INSTITUT DE FORMATION ET DE RECHERCHE INTERDISCIPLINAIRES EN SCIENCES DE LA SANTÉ ET DE L'ÉDUCATION

(IFRISSE)

Bases en probabilités

ILBOUDO Wendyam Fulbert
Statisticien démographe

Ce cours est adapté de celui de Ter Tiero Elias DAH MD,PhD

Objectifs

1. Définir une probabilité
2. Connaître les règles de calcul de probabilités
3. Définir et comprendre les concepts de:
   - variable aléatoire
   - variable aléatoire discrète
   - Variable aléatoire continue
4. Connaître les lois de distribution de probabilité
   - Discrète
   - Continue

I. Introduction

Position du problème

Dans la pratique courante, l'étude directe d'une population est souvent impossible, parce que:

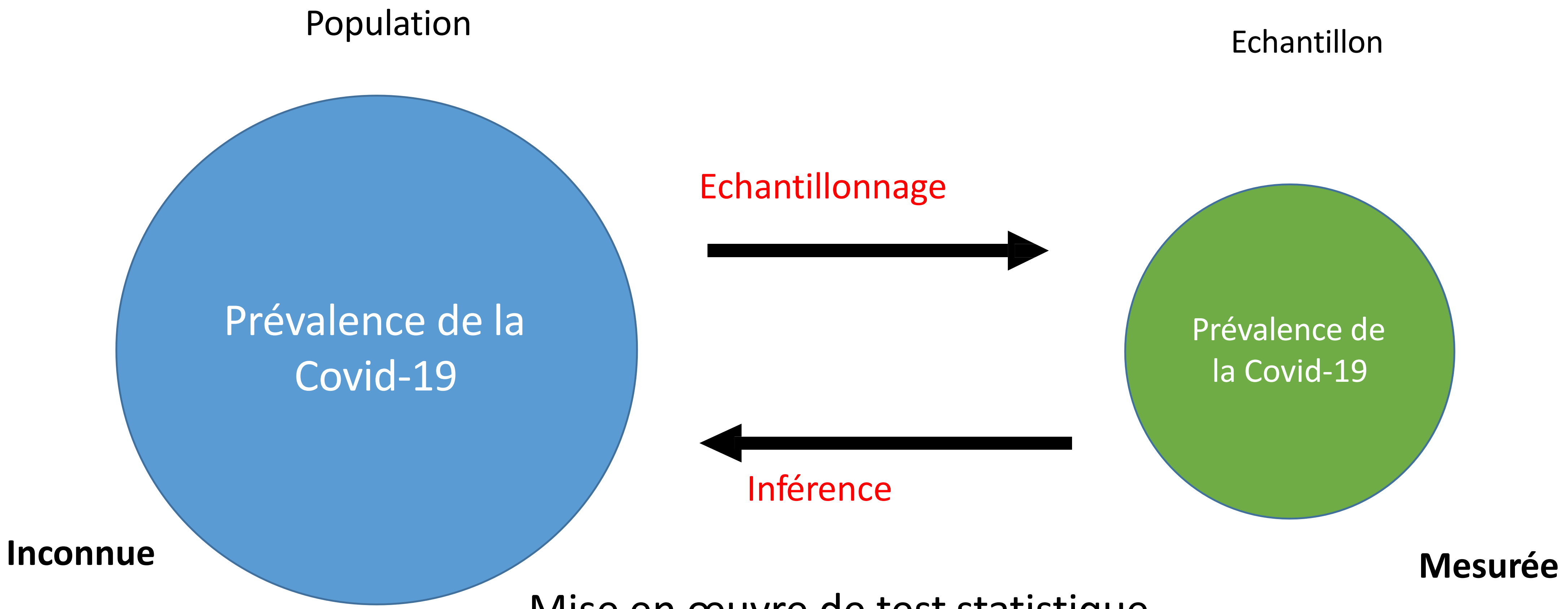
- Taille de la population inconnue
- Taille connue mais trop grande pour permettre une étude exhaustive

Ex: On veut connaître **la prévalence de la Covid-19** (PCR positive au Sars-CoV-2) dans la population de Ouahigouya.

Il est évident que pour des questions logistiques et de coûts, il ne sera pas possible de considérer la totalité des habitants de Ouahigouya pour une telle étude.

Un moyen pratique et efficace d'y parvenir consiste à mettre en œuvre une démarche qui consistera :

- extraire de la population un échantillon aléatoire représentatif (échantillonnage)
- mesurer la Covid-19 (infection par le Sars-CoV-2, et très certainement d'autres variables jugées pertinentes comme l'âge, le sexe, le poids, la taille, etc., chez les sujets de l'échantillon (recueil des données)
- décrire les variables recueillies (statistique descriptive)
- induire ou étendre les résultats observés dans l'échantillon sur la population (statistique inférentielle)



Mise en œuvre de test statistique nécessitant des calculs de **probabilités**

II. Probabilités

Probabilité (1)

La notion intuitive de probabilité d'un événement nous est sûrement plus familière que **la théorie des probabilités**. En général, nous comprenons assez facilement que la probabilité d'un événement traduit les chances de cet événement de se produire, que cette probabilité varie entre 0 et 1 (ou entre 0% et 100%), qu'un événement de probabilité nulle n'a aucune chance de se produire, qu'un événement de probabilité 1 a 100% de chances de se produire.

La notion de probabilité intervient lorsque nous sommes dans des situations d'incertitude. Quelle chance avons-nous de réussir à un examen ? Une personne a un bus à prendre à 11h, elle doit auparavant passer à la banque puis aller poster un colis à la poste, quelle est la probabilité qu'elle puisse prendre son bus en sortant de chez elle à 9h ? Dans ces deux cas nous ne pouvons pas fournir une réponse avec certitude, car elle peut dépendre de facteurs que nous ne maîtrisons pas totalement. Par contre, nous pouvons énumérer dans chaque cas, les différents résultats possibles : réussite ou échec pour le premier cas, prendre son bus à 11h ou le rater pour le second cas.

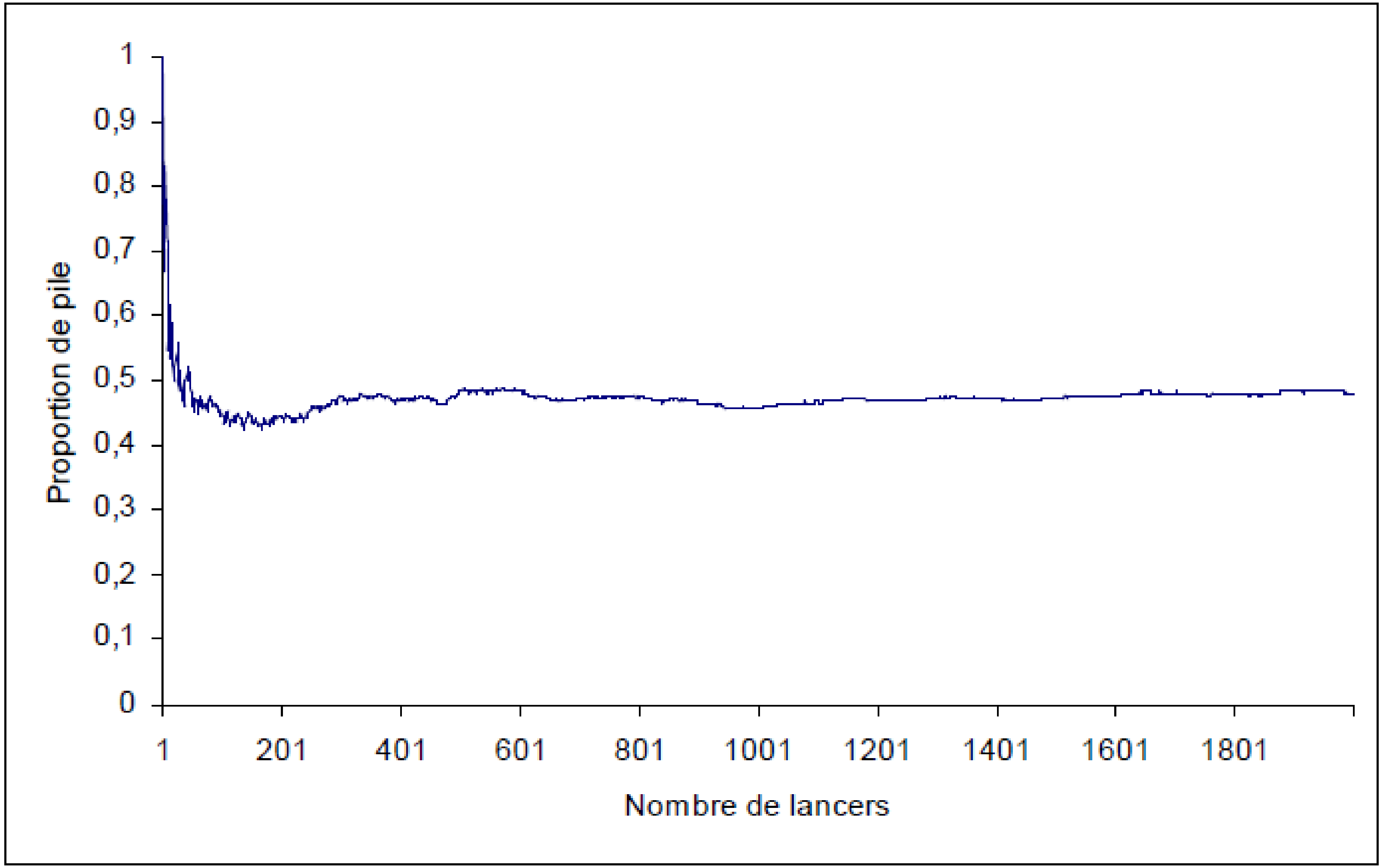
Dans le langage de tous les jours, nous utilisons les termes « probabilité », « il est probable », « probablement » sans y accorder une signification particulière. En statistique, le mot « probabilité » est un terme technique qui est défini et utilisé dans un sens bien particulier en relation avec des événements et des expériences qui entraînent une part d'incertitude. La théorie des probabilités nous permet d'établir des plans d'expériences aléatoires pour effectuer un choix entre les différents résultats possibles de l'expérience.

