

ESIEE

AMIENS

rejoint

UniLaSalle
Terre & Sciences



2023 - 2024

I4 - FISA

Statistiques et Fiabilité des Systèmes

Responsable du cours : Sonia LAHLEB

Auteur : Patrick DROUOT

Contact : sonia.lahleb@unilasalle.fr

ETABLISSEMENT CONSULAIRE SOUS TUTELLE DU MINISTERE DE L'INDUSTRIE



HABILITE PAR LA COMMISSION DES TITRES D'INGENIEUR
ET MEMBRE DE LA CONFERENCE DES GRANDES ECOLES



14 quai de la Somme – BP 10100

80082 AMIENS CEDEX 2

Tél. : 03.22.66.20.00 – Fax : 03.22.66.20.10

<http://www.esice-amiens.fr>



Chapitre 1 - Statistiques descriptives.

Le besoin « statistiques » de posséder des données chiffrées est très anciens. La science statistique semble exister dès la naissance des premières structures sociales. D'ailleurs, les premiers textes écrits retrouvés étaient des recensements du bétail, des informations sur son cours et des contrats divers. On a ainsi trace de recensements en Chine au XXIII^e siècle av. J.-C.¹

D'ailleurs, il semblerait selon l'historien Meyer que l'origine du mot « statistique » appartiendrait au langage administratif français colbertien. Il aurait été utilisé pour la première fois, par Claude Bouchu, intendant de Bourgogne, dans une « Déclaration des biens, charges, dettes et statistiques des communautés de la généralité de Bourgogne de 1666 à 1669 »²

Pour l'instant, les dictionnaires attribuent le mot « statistique » au latin *statisticum* (qui a trait à l'État), qui aurait été germanisé par Schmeitzel en 1749.

La statistique appliquée peut se diviser en deux branches :

- la statistique descriptive : elle concerne les méthodes de recueil, description, visualisation et le résumé des données pouvant être présentés sous la forme de nombres ou de graphiques ;
- l'inférence statistique : la génération des modèles et de prédictions relatives aux phénomènes étudiés, tenant compte de l'aspect aléatoire et de l'incertitude des observations.

I. Quelques définitions.

1. Qu'est ce que la statistiques ?

Les statistiques sont aujourd'hui utilisées dans de nombreux domaines d'activités :

- Les médias qui nous abreuvent de résultats de sondages statistiques...
- L'économie : prospective, sondage, estimation, marketing, calcul de risque, etc.
- La santé : médecine et pharmacologie.
- L'industrie : fiabilité et contrôle de qualité.

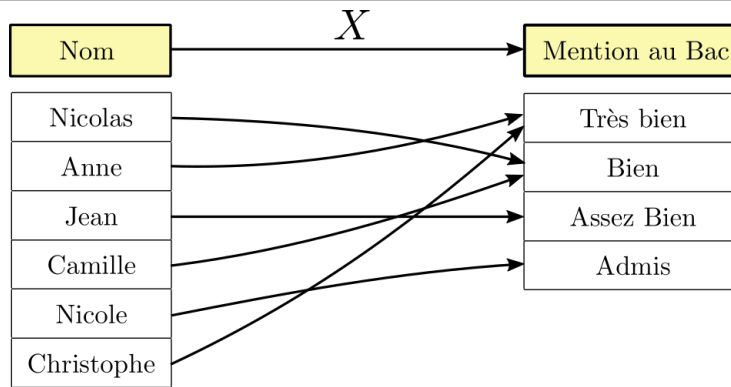
Nous allons voir cette année, que la statistique va nous permettre de faire des propositions en quantifiant leur niveau d'incertitude.

2. Vocabulaire.

- On appelle **population** les objets ou les personnes qu'on étudie.
- Chaque objet ou personne étudié s'appelle un **individu** ou une **unité statistique**.
- La chose étudiée est appelé le **caractère** .
- L'application qui à chaque individu associe sa modalité est appelée la **variable statistique**. Elle est l'expression mathématique du caractère.

1. wikipedia : Histoire des statistiques.

2. Claude Bouchu et le mot « statistique », par Dominique Pepin : <https://hal.archives-ouvertes.fr/hal-00986920>



X est une variable **qualitative**.

- L'ensemble des valeurs prises par le caractère s'appelle des **modalités**.
Les modalités de la variable statistiques X sont : **Très Bien, Bien, Assez Bien, Admis**.
- Les mesures effectuées sur les individus de la population s'appelle des **observations**.
- La séries des observations s'appelle une **série statistique**.

Mentions (x_i)	Admis	Assez Bien	Bien	Très bien
Effectifs (n_i)	1	1	2	2

En numérotant, à partir de 1, les mentions par ordre d'excellence, la série s'écrit :

$$x_1 = \text{« Admis »}, x_2 = \text{« Assez Bien »}, x_3 = \text{« Bien »}, \text{ et }, x_4 = \text{« Très Bien »}.$$

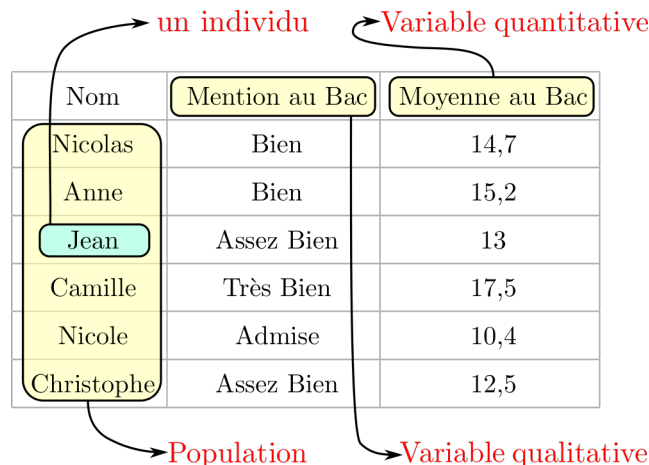
$$n_1 = 1, n_2 = 1, n_3 = 2, \text{ et } n_4 = 2.$$

La série statistique est la donnée des (x_i, n_i) : $(x_1, n_1), (x_2, n_2), (x_3, n_3)$, et (x_4, n_4) .

D'un point de vu statistique, l'identité des individus **n'a aucune importance**.

Une variable statistique est dite :

- **quantitative** lorsqu'elle est mesurée par un nombre (Notes des étudiants à l'examen de statistiques, durée de vie d'un téléphone portable, etc.). La série des « mentions au bac » n'est pas quantitative car les x_i ne se sont pas des nombres.
- **qualitative** lorsque les modalités (ou les valeurs) qu'elle prend sont désignées par des noms. Par exemples, les modalités de la variable « *genre* » sont : Masculin et Féminin ; celles de la variable « *couleur des yeux* » sont Bleu, Marron, Noir, Vert, etc.



Quel sens pourrait-on donner à « la moyenne des mentions au baccalauréat » (Caractère qualitatif) ?
Par contre, il est facile de calculer la moyenne des moyennes au baccalauréat de ce groupe de bacheliers.

Nous n'étudierons que des caractères quantitatifs, car il est difficile de faire des mathématiques sans données numériques.

Parmi les variables quantitatives, on distingue :

- les variables quantitatives **discrètes** : elles ne prennent que des valeurs isolées (la note à l'examen de statistique). Elle peuvent prendre une infinité de valeurs, mais toutes, isolées (le nombre d'étoiles par galaxie).
- les variables quantitatives **continues** : elles peuvent prendre toutes les valeurs dans un intervalle (la circonférence d'une vis en millimètres).

II. Mesures de tendance centrale.

1. Calcul de la moyenne.

Etude n° 1 : Sur un tronçon d'autoroute, les gendarmes ont mesuré la vitesse de 45 voitures :

125 133 129 115 141 117 128 138 152 128 127 105 131 117 163
115 123 144 104 101 112 145 110 118 114 123 102 124 138 126
147 115 120 112 109 124 125 107 113 108 136 116 122 131 132

- La population étudiée est : **les 45 voitures contrôlées ;**
- le caractère étudié (variable aléatoire) : **la vitesse.**
- la variable aléatoire est-elle continue ? **Non, elle est discrète ;**
- L'effectif total : **45.**

La moyenne des vitesses mesurées est : $\frac{125 + 133 + 129 + \dots + 132}{45} = \frac{5565}{45} \simeq 123,7 \text{ km/h}$



Définition:

En statistique, lorsqu'il y a beaucoup de données, on les regroupe dans des intervalles appelés **classes**.

Classe de vitesses	[100,110[[110,120[[120,130[[130,140[[140,150[[150,170[
Centre des classes (x_i)	105	115	125	135	145	160
Nombre de voitures (n_i)	7	12	13	7	4	2



Hypothèses de travail avec des données regroupées en classes.

Les données de chaque classes seront supposées différentes et réparties **uniformément** à l'intérieur de la classe. C'est la raison pour laquelle on représente une classe par son **centre** (la moyenne de ses extrémités).

La moyenne des vitesses est :

$$\frac{7 \times 105 + 12 \times 115 + 13 \times 125 + 7 \times 135 + 4 \times 145 + 2 \times 160}{45} = \frac{5585}{45} \simeq 124,1 \text{ km/h}$$

L'erreur est due à l'hypothèse de travail du regroupement en **classes**.

**Construction de classes**

Le nombres de classes va dépendre de :

- la taille de la population : « Trop de classes » dilue l'information et « peu de classes » la concentre trop ;
- sa répartition : souvent pour les valeurs extrêmes, la concentration diminue et l' **amplitude** des classes augmente.

☞ L'amplitude de la dernière classe de vitesses [150,170] est : $170 - 150 = 20$ km/h

☞ L'amplitude de la deuxième classe de vitesses [110,120] est : $120 - 110 = 10$ km/h

2. Qu'est-ce que la moyenne ?**a. Son calcul :****Notations :**

(x_i) : la série statistique brute : les modalités n'ont pas été regroupées.

(x_i, n_i) : la série statistique où l'on a regroupé les individus ayant la même modalité.

x_i : i^{me} valeur (modalité) de la variable statistique x .

n_i : Effectif de la valeur x_i .

N : Effectif total $N = \sum_i n_i$

\bar{x} : la moyenne de la série statistiques (x_i) , $\bar{x} = \frac{1}{N} \sum_i x_i$

la moyenne de la série statistiques (x_i, n_i) , $\bar{x} = \frac{1}{N} \sum_i n_i x_i$

Exemple n° 1 : Considérons les données brutes suivant :

2 - 5 - 2 - 9 - 9 - 2 - 7 - 5 - 2 - 5

Il y a deux façons de les représenter :

- La première, on parle de la série statistique $x = (x_i) : x_1 = 2, x_2 = 5, x_3 = 2, \dots, x_{101} = 5$

Les n_i sont tous égaux à 1. On écrit $x = (2, 5, 2, 9, 9, 2, 7, 5, 2, 5)$

$$\sum_{i=1}^{10} x_i = 2 + 5 + 2 + 9 + \dots + 5 = 48 \text{ et } \bar{x} = \frac{48}{10} = 4,8$$

- La seconde, on parle de la série statistique $y = (y_i, n_i) :$

$$(y_1, n_1) = (2, 4), (y_2, n_2) = (5, 3), (y_3, n_3) = (9, 2), \text{ et } (y_4, n_4) = (7, 1)$$

On écrit $y = ((2, 4), (5, 3), (9, 2), (7, 1))$

$$\sum_{i=1}^4 n_i x_i = 4 \times 2 + 3 \times 5 + 2 \times 9 + 1 \times 7 = 48 \text{ et } \bar{x} = \frac{48}{10} = 4,8$$



Notations des séries statistiques :

$$(x - b) = (x_1 - b, x_2 - b, x_3 - b, \dots) \text{ et } ay = ((ax_1, n_1), (ax_2, n_2), (ax_3, n_3), \dots)$$

Exemple n° 2 : $(x - 3) = (-1, 2, -1, 6, 6, -1, 4, 2, -1, 2)$ et $2y = ((4, 4), (10, 3), (18, 2), (14, 1))$

b. Propriétés de la moyenne.



Propriété

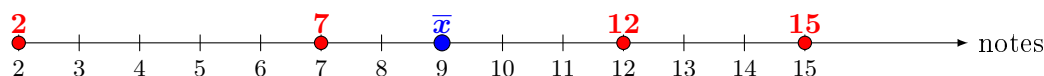
Etant donné une population \mathcal{P} . L'application qui à une série statistique de la population \mathcal{P} associe sa moyenne est linéaire :

$$\overline{a \times x} = a \times \bar{x} \quad , \quad \overline{x + y} = \bar{x} + \bar{y} \quad \text{et} \quad \overline{x + b} = \bar{x} + b$$

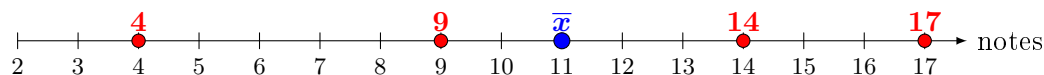
où a et b sont deux constantes réelles.

Exemple n° 3 : Considérons les quatre notes suivantes : 2, 7, 12, et 15.

- La moyenne de ces notes est égale à **9**



- Translatons toutes ces notes de 2 points vers la droites :



La moyenne **a été augmentée de 2 points**. On écrit $\overline{x + 2} = \bar{x} + 2$



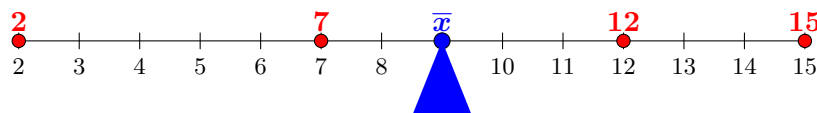
Démonstration

Soient (x_i) et (y_i) deux séries statistiques issues d'une **même population** d'effectif total N , et a un nombre réel.

$$\bullet \quad \overline{ax} = \frac{1}{N} \sum_{i=1}^N ax_i = \frac{a}{N} \sum_{i=1}^N x_i = a\bar{x}$$

$$\bullet \quad \overline{x + y} = \frac{1}{N} \sum_{i=1}^N (x_i + y_i) = \frac{1}{N} \left(\sum_{i=1}^N x_i + \sum_{i=1}^N y_i \right) = \frac{1}{N} \sum_{i=1}^N x_i + \frac{1}{N} \sum_{i=1}^N y_i = \bar{x} + \bar{y}$$

Interprétation mécanique : La moyenne est le point d'équilibre des notes :



c. La moyenne définie par les écarts :

Considérons la série statistiques à k modalités : $(x_i, n_i)_{i=1, \dots, k}$ où $N = \sum_{i=1}^k n_i$

Etant donné un nombre réel a :

- $a - x_i$ est l'écart entre la i^{me} modalité du caractère et a ;
- $a - x_i < 0$ signifie que x_i est **supérieur** à a ;
- $a - x_i > 0$ signifie que x_i est **inférieur** à a ;
- $f(a) = \sum_{i=1}^k n_i(a - x_i)$ est la somme de tous ces écarts pondérés par leur effectif.

Déterminons la valeur de a pour laquelle f s'annule ?

$$\begin{aligned} f(a) &= 0 \\ \sum_{i=1}^k n_i(a - x_i) &= 0 \\ \left(\sum_{i=1}^k n_i a \right) - \left(\sum_{i=1}^k n_i x_i \right) &= 0 \\ a \sum_{i=1}^k n_i &= \left(\sum_{i=1}^k n_i x_i \right) \\ a &= \frac{1}{N} \sum_{i=1}^k n_i x_i = \bar{x} \end{aligned}$$

Ainsi, la moyenne peut être définie comme étant la valeur centrale qui annule les écarts :

la somme des écarts à gauche de la moyenne est égale à celle de droite.

Exemple n° 4 : Considérons les quatre notes suivantes : 2, 7, 12, et 15. La moyenne de ces notes est égale à **9**

Etudions les écarts à la moyenne :

Notes	2	7	12	15	Total
Écarts à la moyenne	-7	-2	3	6	0

 **Propriété**

La somme des **écarts** à la moyenne est nulle.

III. Mesures de dispersion.

1. De la moyenne à la variance

La somme des écarts étant nuls, pour mesurer comment se dispersent les valeurs autour de la moyenne, on va les élever au carré.

Définition:

La somme des carrés des écarts à la moyenne, notée V ou S^2 , est appelée la **variance**:

$$V(x) = \frac{1}{N} \sum_i n_i (\bar{x} - x_i)^2$$

Considérons maintenant un nombre réel t et posons $g(t) = \frac{1}{N} \sum_i n_i (t - x_i)^2$.

Pour $t = \bar{x}$, on sait que $g(t)$ est la variance.

g est un polynôme du second degré en t : $g(t) = \frac{1}{N} \sum_i n_i (t^2 - 2tx_i + x_i^2) = \frac{1}{N} \left(Nt^2 - 2t \sum_i n_i x_i + \sum_i n_i x_i^2 \right)$.

Donc, g n'a pas de maximum, son coefficient dominant $a = 1 > 0$, mais un minimum atteint en

$$-\frac{b}{2a} = -\frac{-\frac{2}{N} \sum_i n_i x_i}{2} = \bar{x}$$

Ainsi, la moyenne est la valeur qui n'augmente pas « artificiellement » la somme des carrés des écarts.

Propriété

La variance est la plus petite somme des carrés des écarts.

Définition:

La racine carrée de la variance est appelée l'**écart-type** et est noté σ ou S :

$$\sigma = \sqrt{\frac{1}{N} \sum_i n_i (\bar{x} - x_i)^2}$$

Remarque : La variance peut aussi être notée σ^2 , mais ce ne sera pas le cas dans ce cours.

2. Propriété de la variance.

Propriété

Soient a et b deux nombres réels, (x_i) une série statistique :

$$V(ax + b) = a^2 V(x) \text{ et } \sigma_{ax+b} = |a| \sigma_x$$

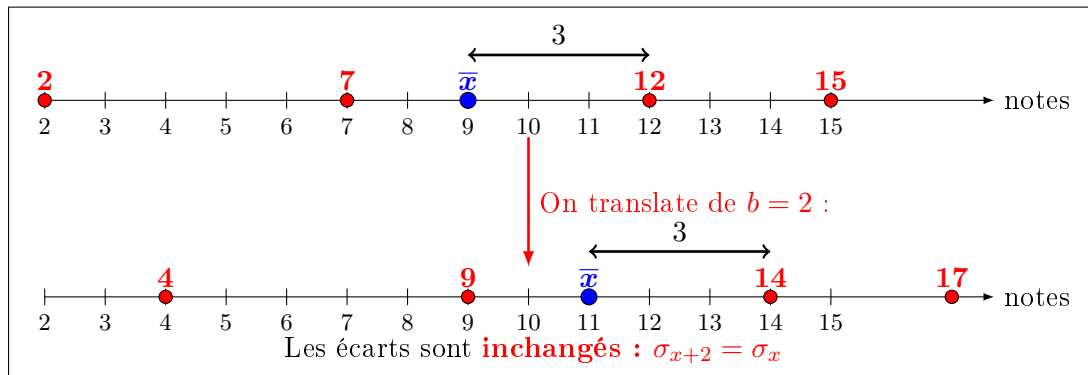
Exemple n° 5 : Considérons les quatre notes suivantes : 2, 7, 12, et 15.

- $S^2 = \frac{(9-2)^2 + (9-7)^2 + (9-12)^2 + (9-15)^2}{4} = \frac{98}{4} = 24,5$ et donc $\sigma = \sqrt{24,5} \simeq 5$

- Translatons toutes les notes de 2 points vers la droites, la moyenne est $9+2=11$ et :

$$S^2 = \frac{(11-4)^2 + (11-9)^2 + (11-14)^2 + (11-17)^2}{4} = \frac{98}{4} = 24,5$$

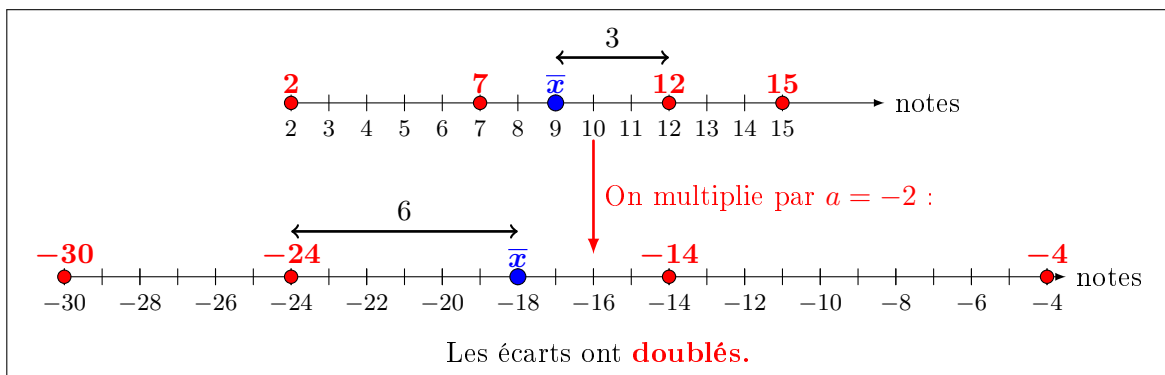
On retrouve les mêmes écarts.



- Multiplions toutes les notes par -2 , la moyenne est $-2 \times 9 = -18$

$$S^2 = \frac{(-18 - (-4))^2 + (-18 - (-14))^2 + (-18 - (-24))^2 + (-18 - (-30))^2}{4} = \frac{392}{4} = 98$$

et donc la variance a été multipliée par $(-2)^2 = 4$



Théorème de König-Huygens (XVII^e)



La variance d'une série statistique (x_i, n_i) est :

$$V(x) = \overline{x^2} - \bar{x}^2 = \frac{1}{N} \sum_i n_i x_i^2 - \left(\frac{1}{N} \sum_i n_i x_i \right)^2$$

Démonstration

$$V(x) = \frac{1}{N} \sum_i n_i (\bar{x} - x_i)^2 = \overline{(\bar{x} - x_i)^2} = \overline{\bar{x}^2 - 2\bar{x}x_i + x_i^2} = \overline{\bar{x}^2} - 2\overline{\bar{x}x_i} + \overline{x_i^2}$$

\bar{x} est une constante donc :

- $\overline{\bar{x}^2} = \bar{x}^2$;

- $\overline{\bar{x}x_i} = \bar{x} \overline{x_i} = \bar{x}^2$

- $\overline{x_i^2}$ se note $\overline{x^2}$

$$\text{Donc, } V(x) = \bar{x}^2 - 2\bar{x}^2 + \overline{x^2} = \overline{x^2} - \bar{x}^2$$

Si vous avez du mal à comprendre cette démonstration, reprenez-la en remplaçant \bar{x} par m .

Exemple n° 6 : Appliquons la formule de König-Huygens pour calculer la l'écart-type de la série statistique (x_i, n_i) suivante :

(3; 4), (7; 2), (-2; 1), et (9; 3)

- $\bar{x} = \frac{4 \times 3 + 2 \times 7 + 1 \times (-2) + 3 \times 9}{4 + 2 + 1 + 3} = \frac{51}{10} = 5,1$

- $\overline{x^2} = \frac{4 \times 3^2 + 2 \times 7^2 + 1 \times (-2)^2 + 3 \times 9^2}{4 + 2 + 1 + 3} = \frac{381}{10} = 38,1$

Ainsi, $S^2 = \overline{x^2} - \bar{x}^2 = 38,1 - 5,1^2 = 12,09$ et donc $\sigma = \sqrt{12,09} \simeq 3,48$



Définition:

Soient $x = (x_i)$ et $y = (y_i)$ deux séries statistiques issues d'une même population d'effectif total N .

La **covariance** des séries x et y , notée **cov(x, y)** est

$$\text{cov}(x, y) = \frac{1}{N} \sum_{i=1}^N (\bar{x} - x_i)(\bar{y} - y_i)$$

Remarque :

- $\text{cov}(x, x) = V(x)$
- Il n'y a pas de formule avec les n_i car, pour un indice i donné, les modalités x_i et les y_i n'ont pas forcément le même effectif.




Propriété

Soient $x = (x_i)$ et $y = (y_i)$ deux séries statistiques issues d'une même population d'effectif total N .

- $\text{cov}(x, y) = \overline{(\bar{x} - x_i)(\bar{y} - y_i)}$

- $\text{cov}(x, y) = \overline{x y} - \bar{x} \bar{y} = \left(\frac{1}{N} \sum_{i=1}^N x_i y_i \right) - \bar{x} \bar{y}$

 **Propriété**

Soient a et b deux nombres réels, (n_i, x_i) et (n_i, y_i) deux séries statistiques issues d'une même population.

$$V(x + y) = V(x) + V(y) + 2cov(x, y)$$

Exemple n° 7 : On a mesuré les longueurs en millimètres d'un échantillon de 100 tiges d'aciers à la sortie d'une machine automatique. On a trouvé les résultats suivants :

Longueur des tiges (en mm)	Nombre de tiges	centre des classes	effectifs partiels × centre	effectifs partiels × centre ²
[120 ; 130[10	125	1250	156250
[130 ; 140[20	135	2700	364500
[140 ; 150[38	145	5510	798950
[150 ; 160[25	155	3875	600625
[160 ; 170[7	165	1155	190575
Total	100		14490	2110900

- La moyenne $\bar{x} = \frac{14490}{100} = 144,9$ mm
- La moyenne des carrés : $\overline{x^2} = \frac{1}{N} \sum_{i=1}^5 n_i x_i^2 = \frac{2110900}{100} = 21109$
- La variance : $S^2(x) = \overline{x^2} - \bar{x}^2 = 21109 - 144,9^2 = 112,99$
- L'écart-type : $S(x) = \sqrt{S^2(x)} = \sqrt{112,99} \simeq 10,630$ mm

IV. Courbes de régression et corrélation.

Lorsque deux variables sont examinées, on cherche souvent à déterminer, s'il existe, un lien entre elles. Ce lien, s'il existe, peut être linéaire, quadratique, exponentiel, etc.

Pour cette recherche de lien, la méthode la plus utilisée est celle des moindres carrés.

Précisons cependant que les liens dont on parle sont strictement algébrique, et qu'ils ne donnent aucune information sur l'existence d'une dépendance entre les variables.

C'est la raison pour laquelle, lorsque deux variables sont indépendantes, leur corrélation est nulle, alors que la réciproque est fautive :

La corrélation nulle de deux variables n'entraîne pas leur indépendance.

