasse pas 2% ?

Estimation par intervalle de confiance

- ***p*** représente la proportion des votants en faveur de A dans toute la population des électeurs.
- Une estimation ponctuelle de p est donnée, grâce à la réalisation de l'E.A.,par la fréquence.
- Condition du T.C.L?

Estimation par intervalle de confiance

2- Précision d'une estimation par intervalle de confiance: Pour un **α** donné, on veut déterminer la taille nécessaire de l'échantillon pour atteindre une certaine précision de l'estimation.

❖ **Longueur (l'amplitude) $I_C(p)$ noté $L_{(p)}$**

- **$L_{(p)}$ = la borne supérieur – la borne inférieur= b - a**

$$L_{(p)} = 2\, t_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}}$$

Estimation par intervalle de confiance

- **La marge d'erreur ou l'erreur d'estimation (précision) noté E:** c'est la demi longueur de l'intervalle de confiance.

$$E_{(p)} = \frac{L_{(p)}}{2} = t_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}}$$

Remarque: Plus l'erreur d'estimation est petite et plus la précision est grande.

- **Détermination de la taille de l'échantillon** (n): On suppose préalablement que le T.C.L est vérifié. Si (n) est inconnue donc f est aussi inconnue !

Estimation par intervalle de confiance

- Ainsi; $E_{(p)} = \boldsymbol{t_{1-\frac{\alpha}{2}}}\sqrt{\dfrac{\boldsymbol{p(1-p)}}{\boldsymbol{n}}}$

1er Cas : on dispose d'une information sur (p): par exemple (p=p*)

D'où :

$$E_{(p)} \leq E_p^* \Leftrightarrow \boldsymbol{t_{1-\frac{\alpha}{2}}}\sqrt{\frac{p^*(\boldsymbol{1}-p^*)}{\boldsymbol{n}}} \leq E_{(p)}^*$$

$$\Leftrightarrow t_{1-\frac{\alpha}{2}} \frac{\sqrt{p^* q^*}}{\sqrt{n}} \leq E_{(p)}^*$$

Estimation par intervalle de confiance

- Donc $\Leftrightarrow \sqrt{n} \geq \dfrac{t_{1-\frac{\alpha}{2}}\sqrt{p^* q^*}}{E^*_{(p)}}$

$$n \geq \left(\frac{t_{1-\frac{\alpha}{2}}\sqrt{p^* q^*}}{E^*_{(p)}}\right)^2 \quad avec: q^* = 1 - p^*$$

- ***<u>2eme Cas : Aucune information sur (p)</u>***

Dans ce cas, on pose p=0,5. Cette valeur permet d’avoir l’intervalle de confiance de p **le plus large.**

- ***<u>2eme Cas : Aucune information sur (p)</u>***

Dans ce cas, on pose p=0,5. Cette valeur permet d’avoir l’intervalle de confiance de p **le plus large.**

$$n \geq \left(\frac{t_{1-\frac{\alpha}{2}}}{2E^{*}_{(p)}} \right)^{2}$$

Estimation par intervalle de confiance

II- Intervalle de Confiance pour la moyenne: $I_C(\mu)$

Pour effectuer une estimation par intervalle de la moyenne d'une population, l'écart type de la population ou celui de l'échantillon permettent de calculer la marge d'erreur.

Estimation par intervalle de confiance

$I_C(\mu)$ dans le cas d'une population Normalement distribuée: Dans ce cas la V.A. parente X suit une loi normale de moyenne**μ** et de variance σ^2

- Or, on sait que est un bon estimateur de la moyenne**μ.**

$$I_\mu = \left[\bar{x} - t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{x} + t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

- Cet intervalle n'a de sens que si σ^2 est connue.

Estimation par intervalle de confiance

Si σ^2 est inconnue, alors un intervalle de confiance au niveau **1- α** pour **μ** est :

$$I_\mu = \left[\bar{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} ; \bar{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$$

avec: $s^2 = quasi\ variance; s = \sqrt{s^2}$

Estimation par intervalle de confiance

- Précisions:

$$t_{1-\frac{1}{\alpha}} \begin{cases} table\ de\ N(0;1) \\ \quad ou \\ table\ Stud_{(n-1)} \end{cases}$$

- avec:

$$P\left(U < t_{1-\frac{\alpha}{2}}\right) = \pi\left(t_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$$

- Ou

$$P\left(U < Stud_{(n-1)}\right) = 1 - \frac{\alpha}{2}$$

Estimation par intervalle de confiance

- **L'amplitude**

$$L_{(\mu)} = 2\, t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \qquad\qquad L_{(\mu)} = 2t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

- **La marge d'erreur**

$$E_{(\mu)} = \frac{L_{(\mu)}}{2} = t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad ou \quad E_{(\mu)} = t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \quad si\ \sigma^2\ inconnue$$

Estimation par intervalle de confiance

- **La taille de l'échantillon**

$$n \geq \left(\frac{t_{1-\frac{\alpha}{2}}\sigma}{E_{\mu}^{*}}\right)^{2}$$

- Ou

$$n \geq \left(t_{1-\frac{\alpha}{2}}\frac{s}{E_{\mu}^{*}}\right)^{2} \Rightarrow si\ \sigma^{2}\ est\ inconnue$$

Estimation par intervalle de confiance

III- Intervalle de Confiance pour la variance: $I_C(s2)$

$$I_{\sigma^2} = \left[\frac{(n-1)s^2}{k_1} ; \frac{(n-1)s^2}{k_2}\right]$$

Avec: $P\left(\chi^2_{(n-1)} > k_1\right) = \frac{\alpha}{2}; P\left(\chi^2_{(n-1)} > k_2\right) = \frac{\alpha}{2}$