Probabilité (2)

Dans une expérience aléatoire, l'inventaire des résultats possibles est connu sans que l'on sache évidemment quel sera le résultat finalement obtenu, et à chacun des résultats est attribuée une chance ou une probabilité d'apparition.

Par exemple, si on lance une pièce de monnaie une fois, les résultats possibles sont pile ou face et la probabilité d'obtenir pile est 1/2. La question que l'on peut se poser est : pourquoi 1/2 et pas 1/3 ? Pour répondre à cette question, considérons une expérience qui consiste à lancer la pièce un grand nombre de fois. On peut observer que la **fréquence relative** de l'obtention de pile (nombre de résultats pile sur nombre de lancers) varie de façon relativement stable autour de 1/2 au fur et à mesure que l'on augmente le nombre de lancers. Plus le nombre de lancers est élevé, plus la fréquence relative de pile sera proche de la probabilité 1/2, comme on peut le voir dans le graphique 1, et serait d'ailleurs égale à cette probabilité si on pouvait lancer la pièce un nombre infini de fois (c'est ce que l'on appelle en mathématique, la **loi des grands nombres** qui est une notion primordiale).

Proportion de piles obtenues en lançant une pièce de monnaie en fonction du nombre de lancers



Probabilité d'évènement

- est la proportion de fois où cet événement se produirait si on répétait à l'infini les conditions où il peut se produire.
- Est une expression quantifiée de la prévision de sa survenue.
- Elle est définie en termes de **fréquence relative.**
- Une probabilité se situe toujours entre 0 et 1 (ou 0 et 100 en pourcentage).
  - Pr (A ne se réalise pas) = 1 – P(A se réalise)

$$\textbf{Pr (A)=} \frac{nombre\ de\ fois\ où\ A\ se\ produit}{nombre\ total\ de\ fois\ où\ A\ peut\ se\ produire}$$

Exemple fil rouge

Lors d’un mariage, 158 personnes sont invitées. Plusieurs mets sont proposés et tous sont consommés par les convives. Cependant, 99 cas d’intoxication alimentaire à type de diarrhées + vomissements sont observés.

La probabilité d’intoxication d’un convive choisi au hasard parmi les 158 est :

La probabilité de ne pas observer d’intoxication alimentaire est :

<u>Exemple fil rouge</u>

Lors d'un mariage, 158 personnes sont invitées. Plusieurs mets sont proposés et tous sont consommés par les convives. Cependant, 99 cas d'intoxication alimentaire à type de diarrhées + vomissements sont observés.

La probabilité d'intoxication d'un convive choisi au hasard parmi les 158 est :

$\frac{99}{158}$**= 0,63 ou 63%.**

La probabilité de ne pas observer d'intoxication alimentaire est :

1 – 0,63 = 0,27 ou 27%.

Probabilité (3)

Chaque résultat possible est une éventualité ; l'ensemble des éventualités constitue l'**espace fondamental** *E*.
Dans la suite, on notera $E=\{e_1, e_2, \ldots, e_n\}$ un espace fondamental constitué de n éventualités.
Un **événement** est un sous-ensemble de l'espace fondamental ; il est constitué d'éventualités et lorsqu'il est réduit à une seule éventualité, il est appelé événement élémentaire.
On dit qu'un événement s'est produit si le résultat de l'expérience appartient à l'ensemble définissant l'événement.

A titre d'exemple, supposons que l'on s'intéresse au statut matrimonial dans une population et considérons une expérience qui consiste à choisir une personne au hasard dans cette population. L'espace fondamental associé à cette expérience est donné par *E* = {*célibataire, vivant en couple, marié, séparé, divorcé, veuf*}. Chaque éventualité définit un événement élémentaire. L'événement « être ou avoir été marié » est défini par le sous-ensemble *A* = {*marié, séparé, divorcé, veuf*} ; il est réalisé si la personne choisie est soit mariée, soit séparée, soit divorcée, soit veuve.

Probabilité (4)

Probabilité conditionnelle

Il s ’agit de la probabilité qu’un événement A se réalise sachant qu’un événement B s’est déjà réalisé.

Elle est noté **Pr(A/B)** et est définie comme suit :

$$\textbf{Pr (A/B)=} \quad \frac{nombre\ de\ fois\ où\ A\ et\ B\ se\ réalisent\ conjointement}{nombre\ de\ fois\ où\ B\ se\ réalise}$$

<u>Revenons à l’exemple fil rouge :</u>

Parmi les nombreux mets proposés, il y’ avait du riz gras. 133 convives déclarent en avoir consommé. Parmi eux, 97 ont eu des signes d’intoxication alimentaire.

Pr (intoxication/ayant consommé le riz gras) $=$

NB: 2 événements A et B sont **indépendants** si la réalisation de l’un n’influe pas sur celle de l’autre.

Dans un tel cas, **Pr (A/B)= Pr (A)**

Probabilité (5)

Probabilité conditionnelle

Il s ’agit de la probabilité qu’un événement A se réalise sachant qu’un événement B s’est déjà réalisé.

Elle est noté **Pr(A/B)** et est définie comme suit :

$$\textbf{Pr (A/B)=} \quad \frac{nombre\ de\ fois\ où\ A\ et\ B\ se\ réalisent\ conjointement}{nombre\ de\ fois\ où\ B\ se\ réalise}$$

<u>Revenons à l’exemple fil rouge :</u>

Parmi les nombreux mets proposés, il y’ avait du riz gras. 133 convives déclarent en avoir consommé. Parmi eux, 97 ont eu des signes d’intoxication alimentaire.

Pr (intoxication/ayant consommé le riz gras) $= \frac{\mathbf{97}}{\mathbf{133}}$ **= 0,73**

NB: 2 événements A et B sont **indépendants** si la réalisation de l’un n’influe pas sur celle de l’autre.

Dans un tel cas, **Pr (A/B)= Pr (A)**

Probabilité (6)

Evénements complexes

Ils sont définis par des combinaisons particulières, par exemple A et B, noté $A \cap B$, ou des alternatives particulières A ou B noté $A \cup B$.

Pr ($A \cap B$) **= probabilité que A et B se réalisent conjointement (ou simultanément)**

Si A et B ne peuvent se réaliser conjointement, on dira qu'ils sont **incompatibles** ou mutuellement exclusifs.

Exemple

Dans un lancer de dé, l'espace fondamental est défini par $E = \{1, 2, 3, 4, 5, 6\}$; l'événement « sortie d'un nombre pair » est le sous-ensemble noté $A = \{2, 4, 6\}$ et l'événement « sortie du 5 » est l'événement élémentaire $B = \{5\}$. Les deux événements A et B sont incompatibles.

On peut aussi s'intéresser à la probabilité complexe d'un événement défini par la réalisation de A seul, de B seul, ou des 2 = **Pr (**$A \cup B$**).**

Le calcul efficace de probabilités associées à un événement fait appel aux règles multiplicatives & additives

Probabilité (7)

Règle multiplicative

- Applicable si A et B sont des événements indépendants

$$\mathbf{Pr}\ (A \cap B) = \mathbf{Pr\ (A) \times Pr\ (B)}$$

<u>Exemple :</u> Les effets secondaires d'un médicament surviennent chez 10% des patients. Si un médecin traite 2 patients avec ce médicament, quelle est la probabilité pour ces 2 patients d'avoir conjointement des effets secondaires?