1. Recherche graphique : le nuage de points

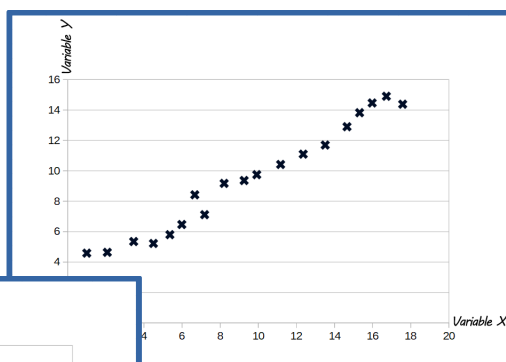


Définition:

On considère deux séries statistiques (x_i) et (y_i) définies sur une même population.
On appelle **nuage de points** l'ensemble des points de coordonnées (x_i, y_i)

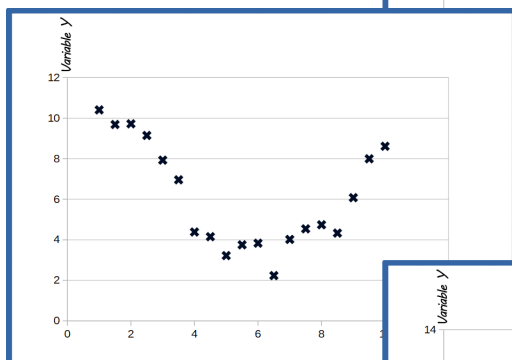
Régression linéaire :

$$Y = aX + b$$

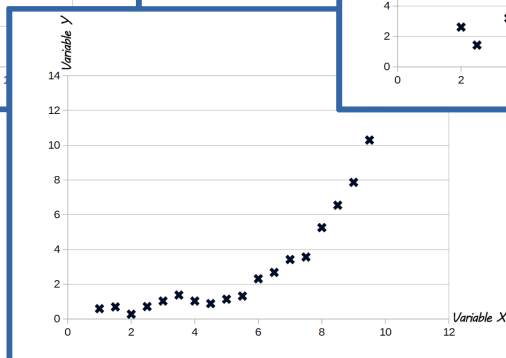
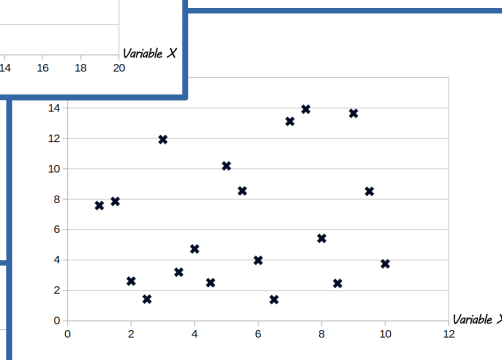


Régression quadratique :

$$Y = aX^2 + bX + c$$



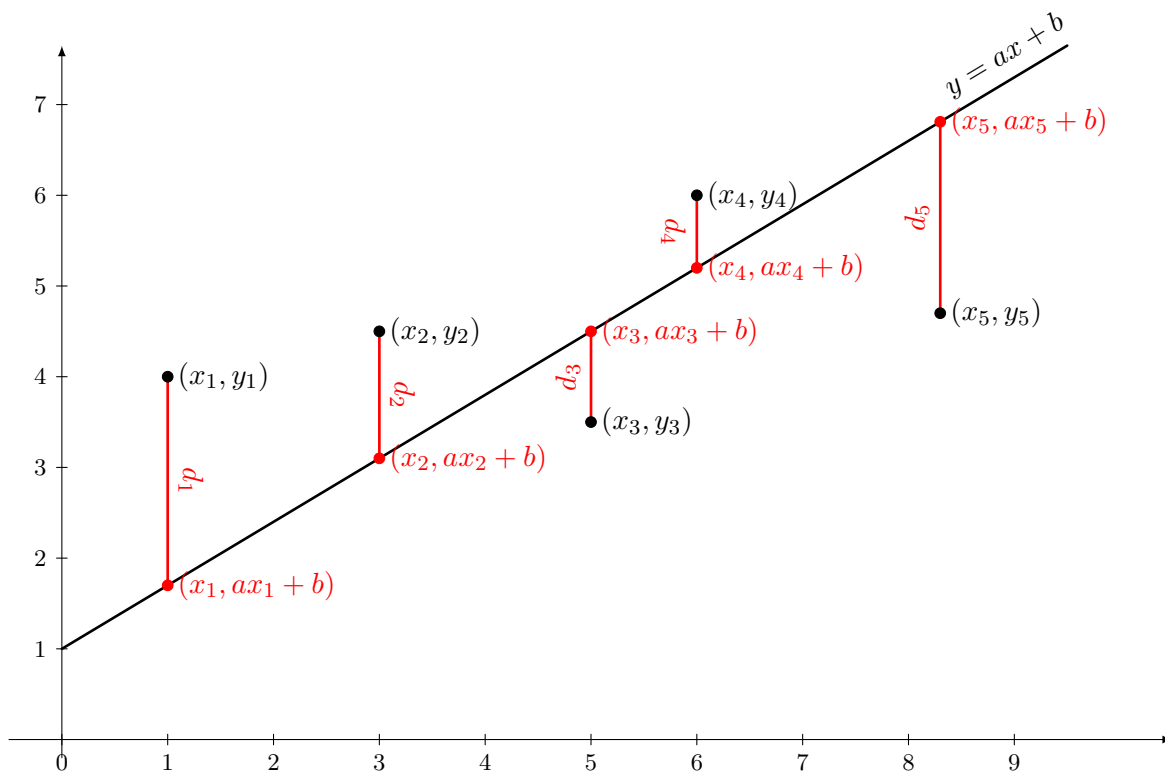
Il ne semble pas exister de lien entre X et Y



Régression exponentielle : $Y = Ar^X$

2. Recherche algébrique : La méthode des moindres carrés.

a. Le cas linéaire :



On cherche la droite d'équation $Y = aX + b$ telles que $\sum_i d_i^2 = \sum_i (ax_i + b - y_i)^2$ soit minimum, d'où le nom de la méthode...



Définition:

On appelle droite de régression par la méthode des moindres carrés, la droite d'équation : $y = ax + b$ où $a = \frac{\text{cov}(x, y)}{V(x)}$ et $b = \bar{y} - a\bar{x}$.
On parle alors de régression **linéaire**.



Propriété

Soit (x_i, y_i) un nuage de points. Notons $y = ax + b$ l'équation réduite de la droite de régression.

- La somme $\sum_i (ax_i + b - y_i)^2$ est **minimum**.
- Le point de coordonnées $(\bar{x}, \bar{y})^*$ appartient à cette droite.

* est appelé le **point moyen** du nuage de points.



Définition:

On appelle **coefficient de corrélation**, noté ρ , le nombre $\frac{cov(x, y)}{\sigma_x \sigma_y}$

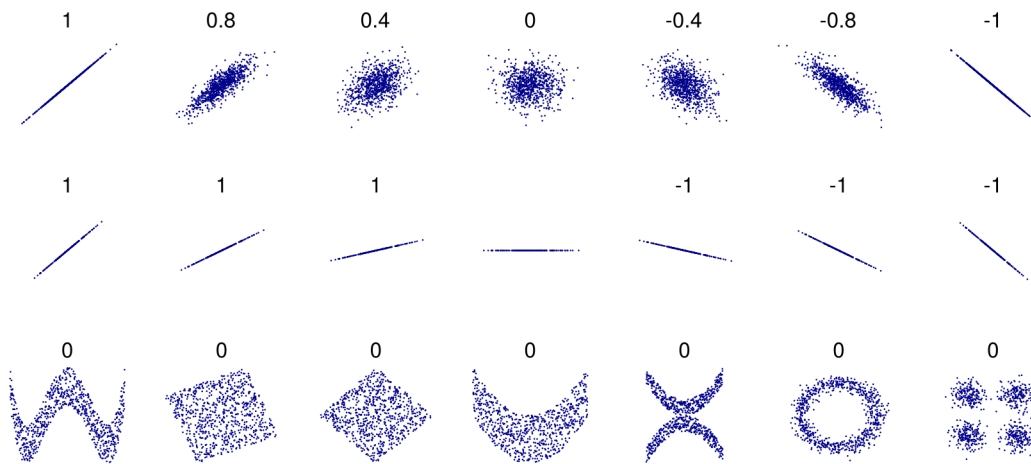


Propriété

Soit ρ le coefficient de corrélation d'un nuage de points.

- Si $\rho = 1$ les points sont alignés suivant une droite ascendante.
- Si $\rho = -1$ les points sont alignés suivant une droite descendante.
- Plus $|\rho|$ est proche de 1, plus les points du nuages sont alignés.

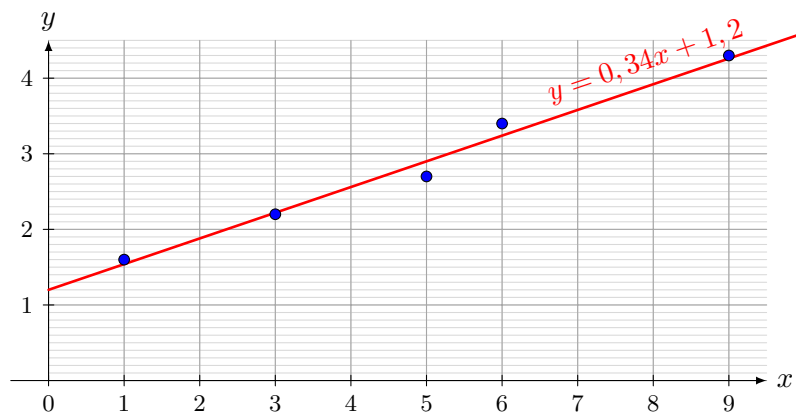
Exemple n° 8 : Coefficients de corrélations de différents nuages de points :



Exemple n° 9 :

x_i	1	3	5	6	9
y_i	1,6	2,2	2,7	3,4	4,3

On cherche à exprimer y en fonction de x . Commençons par construire le nuage de points :



Il semble au vu du nuage de points que **les points sont presque alignés**

On va donc chercher **l'équation de la droite de régression linéaire par la méthode des moindres carrés** :

						Total
x_i	1	3	5	6	9	24
y_i	1,6	2,2	2,7	3,4	4,3	14,2
x_i^2	1	9	25	36	81	152
y_i^2	2,56	4,84	7,29	11,56	18,49	44,74
$x_i y_i$	1,6	6,6	13,5	20,4	38,7	80,8

- $\bar{x} = \frac{1}{N} \sum_i x_i = \frac{24}{5} = 4,8$
- $\overline{x^2} = \frac{1}{N} \sum_i x_i^2 = \frac{152}{5} \simeq 30,4$
- $S^2(x) = \overline{x^2} - \bar{x}^2 = 30,4 - 4,8^2 = 7,36$
- $\sigma_x = \sqrt{S^2(x)} \simeq 2,71$
- $\bar{y} = \frac{1}{N} \sum_i y_i = \frac{14,2}{5} = 2,84$
- $\overline{y^2} = \frac{1}{N} \sum_i y_i^2 = \frac{44,74}{5} = 8,948$
- $S^2(y) = \overline{y^2} - \bar{y}^2 = 8,948 - 2,84^2 \simeq 0,88$
- $\sigma_y = \sqrt{S^2(y)} \simeq 0,94$
- $\overline{xy} = \frac{80,8}{5} \simeq 16,16$
- $cov(x, y) = \overline{xy} - \bar{x} \times \bar{y} = 16,16 - 4,8 \times 2,84 \simeq 2,528$

On a donc $y = at + b$ où

$$\begin{cases} a = \frac{cov(x, y)}{S^2(x)} = \frac{2,528}{7,36} \simeq 0,34 \\ b = \bar{y} - a \times \bar{t} = 2,84 - 0,34 \times 4,8 \simeq 1,20 \end{cases}$$

La droite de régression a pour équation **$y = 0,34x + 1,2$**

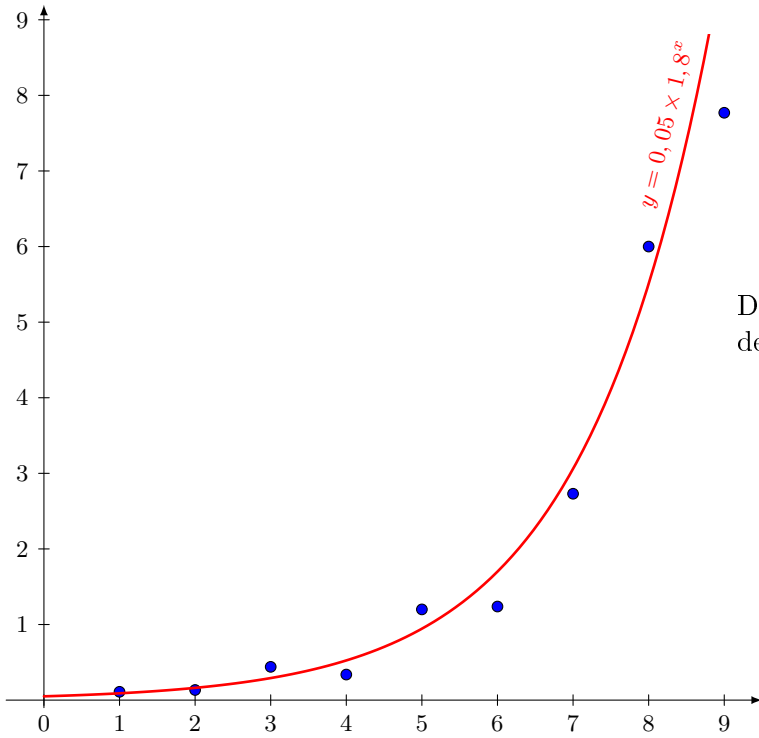
Le coefficient de corrélation doit être proche de **1**, calculons-le :

$$\rho = \frac{cov(x, y)}{\sigma_x \sigma_y} = \frac{2,528}{2,71 \times 0,95} \simeq 0,992$$

b. Le cas exponentiel :

On a un nuage de points dont la répartition semble suivre une courbe de la forme $y = ar^x$ où a et r sont deux constantes positives non nulles.

Exemple n° 10 :



$$y = ar^x \iff \ln(y) = \ln(a) + x \ln(r)$$

Donc, on détermine la droite de régression du nuage de points $(x_i, \ln(y_i))$:

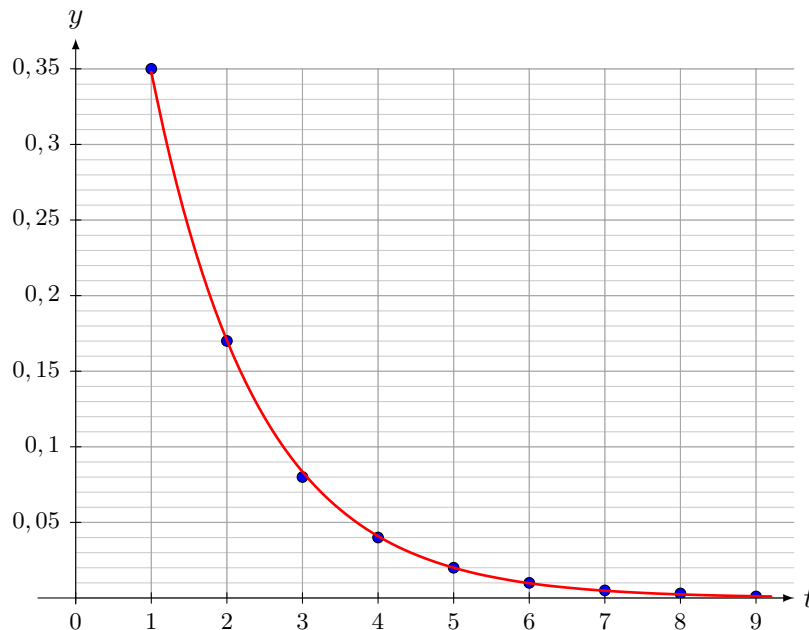
$$\begin{aligned} \ln(y) &= Ax + B \\ y &= e^{Ax+B} \\ y &= (e^A)^x \times e^B \end{aligned}$$

Exemple n° 11 : On étudie la fonction de fiabilité d'un système :

durée de vie t en années	1	2	3	4	5	7	9
Fiabilité y	0,35	0,17	0,08	0,04	0,02	0,005	0,001

Interprétation : La probabilité que le système soit fiable pendant 3 ans est de **8%**

On cherche à exprimer y en fonction de t . Commençons par construire le nuage de points :



Il semble au vu du nuage de points que $y = A \times r^t$ avec $0 < r < 1$.

On va donc chercher l'équation de la droite de régression qui exprime $\ln(y)$ en fonction de t :

Fiabilité y_i	durée de vie t_i	$\ln(y_i)$	t_i^2	$\ln(y_i)^2$	$t_i \ln(y_i)$
0,35	1	-1,05	1	1,10	-1,05
0,17	2	-1,77	4	3,13	-3,54
0,08	3	-2,53	9	6,40	-7,59
0,04	4	-3,22	16	10,37	-12,88
0,02	5	-3,91	25	15,29	-19,55
0,005	7	-5,30	49	28,09	-37,1
0,001	9	-6,91	81	47,75	-62,19
Total	31	-24,69	185	112,13	-143,9

- $\bar{t} = \frac{1}{N} \sum_i t_i = \frac{31}{7} \simeq 4,43$
- $S^2(t) = \bar{t^2} - \bar{t}^2 = 26,43 - 4,43^2 \simeq 6,81$
- $\overline{\ln(y)} = \frac{-24,69}{7} \simeq -3,53$
- $S^2(\ln(y)) = 16,02 - (-3,53)^2 \simeq 3,56$
- $\overline{t \ln(y)} = \frac{-143,9}{7} \simeq -20,56$
- $cov(t, \ln(y)) = \overline{t \ln(y)} - \bar{t} \times \overline{\ln(y)} = -20,56 - 4,43 \times (-3,53) \simeq -4,92$

$$\bar{t^2} = \frac{1}{N} \sum_i t_i^2 = \frac{185}{7} \simeq 26,43$$

$$\sigma_t = \sqrt{S^2(t)} \simeq 2,61$$

$$\overline{\ln(y)^2} = \frac{112,13}{7} \simeq 16,02$$

$$\sigma_{\ln(y)} = \sqrt{S^2(\ln(y))} \simeq 1,89$$

On a donc $\ln(y) = at + b$ où

$$\begin{cases} a = \frac{cov(t, \ln(y))}{S^2(t)} = \frac{-4,92}{6,81} \simeq -0,72 \\ b = \overline{\ln(y)} - a \times \bar{t} = -3,53 + 0,72 \times 4,43 \simeq -0,34 \end{cases}$$

Ainsi, $\ln(y) = -0,72t - 0,34$

$$y = e^{-0,72t-0,34} = (e^{-0,72})^t \times e^{-0,34} = 0,71 \times 0,49^t$$

Chapitre 2 - Probabilités.

I. Rappels

Soient A , B , et C trois événements d'un univers Ω , on a les propriétés suivantes :

- $P(\emptyset) = 0$, $P(\Omega) = 1$, et, $0 \leq P(A) \leq 1$
- $P(\bar{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ et $P(A \cup B) = P(A) + P(B)$ si A et B sont incompatibles.
- Si $P(B) \neq 0$ alors la probabilité de A sachant B est $P_B(A) = \frac{P(A \cap B)}{P(B)}$

II. Evénements indépendants

Subjectivement, on pourrait dire que A est indépendant de B si $P_B(A) = P(A)$.
Mais dans ce cas, $P_A(B)$ est-elle égale à $P(B)$?

$$\begin{aligned}
 P_B(A) = P(A) &\iff P(A) = \frac{P(A \cap B)}{P(B)} \\
 &\iff P(A \cap B) = P(A)P(B) \\
 &\iff P(B) = \frac{P(A \cap B)}{P(A)} \\
 &\iff P_A(B) = P(B)
 \end{aligned}$$



Définition:

Deux événements A et B sont **indépendants** si $P(A \cap B) = P(A)P(B)$.

Exemple n° 12 : Lançons deux dés équilibrés, de couleurs différentes. Considérons les événements suivants :

- Q : « le 1^{er} dé, le dé vert, donne 4. »
- T : « le 2^e dé, le dé rouge, donne 3. »
- S : « la somme des deux dés donne 7. »

L'univers Ω est constitués des **36** couples (a, b) où $a, b \in \{1, 2, 3, 4, 5, 6\}$

Etudions l'indépendance des événements suivants :

1. Etudions l'indépendance de S et Q :

- $P(Q) = \frac{1}{6}$
- $P(S) = \frac{\#\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}}{36} = \frac{6}{36} = \frac{1}{6}$

$$\bullet P(Q \cap S) = \frac{\#\{(4,3)\}}{36} = \frac{1}{36}$$

$P(Q) \times P(S) = P(Q \cap S)$: les événements S et Q sont indépendants.

2. Etudions l'indépendance de S et T :

$$P(S) = \frac{1}{6}, P(T) = \frac{1}{6}, \text{ et } P(S \cap T) = \frac{\#\{(4,3)\}}{36} = \frac{1}{36}$$

$P(T) \times P(S) = P(S \cap T)$: les événements S et T sont indépendants.

3. Etudions l'indépendance de S et $Q \cap T$:

$$\bullet P(Q \cap T) = \frac{\#\{(4,3)\}}{36} = \frac{1}{36}$$

$$\bullet P(S \cap (Q \cap T)) = P(Q \cap T) = \frac{1}{36}$$

$$P(Q \cap T) \times P(S) = \frac{1}{36} \times \frac{1}{6} = \frac{1}{216}, \text{ donc } P(S \cap (Q \cap T)) \neq P(Q \cap T) \times P(S) :$$

les événements $Q \cap T$ et S ne sont pas indépendants.



Théorème

Soient A et B deux événements. Si A et B sont indépendants, alors il est en de même de \bar{A} et \bar{B} , et \bar{A} et B , et de A et \bar{B} .

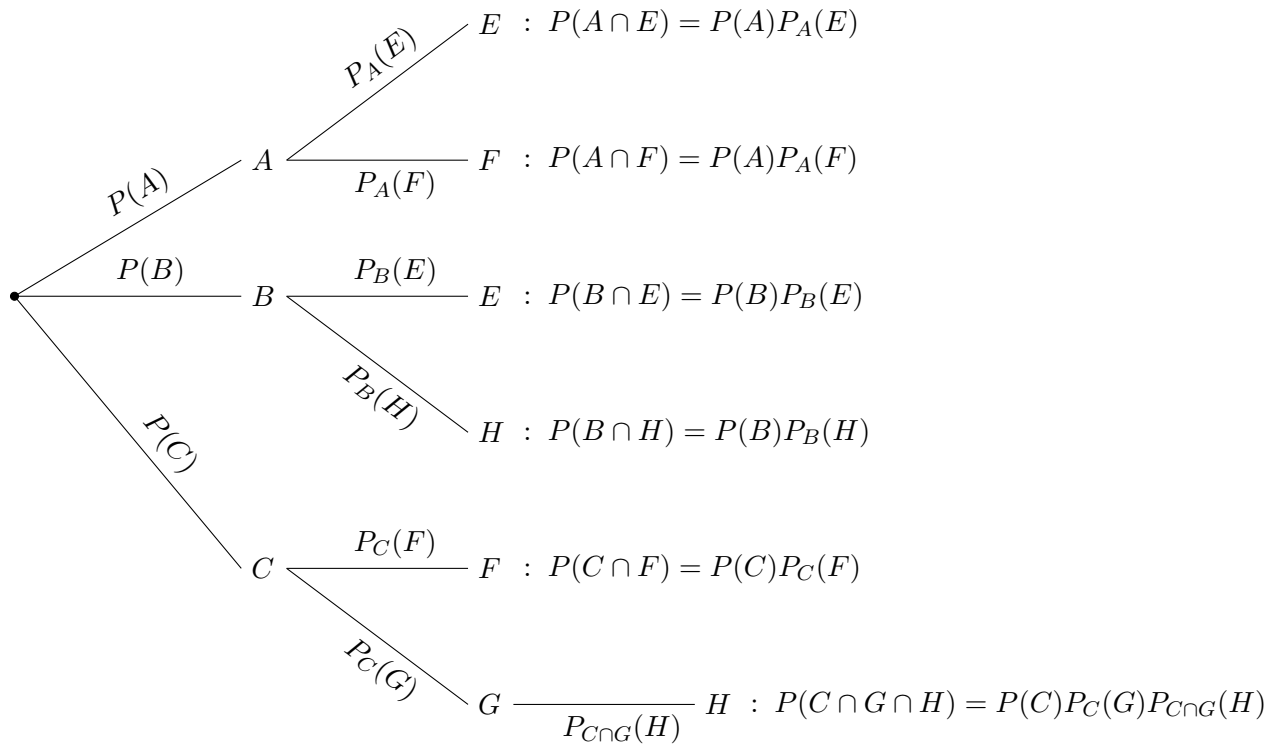


Démonstration

On a $A = (A \cap \bar{B}) \cup (A \cap B)$ et cette réunion est disjointe, d'où par additivité :

$$\begin{aligned} P(A) &= P(A \cap \bar{B}) + P(A \cap B) \\ P(A \cap \bar{B}) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) \\ &= P(A)P(\bar{B}) \end{aligned}$$

III. Arbres pondérés



Propriété

La somme des probabilités issues d'un même nœud est égale à 1.

Exemple n° 13 :

- Au nœud racine : $P(A) + P(B) + P(C) = 1$
- Au nœud C : $P_C(F) + P_C(G) = 1$.

Propriété

La probabilité d'un chemin est le produit des probabilités portées par ses branches.

Exemple n° 14 :

- Au chemin $A \cap F$: $P(A \cap F) = P(A)P_A(F)$
- Au chemin $C \cap G \cap H$: $P(C \cap G \cap H) = P(C)P_C(G)P_{C \cap G}(H)$

Formule des probabilités totales :

la probabilité d'une feuille est la somme des probabilités des chemins menant à cette feuille.