On peut supposer que dans cet exemple, les 2 évènements sont indépendants. En effet, il est peu probable que le développement d'effets secondaires chez un des patients puisse affecter la survenue d'effets secondaires chez l'autre.

Pr (d'apparition d'effets secondaires chez les 2 patients) =

Probabilité (8)

Règle multiplicative

- Applicable si A et B sont des événements indépendants

$$\mathbf{Pr}\ (A \cap B) = \mathbf{Pr\ (A)\ x\ Pr\ (B)}$$

<u>Exemple :</u> Les effets secondaires d'un médicament surviennent chez 10% des patients. Si un médecin traite 2 patients avec ce médicament, quelle est la probabilité pour ces 2 patients d'avoir conjointement des effets secondaires?

On peut supposer que dans cet exemple, les 2 évènements sont indépendants. En effet, il est peu probable que le développement d'effets secondaires chez un des patients puisse affecter la survenue d'effets secondaires chez l'autre.

Pr (d'apparition d'effets secondaires chez les 2 patients) = 0,1 x 0,1 = 0,01

Probabilité (9)

Règle additive

- Applicable si l’événement complexe peut être réalisé de plusieurs manières
  - S’il s’agit de la réalisation de A ou B et que ceux-ci sont compatibles

$$\boxed{\mathbf{Pr\ (A \cup B) = Pr\ (A) + Pr\ (B) - Pr\ (A \cap B)}}$$

  - Si A et B sont incompatibles

$$\boxed{\mathbf{Pr\ (A \cup B) = Pr\ (A) + Pr\ (B)}}$$

<u>Exemple</u> (sur les effets secondaires). Quelle est la probabilité qu’au moins un des 2 patients ait des effets secondaires?

Ces 2 événements n’étant pas incompatibles

Pr (au moins un des 2 patients a un effet secondaire) =

Probabilité (10)

Règle additive

- Applicable si l'événement complexe peut être réalisé de plusieurs manières
  - S'il s'agit de la réalisation de A ou B et que ceux-ci sont compatibles

$$\boxed{\mathbf{Pr\ (A \cup B) = Pr\ (A) + Pr\ (B) - Pr\ (A \cap B)}}$$

  - Si A et B sont incompatibles

$$\boxed{\mathbf{Pr\ (A \cup B) = Pr\ (A) + Pr\ (B)}}$$

<u>Exemple</u> (sur les effets secondaires). Quelle est la probabilité qu'au moins un des 2 patients ait des effets secondaires?

Ces 2 événements n'étant pas incompatibles

Pr (au moins un des 2 patients a un effet secondaire) = 0,1 + 0,1 – 0,01 = 0,19

Quelques notations en probabilité

E	est l'événement certain, toujours réalisé ;
$\varnothing$	désigne l'événement impossible, jamais réalisé ;
$A \subset B$	signifie que la réalisation de A entraîne celle de B ;
$A \cup B$	correspond à la réalisation de A, de B ou des deux ;
$A \cap B$	correspond à la réalisation simultanée de A et de B ;
$\overline{A}$	est l'événement contraire à A ;
$A \cap B = \varnothing$	signifie que les deux événements A et B sont incompatibles.

Probabilité (12)

Exemple

Considérons une grande population d'hommes qui ont été classés en fonction du statut tabagique (fumeur ou non fumeur) et d'une maladie respiratoire chronique (malade ou non malade). Dans cette population, on sait que 5% des hommes ont une maladie respiratoire chronique (MRC) et ne fument pas, 15% ont une MRC et fument, 50% n'ont pas une MRC et ne fument pas, et 30% n'ont pas une MRC et fument. On choisit un homme au hasard dans la population et on considère les événements suivants:

A = {*l'homme choisi a une MRC et fume*}
B = {*l'homme choisi a une MRC et ne fume pas*}
C = {*l'homme choisi n'a pas une MRC et fume*}
D = {*l'homme choisi n'a pas une MRC et ne fume pas*}

Ici, l'ensemble fondamental est $E = \{A, B, C, D\}$, avec $\Pr[E] = 1$ (Pr pour probabilité).
De plus $\Pr[A] = 0{,}15$, $\Pr[B] = 0{,}05$, $\Pr[C] = 0{,}30$, $\Pr[D] = 0{,}50$. Aucun des événements *A*, *B*, *C* ou *D* ne peut se réaliser en même temps lorsqu'un homme est choisi, ils sont donc incompatibles deux à deux, et par exemple $\Pr[A \cup B] = 0{,}15 + 0{,}05 = 0{,}20$.

III. Variables aléatoires

Variable aléatoire (1)

Une variable dont la valeur est déterminée en fonction du résultat d'une expérience aléatoire est appelée **variable aléatoire**. Une variable aléatoire traduit donc une situation liée au hasard dans laquelle les résultats possibles sont des nombres. Les variables quantitatives prennent des valeurs qui sont déjà des nombres. Les variables qualitatives prennent des modalités (comme par exemple homme ou femme pour la variable sexe). La notion de variable aléatoire correspond au codage de ces modalités à l'aide de nombres.
Une variable aléatoire permet donc d'établir une correspondance entre les résultats d'une expérience aléatoire et des valeurs numériques.

Par exemple dans le lancer d'une pièce de monnaie, on peut considérer la correspondance suivante : X(pile) = 1 et X(face) = 0. La variable aléatoire prend alors la valeur 0 ou 1, et on dira alors que X = 1 représente l'événement « obtenir pile » et X = 0 représente l'événement « obtenir face ». De plus, à chaque valeur de X est attribuée une probabilité : Pr[X = 1] = Pr[X = 0] = 1/2.
Si on considère l'expérience qui consiste à choisir au hasard un sujet *s* dans une population, on peut définir les variables aléatoires suivantes :

- X prenant la valeur 1 si *s* est malade, et 0 sinon ;
- Y égale à la taille du sujet *s* ;
- Z égale au poids du sujet *s*.

Variable aléatoire (2)

Pour éviter toute confusion, on note en général par des lettres majuscules les variables aléatoires, et par des lettres minuscules les valeurs prises par les variables aléatoires. Par exemple, la variable aléatoire X (résultat du lancer d'une pièce de monnaie) prend les valeurs x = 0 ou x = 1.

On doit distinguer les variables aléatoires dites discrètes de celles qualifiées de continues, car leur traitement mathématique est différent :

- une **variable aléatoire discrète** prend un ensemble fini ou infini dénombrable de valeurs (ex. : sexe, statut matrimonial, nombre d'enfants dans une famille) ;
- une **variable aléatoire continue** prend un ensemble infini non dénombrable de valeurs (ex. : taille, poids, pression artérielle).

La **loi de probabilité** ou **distribution de probabilité** d'une variable aléatoire est la donnée de la probabilité de survenue de chaque événement possible,
Elle peut être caractérisée par des nombres: **l'espérance mathématique ou moyenne, la variance**. Celles-ci permettent de la « résumer ».

Variable aléatoire (3)

Loi de probabilité d'une variable aléatoire discrète

La loi de probabilité d'une variable aléatoire discrète est la donnée d'un nombre fini ou infini dénombrable de couples (xi, pi), où xi désigne un résultat possible et pi la probabilité de l'événement élémentaire associé à ce résultat.

Les nombres pi = Pr[X = xi] sont compris entre 0 et 1 et leur somme est égale à 1.

La **moyenne**, appelée aussi espérance mathématique et notée *E(X)*, d'une variable aléatoire discrète *X* est définie par

$$\mu = E(X) = \sum_i x_i p_i$$

Cette notion correspond à la valeur moyenne de *X* d'un point de vue probabiliste : c'est ce à quoi on doit s'attendre « en moyenne » lorsqu'on réalise un grand nombre de tirages de valeurs particulières de *X*.

La **variance** d'une variable aléatoire discrète est définie par :

$$\sigma^2 = Var(X) = E(X-\mu)^2 = \sum_i (x_i - \mu)^2 p_i \,.$$

L'écart type s est la racine carrée de la variance. La mesure de l'écart type représente l'ampleur de la déviation par rapport à la moyenne.

Variable aléatoire (4)

Loi de probabilité d'une variable aléatoire discrète

Exemple :
Si on s'intéresse au résultat *X* du lancer d'un dé non pipé, les différentes valeurs prises par *X* sont : xi=i, i=1, 2, …, 6, avec les probabilités Pr[X=xi]=1/6 quel que soit i.

L'espérance mathématique est donc égale à

et la variance est égale à

Variable aléatoire (5)

Loi de probabilité d'une variable aléatoire discrète

Exemple :
Si on s'intéresse au résultat *X* du lancer d'un dé non pipé, les différentes valeurs prises par *X* sont : $x_i=i$, $i=1, 2, \ldots, 6$, avec les probabilités $Pr[X=x_i]=1/6$ quel que soit i.

L'espérance mathématique est donc égale à

$$\mu = 1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=3{,}5$$

et la variance est égale à

$$\sigma^2 = (1-3{,}5)^2\times\frac{1}{6}+(2-3{,}5)^2\times\frac{1}{6}+(3-3{,}5)^2\times\frac{1}{6}+(4-3{,}5)^2\times\frac{1}{6}+(5-3{,}5)^2\times\frac{1}{6}+(6-3{,}5)^2\times\frac{1}{6}$$

$$=\frac{17{,}5}{6}=2{,}92$$

Variable aléatoire (6)

Loi de probabilité d'une variable aléatoire continue

L'application de la définition précédente des lois de probabilité aux variables aléatoires continues pose un problème. En effet, dans le cas d'une variable aléatoire continue, l'ensemble des valeurs possibles est infini et non dénombrable ; on ne peut donc pas préciser les probabilités pi = Pr[X = xi] pour toutes les valeurs. Il faut alors considérer les probabilités que X appartienne à des intervalles de nombres réels tels que] -∞, x] ou [x,+ ∞ [, ou encore [x_1 , x_2].

La définition de la loi de probabilité d'une variable continue repose de façon équivalente sur la notion de fonction de densité de probabilité ou celle de fonction de répartition ; l'une comme l'autre permettent de calculer les probabilités des intervalles définis précédemment.

L'espérance mathématique et la variance peuvent également être calculées ; le sens de ces quantités est la même que dans le cas des variables aléatoires discrètes.

Définition : Une **fonction de répartition** $F(x)$ d'une variable aléatoire absolument continue X relie cette variable à tout nombre réel x de sorte que : $F(x) = \Pr[X \leq x]$.

C'est obligatoirement une fonction numérique non décroissante, continue et telle que :

$F(-\infty) = 0$ et $F(+\infty) = 1$.

Variable aléatoire (7)

Loi de probabilité d'une variable aléatoire continue

F(x) donne la probabilité d'observer une valeur de la variable X inférieure à x. Les événements considérés dans la définition de *F(x)* sont donc de la forme]-∞, *x*] , pour tout *x*, mais la probabilité que *X* appartienne à un intervalle quelconque de la forme $[x_1 , x_2]$ se calcule aisément puisque :

$$\Pr[x_1 \leq X \leq x_2] = \Pr[X \leq x_2] - \Pr[X \leq x_1] \quad = F(x_2) - F(x_1)\ .$$

En particulier, si on pose $x_1 = x$ et $x_2 = x+dx$, $\Pr[x \leq X \leq x + dx] = F(x + dx) - F(x)$, et si l'on choisit *dx* de plus en plus petit, on obtient à la limite :

$$\lim_{dx \to 0} \Pr[x \leq X \leq x + dx] = \lim_{dx \to 0} F(x + dx) - F(x) = 0\ .$$

Ce qui montre que la probabilité attribuée à chaque point *x* est nulle, car il est impossible de tomber exactement sur cette valeur. Par exemple, aucune personne ne pèse exactement 60,7234013218… kilogrammes. Par conséquent, on peut écrire indifféremment

$$\mathbf{\Pr[x_1 \leq X \leq x_2] = \Pr[x_1 \leq X < x_2] = \Pr[x_1 < X < x_2]}$$

$$\mathbf{\Pr\ [X \geq x_1] = 1 - P[X < x_1]}$$

Variable aléatoire (8)

Loi de probabilité d'une variable aléatoire continue

Définition : La **fonction de densité** de probabilité *f(x)* au point *x* est définie par $f(x) = \lim_{dx \to 0} \frac{\Pr[x \leq X \leq x + dx]}{dx}$.

Intuitivement, on peut donc considérer que dans un intervalle de longueur *dx*, *f(x)dx* joue le rôle des pi définies pour les variables aléatoires discrètes.

Il faut noter que la fonction de densité de probabilité n'est pas une probabilité, elle peut par exemple prendre des valeurs supérieures à 1. Par contre l'aire sous la courbe d'une densité de probabilité est égale à 1, ce qui se traduit par la propriété mathématique suivante :

$$\int_{-\infty}^{+\infty} f(x)dx = 1,$$

qui se lit « **intégrale de la fonction *f(x)* entre** -∞ **et** + ∞ **égale à 1** ». La fonction de densité de probabilité *f(x)* correspond à la dérivée de la fonction de répartition *F(x)*. Donc, au même titre que la fonction de répartition *F(x)*, la fonction de densité de probabilité permet de calculer des probabilités :

$$\Pr[x_1 \leq X \leq x_2] = \int_{x_1}^{x_2} f(x)dx.$$

La **moyenne** d'une variable aléatoire continue est définie par : $\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx$

La **variance** d'une variable aléatoire continue est calculée par : $\sigma^2 = Var(X) = E(X-\mu)^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$

Loi de probabilité d'une variable aléatoire continue

Définition : La **fonction de densité** de probabilité *f(x)* au point *x* est définie par :

$$f(x) = \lim_{dx \to 0} \frac{\Pr[x \le X \le x + dx]}{dx}.$$

Intuitivement, on peut donc considérer que dans un intervalle de longueur *dx*, *f(x)dx* joue le rôle des pi définies pour les variables aléatoires discrètes.

Il faut noter que la fonction de densité de probabilité n'est pas une probabilité, elle peut par exemple prendre des valeurs supérieures à 1. Par contre l'aire sous la courbe d'une densité de probabilité est égale à 1, ce qui se traduit par la propriété mathématique suivante :

$$\int_{-\infty}^{+\infty} f(x)dx = 1,$$

qui se lit « **intégrale de la fonction *f(x)* entre -∞ et + ∞ égale à 1** ». La fonction de densité de probabilité *f(x)* correspond à la dérivée de la fonction de répartition *F(x)*. Donc, au même titre que la fonction de répartition *F(x)*, la fonction de densité de probabilité permet de calculer des probabilités :

$$\Pr[x_1 \le X \le x_2] = \int_{x_1}^{x_2} f(x)dx.$$

La **moyenne** d'une variable aléatoire continue est définie par :

$$\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

La **variance** d'une variable aléatoire continue est calculée par :

$$\sigma^2 = Var(X) = E(X-\mu)^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$$

IV. Lois de probabilités

Quelques distributions théoriques

Position du problème

- Toute variable aléatoire suit une loi de probabilité
- Cependant, il y a une méconnaissance de la distribution exacte des variables dans une population ou un échantillon
- Nécessité de se référer à un modèle: « loi de distribution de probabilités »

Existence de plusieurs lois

- Lois de probabilité discrètes
  - Loi de Bernoulli
  - Loi binomiale
  - Loi de poisson
- Lois de probabilité continues :
  - Loi de Laplace-Gauss ou loi normale
  - Loi dérivées de loi normale (loi de Khi2, loi de Student..)

Lois de probabilités discrètes

Loi de Bernouilli

On dit qu'une variable aléatoire *X* suit une loi de Bernoulli de paramètre π , si *X* prend deux valeurs **1** ou

0 avec les probabilités π = Pr[*X* =1] et 1- π = Pr[*X* = 0] .

La valeur *X* = 1 correspond à l'événement « succès » ou à l'événement qu'on étudie (par exemple malade) et *X* = 0 à « échec » (ou

On a les propriétés suivantes :

Soit X une variable aléatoire qui suit une loi de Bernoulli de paramètre π , et on note $X \sim Bernoulli(\pi)$

$$E(X) = \Pr(X = 1) = \pi$$
$$Var(X) = \pi(1 - \pi)$$

On peut alors noter que l'espérance mathématique d'une variable de Bernoulli est égale à la probabilité de l'événement « succès ». La loi de Bernoulli intervient à chaque fois qu'on étudie une variable binaire (malade-non malade, exposé-non exposé, décédé-vivant) et elle joue donc un rôle important en épidémiologie.

Exemple : Si on s'intéresse à la prévalence π d'une maladie dans une population, on peut associer à chaque sujet *i* de la population une variable de Bernoulli *Xi*, avec π = Pr[*Xi* =1] la probabilité que le sujet *i* soit malade.

Lois de probabilités discrètes

Loi binomiale (1)

Une variable binomiale est définie comme la somme de ***n* variables aléatoires de Bernoulli indépendantes et de même paramètre π .** Elle correspond au nombre de « succès » après *n* épreuves de Bernoulli indépendantes de même paramètre π.