Exemple n° 15 : Il y a deux chemins menant à la feuille F : $A \cap F$ et $C \cap F$

$$P(F) = P(A \cap F) + P(C \cap F) = P(A)P_A(F) + P(C)P_C(F)$$

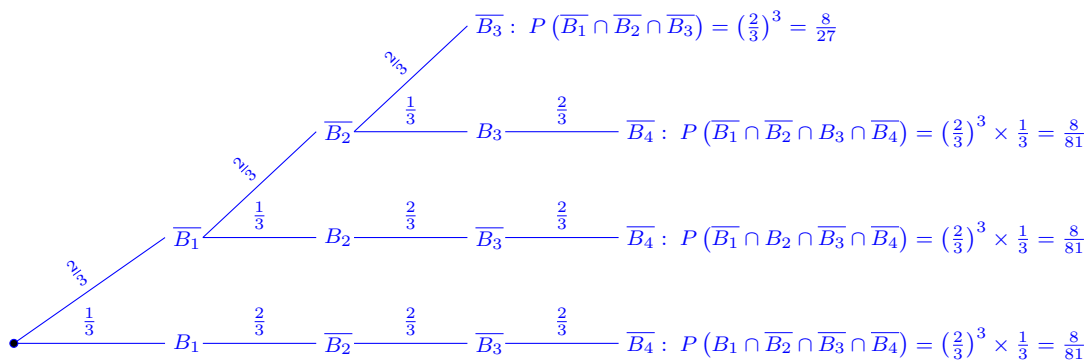
Exercice n° 1: Aline adore la confiture aux prunes de sa grand-mère. Quand elle tartine sa tranche de brioche, au petit déjeuner, la cuillerée de confiture va directement dans sa bouche une fois sur trois. Pour obtenir une tranche de brioche correctement recouverte de confiture, il faut y étaler le contenu de trois cuillers.

On note les événements :

- B_i : « la $i^{\text{ème}}$ cuillerée de confiture va directement dans sa bouche » ;
- A : « Aline a avalé directement au moins deux cuillerées de confiture lorsque la tranche de brioche sera prête à être consommée. »

Calcule la probabilité de l'événement A .

Corrigé : Construisons l'arbre de probabilités de \bar{A} :



$$P(A) = 1 - P(\bar{A}) = 1 - \left(\frac{8}{27} + 3 \times \frac{8}{81}\right) = 1 - \left(\frac{8}{27} + \frac{8}{27}\right) = \frac{11}{27}$$

IV. Variables aléatoires discrètes

Considérons l'expérience aléatoire suivante : On tire au hasard dans une urne contenant une boule rouge R , une boule verte V , et une boule bleue B . Remettons-la dans l'urne et effectuons un second tirage. On a tiré deux boules.

		2 nd tirage		
		R	V	B
1 ^{er} tirage	R	(R, R)	(R, V)	(R, B)
	V	(V, R)	(V, V)	(V, B)
	B	(B, R)	(B, V)	(B, B)

On est en situation d'équiprobabilité donc la probabilité d'avoir au moins une boule bleue est :

$$P(\{(R, B), (V, B), (B, R), (B, V), (B, B)\}) = \frac{5}{9}$$

Complétons l'énoncé :

- Pour chaque boule rouge tirée, on gagne 6€.
- Pour chaque boule verte tirée, on gagne 2€.
- Pour chaque boule bleue tirée, on perd 8€.

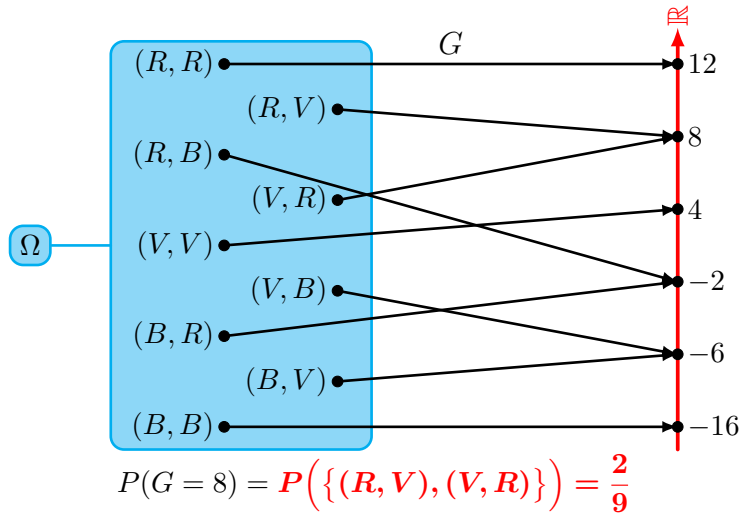
Notons G la variable aléatoire qui à un tirage de deux boules associe le gain du joueur :

$$G: \quad \Omega \quad \longrightarrow \quad \mathbb{R}$$

$$(R, B) \quad \longmapsto \quad -2$$

Les valeurs prises par G sont $G(\Omega) = \{ -16, -6, -2, 4, 8, 12 \}$

On a la situation suivante :



L'espérance de la variable aléatoire G est :

$$\Rightarrow E(G) = 12 \times \frac{1}{9} + 8 \times \frac{2}{9} + 4 \times \frac{1}{9} + (-2) \times \frac{2}{9} + (-6) \times \frac{2}{9} + (-16) \times \frac{1}{9} = 0$$

Ce qui signifie que si on répète l'expérience aléatoire un « grand nombre de fois », la **moyenne** des gains est **nulle**. Autrement dit, ce jeu est **équitable** sur un « grand nombre » de parties.

$$E(G^2) = 12^2 \times \frac{1}{9} + 8^2 \times \frac{2}{9} + 4^2 \times \frac{1}{9} + (-2)^2 \times \frac{2}{9} + (-6)^2 \times \frac{2}{9} + (-16)^2 \times \frac{1}{9} = \frac{624}{9} \simeq 69,3$$

$$\Rightarrow V(G) = E(G^2) - E(G)^2 = 69,3 - 0^2 = 69,3 \text{ et } \sigma_G = \sqrt{69,3} \simeq 8,3$$

Finalement, grande différence entre les statistiques descriptives et les probabilités, c'est que dans le premier cas, l'expérience aléatoire a déjà eu lieu.

- En statistique descriptive, on prélève un échantillon de 100 parties, on mesure la moyenne des gains.
- En probabilités, on « espère » une moyenne nulle.

En mélangeant, ces deux branches des mathématiques, on va estimer la moyenne des gains sur 100 parties, on fera alors des statistiques **inférentielles**.


V. Probabilités et statistiques.

Les variables aléatoires X et Y sont définies sur un même univers Ω tels que :

$$X(\Omega) = \{x_1, x_2, \dots, x_N\} \text{ et } Y(\Omega) = \{y_1, y_2, \dots, y_N\}$$

Probabilités	Statistiques
$E(X) = \sum_{i=1}^N x_i P(X = x_i)$	$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$
E est linéaire	La moyenne est linéaire
$V(X) = \sum_{i=1}^N (E(X) - x_i)^2 P(X = x_i)$ $= E[E(X) - X]$ $= E(X^2) - E(X)^2$	$V(x) = \frac{1}{N} \sum_{i=1}^N (\bar{x} - x_i)^2$ $= \overline{(x - \bar{x})^2}$ $= \overline{x^2} - \bar{x}^2$
$Cov(X, Y) = E[(E(X) - X)(E(Y) - Y)]$ $= E(XY) - E(X)E(Y)$	$Cov(x, y) = \overline{(x - \bar{x})(y - \bar{y})}$ $= \overline{xy} - \bar{x}\bar{y}$
$V(aX + bY) = a^2V(X) + b^2V(Y) + 2abCov(X, Y)$	$V(ay + by) = a^2V(x) + b^2V(y) + 2abCov(x, y)$

Où $E(XY) = \sum_{i,j=1}^N x_i y_j P((X = x_i) \cap (Y = y_j))$

 **Théorème**
 Si X et Y sont deux variables indépendantes alors :

$$E(XY) = E(X)E(Y) ; Cov(X, Y) = 0 ; V(X + Y) = V(X) + V(Y)$$

Exemple n° 16 : Considérons le couple (X, Y) dont la loi est définie par le tableau ci-dessous :

Y \ X	-1	0	1	Total
-1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{4}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
Total	$\frac{5}{16}$	$\frac{3}{8}$	$\frac{5}{16}$	1

$$P((X = -1) \cap (Y = 0)) = \frac{1}{16}$$

$$P((X = 1) \cap (Y = 1)) = \frac{1}{8}$$

$$P(X = 1) = \frac{5}{16}$$

$$P(Y = 1) = \frac{3}{8}$$

Donc, $P((X = 1) \cap (Y = 1)) \neq P(X = 1)P(Y = 1)$

X et Y **ne sont pas indépendantes.**

$$\Rightarrow E(X) = -1 \times \frac{5}{16} + 0 \times \frac{3}{8} + 1 \times \frac{5}{16} = 0$$

$$\Rightarrow E(Y) = -1 \times \frac{3}{8} + 0 \times \frac{1}{4} + 1 \times \frac{3}{8} = 0$$

Ces deux espérances sont **nulles car les lois de X et de Y sont symétriques par rapport à 0.**

$$\Rightarrow E(XY) = (-1)(-1) \times \frac{1}{8} + (1)(-1) \times \frac{1}{8} + (-1)(1) \times \frac{1}{8} + (1)(1) \times \frac{1}{8} = 0$$

$$\Rightarrow Cov(X, Y) = E(XY) - E(X)E(Y) = 0$$



Une covariance nulle n'entraîne pas l'indépendance.

Chapitre 3 - Loi de probabilités.

I. Les lois discrètes.1. Loi de Bernoulli

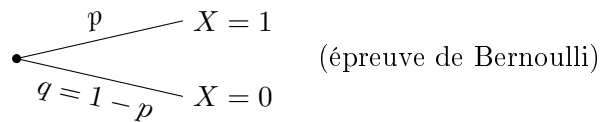
Soit A un événement d'un univers Ω .

Soit X la variable aléatoire :

$$X: \Omega \longrightarrow \{0; 1\}$$

$$\omega \longmapsto \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon } (\omega \in \bar{A}) \end{cases}$$

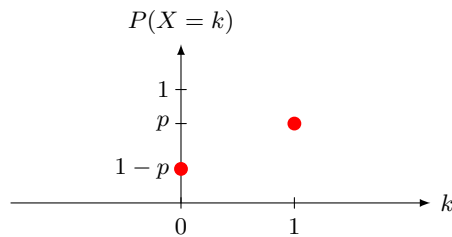
On dit que la variable aléatoire X suit une loi de Bernoulli $\mathcal{B}(p)$ où $p = P(A) = P(X = 1)$:



$$E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) = 0 \times (1 - p) + 1 \times p = p$$

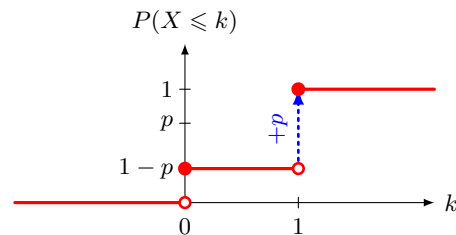
$$E(X^2) = 0^2 \times P(X = 0) + 1^2 \times P(X = 1) = p$$

$$V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1 - p) = pq \text{ donc } \sigma(X) = \sqrt{V(X)} = \sqrt{pq}$$

Loi de X 

$$P(X = -1) = 0; P(X = 0) = 1 - p$$

$$P(X = 0, 4) = 0; P(X = 1, 3) = 0$$

Fonction de répartition F_X 

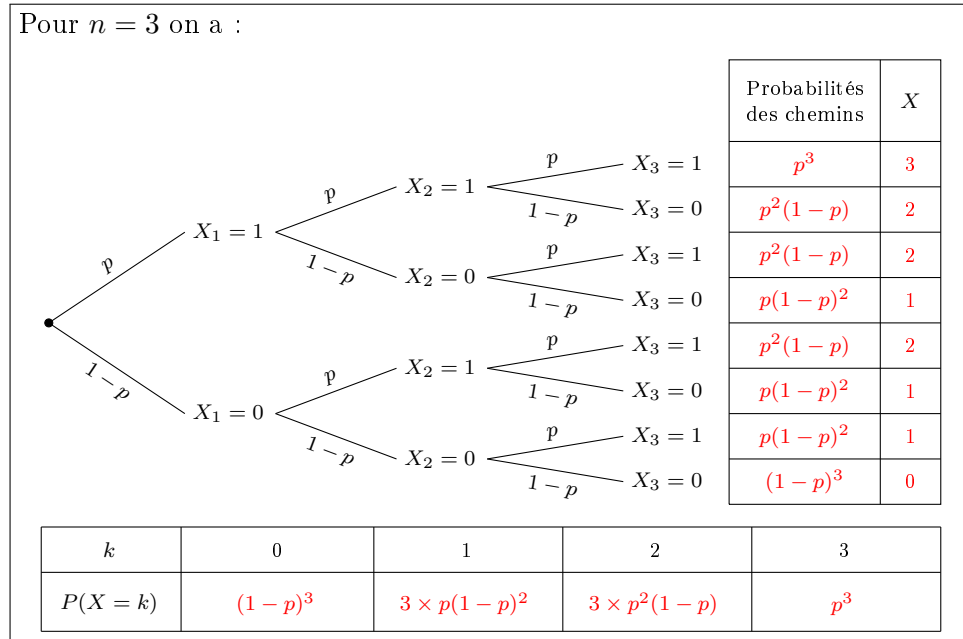
$$F_X(-1) = P(X \leq -1) = 0$$

$$F_X(0, 3) = P(X \leq 0, 3) = P(X = 0) = 1 - p$$

$$F_X(1, 3) = P(X \leq 1, 3) = P(X = 0) + P(X = 1) \\ = (1 - p) + p = 1$$

2. Loi binomiale

On répète successivement n épreuves de Bernoulli **indépendantes** où l'on note à chaque fois la réalisation ou pas d'un événement A , de probabilité $p = P(A)$. A chaque épreuve de Bernoulli, on associe la variable aléatoire X_i , et on pose $X = X_1 + X_2 + \dots + X_n$



On dit que la variable aléatoire X suit une loi de Binomiale $\mathcal{B}(n, p)$:

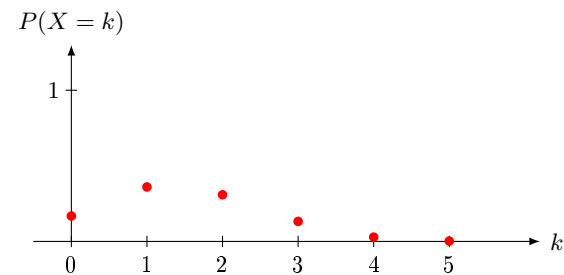
$$P(X = k) = \underbrace{\binom{n}{k}}_{\substack{\text{nb de façons de} \\ \text{réaliser } k \text{ événements } A \\ \text{parmi } n \text{ épreuves.}}} \times p^k(1-p)^{n-k}$$

$$E(X) = E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n) = p + p + \dots + p = np$$

$$V(X) = \sum_{i=1}^n V(X_i) = npq \text{ car les } X_i \text{ sont } \mathbf{\textit{indépendantes}} \text{ donc } \sigma(X) = \sqrt{npq}$$

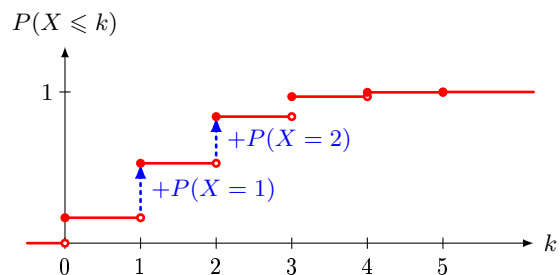
Loi de X où $X \sim \mathcal{B}(5; 0,3)$

Fonction de répartition F_X



$$P(X = 2) = \binom{5}{2} \times 0,3^2 \times 0,343$$

$$= \frac{5 \times 4}{2} \times 0,09 \times 0,7^3 = 0,3087$$

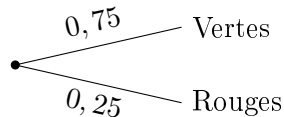


$$F_X(2, 7) = P(X \leq 2, 7)$$

$$= P(X = 0) + P(X = 1) + P(X = 2)$$

Exemple n° 17 : On considère une urne contenant 30 boules vertes et 10 boules rouges. On prélève 8 boules. Pour chacune d'elles, on note sa couleur et on la remet dans l'urne avant d'ôter la suivante. On dit effectuer un tirage de 8 boules **avec remise**. On note X la variable aléatoire qui compte le nombre de boules vertes tirées.

- Les tirages de chaque boules sont indépendants les uns des autres car : **dans l'urne la proportion des boules vertes par rapport au rouges ne change pas.**
- La variable aléatoire X suit une loi binomiale $\mathcal{B}(8; \frac{30}{40})$ car l'expérience aléatoire consiste à répéter successivement de manière **indépendantes** l'épreuve de Bernoulli suivante :



- $P(X = 3) = \binom{8}{3} \times 0,75^3 \times 0,25^5 = \frac{8 \times 7 \times 6}{3 \times 2} \times 0,75^3 \times 0,25^5 = 56 \times 0,75^3 \times 0,25^5 \simeq 0,0231$
- $E(X) = np = 8 \times 0,75 = 6$, $V(X) = npq = 8 \times 0,75 \times 0,25 = 1,5$, et $\sigma_X = \sqrt{1,5} \simeq 1,225$



Tirage avec remise

Une urne contient N boules dont b boules blanches. On tire n boules avec remise, et on note X la variable aléatoire qui compte le nombre de boules blanches dans ce tirage.

- la proportion de boules blanches dans l'urne est $p = \frac{b}{N}$
- la variable aléatoire X suit une loi binomiale $\mathcal{B}(n, p)$

Remarque : Np est **le nombre de boules blanches contenue dans l'urne.**

Exercice n° 2: Un avion peut accueillir 20 personnes. Des statistiques montrent que 25% clients ayant réservé ne viennent pas. Soit X la variable aléatoire : « nombre de clients qui viennent après réservation parmi 20 » .

1. Quelle est la loi de X ?
2. Quelle est son espérance, son écart-type ?
3. Quelle est la probabilité pour que X soit égal à 15 ?

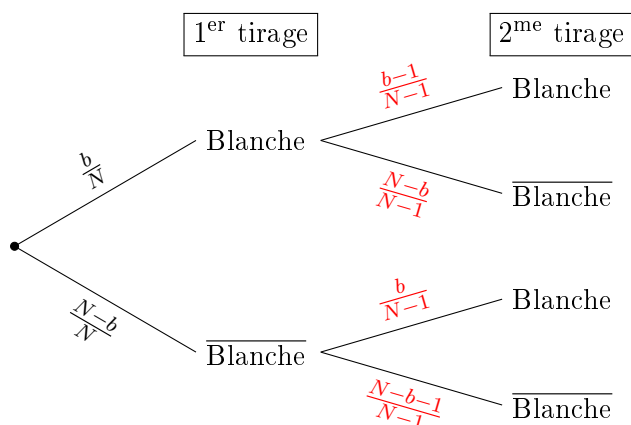
Corrigé :

1. X suit une loi $\mathcal{B}(20; 0,75)$
2. $E(X) = np = 20 \times 0,75 = 15$ et $\sigma_X = \sqrt{npq} = \sqrt{15 \times 0,25} = \sqrt{3,75} \simeq 1,936$
3. $P(X = 15) = \binom{20}{15} \times 0,75^{15} \times 0,25^5 = 15504 \times 0,75^{15} \times 0,25^5 \simeq 0,202$

3. Loi hypergéométrique

Une urne contient N boules dont b boules blanches. On tire n boules de l'urne, mais cette fois-ci, sans remise, et on note X la variable aléatoire qui compte le nombre de boules blanches dans ce tirage.

On n'est plus dans une répétition d'une même épreuve de Bernoulli :



Les tirages successifs sont ici dépendants puisque la composition de l'urne est différente après chaque tirage.

Notons X le nombre de boules blanches tirées. On démontre que $P(X = k) = \frac{\binom{b}{k} \times \binom{N-b}{n-k}}{\binom{N}{n}}$.



Tirage sans remise

Une urne contient N boules dont b boules blanches. On tire n boules **sans** remise, et on note X la variable aléatoire qui compte le nombre de boules blanches dans ce tirage.

- La proportion initiale de boules blanches dans l'urne est $p = \frac{b}{N}$, et $q = 1 - p$ est la proportion initiale des boules qui ne sont pas blanches.
- La variable aléatoire X suit une loi hypergéométrique $\mathcal{H}(n, N, p)$.
- $E(X) = np$ et $Var(X) = npq \left(\frac{N-n}{N-1} \right)$.

Exercice n° 3 : On prend au hasard, en même temps, trois ampoules dans un lot de 15 dont 5 sont défectueuses. On note X le nombre d'ampoules défectueuses.

1. Calcule la probabilité des événements suivants

- A : « exactement une ampoule est défectueuse » .
- B : « au moins une ampoule est défectueuse » ;
- C : « les 3 ampoules sont défectueuses » ;

2. Calcule l'espérance puis l'écart-type de X .

Corrigé : X suit une loi hypergéométrique $\mathcal{H}\left(3, 15, \frac{1}{3}\right)$. On sait que $P(X = k) = \frac{\binom{b}{k} \times \binom{N-b}{n-k}}{\binom{N}{n}}$ où $b = pN = 5$

$$P(X = k) = \frac{\binom{5}{k} \times \binom{10}{3-k}}{\binom{15}{3}} = \frac{\binom{5}{k} \times \binom{10}{3-k}}{455}$$

1.
 - $P(A) = P(X = 1) = \frac{\binom{5}{1} \times \binom{10}{2}}{455} = \frac{5 \times 45}{455} = \frac{45}{91} \simeq 0,494$
 - $P(B) = P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{5}{0} \times \binom{10}{3}}{455} = 1 - \frac{1 \times 120}{455} = \frac{67}{91} \simeq 0,736$
 - $P(C) = P(X = 3) = \frac{\binom{5}{3} \times \binom{10}{3-3}}{455} = \frac{10}{455} = \frac{2}{91} \simeq 0,0220$
2. $E(X) = np = 3 \times \frac{1}{3} = 1$ et $\sigma_X = \sqrt{npq \left(\frac{N-n}{N-1} \right)} = \sqrt{3 \times \frac{1}{3} \times \frac{2}{3} \left(\frac{12}{14} \right)} = \sqrt{\frac{4}{7}} \simeq 0,756$

II. Les lois continues.



Définition:

Une fonction de **densité** ou de distribution de probabilité est une fonction réelle f qui satisfait aux deux conditions suivantes :

$$f \text{ est positive et } \int_{-\infty}^{+\infty} f(x) dx = 1$$



Définition:

On dit qu'une variable aléatoire X a pour densité ou **distribution** de probabilité la fonction f , si pour tous nombres réels a et b tels que $a \leq b$:

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

On dit alors que la variable aléatoire X est **continue**.



Définition:

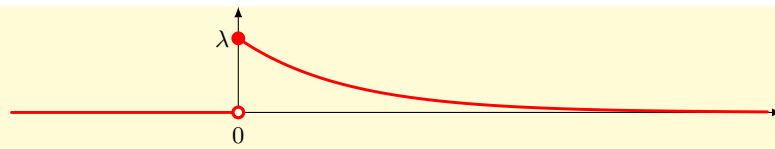
L'**espérance** d'une variable aléatoire continue X , notée $E(X)$ est $\int_{-\infty}^{+\infty} xf(x) dx$

1. Loi exponentielle.



Définition:

Soit λ un nombre réel **strictement positif**. On dit qu'un variable aléatoire X suit une loi exponentielle de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$, si sa densité de probabilité est : $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$



On calcule l'espérance en intégrant par parties :

$$E(X) = \lambda \int_0^{+\infty} x e^{-\lambda x} dx = \left[-x e^{-\lambda x} \right]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx = \frac{1}{\lambda}$$

On calcule de même en intégrant par parties :

$$E(X^2) = \lambda \int_0^{+\infty} x^2 e^{-\lambda x} dx = \left[-x^2 e^{-\lambda x} \right]_0^{+\infty} + 2 \int_0^{+\infty} x e^{-\lambda x} dx = \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2}$$

$$\text{D'où } V(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Théorème

Si une variable aléatoire X suit une loi exponentielle de paramètre $\lambda > 0$, alors :

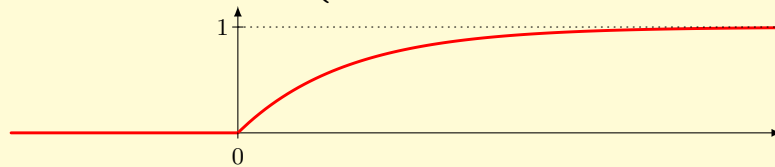
$$E(X) = \frac{1}{\lambda} \text{ et } V(X) = \frac{1}{\lambda^2}$$

Pour $x \geq 0$, la fonction de répartition F est définie par : $F(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$

Théorème

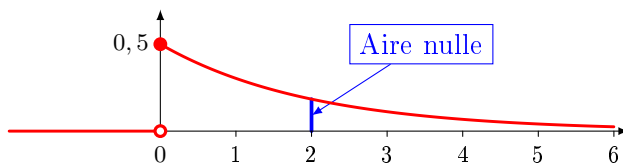
La fonction de répartition d'une variable aléatoire suivant une loi exponentielle de paramètre $\lambda > 0$, est :

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

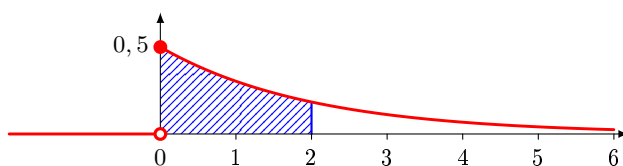


Exemple n° 18 : La variable aléatoire T suit une loi $\mathcal{E}(0,5)$.