Supposons qu'on s'intéresse à une population de taille infinie dans laquelle il y a une proportion p d'individus possédant un certain caractère (ex. : malade, fumeur, gaucher), et donc une proportion (1- π) ne la possédant pas (ex. : non malade, non fumeur, droitier). Si on choisit au hasard *n* sujets dans cette population, le nombre *X* de sujets possédant le caractère est une variable aléatoire qui peut prendre les valeurs *0, 1, …, n* et sa loi est la loi binomiale de paramètres ***n*** et **π** , notée B(n, π). Il est important de noter que si la taille de la population est finie, le tirage au sort des sujets doit se faire avec remise pour satisfaire à cette loi.

En résumé, la variable aléatoire binomiale B(n, π) correspond au nombre de sujets possédant un certain caractère dans un échantillon de n sujets tirés au sort au sein d'une population dans laquelle une proportion p de sujets possèdent le caractère ; le tirage au sort se faisant sans remise si la population est de taille infinie, et avec remise si la population est de taille finie.

Loi binomiale (2)

Soit X une variable aléatoire qui suit une loi binomiale $B(n, \pi)$, alors

$$\Pr[X = k] = C_n^k \pi^k (1-\pi)^{n-k} \text{, avec } k = 0,1,2,\ldots,n$$

$$\mu = E(X) = n\pi$$

$$\sigma^2 = Var(X) = n\pi(1-\pi)$$

Le nombre C_n^k correspond au nombre de manières de choisir k éléments (correspondant aux succès) parmi n, et se calcule par la formule : $C_n^k = \dfrac{n!}{k!(n-k)!}$ où $n!$ se lit « factorielle n » et se calcule par : $n!$=1´ 2´3´...´n .
Par convention, 0 ! = 1.

Exemple : Un couple décide d'avoir 3 enfants. En supposant que la probabilité d'avoir un garçon est égale à 0,51, quelle est la probabilité que ce couple ait une fille et deux garçons ?

On s'intéresse alors au nombre *X* de garçons sur 3 naissances. La solution est donnée en considérant que *X* suit une loi binomiale de paramètres n = 3 et π = 0,51 et en calculant :

$$\Pr[X = 2] = C_3^2 (0,51)^2 (1-0,51)^1 = 0,382$$

Loi binomiale (3)

L'expression $\Pr[X = k] = C_n^k \pi^k (1-\pi)^{n-k}$ est à la base du calcul de probabilités d'une variable binomiale ; elle est facilement calculable lorsque n est petit, par exemple n plus petit que 10. Quand n est plus grand, le calcul devient plus ardu et demande plus d'efforts. Pour des valeurs modérées de n, inférieures à 20 ou 25, il existe des tables de la loi binomiale qui donne les valeurs de Pr[*X* = *k*] en fonction des valeurs de n, π et *k* souhaitées. Lorsque n est grand, supérieur à 25 ou 30, on peut utiliser les approximations par la loi de Poisson ou la loi normale, que nous allons définir plus loin.

Lois de probabilités discrètes

Loi de poisson

- Utilisée lorsque l'on étudie un phénomène rare dans certaines conditions
  - Ex: Survenue du cancer de la prostate chez des hommes jeunes, âgés de moins de 50 ans

- Soit ʌ un réel strictement positif et *X* une variable aléatoire. On dit que *X* suit la loi de Poisson de paramètre ʌ, et on note X ~ P (ʌ), lorsque:
  - l'ensemble des valeurs prises par X est l'ensemble de tous les entiers naturels X (Ω) = N
  - Pour tout $k \in \mathbb{N}$, $P(X = k) = e^{-\lambda} \dfrac{\lambda^k}{k!}$.
- Implications de la définition

Soit $\lambda \in \mathbb{R}_+^*$ et X une variable aléatoire.
Si X suit une loi de Poisson de paramètre λ, alors :

$$E(X) = \lambda \qquad V(X) = \lambda \qquad \sigma_X = \sqrt{\lambda}$$

On peut utiliser une **table de la loi de Poisson** pour déterminer les nombres P(*X*=*k*) pour différentes valeurs de ʌ et de *k*.

Loi normale ou loi de Laplace-Gauss (1)

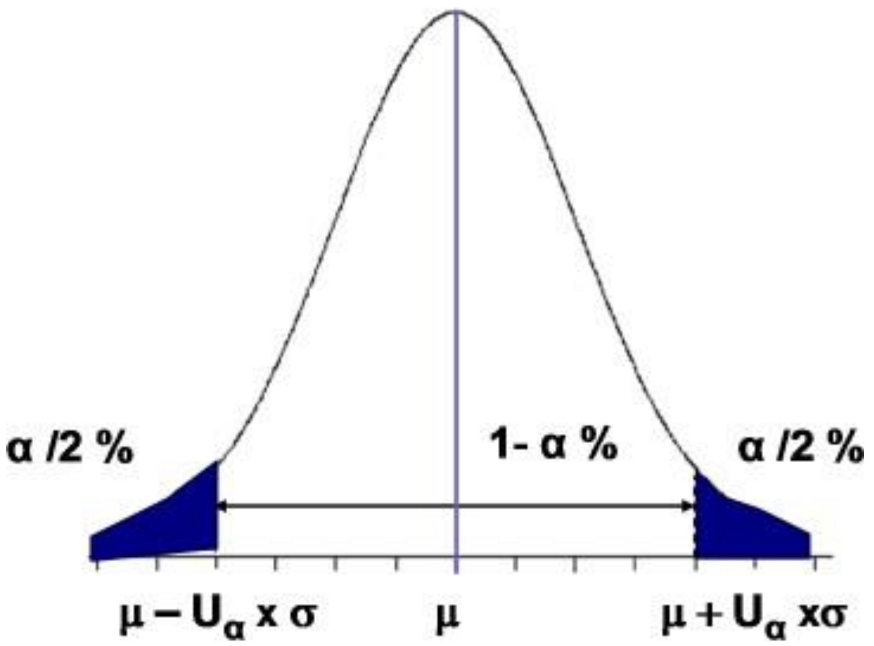
La distribution de Laplace-Gauss est **incontournable en statistique**. Les applications de cette loi sont si répandues qu'elle mérite bien le nom de « loi normale » qui lui est donné.

La loi normale s'applique à des **variables aléatoires continues** pouvant prendre toutes les valeurs réelles possibles entre -∞ et + ∞. Elle est entièrement définie par deux paramètres, la moyenne μ et la variance σ^2 . On dira qu'une variable aléatoire continue *X* suit une loi normale de paramètres μ et σ^2, et on note $X \sim N(\mu,\sigma^2)$, si sa fonction de densité *f(x)* est définie par l'expression mathématique :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Ici π représente le nombre égal à 3,141592… Le tracé de cette fonction correspond à la fameuse courbe en cloche, unimodale et symétrique autour de la moyenne μ comme le montre le graphique suivant.

Loi normale ou loi de Laplace-Gauss (2)



α	u_α
20%	1,28
10%	1,65
5%	1,96
2%	2,33
1%	2,58
0,1%	3,3

Si α =5%, alors 95% des valeurs de la VA X sont comprises sous la cloche, entre **μ - 1,96σ** **et** **μ+1,96σ**

On voit bien que le tracé de la fonction de densité f(x) d'une variable *N* (μ,σ²) dépend de la moyenne μ et de celle de l'écart-type σ.

La fonction de répartition *F(x)* correspondante est la probabilité que la variable aléatoire *X* ait une valeur inférieure ou égale à une quantité *x*. Cette fonction est exprimée par :

$$F(x) = \Pr[X \leq x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right] du$$

Lois de probabilités continues

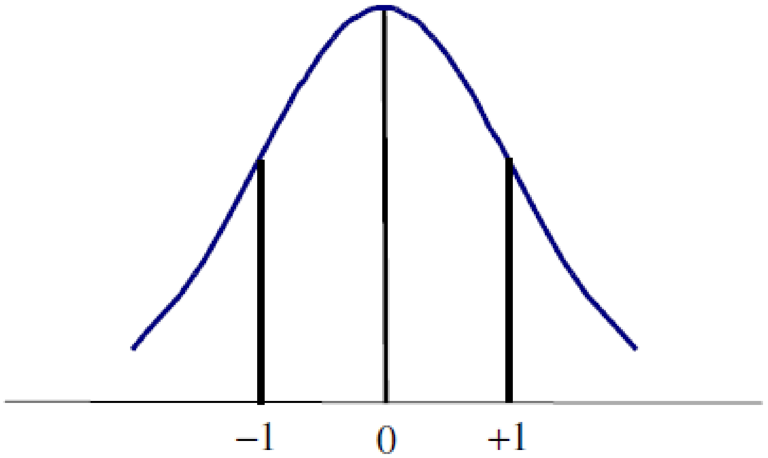
Loi normale ou loi de Laplace-Gauss (3)

Il existe autant de lois normales que de valeurs des paramètres μ et σ^2, mais on peut toujours se ramener par une simple transformation (changement de variable) à la loi normale de moyenne 0 et de variance 1, notée N(0,1) et appelée **loi normale centrée réduite** ou **loi de l'écart-réduit**.

En effet, si $X \sim N(\mu \;\; \sigma^2)$, la nouvelle variable U obtenue par la transformation

$$U = \frac{(X - \mu)}{\sigma}$$

suit une loi normale N(0,1). Le tracé de la fonction de densité de probabilité de la variable aléatoire normale centrée réduite U est donné par le graphique suivant



De façon équivalente, on peut passer d'une variable *U à* une variable $X \sim N(\mu \;\; \sigma^2)$ par la transformation: $X = \mu + \sigma U$

Loi normale ou loi de Laplace-Gauss (4)

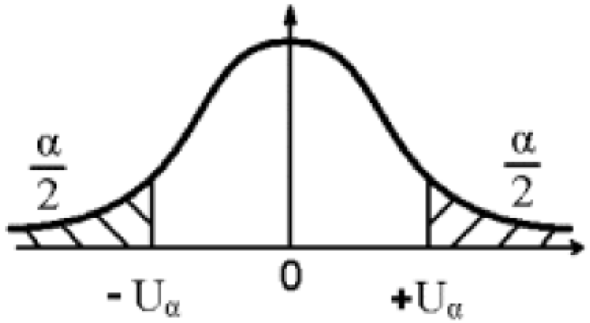
Le calcul de probabilités pour une variable $X \sim N(\mu\ \sigma^2)$ ne peut pas se faire de façon simple et nécessite l'utilisation de méthodes de calculs numériques. Les résultats de ces calculs sont présentés sous forme de tables, appelées « tables de la loi normale ». Fort heureusement, il n'est pas nécessaire d'effectuer les calculs pour toutes les lois normales possibles, il suffit de disposer d'une table pour la loi normale centrée réduite, et grâce à la
transformation $X = \mu + \sigma U$, on peut effectuer des calculs de probabilités pour n'importe quelle loi normale. Nous allons illustrer dans la suite l'utilisation d'une table particulière de la loi $N(0,1)$, appelée « table de l'écart-réduit ».

La table de l'écart-réduit donne pour différentes leurs notées u_α, la probabilité α correspondante d'appartenir à la région grisée sous la courbe (voir table), c'est à dire

$$\alpha = \Pr[U < -u_\alpha] + \Pr[U > u_\alpha] = \Pr[|U| > u_\alpha]$$

Où U suit une loi de $N(0,1)$

Table de la variable normale centrée réduite



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	infini	2,5758	2,3263	2,1701	2,0537	1,9600	1,8808	1,8119	1,7507	1,6954
0,1	1,6449	1,5982	1,5548	1,5141	1,4758	1,4395	1,4051	1,3722	1,3408	1,3106
0,2	1,2816	1,2536	1,2265	1,2004	1,1750	1,1503	1,1264	1,1031	1,0803	1,0581
0,3	1,0364	1,0152	0,9945	0,9741	0,9542	0,9346	0,9154	0,8965	0,8779	0,8596
0,4	0,8416	0,8239	0,8064	0,7892	0,7722	0,7554	0,7388	0,7225	0,7063	0,6903
0,5	0,6745	0,6588	0,6433	0,6280	0,6128	0,5978	0,5828	0,5681	0,5534	0,5388
0,6	0,5244	0,5101	0,4958	0,4817	0,4677	0,4538	0,4399	0,4261	0,4125	0,3989
0,7	0,3853	0,3719	0,3585	0,3451	0,3319	0,3186	0,3055	0,2924	0,2793	0,2663
0,8	0,2533	0,2404	0,2275	0,2147	0,2019	0,1891	0,1764	0,1637	0,1510	0,1383
0,9	0,1257	0,1130	0,1004	0,0878	0,0753	0,0627	0,0502	0,0376	0,0251	0,0125

α	0,001	0,0001	0,00001	0,000001	0,0000001	0,00000001	0,000000001
$U\alpha$	3,2905	3,8906	4,4172	4,8916	5,3267	5,7307	6,1094

Loi normale ou loi de Laplace-Gauss (5)

Comment utiliser la table de la loi normale centrée réduite? (1)

- à l'intérieur du tableau, figurent les valeurs d'une distribution normale centrée réduite (les valeurs dites de $u\alpha$)
- et dans les bandeaux à gauche et au-dessus, figurent les valeurs des probabilités correspondantes α.

On trouve la probabilité α d'observer des valeurs appartenant aux intervalles $]-\infty; -u\alpha[$ et $]+u\alpha; +\infty[$ en additionnant la valeur correspondante de la ligne sur le bandeau à gauche et la valeur correspondante de la colonne sur le bandeau de titre.

Par exemple, comment trouver la probabilité d'être en dessous de $u\alpha = -1$ ou au-dessus de $u\alpha = +1$, soit (–1 écart type) ou (+1 écart type) ?

1ère étape : chercher *uα* à l'intérieur du tableau.

Dans l'exemple, on cherche la valeur « 1 ». Le résultat de la recherche est décevant car la valeur « 1 » (exactement) n'y figure pas. On procède alors par approximation pour trouver la valeur la plus proche. Dans ce cas, c'est 0,9945.

Lois de probabilités continues

Loi normale ou loi de Laplace-Gauss (5)

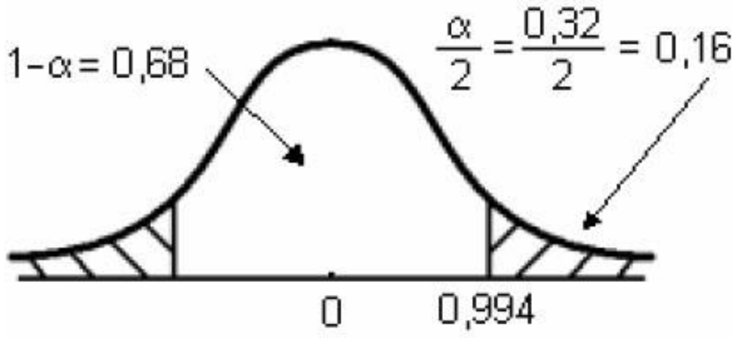
Comment utiliser la table de la loi normale centrée réduite? (2)

2ème étape : chercher la probabilité correspondant à la valeur de *uα* .

Dans l’exemple, on lit : 0,3 en ligne (bandeau de gauche) et 0,02 en colonne (bandeau de titre), la probabilité correspondante α est donc 0,3 + 0,02 = 0,32.

3ème étape : interpréter le résultat obtenu pour α.

Dans l’exemple, pour une variable ayant une distribution normale, la probabilité de s’éloigner d’1 écart type de la moyenne (0) est 0,32 : 0,16 en dessous de « –1 écart type » et 0,16 au-dessus de « +1 écart type », car la courbe est symétrique. Si la distribution théorique de la variable dans la population est normale, on peut toujours dire que la probabilité qu’une valeur soit située entre –1 écart type de la moyenne et +1 écart type de la moyenne est 0,68, car cette probabilité est complémentaire de 0,32 (1 – 0,32 = 0,68).



Lois de probabilités continues

Loi normale ou loi de Laplace-Gauss (6)

Exemple (1)

Dans la population des hommes de 35-40 ans, la concentration moyenne de cholestérol total dans le sang est 1,84 g/l et l'écart type 0,40 g/l. On fait l'hypothèse que la concentration de cholestérol total dans le sang a une distribution normale.

Question : Quelle est la probabilité d'observer une valeur de la concentration de cholestérol total > 2,50 g/l dans le sang si $\mu = 1{,}84$ g/l et $\sigma = 0{,}4$ g/l ?

Réponse : Selon l'hypothèse d'une distribution normale, on peut calculer de combien d'écarts types, la valeur 2,50 g/l est éloignée de la moyenne :

Lois de probabilités continues

Loi normale ou loi de Laplace-Gauss (7)

Exemple (1)

Dans la population des hommes de 35-40 ans, la concentration moyenne de cholestérol total dans le sang est 1,84 g/l et l'écart type 0,40 g/l. On fait l'hypothèse que la concentration de cholestérol total dans le sang a une distribution normale.

Question : Quelle est la probabilité d'observer une valeur de la concentration de cholestérol total > 2,50 g/l dans le sang si $\mu = 1{,}84$ g/l et $\sigma = 0{,}4$ g/l ?