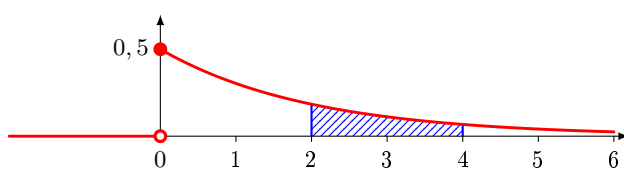
$$\bullet P(T = 2) = \int_2^2 0,5 e^{-0,5x} dx = 0$$



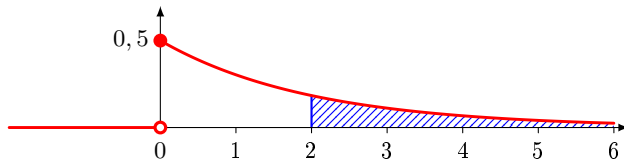
$$\bullet P(T \leq 2) = \int_{-\infty}^2 0,5 e^{-0,5x} dx = F_T(2) = 1 - e^{-0,5 \times 2} = 1 - e^{-1} \simeq 0,632$$



$$\bullet P(2 \leq T \leq 4) = \int_2^4 0,5e^{-0,5x} dx = F_T(4) - F_T(2) = 1 - e^{-2} - (1 - e^{-1}) = e^{-1} - e^{-2} \simeq 0,233$$



$$\bullet P(T \geq 2) = \int_2^{+\infty} 0,5e^{-0,5x} dx = 1 - F_T(2) = 1 - (1 - e^{-1}) = e^{-1} \simeq 0,368$$



Théorème

La loi exponentielle est l'unique loi continue **sans mémoire** : Si X suit une loi exponentielle :

$$\text{pour tout } t > 0 \text{ et tout } h > 0, P_{(X \geq t)}(X \geq t + s) = P(X \geq s)$$

Remarque : Si la durée de votre survie, en années, suivait une loi exponentielle, alors, que vous ayez vécu $t = 4$ ans ou $t = 80$ ans, votre probabilité de vivre encore $s = 20$ ans est la même.

Exercice n° 4: La durée de vie, T , d'un robot, exprimée en années, jusqu'à ce que survienne la première panne est une variable aléatoire qui suit une loi exponentielle de paramètre $\lambda = 0,2$.

1. Quelle est la probabilité qu'un robot n'ait pas eu de panne au cours des deux premières années ?
2. A quel instant t , à un mois près, la probabilité qu'un robot tombe en panne pour la première fois est-elle de 0,5 ?
3. Quelle est la durée de vie moyenne d'un robot ?

Corrigé :

$$1. P(T \geq 2) = e^{-0,2 \times 2} = e^{-0,4} \simeq 0,630$$

$$2. P(T \leq t) = 0,5 \iff 1 - e^{-0,2t} = 0,5 \iff t = \frac{\ln(2)}{0,2} \simeq 3,466$$

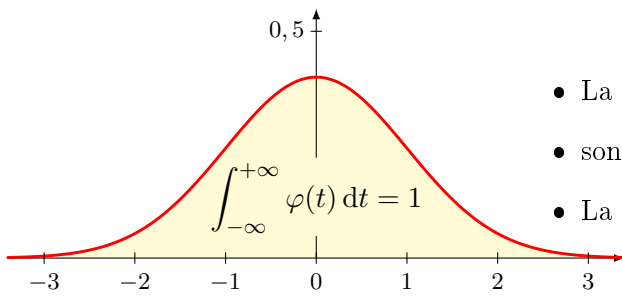
0,46 année durent $0,46 \times 12 = 5,52$ mois. L'instant t cherché est 3 ans et 5 mois à un mois près.

$$3. E(T) = \frac{1}{\lambda} = 5 \text{ ans.}$$

2. La loi normale centrée réduite.

Définition:

On appelle densité de probabilité de **Laplace-Gauss**, la fonction φ définie sur \mathbb{R} par $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

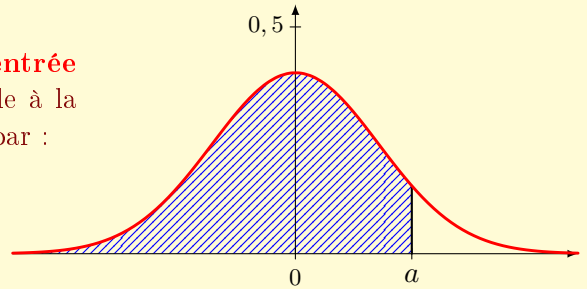


- La fonction φ est paire ;
- son maximum est atteint en 0 ;
- La courbe \mathcal{C}_φ est appelée courbe en cloche ou **courbe de Gauss**.

Définition:

On dit que la variable aléatoire Z suit une loi normale **centrée réduite**, notée $\mathcal{N}(0, 1)$ si sa densité de probabilité est égale à la fonction φ . Sa fonction de répartition, notée Φ , est définie par :

$$\Phi(a) = P(Z \leq a) = \int_{-\infty}^a \varphi(t) dt$$



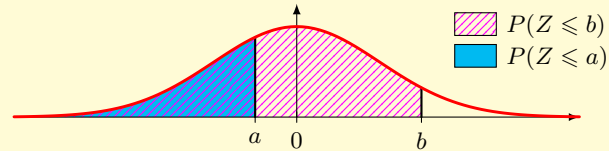
Utilisation de la table de la fonction de répartition de la loi normale centrée réduite :

$$\begin{aligned}
 P(Z \leq 1,27) &\simeq \mathbf{0,8980} & P(Z \leq 0,23) &\simeq \mathbf{0,5910} & P(Z \leq 0,09) &\simeq \mathbf{0,5359} \\
 P(Z \leq 1,5) &\simeq \mathbf{0,9332} & P(Z < 1,5) &\simeq \mathbf{0,9332} & P(Z \geq 1,5) &\simeq \mathbf{1 - 0,9332 = 0,0668}
 \end{aligned}$$

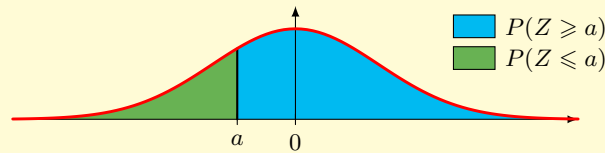
Théorème

Si une variable aléatoire Z suit une loi normale centrée réduite alors pour tout réels a et b , $a < b$ on a :

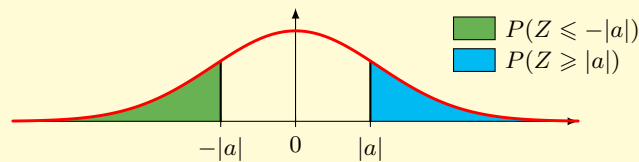
i. $P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$;



ii. $P(Z \geq a) = 1 - P(Z \leq a)$;



iii. $P(Z \leq -|a|) = 1 - P(Z \leq |a|)$

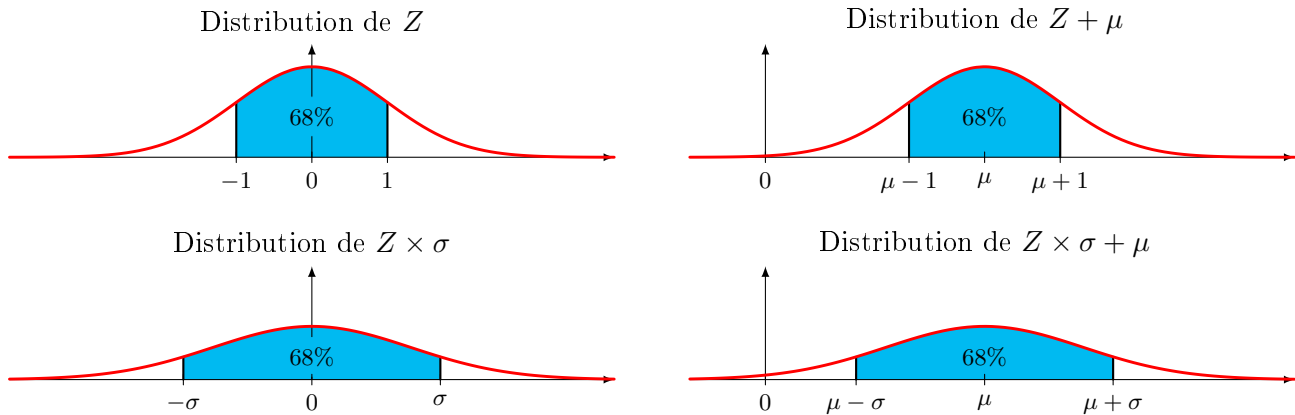


Exemple n° 19 : Z suit une loi normale centrée réduite.

- $P(Z \geq 1,27) \stackrel{i.}{=} 1 - P(Z \leq 1,27) = 1 - \Phi(1,27) \simeq 1 - 0,8980 = 0,102$
- $P(-1 \leq Z \leq 1) \stackrel{i.}{=} P(Z \leq 1) - P(Z \leq -1) \stackrel{iii.}{=} P(Z \leq 1) - (1 - P(Z \leq 1)) = 2P(Z \leq 1) - 1$
 $\simeq 2 \times 0,8413 - 1 = 0,6826$
- $P(Z \geq -0,74) \stackrel{i.}{=} 1 - P(Z \leq -0,74) \stackrel{iii.}{=} 1 - (1 - P(Z \leq 0,74)) = P(Z \leq 0,74) \simeq 0,7704$
- $P(Z = 0,85) = 0$

3. La loi normale $\mathcal{N}(\mu, \sigma)$.

Soit Z une variable aléatoire suivant une loi normale centrée réduite.



Théorème

Soient Z une variable aléatoire normale centrée et réduite, et σ un nombre réel strictement positif. La variable aléatoire $X = Z \times \sigma + \mu$ suit la loi de densité (distribution) :

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{t-\mu}{\sigma}\right)^2}$$

Son espérance $E(X) = \mu$ et son écart-type $\sigma_X = \sigma$.

Définition:

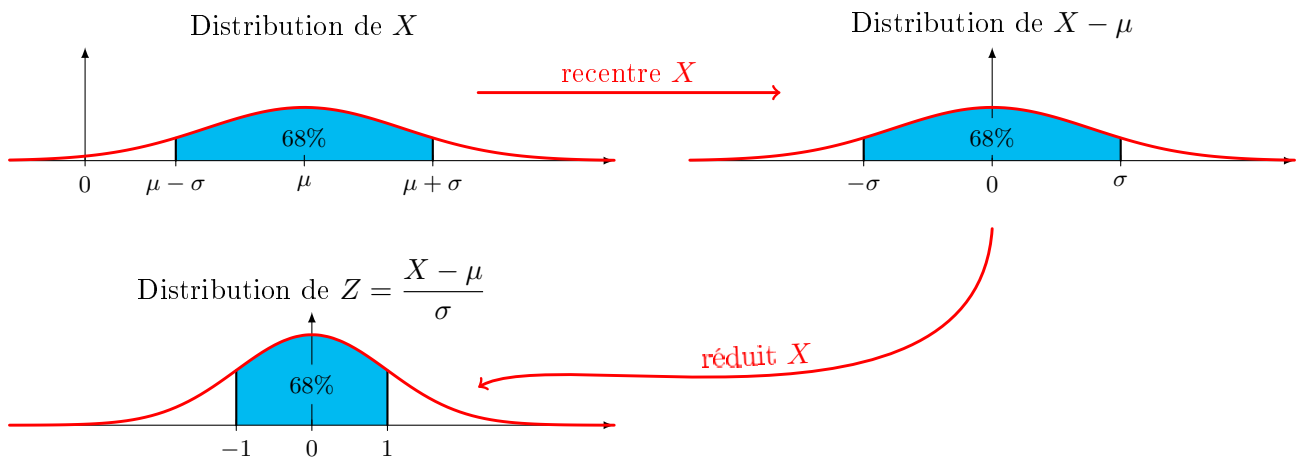


Carl Friedrich Gauss

Etant donné un nombre réel σ strictement positif et un réel μ . La **loi normale** de centre μ et d'écart-type σ , notée $\mathcal{N}(\mu, \sigma)$ est la loi dont la densité de probabilité est

$$\begin{aligned} f: \mathbb{R} &\longrightarrow [0; +\infty] \\ t &\longmapsto \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{t-\mu}{\sigma}\right)^2} \end{aligned}$$

Etant donné une variable aléatoire X suivant une loi $\mathcal{N}(\mu, \sigma)$:



Théorème

Si la variable aléatoire X suit une loi normale $\mathcal{N}(\mu, \sigma)$, alors la variable aléatoire $Z = \frac{X - \mu}{\sigma}$ suit une loi normale centrée réduite.

Exemple n° 20 : X suit une loi normale $\mathcal{N}(5, 2)$.

$$P(X \leq 8) = P\left(\frac{X - 5}{2} \leq \frac{8 - 5}{2}\right) = P(Z \leq 1,5) = \Phi(1,5) \simeq 0,9332 \text{ où } Z \sim \mathcal{N}(0, 1).$$

Théorème

Si les variables aléatoires X_1 et X_2 suivent des lois normales indépendantes alors :

- $X_1 + X_2$ suit une loi normale ;
- pour tout réel a , la variable aléatoire aX_1 suit une loi normale.

Exemple n° 21 : X suit une loi normale $\mathcal{N}(2, 3)$, Y suit une loi normale $\mathcal{N}(5, 1)$, et X et Y sont indépendantes.

1. La loi de $X + Y$:

- | | |
|--|--|
| • $E(X + Y) = E(X) + E(Y) = 5$ | • $\sigma_{X+Y} = \sqrt{4} = 2$ |
| • $V(X + Y) = V(X) + V(Y) = 4$
car X et Y sont indépendantes | • $(X + Y)$ suit une loi normale $\mathcal{N}(5, 2)$
car X et Y sont indépendantes |

2. La loi de $2X - 3Y$:

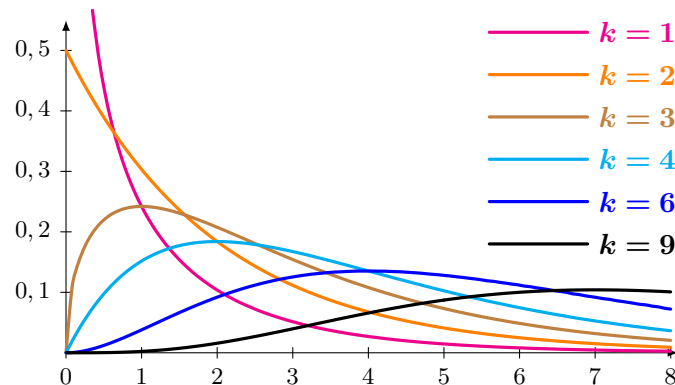
- $E(2X - 3Y) = E(2X) + E(-3Y) = 2 \times 2 + (-3) \times 5 = -11$
- $V(2X - 3Y) = (2^2)V(X) + (-3)^2V(Y) = 4 \times 3 + 9 \times 1 = 21$ car X et Y sont **indépendantes**
- $\sigma_{2X-3Y} = \sqrt{21}$
- $(2X - 3Y)$ suit une loi normale $\mathcal{N}(-11, \sqrt{21})$ car X et Y sont **indépendantes**

4. Loi du χ^2

Définition:
 Soient k variables aléatoires indépendantes X_i suivant une loi normale d'espérance μ_i et écart-type σ_i . Par définition la variable aléatoire

$$Y = \sum_{i=1}^k \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2$$

suit une loi du χ^2 à k degrés de liberté.



- Son espérance vaut k et n'est définie que pour $k \geq 2$;
- Sa variance vaut $2k$ et n'est définie que pour $k \geq 3$.

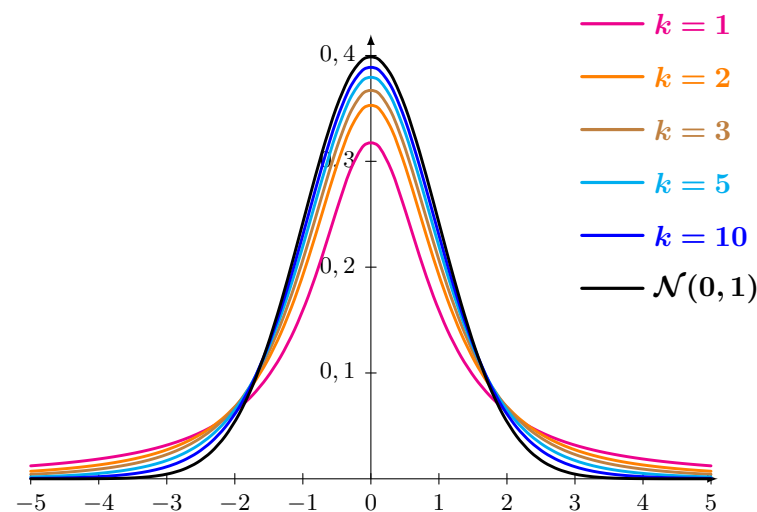
Remarque : En pratique, lorsque $k \geq 100$, on approche la loi du χ^2 à k degré de liberté par la loi normale $\mathcal{N}(k, \sqrt{2k})$.

5. Loi de Student

Définition:
 Soit Z une variable aléatoire de loi normale centrée et réduite et soit U une variable indépendante de Z et distribuée suivant la loi du χ^2 à k degrés de liberté. Par définition la variable

$$T = \frac{Z}{\sqrt{U/k}}$$

suit une loi de Student à k degrés de liberté.



- Son espérance vaut 0 et n'est définie que pour $k \geq 2$;
- Sa variance vaut $\frac{k}{k-2}$ et n'est définie que pour $k \geq 3$.

Lorsque k tend vers $+\infty$ la loi de Student tend vers la loi normale centrée réduite.

Remarque : En pratique lorsque $k \geq 30$, on approche la loi de Student à k degrés de liberté par une loi normale centrée réduite.



William Sealy Gosset



connu sous le pseudonyme **Student** est un statisticien anglais (1876 – 1937) qui inventa la distribution ne portant pas son nom.

Il était un employé de la brasserie Guinness qui lui demanda d'utiliser un pseudonyme pour diverses raisons. *Peut-être prétendait-il que la qualité de leurs produits était improbable...*

Chapitre 4 - Statistiques inférentielles : estimation par intervalles de confiances.

L' **inférence statistique** est l'ensemble des méthodes mathématiques utilisées lors d'un sondage sur un échantillon afin de prédire le comportement d'une population, et ce, avec un contrôle sur l'erreur.

Ces méthodes se déclinent en deux catégories :

- l'estimation par **intervalle de confiance** ;
- les **tests d'hypothèses**.

Ces méthodes reposent essentiellement sur le théorème de la LIMITE CENTRALE.

I. Variables d'échantillonnages

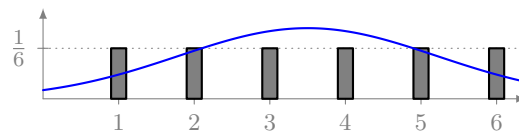
1. Le théorème de la limite centrale

Etude n° 2 : Somme répétée de variables indépendantes suivant une loi $\mathcal{U}(6)$.

On lance n dés cubiques, on note pour chacun d'eux X_n le résultat. On pose $S_n = \sum_{i=1}^n X_i$ la somme de tous les résultats.

Pour $n = 1$:

k	1	2	3	4	5	6
$P(S_1 = k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



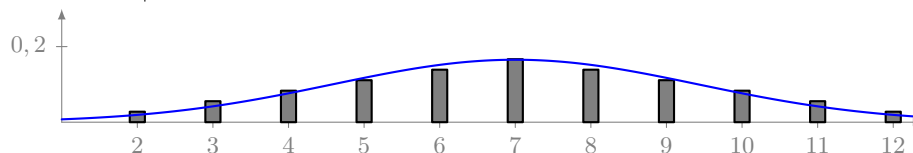
$$E(S_1) = E(X_1) = (1 \times \frac{1}{6}) + (2 \times \frac{1}{6}) + (3 \times \frac{1}{6}) + (4 \times \frac{1}{6}) + (5 \times \frac{1}{6}) + (6 \times \frac{1}{6}) = \frac{21}{6} = \frac{7}{2} = 3,5$$

$$\begin{aligned} V(S_1) &= V(X_1) = E(X_1^2) - E(X_1)^2 \\ &= (1^2 \times \frac{1}{6}) + (2^2 \times \frac{1}{6}) + (3^2 \times \frac{1}{6}) + (4^2 \times \frac{1}{6}) + (5^2 \times \frac{1}{6}) + (6^2 \times \frac{1}{6}) - (\frac{7}{2})^2 \\ &= \frac{91}{6} - \frac{49}{4} = \frac{35}{12} \simeq 2,917 \end{aligned}$$

$$\sigma(S_1) = \sqrt{V(S_1)} \simeq 1,708$$

Pour $n = 2$:

k	2	3	4	5	6	7	8	9	10	11	12
$P(S_2 = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



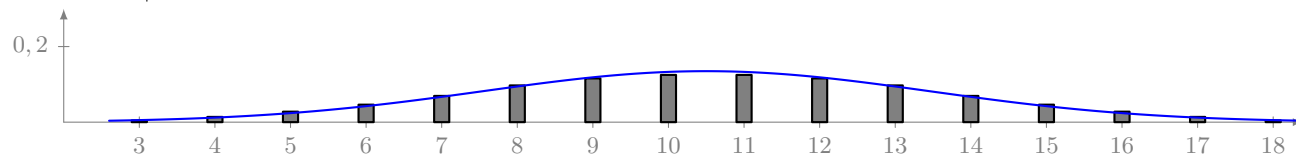
$$E(S_2) = E(X_1 + X_2) = 3,5 + 3,5 = 7$$

$$V(S_2) = V(X_1 + X_2) = V(X_1) + V(X_2) = 2 \times \frac{35}{12} \simeq 5,833$$

$$\sigma(S_2) = \sqrt{V(S_2)} \simeq 2,415$$

Pour $n = 3$:

k	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$P(S_3 = k)$	$\frac{1}{6^3}$	$\frac{3}{6^3}$	$\frac{6}{6^3}$	$\frac{10}{6^3}$	$\frac{15}{6^3}$	$\frac{21}{6^3}$	$\frac{25}{6^3}$	$\frac{27}{6^3}$	$\frac{27}{6^3}$	$\frac{25}{6^3}$	$\frac{21}{6^3}$	$\frac{15}{6^3}$	$\frac{10}{6^3}$	$\frac{6}{6^3}$	$\frac{3}{6^3}$	$\frac{1}{6^3}$



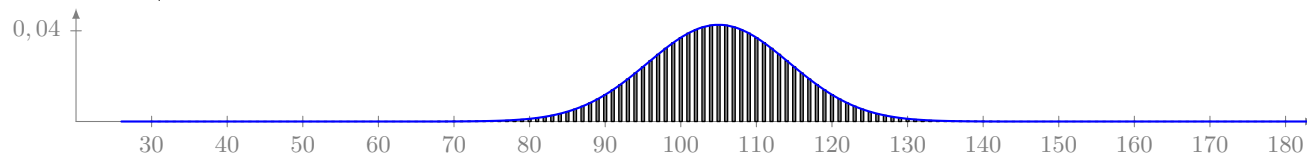
$$E(S_3) = E(X_1 + X_2 + X_3) = 3 \times 3,5 = 10,5$$

$$V(S_3) = V(X_1 + X_2 + X_3) = 3 \times V(X_1) = 3 \times \frac{35}{12} = 8,75$$

$$\sigma(S_3) = \sqrt{V(S_3)} \simeq 2,958$$

Pour $n = 30$:

k	30	31	32	33	34	...	90	...	100	...	105	...	110	...	140	...	180
$P(S_{30} = k)$	$\frac{1}{6^{30}}$	$\frac{30}{6^{30}}$	$\frac{465}{6^{30}}$	$\frac{4960}{6^{30}}$	$\frac{40920}{6^{30}}$...	0,0119	...	0,03688	...	0,04242	...	0,03688	...	0,000031025	...	$\frac{1}{6^{30}}$



$$E(S_{30}) = E(X_1 + \dots + X_{30}) = 30 \times 3,5 = 105$$

$$V(S_{30}) = V(X_1 + \dots + X_{30}) = 30 \times V(X_1) = 30 \times \frac{35}{12} = 87,5$$

$$\sigma(S_{30}) = \sqrt{V(S_{30})} \simeq 9,354$$

Théorème de la limite centrale



Pierre-Simon de Laplace publie en 1812, le théorème de Laplace, appelé aujourd'hui théorème central limite...

Soient X_1, X_2, \dots, X_n des variables aléatoires **indépendantes**, suivant toutes la **même loi**, admettant une moyenne μ et un écart-type $\sigma \neq 0$ alors pour n **suffisamment grand** :

- la loi normale $\mathcal{N}(n\mu, \sigma\sqrt{n})$ est une bonne approximation de la variable aléatoire $S_n = \sum_{i=1}^n X_n$;
- la loi normale $\mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ est une bonne approximation de la variable aléatoire $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_n$.

Remarque :

- i. « n suffisamment grand » sera précisé par la suite, en général, $n \geq 30$ est suffisant.
- ii. A défaut de démontrer ce théorème, on peut en expliquer les paramètres :

- $E(S_n) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n \mu = n\mu$;

Et donc, $E(\bar{X}_n) = E\left(\frac{1}{n}S_n\right) = \mu$.

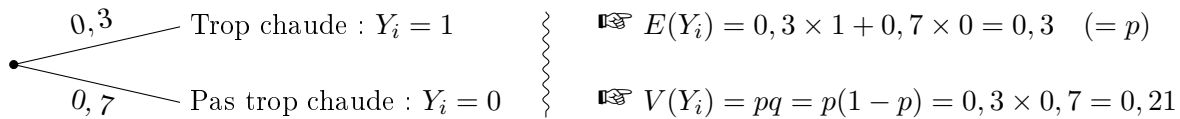
- $V(S_n) = V\left(\sum_{i=1}^n X_i\right) \stackrel{\text{indépendance}}{=} \sum_{i=1}^n V(X_i) = \sum_{i=1}^n \sigma^2 = n\sigma^2$, donc l'écart-type de S_n est $\sqrt{n}\sigma$.

Ainsi, $V(\bar{X}_n) = V\left(\frac{1}{n}S_n\right) = \left(\frac{1}{n}\right)^2 V(S_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$. L'écart-type de \bar{X}_n est $\frac{\sigma}{\sqrt{n}}$.

2. Application à la loi binomiale.

Etude n° 3 : A la sortie d'une machine-outil, 30% des pièces sont trop chaudes pour être assemblées. Sur un échantillon de 30 pièces, on note S la variable aléatoire donnant le nombre de pièces trop chaudes.

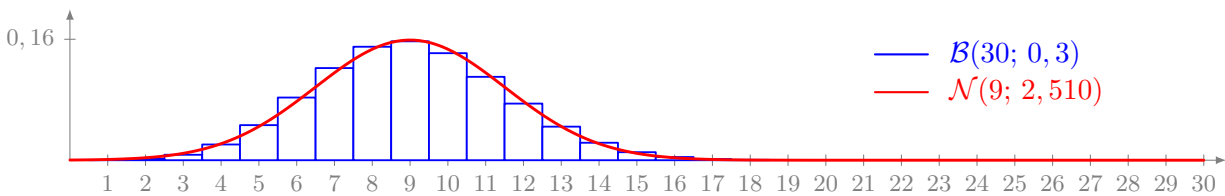
A chacune des 30 pièces, on peut associer une variable aléatoire Y_i et l'épreuve de Bernoulli suivante :



Comme les Y_i sont indépendants, la variable S suit une loi binomiale $\mathcal{B}(n; p) = \mathcal{B}(30; 0,3)$:

- Son espérance est $np = 30 \times 0,3 = 9$;
- Sa variance est $npq = 9 \times 0,7 = 6,3$;
- Son écart-type est $\sqrt{npq} = \sqrt{6,3} \simeq 2,510$.

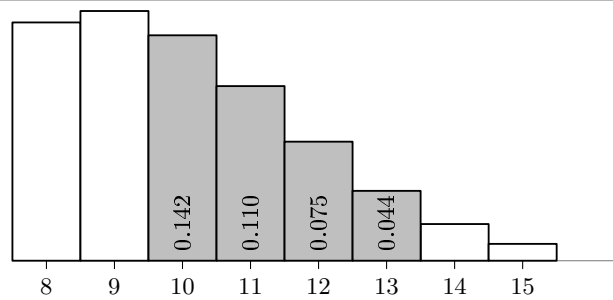
Or, S est une somme de variables aléatoires indépendantes suivant la même loi : $S = Y_1 + Y_2 + \dots + Y_n$.
Donc, d'après le théorème de la limite centrale, S peut être approchée par la loi normale $\mathcal{N}(9; 2,510)$.



Calculons $P(10 \leq S \leq 13)$

- Avec la loi binomiale $\mathcal{B}(30; 0,3)$:

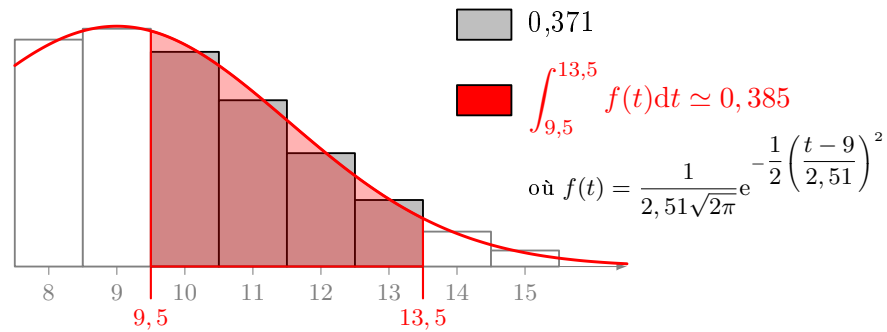
k	10	11	12	13
$P(S = k)$	0,142	0,110	0,075	0,044



$$P(10 \leq S \leq 13) = 0,142 + 0,11 + 0,075 + 0,044 = 0,371$$

Les rectangles étant le largeur 1, $P(10 \leq S \leq 13)$ est l'aire grisée.

- Avec l'approximation normale $\mathcal{N}(9; 2,510)$:

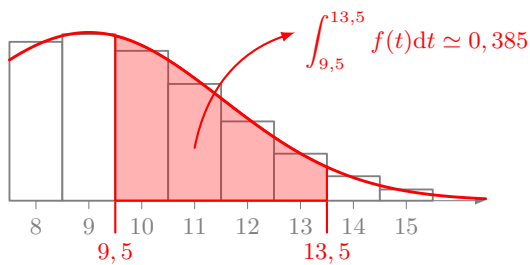


$$P(10 \leq S \leq 13) \simeq 0,385$$

Remarque : L'erreur d'approximation de la loi binomiale par la loi normale est

$$|0,385 - 0,371| = 0,014 = 1,4\%$$

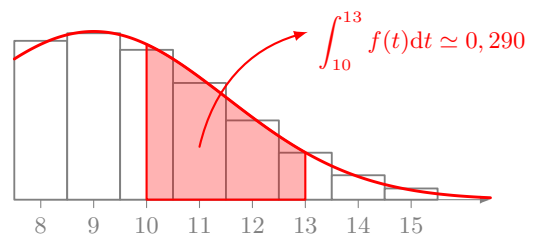
Avec correction de continuité :



L'erreur d'approximation est de :

$$|0,385 - 0,371| = 0,014 = 1,4\%.$$

Sans correction de continuité :



L'erreur d'approximation est de :

$$|0,290 - 0,371| = 0,081 = 8,1\%.$$



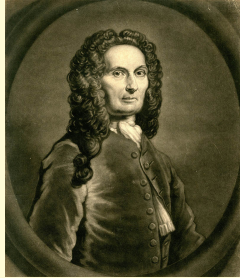
Correction de continuité

Si X suit une loi binomiale et Z une loi normale de même espérance et de même écart-type (non nul), alors, on peut approcher :

- $P(X = k)$ par $P(k - \frac{1}{2} \leq Z \leq k + \frac{1}{2})$;
- $P(a \leq X \leq b)$ par $P(a - \frac{1}{2} \leq Z \leq b + \frac{1}{2})$;
- $P(a \leq X)$ par $P(a - \frac{1}{2} \leq Z)$;
- $P(X \leq b)$ par $P(Z \leq b + \frac{1}{2})$.



Approximation d'une loi binomiale par une loi normale



En pratique, on pourra faire l'approximation d'une loi binomiale par une loi normale de même espérance et de même écart-type, lorsque l'on aura les conditions suivantes sont satisfaites :

$$n \geq 30, np \geq 5, \text{ et } nq \geq 5$$

De Moivre est le premier, en 1733, à faire apparaître la loi normale comme loi limite d'une loi binomiale.

Remarque : dans notre étude, les conditions sont satisfaites :

$$n = 30 \geq 30, np = 30 \times 0,3 = 9 \geq 5 \text{ et } nq = 30 \times 0,7 = 21 \geq 5$$

Mais que signifie $np = 9$? np est l'espérance de S . Donc, si on répète 30 fois l'expérience aléatoire de l'étude n° 2, ou si on prend un échantillons de 30 expériences aléatoires de l'étude n° 2, alors en moyenne $np = 9$ pièces seront trop chaudes, et $nq = 21$ ne le seront pas.

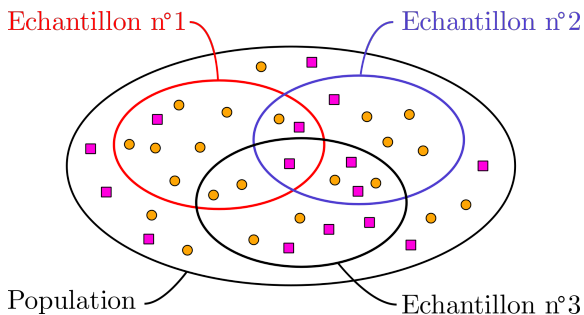
II. Intervalles de confiance pour une proportion.

Objectif : Nous cherchons des informations sur une population à partir de l'étude d'un échantillon. Ce type de situation se rencontre fréquemment dans l'industrie. Il n'est pas possible en général d'étudier la population entière, car cela pourrait prendre trop de temps, reviendrait trop cher, ou serait aberrant dans le cas de contrôle de qualité entraînant la destruction des pièces.

Les statistiques descriptives donnent une réponse ponctuelle à cet objectif. On calcule, par exemple, la moyenne sur l'échantillon, et on estime qu'il est à peu près le même sur la population.

Les statistiques inférentielles offrent une réponse plus rigoureuse avec un contrôle sur l'erreur.

Etude n° 4 : Regardons cela de plus près. Considérons une population de carrés et de ronds, dont on cherche la proportion p de ronds à partir d'échantillons :



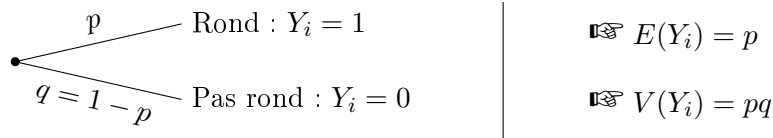
	Ronds	Carrés	Total	Proportion de ronds
Echantillon n° 1	9	3	12	0,75
Echantillon n° 2	7	5	12	0,58
Echantillon n° 3	6	6	12	0,50
Population	22	15	37	0,59

On constate qu'aucun des échantillons ne donne la proportion exacte de ronds. En particulier, l'échantillon n° 1 n'est pas **représentatif** de la population.

1. Modélisation de la situation

On suppose que les échantillons aient tous **le même** nombre $n = 50$ d'individus et que les conditions soient réunies pour utiliser le théorème de la limite centrale, en particulier $n \geq 30$.
 Notre objectif est d'estimer la proportion p de la population.

Dans un échantillon de n individus, chacun d'eux se comporte comme l'épreuve de Bernoulli suivante :

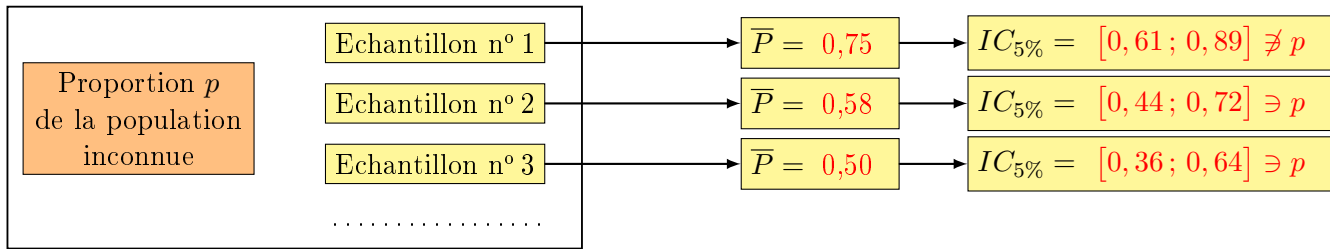


Le nombre d'individus ronds dans un échantillon est la variable aléatoire $S = \sum_{i=1}^n Y_i$. S suit une loi $\mathcal{B}(n, p)$.

La proportion d'individus ronds dans un échantillon est la variable aléatoire $\bar{P} = \frac{S}{n} = \frac{1}{n} \sum_{i=1}^n Y_i$.

$$\Leftrightarrow E(\bar{P}) = \frac{1}{n} E(S) = \frac{np}{n} = p \quad \Leftrightarrow V(\bar{P}) = \frac{1}{n^2} V(S) = \frac{npq}{n^2} = \frac{pq}{n}$$

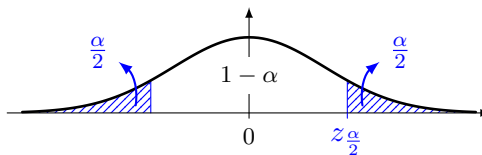
Ainsi, d'après le théorème de la limite centrale, \bar{P} suit approximativement une loi $\mathcal{N}\left(p; \sqrt{\frac{pq}{n}}\right)$



La variable aléatoire $Z = \frac{\bar{P} - p}{\sqrt{\frac{pq}{n}}}$ suit une loi $\mathcal{N}(0; 1)$. Pour un α donné, il existe un $z_{\frac{\alpha}{2}}$ tel que :

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{P} - p}{\sqrt{\frac{pq}{n}}} \leq z_{\frac{\alpha}{2}}\right) \geq 1 - \alpha$$

$$P\left(\bar{P} - z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}} \leq p \leq \bar{P} + z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}\right) \geq 1 - \alpha$$



Prenons $\alpha = 5\%$, on sait que $z_{\frac{\alpha}{2}} = 1,96$ et on calcule $z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}} = 1,96 \sqrt{\frac{0,59 \times 0,41}{50}} \simeq 0,14$

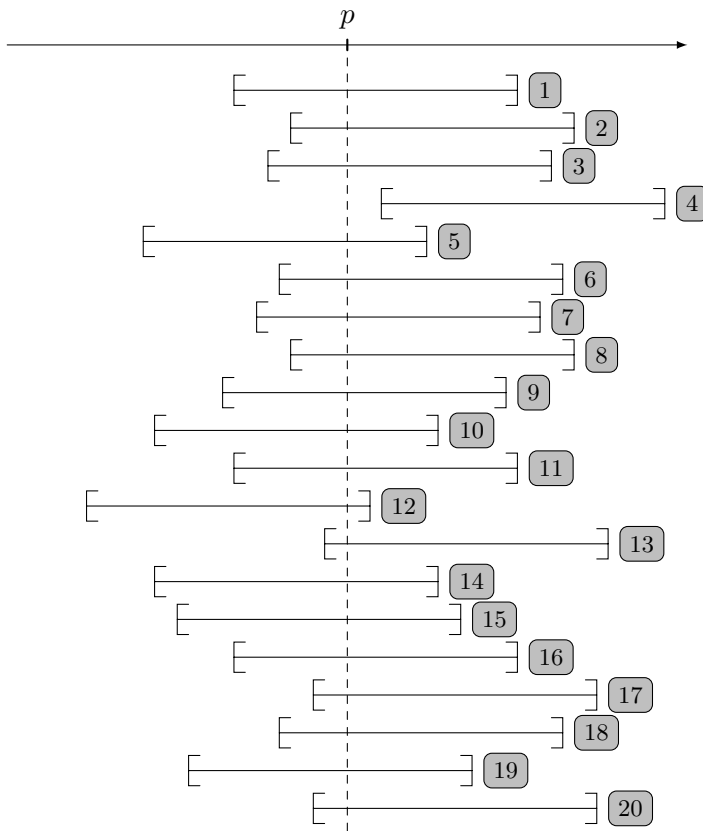
$$IC_{5\%} = \left[\bar{P} - z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}, \bar{P} + z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}\right] \simeq [\bar{P} - 0,14, \bar{P} + 0,14]$$

$IC_{5\%}$ est un **intervalle de confiance**. Ce n'est plus une estimation ponctuelle, mais par intervalle.

- Si $\bar{P} = 0,75$ alors $IC_{5\%} = [0,75 - 0,14; 0,75 + 0,14] = [0,61; 0,89]$
 On constate que la proportion cherchée $p = 0,59 \notin [0,61; 0,89]$.
 L'échantillon n° 1 n'est pas **représentatif** de la population.
- Si $\bar{P} = 0,58$ alors $IC_{5\%} = [0,58 - 0,14; 0,58 + 0,14] = [0,44; 0,72]$
 On constate que la proportion cherchée $p = 0,59 \in [0,44; 0,72]$.
 L'échantillon n° 2 est **représentatif** de la population.

- Si $\bar{P} = 0,50$ alors $IC_{5\%} = [0,50 - 0,14; 0,50 + 0,14] = [0,36; 0,64]$

On remarque que le p n'est pas dans l'intervalle de confiance du 1^{er} échantillon, mais dans ceux des deux autres. Le premier intervalle n'était pas représentatif de la population. L'estimation est fautive pour cet échantillon. On avait $\alpha = 5\%$ de chance de choisir un échantillon dont l'intervalle de confiance ne contienne pas p , autrement dit, $1 - \alpha = 95\%$ de chances de choisir un échantillon dont l'intervalle de confiance contienne p . $1 - \alpha$ est le niveau de confiance.



$\alpha = 5\%$ d'erreur signifie qu'en moyenne sur 20 échantillons prélevés au hasard, un seul intervalle de confiance ne contient pas la proportion p .

Ici, sur 20 simulations d'échantillons, seule l'intervalle de confiance calculé à partir de l'échantillon n° 4 ne convient pas.

aux idées fausses sur une éventuelle localisation de p dans l'intervalle de confiance : la proportion p peut-être n'importe où dans l'intervalle.

Problème : Pour calculer les bornes des intervalles de confiances, on a utilisé p qu'on ne connaît pas puisque qu'on cherche à l'estimer !

En pratique, on a un seul échantillon, donc une seule réalisation de la variable aléatoire \bar{P} . Cette réalisation, on la note avec un chapeau : \hat{p} et on l'utilise à la place de p pour calculer les bornes de l'intervalle de confiance.

Dans notre étude, pour le deuxième échantillon on a observé $\hat{p} = 0,58$: $z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1,96 \sqrt{\frac{0,58 \times (1-0,58)}{50}} \simeq 0,14$

On retrouve le même intervalle de confiance : $IC_{5\%} = [0,44; 0,72]$


Questions : Supposons que la population soit constituée de $N = 21400$ individus (effectif total) et qu'on ait prélevé l'échantillon n° 2 pour faire notre étude statistique ($n = 50$).

1. Combien de ronds estime-t-on avoir dans la population avec un niveau de confiance de 95% ? On a trouvé $IC_{5\%} = [0,44, 0,72]$ donc on estime, au niveau de confiance 95%, qu'il y a entre $0,44 \times 21400 = 9416$ et $0,72 \times 21400 = 15408$ individus ronds.

2. On reprend l'étude avec un niveau de confiance de 99%.

- (a) Détermine le nouvel intervalle de confiance ? Le seul changement, c'est la valeur de α donc de $z_{\frac{\alpha}{2}}$.
 $\alpha = 0,01 \implies z_{\frac{\alpha}{2}} = 2,58$.
 $z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} = 2,58 \sqrt{\frac{0,58 \times (1-0,58)}{50}} \simeq 0,18$ donc $IC_{1\%} = [0,58 - 0,18 ; 0,58 + 0,18] = [0,40 ; 0,76]$
- (b) Combien de ronds estime-t-on avoir dans la population avec un niveau de confiance de 99% ? On estime, au niveau de confiance 99%, qu'il y a entre $0,40 \times 21400 = 8560$ et $0,76 \times 21400 = 16264$ individus ronds.
- (c) Pourquoi l'intervalle $IC_{1\%}$ est-il plus grand ? L'intervalle $IC_{1\%}$ est plus grand, c'est normal, on a moins de chance de se tromper.


2. Intervalles de confiances d'une proportion.

 **Notations :**

Population :
 p : proportion à estimer
 N : effectif total

Echantillon :
 \hat{p} : proportion observée sur l'échantillon
 n : effectif de l'échantillon

$IC_{\alpha\%} = [\hat{p} - \dots ; \hat{p} + \dots]$
 $P(p \notin IC_{\alpha}) = \alpha$

 **Théorème : intervalle de confiance d'une proportion.**
 Si $n \geq 30$, $n\hat{p} \geq 5$ et $n(1 - \hat{p}) \geq 5$, alors l'intervalle de confiance au risque de α est :

$$IC_{\alpha\%} = \left[\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} , \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right]$$

La probabilité que p soit dans cet intervalle est $1 - \alpha$ (niveau de confiance).

Exercice n° 5: Dans une boîte contenant 500 vis, parmi 42 tirées au hasard, 17 sont à têtes plates.

- Les conditions d'application du théorème sont-elles vérifiées ?
 $n \geq 30$, $n\hat{p} = 17 \geq 5$ et $n\hat{q} = 42 - 17 = 25 \geq 5$. Elles le sont.
- Estimez le nombre de vis à têtes plates dans cette boîte avec un risque $\alpha = 1\%$.

Il s'agit d'un tirage avec remise d'où $IC_{\alpha\%} = \left[\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} ; \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right]$

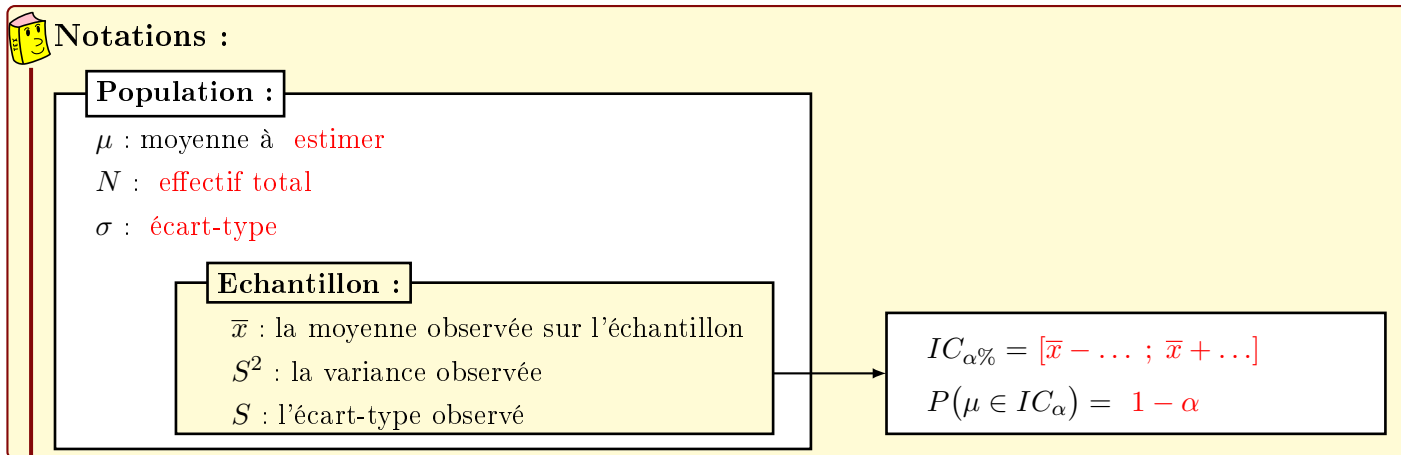
$\hat{p} = \frac{17}{42} \simeq 0,405$. Table de l'écart réduit : $z_{\frac{\alpha}{2}} = 2,576$

$$IC_{1\%} = \left[0,405 - 2,576 \sqrt{\frac{0,405 \times 0,595}{42}} ; 0,405 + 2,576 \sqrt{\frac{0,405 \times 0,595}{42}} \right]$$

$$IC_{1\%} = [0,405 - 0,195 ; 0,405 + 0,195] = [0,210 ; 0,600]$$

On estime au risque de 1%, d'avoir entre 105 et 300 vis à têtes plates dans cette boîte.

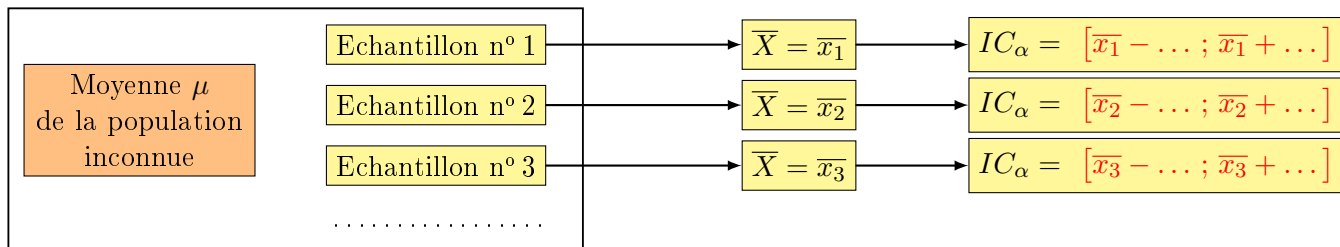
III. Intervalles de confiances pour une moyenne.



1. L'écart-type σ de la population est connu.

Etude n° 5 : Supposons que l'on cherche la taille moyenne d'une population d'effectif total N . On prélève un échantillon de n individus. Chaque individu de l'échantillon, numéroté de 1 à n donne sa taille x_i . Sur cet échantillon, on a une estimation ponctuelle de la moyenne des tailles : $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

D'un point de vue probabiliste, notons X_k la variable aléatoire qui à au k-ième individu d'un échantillon donne sa taille. $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ est la variable aléatoire qui à un échantillon associe sa moyenne :



Cas n° 1 : l'échantillon est **grand** ($n \geq 30$).

On sait que $E(\bar{X}) = \mu$ et $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$, donc, en utilisant le théorème de la limite centrale, on obtient :

$$\bar{X} \text{ suit approximativement une loi } \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Il s'en suit : $P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$ soit $P\left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$

On obtient donc l'intervalle de confiance : $IC_{\alpha} = \left[\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$

Cas n° 2 : l'échantillon est **petit** ($n < 30$) et X suit une loi normale.

Si on sait que les valeurs sont **normalement** distribuées, autrement dit, X suit une loi normale, alors la somme de lois normales indépendantes étant une loi normale, \bar{X} suit une loi normale et on retrouve :

$$IC_{\alpha} = \left[\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

Cas n° 3 : l'échantillon est **petit** ($n < 30$) et X suit une loi quelconque.

On ne peut rien dire sur la loi de \bar{X} ... Mais nous pouvons nous servir de l'inégalité de Bienaymé-Tchebychev :

$$P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2} \text{ pour tout } a > 0$$

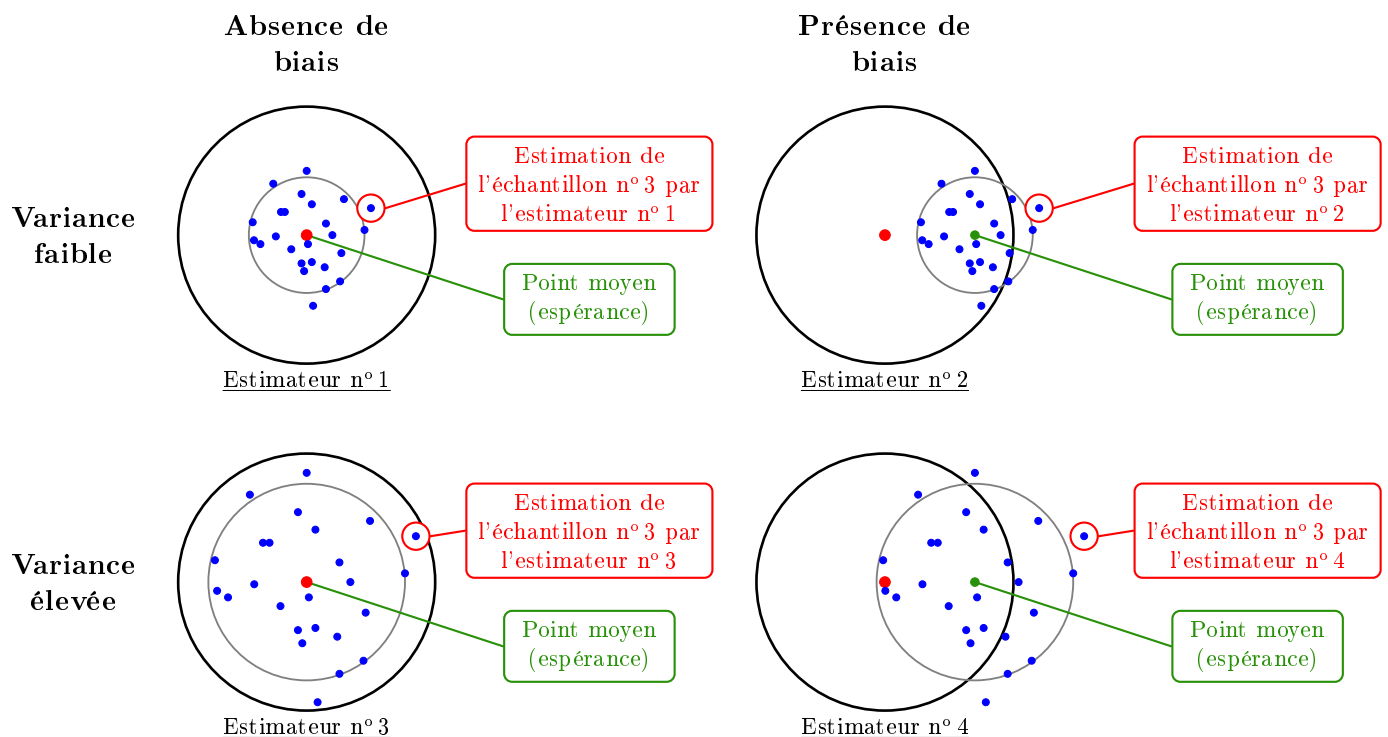
qui appliquée à \bar{X} en prenant $a^2 = \frac{V(\bar{X})}{\alpha}$ donne $P\left(\bar{X} - \frac{\sigma_{\bar{X}}}{\sqrt{\alpha}} \leq \mu \leq \bar{X} + \frac{\sigma_{\bar{X}}}{\sqrt{\alpha}}\right) \geq \alpha$

et permet de construire l'intervalle de confiance :

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

2. L'écart-type σ de la population est inconnu.

Les qualités d'un estimateur dépendent de la formule qu'on utilise pour le calculer et de la façon dont l'échantillon a été choisi. Dans ce cours, on supposera toujours que les échantillons sont bien constitués. Le schéma ci-dessous montre les valeurs prises par quatre estimateurs sur 25 échantillons différents. Chaque point correspond à la valeur prise par l'un des estimateurs sur un échantillon. Le point central, en rouge, étant la valeur qu'on cherche à estimer.