Réponse : Selon l'hypothèse d'une distribution normale, on peut calculer de combien d'écarts types, la valeur 2,50 g/l est éloignée de la moyenne :

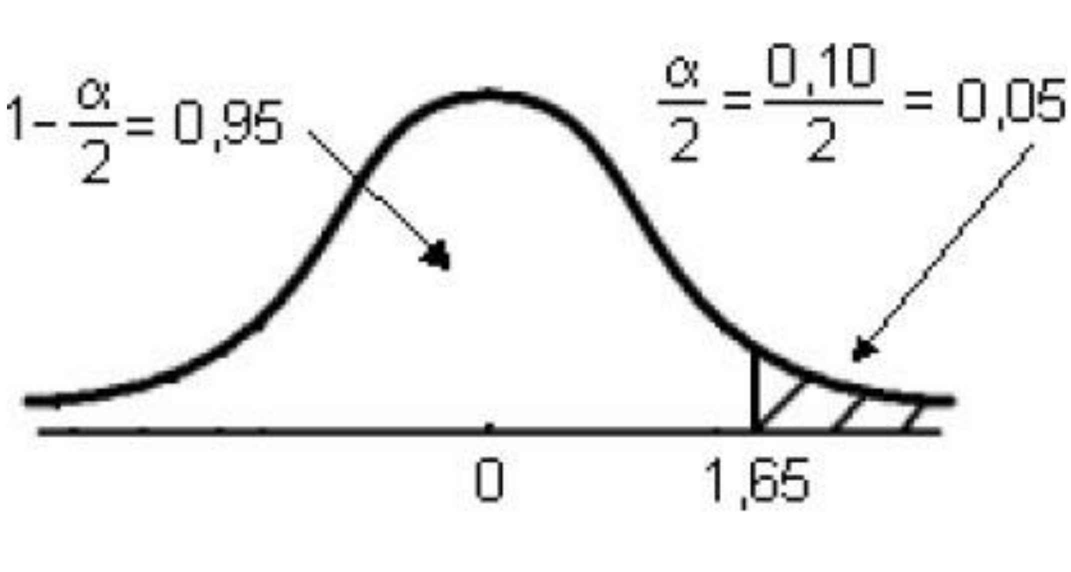
$$u_\alpha = \frac{(2{,}5 - 1{,}84)}{0{,}4} = 1{,}65$$

La valeur 2,50 g/l est éloignée de la moyenne de 1,65 écart-types. Dans la table de la loi normale centrée réduite, on trouve que la probabilité d'être en dehors de 1,65 écarts types (en dessous ou au-dessus) est $\alpha = 0{,}10$. On en déduit que la probabilité de se situer au-dessus de 1,65 écarts types est donc $\alpha/2 = 0{,}10/2 = 0{,}05$. Autrement dit, 5% de la population a, en théorie, une concentration de cholestérol total dans le sang au-dessus de cette valeur de 2,50 g/l.

Lois de probabilités continues

Loi normale ou loi de Laplace-Gauss (8)

Exemple (2)



Loi normale ou loi de Laplace-Gauss ()

Quelques informations importantes concernant la distribution des valeurs d'une variable normale sont résumées dans le tableau suivant :

	Probabilité d'être	
Intervalle	dans l'intervalle	en dehors de l'intervalle
Moyenne ± 1 écart type	0,683	0,317
Moyenne ± 2 écarts types	0,954	0,046
Moyenne ± 3 écarts types	0,997	0,003

On note que 95% environ des valeurs d'une distribution normale se situent dans un intervalle de chaque côté de la moyenne délimité par $\mu - 2$ écarts types et $\mu + 2$ écarts types. Pour être exact, 95% des valeurs d'une distribution normale se situent dans un intervalle de chaque côté de la moyenne délimité par $\mu \pm 1{,}96$ écarts types.

Loi normale ou loi de Laplace-Gauss (10)

Résumé

Si X suit une loi normale, de moyenne μ et d'ecart-type σ, on dit que **X suit N (μ,σ)**

A partir de cette loi, on peut

- Estimer des probabilités que X se trouve entre 2 valeurs Borne fixée = |μ – Uα.σ|
  si l'on connait μ et σ on en déduit Uα
  la probabilité α correspondant a Uα se trouve dans la table (loi normale centrée réduite)

- Identifier un intervalle de valeurs contenant X avec une probabilité donnée
  - Probabilité fixée : l'intervalle doit contenir 1- α % des valeurs
    La table de la loi normale donne Uα correspondant a la probabilité α
  - L'intervalle est alors [μ – Uα.σ ; μ + Uα.σ]

Lois de probabilités continues

Approximation de la loi binomiale par le loi normale

Une variable aléatoire X binomiale B(n,π) étant définie comme **la somme de n variables de Bernoulli indépendantes de même paramètre π** , il apparaît, d'après le théorème de la limite centrale, que si le nombre n de variables de Bernoulli augmente indéfiniment, la variable binomiale B(n, π) peut être approximée par une variable normale de moyenne n π et de variance n π (1- π).

Une conséquence importante et pratique de ce résultat est qu'il est **possible de calculer avec une bonne approximation les probabilités de variables binomiales en utilisant la loi normale dès que le nombre n est suffisamment grand.** Cette approximation est d'autant meilleure que π est proche de 0,5. Dans la pratique, et dans toute la suite du cours, on considère que n est grand dès lors que les produits n π et n(1- π) sont supérieurs ou égaux à 5.

Lois de probabilités continues

Lois dérivées de la loi normale

- Un certain nombre de lois théoriques sont dérivées de loi normale et jouent un rôle important dans la théorie de l'estimation et la théorie des tests
  - **Loi du Chi-deux ou (Khi-2) ou χ^2**
  - **Loi de Student**

Loi du Chi-deux (χ^2)

Si *U1, U2, …, U*n sont des variables aléatoires normales centrées réduites et indépendantes entre elles, la somme des carrés de ces variables aléatoires définit une nouvelle variable aléatoire, notée X^2, qui est dite suivre une loi χ^2 à n degrés de liberté

$$X^2(\nu) = U_1^2 + U_2^2 + \ldots + U_\nu^2$$

Ainsi, la variable aléatoire X^2 (1) à un degré de liberté (v = 1) n'est autre que le carré d'une loi normale centrée réduite.

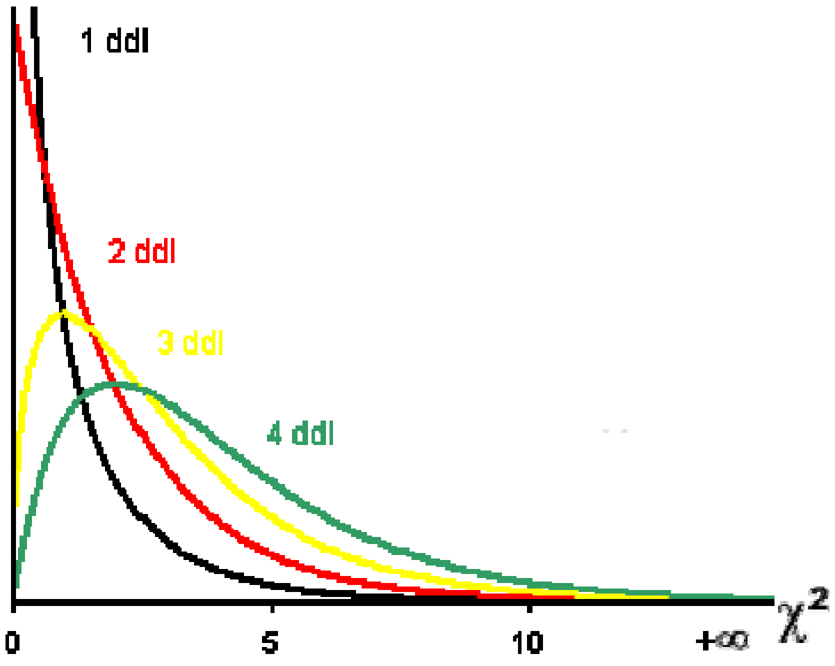
Loi du Chi-deux ($\chi 2$)

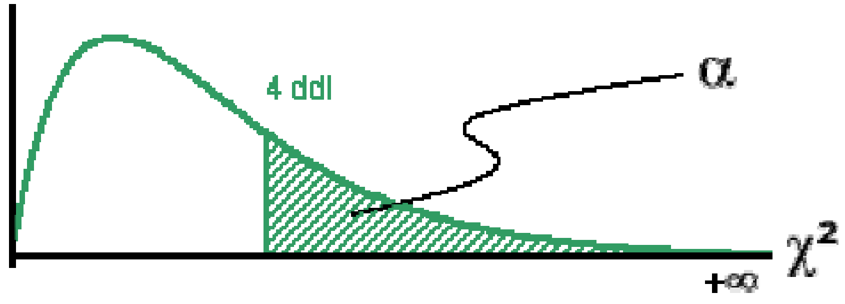
Il existe une table dite « table du $\boldsymbol{\chi^2}$» permettant le calcul de probabilités d'une variable aléatoire X^2. Cette table indique pour des nombres de degrés de liberté allant de 1 à 30, les valeurs du X^2 correspondant aux probabilités $Pr[X^2 \geq \boldsymbol{\chi_\alpha^2}]$:

- la première colonne de la table indique le nombre de degrés de liberté,
- la première ligne indique la valeur de α , et
- l'intersection d'une ligne et d'une colonne fournit la valeur de $\boldsymbol{\chi_\alpha^2}$ recherchée.