Le **point moyen** correspond à la moyenne des 25 estimations.

Chaque cercle gris est centré sur l'espérance de l'estimateur étudié et son rayon est son écart-type.



Définition:

On dit qu'un estimateur est :

- **biaisé** si son espérance sur les échantillons n'est pas égale à la valeur qu'il doit estimer sur la population.
- **efficace** si sa variance sur les échantillons est plus petite que celle de tout autre estimateur.

$S^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$ est un bon estimateur de la variance de la population à partir de l'échantillon, mais il est biaisé : son espérance n'est pas la variance σ^2 de la population, mais $\frac{n-1}{n}\sigma^2$. Donc, on le « corrige » pour qu'il soit sans biais :



Définition:

Etant donné un échantillon (x_k) de taille n d'une population, on appelle **variance corrigée** et **écart-type corrigé** les estimateurs :

$$S_c^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2 \text{ et } S_c = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2}.$$

Remarque : Comme $S^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$, on passe de la variance à la variance corrigée par la formule :

$$S_c^2 = \frac{n}{n-1} S^2$$



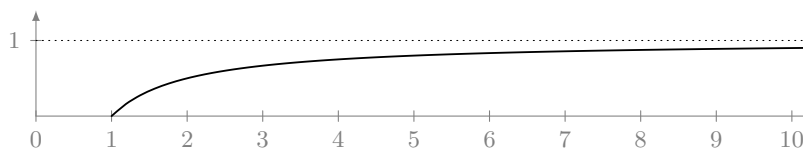
Théorème d'efficacité :

Les estimateurs \bar{X} et S_c de la moyenne et de l'écart-type, sont **non biaisés**:

$$E(\bar{X}) = \mu \text{ et } E(S_c) = \sigma$$

et **efficaces** : leur variance (leur dispersion autour de μ pour \bar{X} et de σ pour S_c) sont plus petites que tout autre estimateur.

Remarque : L'espérance de S^2 (la variance non corrigée) est $\frac{n-1}{n}\sigma^2$. Or $\lim_{n \rightarrow +\infty} \frac{n-1}{n} = 1$:



On constate que lorsque la taille n de l'échantillon croît, $\frac{n-1}{n}$ se rapproche de 1, et donc que $Var(S^2)$ se rapproche de σ^2 , la variance de la population. C'est la raison pour laquelle, en pratique, à partir de $n \geq 30$, des statisticiens ne corrigent plus la variance observée sur l'échantillon (S^2) : $\frac{29}{30} \simeq 0,97$ (une erreur de 3%).

Dans ce cours, la variance sera systématiquement corrigée.

Reprenons l'étude du cas n° 1 :

Variable qui suit
la loi $\mathcal{N}(0, 1)$

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

Comme on ne connaît pas l'écart-type σ , on va le remplacer par l'estimateur corrigé de l'écart-type S_c :

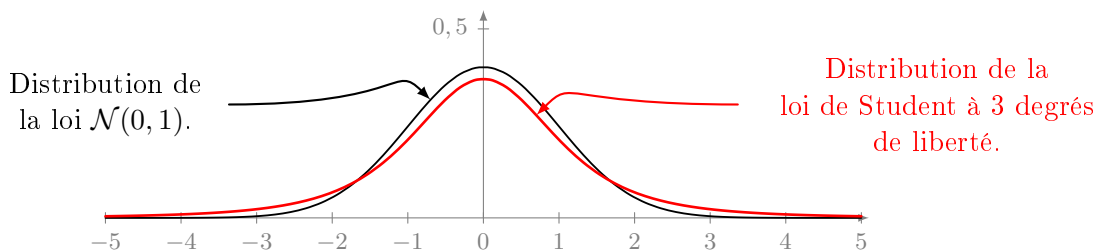
$$P\left(-t_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{S_c}{\sqrt{n}}} \leq t_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

Variable qui suit la loi
de Student à $n - 1$ degrés
de liberté.

Donc, on va retrouver les mêmes intervalles de confiance que précédemment, sauf que σ sera remplacé par S_c et $z_{\frac{\alpha}{2}}$ par $t_{\frac{\alpha}{2}}$ issu de la table de Student.

Mais, comme dès que $n \geq 30$, la distribution de Student à $n - 1$ degrés de liberté se comporte pratiquement comme une loi normale $\mathcal{N}(0, 1)$. On utilisera la loi de Student seulement pour les petits effectifs.

D'ailleurs, la loi de Student est souvent surnommée la « **loi des petits échantillons** ($n < 30$). ».



La loi de Student $\mathcal{T}(3)$ est plus légèrement plus aplatie que la loi normale, son écart-type est donc plus grand ($\sqrt{3} \simeq 1,7$), ce qui traduit la perte d'une information, celle de l'écart-type σ de la population, remplacé par son estimation S_c .

3. Synthèse.

Conditions	Intervalle de confiance d'une proportion.
Grand échantillon : $n \geq 30$ $n\hat{p} \geq 5$ $n(1 - \hat{p}) \geq 5$	$IC_{\alpha} = \left[\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right]$

a. L'écart-type σ de la population est connu.

Conditions	Intervalle de confiance d'une moyenne où l'écart-type σ de la population est connu .
Grand échantillon : $n \geq 30$	$IC_\alpha = \left[\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$
Conditions	Intervalle de confiance d'une moyenne sur un petit échantillon où l'écart-type σ de la population est connu et \mathbf{X} suit une loi normale .
Petit échantillon : $n < 30$	$IC_\alpha = \left[\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$
Conditions	Intervalle de confiance d'une moyenne sur un petit échantillon où l'écart-type σ de la population est connu et \mathbf{X} suit une loi inconnue .
Petit échantillon : $n < 30$	$IC_\alpha = \left[\bar{x} - \frac{1}{\sqrt{\alpha}} \frac{\sigma}{\sqrt{n}} ; \bar{x} + \frac{1}{\sqrt{\alpha}} \frac{\sigma}{\sqrt{n}} \right]$

b. L'écart-type σ de la population est inconnu :

$$S_c^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2 \text{ et } S_c^2 = \frac{n}{n-1} S^2$$

Conditions	Intervalle de confiance d'une moyenne sur un grand échantillon où l'écart-type σ de la population est inconnu .
Grand échantillon : $n \geq 30$	$IC_\alpha = \left[\bar{x} - z_{\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}} ; \bar{x} + z_{\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}} \right]$
Conditions	Intervalle de confiance d'une moyenne sur un petit échantillon où l'écart-type σ de la population est inconnu et \mathbf{X} suit une loi normale .
Petit échantillon : $n < 30$	$IC_\alpha = \left[\bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{S_c}{\sqrt{n}} ; \bar{x} + t_{\frac{\alpha}{2}, n-1} \frac{S_c}{\sqrt{n}} \right]$
Conditions	Intervalle de confiance d'une moyenne sur un petit échantillon où l'écart-type σ de la population est inconnu et \mathbf{X} suit une loi inconnue .
Petit échantillon : $n < 30$	$IC_\alpha = \left[\bar{x} - \frac{1}{\sqrt{\alpha}} \frac{S_c}{\sqrt{n}} ; \bar{x} + \frac{1}{\sqrt{\alpha}} \frac{S_c}{\sqrt{n}} \right]$

Exemple n° 22 : Le poids d'une variété de laitues est normalement distribué. Un laboratoire de recherche teste un nouvel engrais sur un plan de 600 laitues. Le poids moyen d'un échantillon sans remise de laitues est de 335g. Détermine un intervalle de confiance pour le poids moyen de cette variété de laitues cultivées dans ce nouvel engrais avec un niveau de confiance de 95%.

1. Sachant que l'écart-type est connu : $\sigma = 14,2\text{g}$, et que l'effectif de l'échantillon est $n = 40$

☞ L'échantillon est **grand** $n = 40 \geq 30$, l'écart-type σ de la population est **connu** donc :

$$IC_{5\%} = \left[\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

☞ L'écart-type est connu, donc on utilise la table de la loi de **normale centrée réduite** avec $\alpha = 5\%$:

$$z_{\frac{\alpha}{2}} = \mathbf{1,960}$$

On obtient l'intervalle de confiance :

$$IC_{5\%} = \left[335 - 1,96 \times \frac{14,2}{\sqrt{40}} ; 335 + 1,96 \times \frac{14,2}{\sqrt{40}} \right] = [330,6 ; 339,4]$$

2. L'écart-type de la population n'étant pas connu, on a dû estimer la variance sur l'échantillon : 214g².

Il compte 22 laitues.

☞ On corrige l'écart-type : $S_c^2 = \frac{n}{n-1} S^2 = \frac{\mathbf{22}}{\mathbf{21}} \times \mathbf{214}$ donc $S_c = \sqrt{\frac{\mathbf{22}}{\mathbf{21}} \times 214} \simeq \mathbf{14,973}$

☞ L'échantillon est **petit** $n = 22 < 30$, l'écart-type de la population est **inconnu**, et la distribution suit une loi de Student donc :

$$IC_{5\%} = \left[\bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{S_c}{\sqrt{n}} ; \bar{x} + t_{\frac{\alpha}{2}, n-1} \frac{S_c}{\sqrt{n}} \right]$$

☞ L'écart-type est inconnu, donc on utilise la table de la loi de **Student** avec $n - 1 = \mathbf{21}$ degrés de liberté, et $\alpha = \mathbf{2,5\%}$ car cette table ne répartit pas l'erreur bilatéralement : $t_{\mathbf{0,025}; \mathbf{21}} = \mathbf{2,080}$

On obtient l'intervalle de confiance :

$$IC_{5\%} = \left[335 - 2,08 \times \frac{14,973}{\sqrt{22}} ; 335 + 2,08 \times \frac{14,973}{\sqrt{22}} \right] = [328,4 ; 341,6]$$

Chapitre 5 - Statistiques inférentielles : Test de validité d'hypothèses.

I. Variables d'échantillonnages

Nous parlerons de tests de comparaisons d'un paramètre à une norme ou une spécification :

Par exemple :

- Lorsqu'un fabricant déclare que la moyenne de durée de vie d'une bougie est de 8h. On testera cette information en la comparant à la moyenne \bar{x} d'un échantillon.
- Lorsque la proportion de pièces défectueuses fabriquées par une machine ne doit pas dépasser 2%. On testera cette information en la comparant à la proportion \hat{p} d'un échantillon.

Nous parlerons de tests de comparaisons entre deux paramètres de même nature, associés à deux populations.

Par exemple :

- En comparant, à l'aide de deux échantillons, la durée moyenne de séchage de deux peintures de marques différentes.
- En comparant, à l'aide de deux échantillons, la proportion de réussite au baccalauréat entre les filles et les garçons.

1. Etude du déroulement d'un test d'hypothèses de comparaison de moyenne.

Soit une population pour laquelle, théoriquement la moyenne μ de la variable X est égale à μ_0 . On va tester l'exactitude de cette valeur théorique. Pour ce faire, on va la comparer à la moyenne m prise par X sur un échantillon.

Nous sommes donc en présence de deux hypothèses possibles :

- L'hypothèse nulle, hypothèse du statu quo, que nous noterons H_0 signifiant qu'il n'y a aucune différence, aucun changement, que « μ_0 est exacte ».
- L'hypothèse **alternative**, que nous noterons H_1 , signifiant qu'un changement est survenu, qu'il y a une différence, que « μ_0 est inexacte ».



Paires d'hypothèses possibles relativement à une moyenne

Test bilatéral :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Test unilatéral à droite :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0$$

Test unilatéral à gauche :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu < \mu_0$$

Exemple n° 23 : Un fabricant déclare que la moyenne de durée de vie de ses bougies est de 8h. Une association de consommateur se demande si cette durée est exacte. On suppose l'écart-type σ connu.

Soit X la durée de vie d'une bougie :

$$\begin{cases} H_0 : \mu = 8 \text{ (hypothèse nulle)} \\ H_1 : \mu \neq 8 \text{ (hypothèse alternative)} \end{cases}$$

Pour confronter l'hypothèse nulle, $\mu = \mu_0$, on choisit au hasard un échantillon, de taille $n \geq 30$, on y calcule la moyenne \bar{x} afin de la comparer à $\mu_0 = 8$.

Première ébauche d'une règle de décision :

- Si la moyenne observée \bar{x} est « près » de la valeur $\mu_0 = 8$, alors on accepte l'hypothèse H_0 .
- Si la moyenne observée \bar{x} est « loin » de la valeur $\mu_0 = 8$, alors on rejette l'hypothèse H_0 .

Mais, que signifient « \bar{x} est près de $\mu_0 = 8$ » ou « \bar{x} est loin de $\mu_0 = 8$ » ?

Plaçons nous dans le cas où l'hypothèse H_0 est vraie :

On a vu dans le chapitre précédent que \bar{X} suit approximativement une loi $\mathcal{N}\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$.

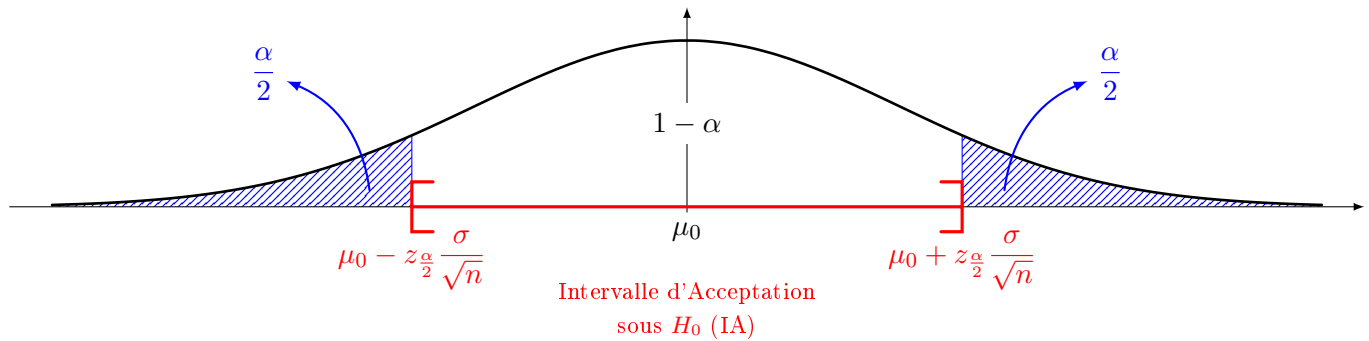
Pour un α donné, il existe un $z_{\frac{\alpha}{2}}$ tel que :

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha \text{ soit } P\left(\mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Autrement dit, pour un $\alpha \in]0, 1[$ donné, la probabilité que la moyenne \bar{x} observée sur l'échantillon soit dans l'intervalle **d'acceptation de (H_0)** :

$$\mathbf{IA} = \left[\mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

est égale à $1 - \alpha$.



IA n'est pas un intervalle de confiance. On a supposé que l'hypothèse H_0 était vraie. Donc que la moyenne dans la population était bien μ_0 , dans notre exemple $\mu_0 = 8$. Lorsqu'on prend un échantillon, il est « plus ou moins représentatif » : la moyenne observée sur l'échantillon est « plus ou moins proche » de celle de la population. Cette variation, cette fluctuation de la moyenne observée autour de la moyenne μ_0 est appelée la fluctuation d'échantillonnage. L'intervalle d'acceptation IA est un intervalle de **fluctuation d'échantillon**.

Si on choisit, par exemple, $\alpha = 5\%$. On sait que 95% des moyennes observées dans les échantillons seront comprise dans l'intervalle IA = $[\mu_0 - \dots, \mu_0 + \dots]$. On peut donc émettre un jugement en décidant que, si dans notre échantillon, la moyenne observée \bar{x} n'est pas dans IA, alors l'hypothèse H_0 est fausse. Autrement dit, que la moyenne dans la population n'est pas μ_0 . Mais, il se peut que l'échantillon prélevé ne soit pas représentatif (5% de chance), et que l'hypothèse H_0 soit quand même vraie. Le choix de la valeur de α revient à décider de ce qui relève de la fluctuation d'échantillon ou pas. C'est un jugement de signification du test, et il est arbitraire.

Deuxième ébauche d'une règle de décision :

- Si $\bar{x} \in \text{IA (intervalle d'acceptation)}$, alors on accepte l'hypothèse H_0 .
- Si $\bar{x} \notin \text{IA}$, alors on rejette l'hypothèse H_0 .



Définition:

α est appelé le **seuil de signification du test**

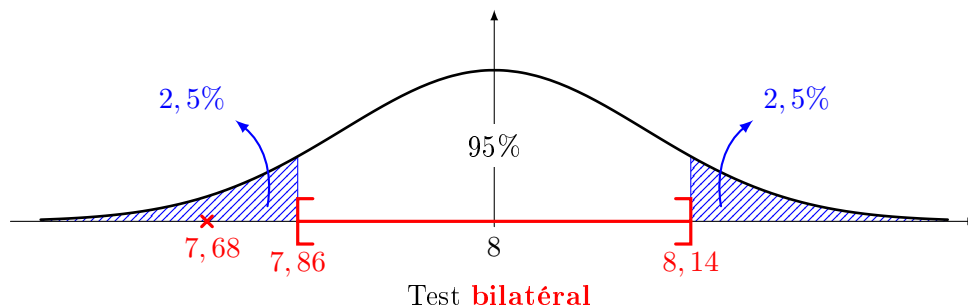
Dans notre exemple, le seuil de signification signifie que la moyenne observée \bar{x} a moins de 5% de chances d'être obtenu par hasard.

Supposons, dans notre exemple, que $\sigma = 0,5h$, que l'échantillon soit de $n = 50$ bougies, et que la moyenne observée sur l'échantillon soit $\bar{x} = 7,68h$. On fixe $\alpha = 5\%$ soit $z_{\frac{\alpha}{2}} = 1,96$ (table de l'écart réduit de la loi normale centrée réduite).

On a : $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1,96 \times \frac{0,5}{\sqrt{50}} \simeq 0,14$

$$\left[\mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right] = [8 - 0,14 ; 8 + 0,14] = [7,86 ; 8,14]$$

$\bar{x} = 7,68 \notin [7,86 ; 8,14]$ donc, on rejette H_0 .



Formulons autrement cette règle de décision :

$$\bar{x} \in \left[\mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right] \iff \left[\begin{array}{c} \text{Diagram showing } \mu_0 \text{ on the x-axis, } \bar{x} \text{ to its right, and a bracketed interval } [\mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}] \text{ containing } \bar{x}. \end{array} \right]$$

$$\iff |\bar{x} - \mu_0| \leq z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow \frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}$$

Règle de décision :

- Si $\frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}$, alors on accepte l'hypothèse H_0 .
- Si $\frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}} > z_{\frac{\alpha}{2}}$, alors on rejette l'hypothèse H_0 .

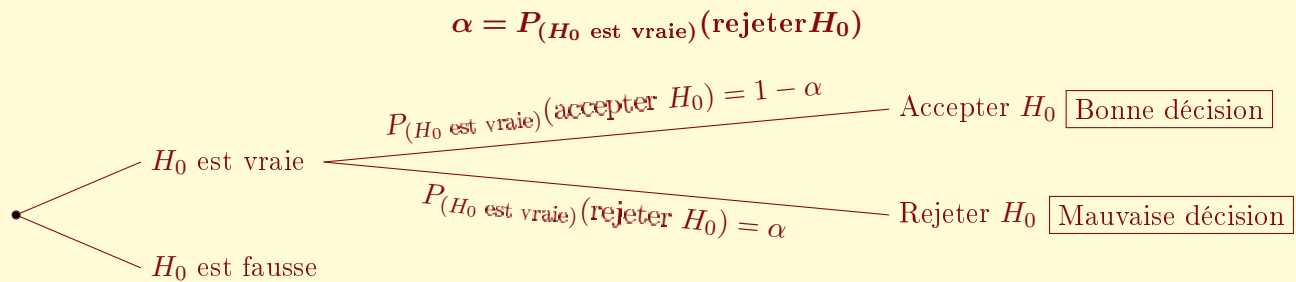
Reprenons notre exemple :

$$\frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}} = \frac{|7,68 - 8|}{\frac{0,5}{\sqrt{50}}} \simeq 4,53 > z_{\frac{\alpha}{2}} = 1,96 \text{ donc, on rejette } H_0.$$



Définition:

Lors d'un test, on commet une erreur de **1^{er} espèce** lorsqu'on décide de rejeter l'hypothèse H_0 alors qu'elle est vraie. La probabilité de cette erreur est notée α :



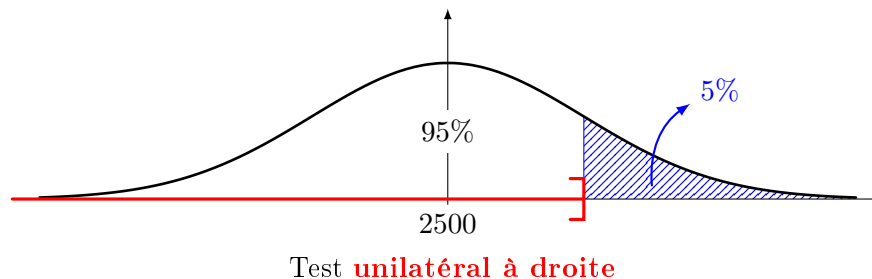
Exemple n° 24 : La durée de vie d'une certaine marque de lampe était de 2500h en 2015. Suite à des avancées technologiques, on pense que la durée de vie de la nouvelle génération de ces lampes a augmenté. On sait que la durée de vie de ces lampes suit une loi normale d'écart-type $\sigma = 30$ h.

Soit X la durée de vie d'une lampe :

$$\begin{cases} H_0 : \mu = 2500 \text{ (hypothèse nulle)} \\ H_1 : \mu > 2500 \text{ (hypothèse alternative)} \end{cases}$$

Pour confronter l'hypothèse nulle, $\mu = 2500$, on choisit au hasard un échantillon de 64 lampes, on y calcule la moyenne \bar{x} de leurs durées de vie qui est égale à 2584 heures.

On va tester ces hypothèses avec un seuil de signification de 5%.



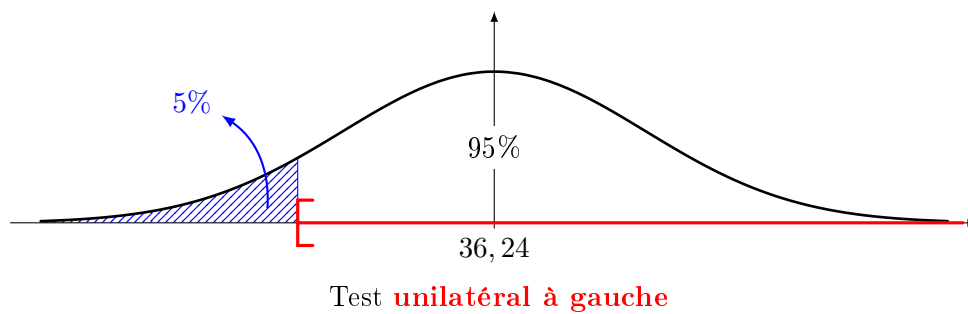
Dans la table de l'écart réduit de la loi normale centrée réduite, on va prendre $\alpha = 10\%$, on trouve $z_\alpha = 1,645$.

$$\frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}} = \frac{|2584 - 2500|}{\frac{30}{\sqrt{64}}} = 22,4 > z_\alpha = 1,645 \text{ donc, on accepte } H_1.$$

Exemple n° 25 : La durée moyenne d'une opération par un robot de soudage était de 36,24 secondes. Après un réglage, sur un échantillon de 42 opérations, la durée moyenne est de 36,16 secondes. Sachant que le temps d'une opération suit une loi normale d'écart-type $\sigma = 0,15$ seconde.

Soit X la durée de vie d'une lampe : $H_0 : \mu = 36,24$ (hypothèse nulle)
 $H_1 : \mu < 36,24$ (hypothèse alternative)

On va tester ces hypothèses avec un seuil de signification de 5%.



Dans la table de l'écart réduit de la loi normale centrée réduite, on va prendre $\alpha = 10\%$, on trouve $z_\alpha = 1,645$.

- Si on ne mettait pas de valeur absolue, le test unilatéral à gauche serait :

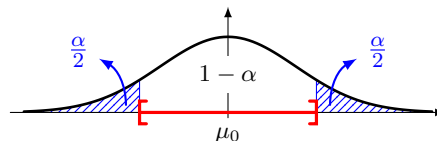
$$\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{36,16 - 36,24}{\frac{0,15}{\sqrt{42}}} = -3,456 < -z_\alpha = -1,645 \text{ donc, on accepte } H_1.$$

- Mais, le test utilise une valeur absolue, ce qui évite les erreurs de signes :

$$\frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}} = \frac{|36,16 - 36,24|}{\frac{0,15}{\sqrt{42}}} = 3,456 > z_\alpha = 1,645 \text{ donc, on accepte } H_1.$$

II. Synthèse graphique

Test bilatéral :



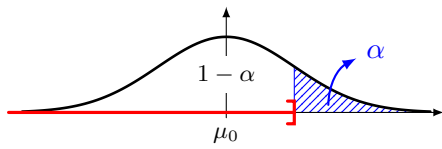
$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Test unilatéral à droite :

$$H_0 : \mu = \mu_0$$

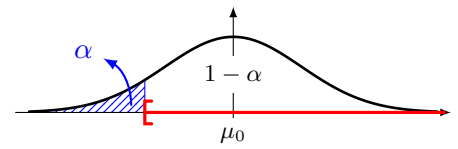
$$H_1 : \mu > \mu_0$$



Test unilatéral à gauche :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu < \mu_0$$



! Tests unilatéraux

- Pour les tests unilatéraux, on utilise la table de l'écart réduit de la loi normale centrée réduite en doublant la valeur de α car on met toute l'erreur du même côté.
- Pour le test unilatéral à gauche, on utilise le même protocole que le test unilatéral à droite.

📖 Fixer le seuil de signification α c'est :

- déterminer ce qu'on est prêt à accepter comme probabilité de commettre l'erreur de 1^{er} espèce ;
- déterminer les zones d'acceptation et de rejet de H_0 .

III. Tests d'hypothèses de comparaison de moyennes et de proportions

1. Test de comparaison d'une moyenne μ_0 d'une population à celle \bar{x} de l'un de ses échantillons.

μ_0 est la valeur moyenne de la population.
On la compare à la moyenne \bar{x} observée sur l'un de ses échantillons.
On cherche à vérifier l'exactitude de μ_0 .

Hypothèses bilatérales : $\left\{ \begin{array}{l} H_0 : \mu = \mu_0 \text{ (la moyenne de la population } \mu \text{ est égale à } \mu_0\text{).} \\ H_1 : \mu \neq \mu_0 \end{array} \right.$

X_i est la variable aléatoire qui au i^{e} individu d'un échantillon de taille n associe sa moyenne. $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ est la variable aléatoire qui à un échantillon associe sa moyenne. On note $\sigma_{\bar{X}}$ son écart-type.

$S_c^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2$ est la variance corrigée observée sur l'échantillon prélevé et $S_c = \sqrt{S_c^2}$ est l'écart-type corrigé. La variance et la variance corrigée sont reliées par la formule :

$$S_c^2 = \frac{n}{n-1} S^2$$

Effectif	Conditions d'application	Statistique du test	Seuil de signification (test bilatéral)
Grand échantillon : $n \geq 30$ et σ connu	Aucune	$T = \frac{ \bar{x} - \mu_0 }{\frac{\sigma}{\sqrt{n}}}$	rejet de H_0 si $T > z_{\frac{\alpha}{2}}$
Grand échantillon : $n \geq 30$ et σ inconnu	Aucune	$T = \frac{ \bar{x} - \mu_0 }{\frac{S_c}{\sqrt{n}}}$	rejet de H_0 si $T > z_{\frac{\alpha}{2}}$
Petit échantillon : $n < 30$ et σ connu	X suit une loi normale	$T = \frac{ \bar{x} - \mu_0 }{\frac{\sigma}{\sqrt{n}}}$	rejet de H_0 si $T > z_{\frac{\alpha}{2}}$
Petit échantillon : $n < 30$ et σ inconnu	X suit une loi normale	$T = \frac{ \bar{x} - \mu_0 }{\frac{S_c}{\sqrt{n}}}$	rejet de H_0 si $T > t_{\frac{\alpha}{2}, n-1}$

Exemple n° 26 : Un employé responsable du contrôle de qualité doit tester avec un seuil de signification de 1%, la durée moyenne théorique de vie d'un condensateur au tantale qui serait de 4500 heures. La moyenne obtenue pour un échantillon de 17 condensateurs est de 4158 heures. Que décidera-t-il ?

1. Il sait que la durée de vie de ces condensateurs suit une loi normale d'écart-type 72 heures.

Il s'agit d'un test **bilatéral** de comparaison d'une moyenne théorique μ_0 à celle d'un échantillon de **petite taille** $n = 17 < 30$ où X suit une **loi normale**, et **l'écart-type est connu** :

1 Formulation des hypothèses :

$$\begin{cases} H_0 : \mu = 4500 \\ H_1 : \mu \neq 4500 \end{cases}$$

2 On corrige l'écart-type : Non ! l'écart-type $\sigma = 72$ est connu, c'est une valeur théorique, elle n'est pas estimée.

3 On calcule la statistique du test : $T = \frac{|4158 - 4500|}{\frac{72}{\sqrt{17}}} \simeq \frac{342}{17,463} \simeq 19,58$

4 Règle de décision : L'écart-type est connu, donc on utilise la table de l'écart réduit de la loi **normale centrée réduite** :

$$z_{\frac{\alpha}{2}} = 2,576$$

$$T = 19,58 > z_{\frac{\alpha}{2}} = 2,576 \text{ donc } H_0 \text{ est rejetée.}$$

On peut supposer, au risque de 1^{er} espèce de **1%**, que la durée moyenne théorique **n'est pas de 4500 heures**.

2. Il sait que la durée de vie de ces condensateurs suit une loi normale, mais il ne connaît pas son écart-type. Il estime l'écart-type sur l'échantillon : $S = 92$

Il s'agit d'un test **bilatéral** de comparaison d'une moyenne théorique μ_0 à celle d'un échantillon de **petite taille** $n = 17 < 30$ où X suit une **loi normale**, et **l'écart-type est inconnu** :

- 1 Formulation des hypothèses :
$$\begin{cases} H_0 : \mu = 4500 \\ H_1 : \mu \neq 4500 \end{cases}$$
- 2 On corrige l'écart-type : $S_c^2 = \frac{n}{n-1} S^2$ donc $S_c = \sqrt{\frac{17}{16}} \times 92 \simeq 94,83$
- 3 On calcule la statistique du test : $T = \frac{|4158 - 4500|}{\frac{94,83}{\sqrt{17}}} \simeq \frac{342}{23,00} \simeq 14,87$
- 4 Règle de décision :
L'écart-type étant inconnu, on utilise la table de la loi de **Student**.



La table de Student ne répartissant pas l'erreur de façon bilatérale, on doit diviser l'erreur $\alpha = 1\%$ par deux.

Ainsi, $t_{0,005;16} = 2,921$ et

$T = 14,87 > t_{0,005;16} = 2,921$ donc H_0 est **rejetée**.

On peut supposer, au risque de 1^{er} espèce de 1%, que la durée moyenne théorique **n'est pas de 4500 heures**.

2. Test de comparaison d'une proportion p_0 sur une population à celle \hat{p} observée sur l'un de ses échantillons.

p_0 est la proportion d'une modalité sur une population.
On la compare à la proportion \hat{p} observée sur l'un de ses échantillons.

Hypothèses bilatérales :
$$\begin{cases} H_0 : p = p_0 \text{ (la proportion } p \text{ sur la population est égale à } p_0\text{).} \\ H_1 : p \neq p_0 \end{cases}$$

Effectif	Conditions d'application	Statistique du test	Seuil de signification (test bilatéral)
Grand échantillon : $n \geq 30$	$n_1 \hat{p} \geq 5$ et $n_1 \hat{q} \geq 5$ où $\hat{q} = 1 - \hat{p}$	$T = \frac{ \hat{p} - p_0 }{\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}}$	rejet de H_0 si $T > z_{\frac{\alpha}{2}}$

3. Test de comparaison de deux proportions observées \hat{p}_1 et \hat{p}_2 .

On veut comparer les proportions d'une modalité sur deux populations à partir des proportions observées \hat{p}_1 et \hat{p}_2 sur un échantillon de chacune d'entre elles, les deux échantillons étant indépendants.

Hypothèses bilatérales :
$$\begin{cases} H_0 : p_1 = p_2 \text{ (les proportions sont les mêmes sur les deux populations).} \\ H_1 : p_1 \neq p_2 \end{cases}$$

Première population :

L'échantillon prélevé sur cette population est de taille n_1 . La proportion observée sur cet échantillon est \hat{p}_1 , et on note $\hat{q}_1 = 1 - \hat{p}_1$.

Deuxième population :

L'échantillon prélevé sur cette population est de taille n_2 . La proportion observée sur cet échantillon est \hat{p}_2 , et on note $\hat{q}_2 = 1 - \hat{p}_2$.

On calcule proportion commune aux deux échantillons p_c , pondérée par leur taille :

$$p_c = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} \text{ et } q_c = 1 - p_c$$

Effectif	Conditions d'application	Quantité à calculer	Seuil de signification (test bilatéral)
Grands échantillons : $n_1 \geq 30$ et $n_2 \geq 30$	$n_1\hat{p}_1 \geq 5$ et $n_1\hat{q}_1 \geq 5$ $n_2\hat{p}_2 \geq 5$ et $n_2\hat{q}_2 \geq 5$	$T = \frac{ \hat{p}_1 - \hat{p}_2 }{\sqrt{\frac{p_c q_c}{n_1} + \frac{p_c q_c}{n_2}}}$	Rejet de H_0 si $T > z_{\frac{\alpha}{2}}$

Exemple n° 27 : Le groupe d'hypermarchés Merlan se demande si le changement de direction de ses hypermarchés a amélioré la satisfaction de ses clients. Pour ce faire, il compare deux études de satisfactions, l'une faite avant l'arrivée de la nouvelle direction, et l'autre faite deux mois après.

	Avant	Après
Nombre de clients satisfaits	105	110
Nombre de clients interrogés	124	136

Les taux de satisfactions :

- avant le changement de direction : $\hat{p}_1 = \frac{105}{124} \simeq 84,7\%$
- après le changement de direction : $\hat{p}_2 = \frac{110}{136} \simeq 80,9\%$

Cette différence de pourcentages, peut-elle être considérer comme une simple fluctuation d'échantillon avec un niveau de confiance de 0,95 ?

Il s'agit d'un test bilatéral de comparaison de deux proportions observées :

1 Formulation des hypothèses : $\left\{ \begin{array}{l} H_0 : p_1 = p_2 \text{ (les taux de satisfactions sont les mêmes sur les deux périodes).} \\ H_1 : p_1 \neq p_2 \end{array} \right.$

2 Les conditions sont vérifiées : $\left\{ \begin{array}{l} n_1\hat{p}_1 = 105 \geq 5, n_1\hat{q}_1 = 19 \geq 5, n_2\hat{p}_2 = 110 \geq 5, \text{ et } n_2\hat{q}_2 = 26 \geq 5 \\ \text{les échantillons sont grands : } n_1 = 124 \geq 30 \text{ et } n_2 = 136 \geq 30 \end{array} \right.$

3 Calcul du test :

- La proportion commune : $p_c = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{105 + 110}{260} \simeq 0,827$ et $q_c = 1 - 0,827 = 0,173$

$$\bullet T = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\frac{p_c q_c}{n_1} + \frac{p_c q_c}{n_2}}} = \frac{|0,847 - 0,809|}{\sqrt{\frac{0,827 \times 0,173}{124} + \frac{0,827 \times 0,173}{136}}} \simeq 0,809$$

4 Règle de décision :

La table de l'écart réduit de la loi normale centrée réduite avec $\alpha = 5\%$ (test bilatéral) : $z_{\frac{\alpha}{2}} = 1,960$

$T = 0,809 < 1,96$ donc on accepte H_0 : il n'y a pas de différence sensible de satisfaction, au niveau de confiance de 95%.

4. Test de comparaison des moyennes observées \bar{x}_1 et \bar{x}_2 .

On veut comparer les moyennes d'une même modalité sur deux populations à partir des moyennes observées \bar{x}_1 et \bar{x}_2 sur un échantillon de chacune d'entre elles, les deux échantillons étant indépendants.

Hypothèses bilatérales : $\left\{ \begin{array}{l} H_0 : \mu_1 = \mu_2 \text{ (les moyennes sont les mêmes sur les deux populations).} \\ H_1 : \mu_1 \neq \mu_2 \end{array} \right.$

Première population :

\bar{X}_1 est la variable aléatoire qui à un échantillon de la première population associe sa moyenne.

L'échantillon prélevé sur cette population est de taille n_1 . La moyenne observée sur cet échantillon est \bar{x}_1 .

σ_1 est son écart-type, s'il est connu. Sinon, on calculera S_{1c} son écart-type corrigé.

Deuxième population :

\bar{X}_2 est la variable aléatoire qui à un échantillon de la deuxième population associe sa moyenne.

L'échantillon prélevé sur cette population est de taille n_2 . La moyenne observée sur cet échantillon est \bar{x}_2 .

σ_2 est son écart-type, s'il est connu. Sinon, on calculera S_{2c} son écart-type corrigé.

Les tests reposent sur l'étude de la variable aléatoire $\bar{X}_1 - \bar{X}_2$.

a. Les écart-types σ_1 et σ_2 sont connus.

Effectif	Conditions d'application	Quantité à calculer	Seuil de signification (test bilatéral)
Grands échantillons : $n_1 \geq 30$ et $n_2 \geq 30$	Aucune	$T = \frac{ \bar{x}_1 - \bar{x}_2 }{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	rejet de H_0 si $T > z_{\frac{\alpha}{2}}$
n_1 ou $n_2 < 30$	X_1 et X_2 suivent des lois normales		

b. Les écart-types σ_1 et σ_2 sont inconnus.

Effectif	Ecart-types	Conditions d'application	Quantité à calculer	Seuil de signification (test bilatéral)
Grands échantillons : $n_1 \geq 30$ et $n_2 \geq 30$	Aucune condition	Aucune	$T = \frac{ \bar{x}_1 - \bar{x}_2 }{\sqrt{\frac{S_{1c}^2}{n_1} + \frac{S_{2c}^2}{n_2}}}$	Rejet de H_0 si $T > z_{\frac{\alpha}{2}}$
n_1 ou $n_2 < 30$	σ_1 et σ_2 sont inconnus mais égaux	X_1 et X_2 suivent des lois normales	$T = \frac{ \bar{x}_1 - \bar{x}_2 }{S_p^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	Rejet de H_0 si $T > t_{\frac{\alpha}{2}, n_1+n_2-2}$
$n_1 = n_2 < 30$	σ_1 et σ_2 sont inconnus mais ^(★) $\frac{1}{3} \leq \frac{S_{1c}^2}{S_{2c}^2} \leq 3$	X_1 et X_2 suivent des lois normales		

(★) $S_p = \sqrt{\frac{(n_1 - 1)S_{1c}^2 + (n_2 - 1)S_{2c}^2}{n_1 + n_2 - 2}}$ est la racine carrée de la moyenne pondérée par leur degré de liberté des variances

(★) $\frac{1}{3} \leq \frac{S_{1c}^2}{S_{2c}^2} \leq 3$ signifie que les variances estimées ne sont pas trop différentes (dans un rapport de 3).

Chapitre 6 - Statistiques inférentielles : Test d'ajustement.



Notations :

Etant donné un ensemble fini A , le nombre d'éléments de A est appelé le **cardinal** de A et noté $\#(A)$.

Exemple n° 28 : $\#\left(\left\{\heartsuit, \spadesuit, \bullet\right\}\right) = 3$ et $\#\left(\left\{0, 1, 2, 3, \dots, 10\right\}\right) = 11$

I. Tests d'adéquation d'une distribution à une distribution théorique.

On va comparer une distribution statistique (des données) à des distributions théoriques.

1. Adéquation d'une distribution à une distribution équirépartie.

Lançons un dé 600 fois de suites, et notons les résultats obtenus :

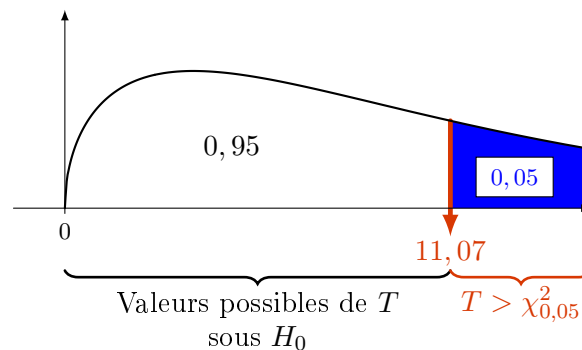
Numéro de la face	1	2	3	4	5	6	Total
Effectifs	90	108	109	93	102	$600 - 502 = 98$	600
Effectifs théorique	100	100	100	100	100	100	600

Le nombre de degrés de liberté est de $6 - 1 = 5$, car lorsqu'on connaît l'effectif de 5 faces, on connaît l'effectif de la **sixième**. On souhaite tester l'hypothèse (H_0) selon laquelle le dé n'est pas truqué, avec un risque $\alpha = 0,05$.

On calcule la variable aléatoire :

$$T = \frac{(90 - 100)^2}{100} + \frac{(108 - 100)^2}{100} + \frac{(109 - 100)^2}{100} + \frac{(93 - 100)^2}{100} + \frac{(102 - 100)^2}{100} + \frac{(98 - 100)^2}{100} = 3,02$$

Sous l'hypothèse H_0 : le dé est parfaitement équilibré, la variable aléatoire T suit une loi du χ^2 à 5 degrés de liberté. D'après la table du χ^2 : $\chi_{0,05;5}^2 \simeq 11,07$.



Comme $T = 3,02 < 11,07$, on considère que le dé n'est pas truqué.

En résumé :

Nous observons des effectifs O_1, O_2, \dots, O_n sur un échantillon.

- 1 On émet l'hypothèse (H_0) que ces données suivent une distribution particulière (normale, exponentielle, etc.).
- 2 En utilisant cette loi supposée, on calcule les effectifs attendus C_1, C_2, \dots, C_n :

Ces effectifs calculés, les C_i , ils doivent être supérieurs ou égaux à 5.

Si ces conditions ne sont pas satisfaites, on peut, si c'est possible, regrouper des observations jusqu'à ce que les effectifs calculés soient suffisamment grands.

- 3 A partir de ces données calculées (C_i) et observées (O_i) on calcule la quantité suivante :

$$T = \sum_{i=1}^n \frac{(O_i - C_i)^2}{C_i}$$

Si l'hypothèse (H_0) est vraie alors T suit une loi du χ^2 de degré de liberté :

$$\text{ddl} = \left(\begin{array}{c} \text{nb de } C_i \text{ utilisés} \\ \text{dans le calcul de } T \end{array} \right) - 1 - \left(\begin{array}{c} \text{nb de paramètres} \\ \text{estimés pour} \\ \text{le calcul des } C_i \end{array} \right)^*$$

- 4 La règle de décision est la suivante : $\left\{ \begin{array}{l} H_0 : T \leq \chi_{\alpha, \text{ddl}}^2 \\ H_1 : T > \chi_{\alpha, \text{ddl}}^2 \end{array} \right.$

(*) Généralement la moyenne ou l'écart-type quand ils sont inconnus.

2. Adéquation d'une distribution à une distribution normale.

Lors de l'étude sur le reboisement dans un secteur donné, 100 arbres sont choisis aléatoirement. A l'aide des données ci-contre, et sachant que le diamètre moyen des arbres de l'échantillon est de 260,4mm et son écart-type corrigé de 30,88mm, peut-on penser avec un seuil fixé à 10% que le diamètre de ces arbres suit une loi normale ?

Diamètres des arbres (en mm)	Nombre d'arbres
[170 ; 190[1
[190 ; 210[6
[210 ; 230[8
[230 ; 250[16
[250 ; 270[34
[270 ; 290[22
[290 ; 310[7
[310 ; 330[4
[330 ; 350[2
Total	100

- 1 Formulation des hypothèses :

H_0 : Le diamètre des arbres suit une loi $\mathcal{N}(260, 4; 30, 88)$
 H_1 : Le diamètre des arbres ne suit pas cette loi.

- 2 On calcule les Fréquences avec la table de la fonction de répartition de la loi normale centrée réduite, puis les effectifs calculés :

Diamètres des arbres (en mm)	Nombre d'arbres O_i	Fréquences calculées	Effectifs calculés C_i
[170 ; 190[1	0,010	1,0
[190 ; 210[6	0,040	4,0
[210 ; 230[8	0,111	11,1
[230 ; 250[16	0,206	20,6
[250 ; 270[34	0,255	25,5
[270 ; 290[22	0,209	20,9
[290 ; 310[7	0,115	11,5
[310 ; 330[4	0,042	4,2
[330 ; 350[2	0,010	1,0
Total	100	1	100

Détails des calculs pour l'intervalle [250 ; 270[:

On note D la variable aléatoire « diamètre des arbres » qui suit une loi $\mathcal{N}(260, 4; 30, 88)$.

$$\begin{aligned}
 P(250 \leq D < 270) &= P(-0,34 \leq Z \leq 0,31) \text{ où } Z \sim \mathcal{N}(0; 1) \\
 &= P(Z \leq 0,31) - P(Z \leq -0,34) \\
 &= P(Z \leq 0,31) + P(Z \leq 0,34) - 1 \\
 &= P(Z \leq 0,31) - P(Z \leq -0,34) = 0,6217 + 0,6331 - 1 = 0,2548 \simeq 0,255
 \end{aligned}$$

L'effectif calculé C_5 est donc $0,255 \times 100 = 25,5$ arbres.

Les effectifs calculés doivent être supérieurs à 5, donc on regroupe les deux premières et les deux dernières classes :

Diamètres des arbres (en mm)	Nombre d'arbres O_i	Fréquences calculées	Effectifs calculés C_i
[170 ; 210[7	0,050	5,0
[210 ; 230[8	0,111	11,1
[230 ; 250[16	0,206	20,6
[250 ; 270[34	0,255	25,5
[270 ; 290[22	0,209	20,9
[290 ; 310[7	0,115	11,5
[310 ; 350[6	0,052	5,2
Total	100	1	100

3 On calcule T

$$T = \frac{(7-5)^2}{5} + \frac{(8-11,1)^2}{11,1} + \frac{(16-20,6)^2}{20,6} + \frac{(34-25,5)^2}{25,5} + \frac{(22-20,9)^2}{20,9} + \frac{(7-11,5)^2}{11,5} + \frac{(6-5,2)^2}{5,2} \simeq 7,47$$

Le nombre de degrés de liberté est $\text{ddl} = \underbrace{\left(\text{nb de } C_i \text{ utilisés dans le calcul de } T \right)}_{7: \text{ un par classe}} - 1 - \underbrace{\left(\text{nb de paramètres estimés pour le calcul des } C_i \right)}_{2: \text{ la moyenne et l'écart-type}}$

$$\text{ddl} = 7 - 1 - 2 = 4$$

4 Règle de décision :

$H_0 : T \leq \chi_{5\%,4}^2$ $\chi_{5\%,4}^2 = 9,488$ donc $T = 7,47 \leq 9,488$ l'hypothèse H_0 est acceptée.
 $H_1 : T > \chi_{5\%,4}^2$ On peut dire au risque de 10% que le diamètre des arbres suit une loi normale.

II. Tests d'indépendance de deux critères.

Claude et Virginie se sont intéressés aux résultats scolaires de 120 étudiants et à leurs habitudes concernant le tabac. Les résultats obtenus sont réunis dans le tableaux suivant :

Usage du tabac	Résultats scolaires			Total
	Excellents	satisfaisants	Médiocres	
Fume Beaucoup	8	16	7	31
Fume Modérément	9	20	8	37
Ne Fume pas	6	34	12	52
Total	23	70	27	120

1 Formulation des hypothèses :

H_0 : Les résultats scolaires et la consommation de tabac **sont indépendants.**
 H_1 : Les résultats scolaires et la consommation de tabac **ne sont pas indépendants.**

2 On calcule les effectifs théoriques. On rappelle que :

Les événements FB et RS sont indépendants $\iff P(FB \cap RS) = P(FB)P(RS)$

Usage du tabac	RE	RS	RM	Total
FB	$FB \cap RE$	$FB \cap RS$	$FB \cap RM$	31
FM	$FM \cap RE$	$FM \cap RS$	$FM \cap RM$	37
FP	$FP \cap RE$	$FP \cap RS$	$FP \cap RM$	52
Total	23	70	27	120

$$\begin{aligned}
 P(FB \cap RS) &= P(FB)P(RS) \\
 &= \frac{31}{120} \times \frac{70}{120} \simeq 0,1507
 \end{aligned}$$

Donc, $\#(FB \cap RS) \simeq 0,1507 \times 120 = 18,1$

$$\#(FB \cap RS) = \frac{\#(FB) \times \#(RS)}{\text{effectif total}}$$

Ainsi, $\#(FM \cap RM) = \frac{37 \times 27}{120} \simeq 8,3$

Usage théorique du tabac	RE	RS	RM
FB	5,9	18,1	7,0
FM	7,1	21,6	8,3
FP	10,0	30,3	11,7

3 On calcule T

$$T = \frac{(8 - 5,9)^2}{5,9} + \frac{(16 - 18,1)^2}{18,1} + \dots + \frac{(12 - 11,7)^2}{11,7} \simeq 3,7 \quad (3,63 \text{ sans arrondi.})$$

Le nombre de degrés de liberté est $ddl = \underbrace{\left(\text{nb de } C_i \text{ utilisés dans le calcul de } T \right)}_{9: \text{ un par } \cap} - 1 - \underbrace{\left(\text{nb de paramètres estimés pour le calcul des } C_i \right)}_{2+2: \text{ colonne et ligne des totaux}} = 9 - 1 - 4 = 4$

Ici, les paramètres estimés sont les probabilités d'évènements de chaque variable, ceux que nous retrouvons en ligne et en colonne. Ainsi, On a dû estimer la ligne des totaux c'est-à-dire les probabilités des évènements RE , RS , et RM . Mais celle de RM n'a pas dû être estimée puisque leur somme est égale à 1.