Ainsi pour un nombre de degrés de liberté égal à 12, et pour une valeur de α égale à 0,05 : $\boldsymbol{\chi_\alpha^2}$ = 21,03 ; ce qui veut dire que la probabilité qu'une variable $X^2(12)$ prenne une valeur supérieure ou égale à 21,03 est égale à 0,05.

Table du Chi-2





ddl \ α	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,0158	0,4549	1,0742	1,6424	2,7055	3,8415	5,4119	6,6349	10,8274
2	0,2107	1,3863	2,4079	3,2189	4,6052	5,9915	7,8241	9,2104	13,8150
3	0,5844	2,3660	3,6649	4,6416	6,2514	7,8147	9,8374	11,3449	16,2660
4	1,0636	3,3567	4,8784	5,9886	7,7794	9,4877	11,6678	13,2767	18,4662
5	1,6103	4,3515	6,0644	7,2893	9,2363	11,0705	13,3882	15,0863	20,5147
6	2,2041	5,3481	7,2311	8,5581	10,6446	12,5916	15,0332	16,8119	22,4575
7	2,8331	6,3458	8,3834	9,8032	12,0170	14,0671	16,6224	18,4753	24,3213
8	3,4895	7,3441	9,5245	11,0301	13,3616	15,5073	18,1682	20,0902	26,1239
9	4,1682	8,3428	10,6564	12,2421	14,6837	16,9190	19,6790	21,6660	27,8767
10	4,8652	9,3418	11,7807	13,4420	15,9872	18,3070	21,1608	23,2093	29,5879
11	5,5778	10,3410	12,8987	14,6314	17,2750	19,6752	22,6179	24,7250	31,2635
12	6,3038	11,3403	14,0111	15,8120	18,5493	21,0261	24,0539	26,2170	32,9092
13	7,0415	12,3398	15,1187	16,9848	19,8119	22,3620	25,4715	27,6882	34,5274
14	7,7895	13,3393	16,2221	18,1508	21,0641	23,6848	26,8727	29,1412	36,1239
15	8,5468	14,3389	17,3217	19,3107	22,3071	24,9958	28,2595	30,5780	37,6978
16	9,3122	15,3385	18,4179	20,4651	23,5418	26,2962	29,6332	31,9999	39,2518
17	10,0852	16,3382	19,5110	21,6146	24,7690	27,5871	30,9950	33,4087	40,7911
18	10,8649	17,3379	20,6014	22,7595	25,9894	28,8693	32,3462	34,8052	42,3119
19	11,6509	18,3376	21,6891	23,9004	27,2036	30,1435	33,6874	36,1908	43,8194
20	12,4426	19,3374	22,7745	25,0375	28,4120	31,4104	35,0196	37,5663	45,3142
21	13,2396	20,3372	23,8578	26,1711	29,6151	32,6706	36,3434	38,9322	46,7963
22	14,0415	21,3370	24,9390	27,3015	30,8133	33,9245	37,6595	40,2894	48,2676
23	14,8480	22,3369	26,0184	28,4288	32,0069	35,1725	38,9683	41,6383	49,7276
24	15,6587	23,3367	27,0960	29,5533	33,1962	36,4150	40,2703	42,9798	51,1790
25	16,4734	24,3366	28,1719	30,6752	34,3816	37,6525	41,5660	44,3140	52,6187
26	17,2919	25,3365	29,2463	31,7946	35,5632	38,8851	42,8558	45,6416	54,0511
27	18,1139	26,3363	30,3193	32,9117	36,7412	40,1133	44,1399	46,9628	55,4751
28	18,9392	27,3362	31,3909	34,0266	37,9159	41,3372	45,4188	48,2782	56,8918
29	19,7677	28,3361	32,4612	35,1394	39,0875	42,5569	46,6926	49,5878	58,3006
30	20,5992	29,3360	33,5302	36,2502	40,2560	43,7730	47,9618	50,8922	59,7022

Lois de probabilités continues

Loi de Student

Soit X^2 une variable aléatoire obéissant à une loi du **χ^2** à v degrés de liberté et U une variable aléatoire normale centrée réduite indépendante, la variable aléatoire

$$T = \frac{U}{\sqrt{X^2/_{\nu}}}$$

suit par définition une loi de Student à ν degrés de liberté.

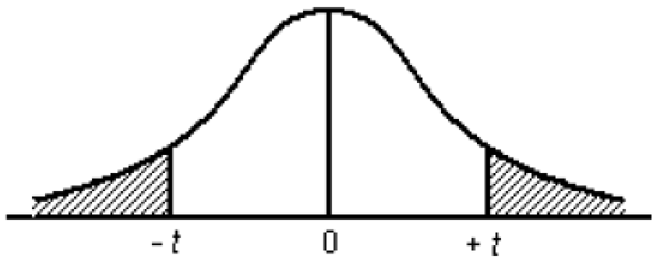
Il existe une table de la loi de Student qui indique, pour un nombre de degrés de liberté allant de 1 à 120, la probabilité d'obtenir une valeur de T à l'extérieur de l'intervalle $[-t_\alpha , + t_\alpha]$. Elle indique donc la valeur $\Pr[|T| > t]$:

- la première colonne de la table correspond au nombre de degrés de liberté,
- la première ligne à quelques valeurs de α ,
- et l'intersection d'une ligne et d'une colonne indique la valeur de t_α .

A titre d'exemple, si v = 8 et α = 0,05 , la valeur t_α lue dans la table est égale à 2,306.

Pour des degrés de liberté supérieurs à 30, on utilise généralement la table de la loi normale centrée réduite, car la loi de Student tend vers la loi normale centrée réduite lorsque son nombre de degrés de liberté tend vers l'infini.

Table de Student (ddl : 1 à 30)



ddl \ α	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,1584	1,0000	1,9626	3,0777	6,3137	12,7062	31,8210	63,6559	636,5776
2	0,1421	0,8165	1,3862	1,8856	2,9200	4,3027	6,9645	9,9250	31,5998
3	0,1366	0,7649	1,2498	1,6377	2,3534	3,1824	4,5407	5,8408	12,9244
4	0,1338	0,7407	1,1896	1,5332	2,1318	2,7765	3,7469	4,6041	8,6101
5	0,1322	0,7267	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8685
6	0,1311	0,7176	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9587
7	0,1303	0,7111	1,1192	1,4149	1,8946	2,3646	2,9979	3,4995	5,4081
8	0,1297	0,7064	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0414
9	0,1293	0,7027	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,1289	0,6998	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	4,5868
11	0,1286	0,6974	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	4,4369
12	0,1283	0,6955	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	4,3178
13	0,1281	0,6938	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	4,2209
14	0,1280	0,6924	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768	4,1403
15	0,1278	0,6912	1,0735	1,3406	1,7531	2,1315	2,6025	2,9467	4,0728
16	0,1277	0,6901	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208	4,0149
17	0,1276	0,6892	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982	3,9651
18	0,1274	0,6884	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784	3,9217
19	0,1274	0,6876	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609	3,8833
20	0,1273	0,6870	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453	3,8496
21	0,1272	0,6864	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314	3,8193
22	0,1271	0,6858	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188	3,7922
23	0,1271	0,6853	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073	3,7676
24	0,1270	0,6848	1,0593	1,3178	1,7109	2,0639	2,4922	2,7970	3,7454
25	0,1269	0,6844	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,7251
26	0,1269	0,6840	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787	3,7067
27	0,1268	0,6837	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707	3,6895
28	0,1268	0,6834	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633	3,6739
29	0,1268	0,6830	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564	3,6595
30	0,1267	0,6828	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500	3,6460

Documents ressources

1. Jean Bouyer. Méthodes statistiques. Médecine-Biologie. 2017.
2. Thierry Ancelle. Statistique-Epidémiologie, 4ème édition. 2017.
3. François Dabis, Jean Claude Desenclos. Epidémiologie de terrain. Méthodes et applications 2017
4. Simone Benazeth, Michel Boniface, Catherine Demarquilly, Virginie Lasserre, Mohamed Lemdani, Ioannis Nicolis. Biomathémathiques. Analyses, algèbre, probabilités, statistiques. Pharmacie, Médecine 1ère & 2ème années, Masson. 2001;
5. Richard F Morton, Richard J Hebel, Robert J, Mc Carter. Epidémiologie et biostatistique. 4ème édition 1998.

Merci pour votre attention