Une approche différente :

	RE	RS	RM
FB			
FM			
FP			

Sur chaque ligne, et chaque colonne, seules deux cases sont libres. Une fois qu'elles ont été estimées, les autres probabilités en découlent, donc :

$$ddl = (\text{nb colonnes} - 1)(\text{nb lignes} - 1)$$

Propriété : Nombre de degrés de liberté

En considérant seulement les intersections, autrement dit, le tableau sans les intitulés ni les totaux, on a :

$$ddl = (\text{nb colonnes} - 1)(\text{nb lignes} - 1)$$

4 Règle de décision : Seuil de signification du test $\alpha = 10\%$.

$$H_0 : T \leq \chi_{10\%,4}^2$$

$$H_1 : T > \chi_{10\%,4}^2$$

$\chi_{10\%,4}^2 = 7,779$ et $T = 3,7 \leq 7,779$, donc l'hypothèse H_0 est acceptée.

On peut dire au risque de 10% que les événements FB et RS sont indépendants.

III. Tests d'homogénéité.

Dans une population formée d'individus répartis en différentes catégories (hommes/femmes, classes d'ages, niveaux socio-économiques, etc...), on observe une variable (durée de vie, présence d'un risque, performances, etc.) et on se demande si ses variations selon les différentes catégories de la population sont simplement dues aux fluctuations d'échantillonnage ou si au contraire elles révèlent des comportements différents de la variable dans chacune de ces catégories.

Pour conclure, on va utiliser une test d'homogénéité, qui revient exactement à faire les mêmes calculs que dans un test d'indépendance.

Exemple n° 29 : Avant le dépôt d'un projet de loi sur la dépénalisation des drogues douces, un sondage est effectué auprès des membres de la majorité et des membres de l'opposition. Les résultats sont les suivants :

	Favorable	Opposé	Abstention
Majorité	90	110	25
Opposition	50	96	29

1 Formulation des hypothèses :

H_0 : Il n'y a pas de différence significative d'opinion sur la dépénalisation

H_1 : Il y a une différence significative d'opinion sur la dépénalisation

2 On calcule les effectifs théoriques : Pour calculer les effectifs théorique, on multiplie les totaux de la ligne et de la colonne correspondant à la case et en divisant par l'effectif total.

	Favorable	Opposé	Abstention	Total
Majorité	90 (78,8)	110 (115,9)	25 (30,4)	225
Opposition	50 (61,3)	96 (90,1)	29 (23,6)	175
Total	140	206	54	400

3 On calcule T

$$T = \frac{(90 - 78,8)^2}{78,8} + \frac{(110 - 115,9)^2}{115,9} + \dots + \frac{(29 - 23,6)^2}{23,6} \simeq 6,6 \quad (6,53 \text{ sans arrondi.})$$

Le nombre de degrés de liberté est $ddl = (3 - 1) \times (2 - 1) = 2$

Remarque : $ddl = \underbrace{\left(\begin{array}{c} \text{nb de } C_i \text{ utilisés} \\ \text{dans le calcul de } T \end{array} \right)}_{6: \text{ un par } \cap} - 1 - \underbrace{\left(\begin{array}{c} \text{nb de paramètres} \\ \text{estimés pour} \\ \text{le calcul des } C_i \end{array} \right)}_{2+1: \text{ colonne et ligne des totaux}} = 6 - 1 - 3 = 2$

4 Règle de décision : Seuil de signification du test $\alpha = 5\%$.

$$\left| \begin{array}{l} H_0 : T \leq \chi_{5\%, 2}^2 \\ H_1 : T > \chi_{5\%, 2}^2 \end{array} \right.$$

$\chi_{5\%, 2}^2 = 5,991$ donc $T = 6,6 > 5,991$ l'hypothèse H_0 est rejetée.

On peut dire au risque de 5% qu'il y a une différence significative d'opinion sur la dépenalisation entre l'opposition et la majorité.

Chapitre 7 - Fiabilité des systèmes.

Dans la vie courante, la fiabilité est la capacité d'un appareil de fonctionner correctement dans le temps.

Ainsi, pour L'Association Française de Normalisation (AFNOR) la **fiabilité** est « *la caractéristique d'un dispositif exprimée par la probabilité que ce dispositif accomplisse une fonction requise dans des conditions d'utilisation données et pour une période de temps déterminée.* »

Dans ce chapitre, nous nous limitons au cas où cette « période donnée » est située avant la première panne ou défaillance, soit après une réparation qui a permis de remettre le dispositif à neuf. Dans chacune de ces deux situations, nous allons étudier comment une telle probabilité peut être obtenue.

Une telle étude est maintenant devenue importante pour des raisons évidentes de qualité, mais essentiel dans les secteurs où se posent des problème de sécurité ou lorsque les réparations sont impossibles.

I. Premières notions de fiabilité.

1. Définitions.

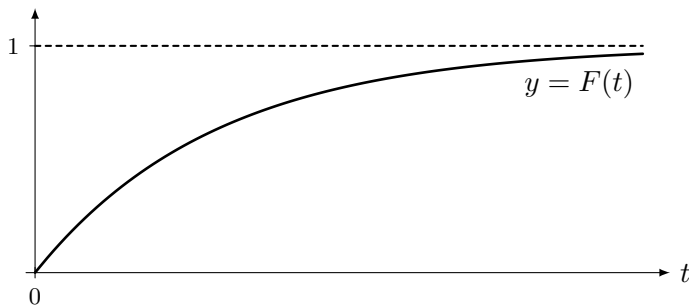
Dans tout ce chapitre,

- nous nous intéressons à un dispositif **pris au hasard** dans une population constituée des dispositifs du même type. Dans le domaine industriel, ce peut être une partie d'une machine, une machine, un réseau de machines, etc.
- On désigne par T la variable aléatoire qui, à tout dispositif choisi au hasard, associe son temps de bon fonctionnement ou sa durée de vie avant une défaillance.

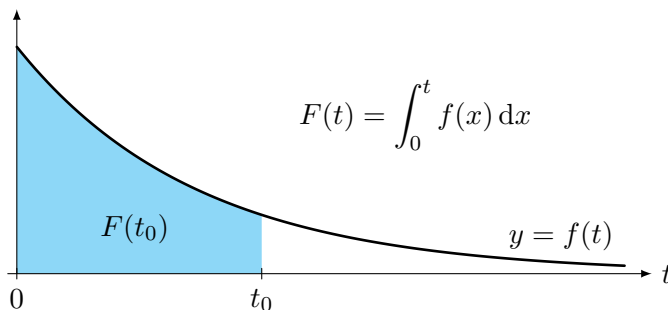


Définition:

On appelle **fonction de défaillance** la fonction F définie pour tout $t \geq 0$ par $F(t) = P(T \leq t)$.



- ☞ $F(t)$ est la probabilité qu'un dispositif prélevé au hasard dans la population ait une défaillance avant l'instant t .
- ☞ $t \geq 0$ car on prélève le dispositif à l'instant $t = 0$.
- ☞ Lorsque t tend vers $+\infty$, la probabilité d'avoir une défaillance devient certaine.



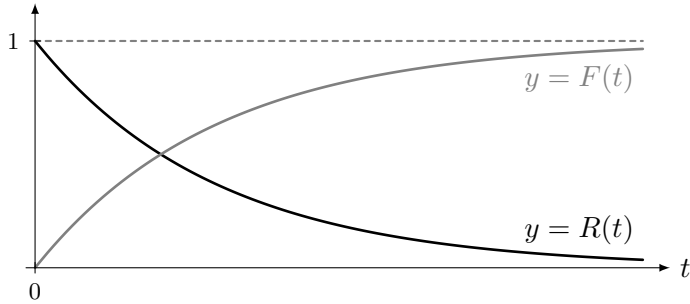
- ☞ F est une fonction répartition. Sa distribution (densité) est notée $f : F' = f$.
- ☞ $t \geq 0$ car on prélève le dispositif à l'instant $t = 0$.

On a alors : $P(T > t) = P(\overline{T \leq t}) = 1 - P(T \leq t) = 1 - F(t)$. Il s'en suit la définition suivante :



Définition:

On appelle **fonction de fiabilité**, la fonction R , définie pour tout $t \geq 0$ par $R(t) = 1 - F(t)$



$R(t)$ représente la probabilité qu'un dispositif choisi au hasard dans la population n'ait pas de défaillance avant l'instant t .

2. Estimation de $F(t)$ et de $R(t)$.

Dans la pratique, pour un dispositif donné, nous ne connaissons pas les valeurs exactes de $F(t)$ et de $R(t)$ pour une valeur donnée de t . Aussi sommes-nous amenés à **estimer** les $F(t)$ et $R(t)$ à partir de valeurs observées sur un échantillon.

Exemple n° 30 : On a mesuré pour 20 aspirateurs du même type le temps en heures écoulé avant la première panne :

Intervalle de temps en heures	[0; 500]]500; 1000]]1000; 1500]]1500; 2000]]2000; 2500]]2500; 3000]]3000; 4000]
Nombre d'appareils	7	4	3	2	2	1	1

On souhaite estimer les valeurs de la fonction de défaillance F suivant les valeurs de t .

On note n_i le nombre de dispositifs défaillants à l'instant t_i et n l'effectif total de l'échantillon.

On peut utiliser 3 méthodes :

1. **Méthode des rangs bruts :**

On calcule les valeurs de F grâce à la formule $F(t_i) = \frac{n_i}{n}$

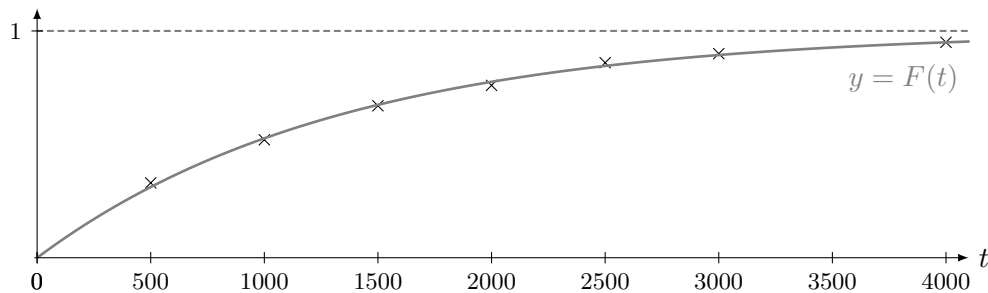
t	500	1000	1500	2000	2500	3000	4000
$F(t)$	$\frac{7}{20} = 0,35$	$\frac{11}{20} = 0,55$	$\frac{14}{20} = 0,70$	0,80	0,90	0,95	1
$R(t)$	$1 - 0,35 = 0,65$	$1 - 0,55 = 0,45$	$1 - 0,70 = 0,30$	0,20	0,10	0,05	0

2. **Méthode des rangs moyens :**

Avec la méthode précédente, la probabilité qu'un dispositif n'ait pas eu de défaillance à l'instant $t = 4000$ est estimée à $R(4000) = 0$: Aucun dispositif n'aurait une durée de vie supérieure à 4000 heures. Ce qui dans la réalité semble exagéré. En effet, si on considère un grand nombre de dispositifs, certains devraient survivre à 4000 heures. Pour remédier à ce problème, en particulier lorsque l'échantillon est petit, on peut prendre

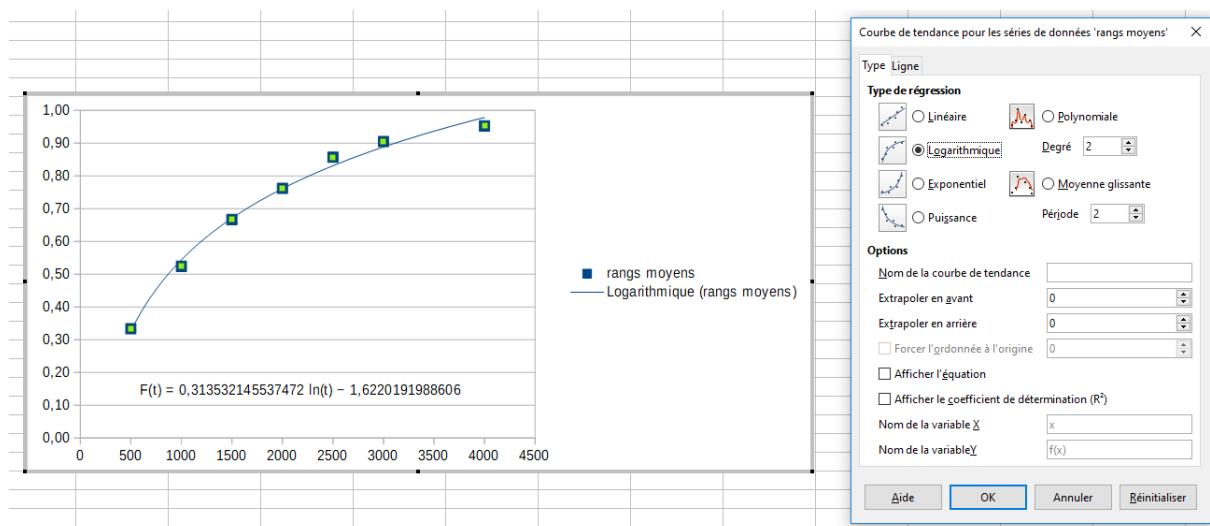
$F(t_i) = \frac{n_i}{n + 1}$:

t	500	1000	1500	2000	2500	3000	4000
$F(t)$	$\frac{7}{21} = 0,33$	$\frac{11}{21} = 0,52$	0,67	0,76	0,86	0,90	0,95



Pour déterminer les images la fonction de défaillance F :

- On peut relier « harmonieusement » les points à la main, et lire les valeurs prises par F graphiquement sur la courbe. On oublie pas de tracer l'asymptote $y = 1$.
- Utiliser la méthode des moindres carrés pour déterminer une régression linéaire, exponentielle, quadratique, etc.
- Utiliser un tableur, et lui demander d'« insérer une courbe de tendance » :



Le tableur de « LibreOffice » propose $F(t) \simeq 0,3135 \ln(t) - 1,622$.

- Reconnaître la fonction de répartition d'une loi de probabilité connue.

Les points semble suivre la fonction de répartition d'une loi **exponentielle** dont l'expression de la fonction de répartition est : $F(t) = 1 - e^{-\lambda t}$ avec $\lambda > 0$. On sait que $F(4000) = 0,95$ donc :

$$1 - e^{-4000\lambda} = 0,95$$

$$-4000\lambda = \ln(0,05)$$

$$\lambda = \frac{-\ln(0,05)}{4000}$$

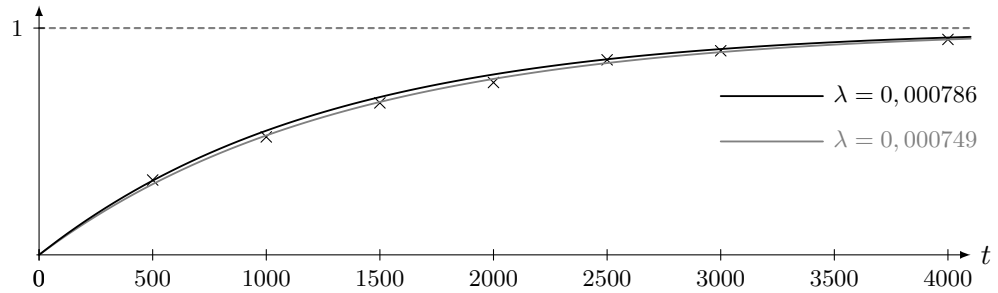
$$\lambda \simeq 0,000749$$

$$\text{d'où } F(t) \simeq 1 - e^{-0,000749t}$$

Ainsi, la probabilité qu'un dispositif n'ait pas de défaillance avant 1233 heures est :

$$R(1233) \simeq 1 - F(1233) = e^{-0,000749 \times 1233} = 60,3\%$$

Remarque : On a choisit de déterminer λ en posant $F(4000) = 0,95$. Ce choix est arbitraire, on aurait pu prendre $F(2500) = 0,86$, ce qui nous aurait conduit à $\lambda \simeq 0,000786$.



Les différentes estimations des fonctions de défaillances suivant les valeurs de λ .

3. Méthode des rangs médians :

Enfin, quand l'échantillon est **petit**, on peut prendre $F(t_i) = \frac{n_i - 0,3}{n + 0,4}$

t	500	1000	1500	2000	2500	3000	4000
$F(t)$	$\frac{6,7}{20,4} \simeq 0,33$	$\frac{10,7}{20,4} \simeq 0,52$	0,67	0,77	0,87	0,92	0,97

3. Taux d'avarie.

a. Approche statistique.

Reprenons notre exemple :

Instants en heures	500	1000	1500	2000	2500	3000	4000
Nombre d'appareils en fonctionnement	$20 - 7 = 13$	9	6	4	2	1	0

- Entre 1000 et 1500 heures, $9 - 6 = 3$ aspirateurs sont tombés en panne sur 9.

Donc, le taux d'**avarie moyen** entre 1000 et 1500 heures est $\frac{3}{9} = 0,33$, autrement dit,

33% des aspirateurs sont tombés en panne dans l'intervalle]1000;1500].

Le taux d'**avarie moyen par unité de temps** entre 1000 et 1500 heures est donc de $\frac{0,33}{500} = 0,00066$ aspirateurs par heure.

- Entre 2000 et 2500 heures, $4 - 2 = 2$ aspirateurs sont tombés en panne sur 4.

Donc, le taux d'**avarie moyen** entre 2000 et 2500 heures est $\frac{2}{4} = 50\%$.

Le taux d'avarie moyen par unité de temps entre 2000 et 2500 heures est de $\frac{0,5}{500} = 0,001$ aspirateurs par heure.

Intervalle de temps en heures	[0; 500]]500; 1000]]1000; 1500]]1500; 2000]]2000; 2500]]2500; 3000]]3000; 4000]
Taux d'avarie moyen	0,35	0,31	0,33	0,33	0,50	0,50	1
Taux d'avarie moyen par heure	0,07%	0,062%	0,066%	0,066%	0,1%	0,1%	0,1%

b. Approche probabiliste.

Notons N l'effectif totale de notre échantillon, et plaçons-nous dans l'intervalle de temps $[t, t + h]$:

- $F(t) \times N$ est le nombre d'aspirateurs défectueux à l'instant t .
- $F(t + h) \times N$ est le nombre d'aspirateurs défectueux à l'instant $t + h$.
- $(F(t + h) - F(t)) \times N$ est le nombre d'aspirateurs défectueux dans l'intervalle de temps $[t, t + h]$.
- $R(t) \times N$ est le nombre d'aspirateurs en fonctionnement à l'instant t .



Définition:

Dans l'intervalle de temps $[t, t + h]$, le **taux d'avarie (ou de défaillance)**

- moyen est $\frac{F(t + h) - F(t)}{R(t)}$
- moyen par unité de temps est $\frac{F(t + h) - F(t)}{h \times R(t)}$
- **instantanée** est $\frac{f(t)}{R(t)}$

Le taux d'avarie instantanée à l'instant t est noté $\lambda(t)$.



Démonstration

Le taux d'avarie instantanée est $\lim_{h \rightarrow 0} \frac{(F(t + h) - F(t)) \times N}{h \times R(t) \times N} = \frac{F'(t)}{R(t)} = \frac{f(t)}{R(t)}$.

On en déduit que :




Propriété

$$\lambda(t) = \frac{f(t)}{R(t)}; \lambda(t) = -\frac{R'(t)}{R(t)}; \text{ et } \lambda(t) = \frac{f(t)}{1 - F(t)}$$

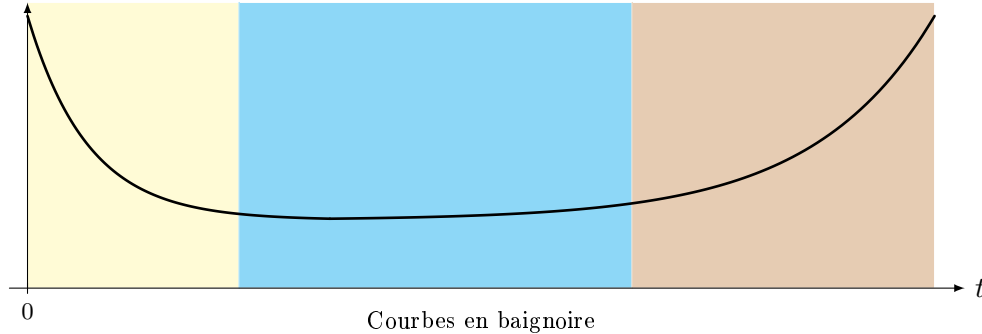
Ces égalités permettent de relier le taux d'avarie instantanée à la fonction de défaillance ou à la fonction de fiabilité.

En résolvant ces égalités qui sont des équations différentielles on obtient :

 **Propriété**


$$R(t) = \exp\left(-\int_0^t \lambda(x) dx\right) \text{ et } F(t) = 1 - \exp\left(-\int_0^t \lambda(x) dx\right)$$

On constate expérimentalement que, pour la plupart des matériels, la courbe représentative du taux d'avarie instantané $t \mapsto \lambda(t)$ a la forme donnée par la figure ci-dessous. Elle est appelée « courbe en baignoire » et comporte trois parties distinctes :



- Pannes précoces** : la période de début de fonctionnement, où le taux d'avarie instantané décroît avec le temps, car les pannes précoces dues à des défauts de fabrication ou de conception sont de moins en moins nombreuses.
- Vie utile** : la période de maturité, ou « vie utile », où le taux d'avarie instantané reste à peu près constant ; pendant cette période, les pannes paraissent dues au hasard.
- Usure** : la période d'usure, où le taux d'avarie instantané augmente avec le temps, car les pannes sont dues à l'usure croissante du matériel.

c. MTBF

 **Définition:**

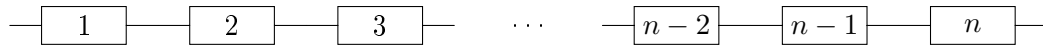
Le **Temps Moyen de Bon Fonctionnement** (Mean Time Between Failure) est l'espérance de T :

$$\text{MTBF} = E(T) = \int_0^{+\infty} t f(t) dt$$

L'espérance, $E(T)$ de la variable aléatoire T représente la **durée de vie moyenne** d'un dispositif avant sa première défaillance.

II. Fiabilité d'un système.

- **Fiabilité d'un système monté en série :**



Pour un système constitués de n composants montés en série (le bon fonctionnement de chacun étant indépendant du bon fonctionnement des autres), on montre que l'on a :

$$R(T) = R_1(t) \times R_2(t) \times \dots \times R_n(t)$$

où R_1, R_2, \dots, R_n sont les fonctions de fiabilités respectives des n composants.

Remarque : En effet, le système est défaillant dès qu'un seul composant est défaillant.



Démonstration

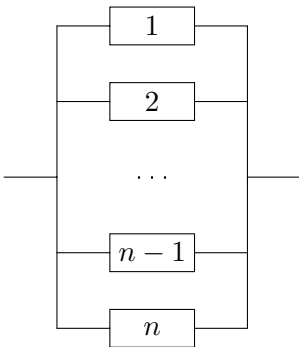
En notant T_i la variable aléatoire durée de vie avant une défaillance du composant i , on a :

$$P(T > t) = P((T_1 > t) \cap (T_2 > t) \cap \dots \cap (T_n > t))$$

Le bon fonctionnement de chacun des composants étant indépendant du bon fonctionnement des autres, on a :

$$P(T > t) = P(T_1 > t) \times P(T_2 > t) \times \dots \times P(T_n > t) = R_1(t) \times R_2(t) \times \dots \times R_n(t)$$

- **Fiabilité d'un système monté en parallèle :**



Pour un système constitués de n composants montés en parallèles (le bon fonctionnement de chacun étant indépendant du bon fonctionnement des autres), on montre que l'on a

$$F(T) = F_1(t) \times F_2(t) \times \dots \times F_n(t)$$

où F_1, F_2, \dots, F_n sont les fonctions de défaillances respectives des n composants. En effet, le système est fonctionnel dès qu'un seul composant est fonctionnel.



Démonstration

Pour que le système monté en parallèles soit défaillant, il faut que tous les composants le soit, donc :

$$P(T \leq t) = P((T_1 \leq t) \cap (T_2 \leq t) \cap \dots \cap (T_n \leq t))$$

Le bon fonctionnement de chacun des composants étant indépendant du bon fonctionnement des autres, donc* le mauvais fonctionnement de chacun des composants étant indépendant du mauvais fonctionnement des autres, et on a :

$$P(T \leq t) = P(T_1 \leq t) \times P(T_2 \leq t) \times \dots \times P(T_n \leq t) = F_1(t) \times F_2(t) \times \dots \times F_n(t)$$

(*) Si A et B sont indépendants alors \overline{A} et \overline{B} aussi.

III. Application avec la loi exponentielle.

Dans un certain nombre de cas, à partir des valeurs numériques de fiabilité ou de défaillance établies grâce à un échantillonnage et par exemple la méthode des rangs, on peut utiliser une loi de probabilité pour décrire les données.

On peut utiliser n'importe quelle loi de probabilité pourvu qu'elle décrive les données. Dans la pratique, on retient en général quatre lois :

- la loi exponentielle ;
- la loi normale ;
- la loi log-normale ;
- la loi de Weibull.

Dans toute cette section, nous allons nous concentrer sur la loi exponentielle. Ainsi, la densité de la variable aléatoire T sera toujours : $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$ avec $\lambda > 0$.

Cette loi concerne tous les matériels pendant une durée de leur « **vie utile** » (voir la courbe en baignoire) et les matériels électroniques pendant presque toute leur vie.

Propriété

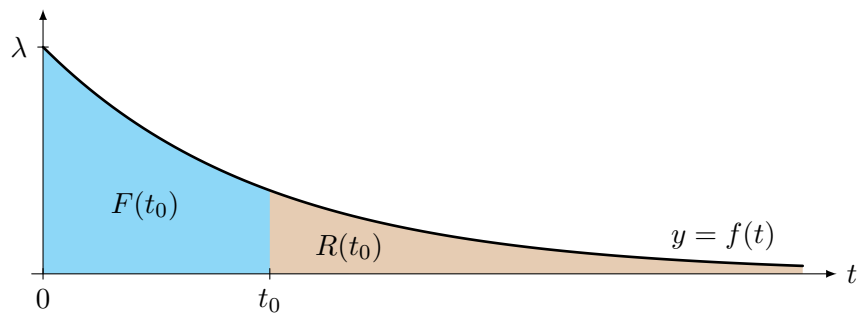
La loi exponentielle est la loi suivie par la variable aléatoire T lorsque le taux d'avarie est **constant**. Autrement dit, on a :

$$\forall t \geq 0, \lambda(t) = \lambda \iff \forall t \geq 0, f(t) = \lambda e^{-\lambda t}$$

où λ est une constante réelle strictement positive.

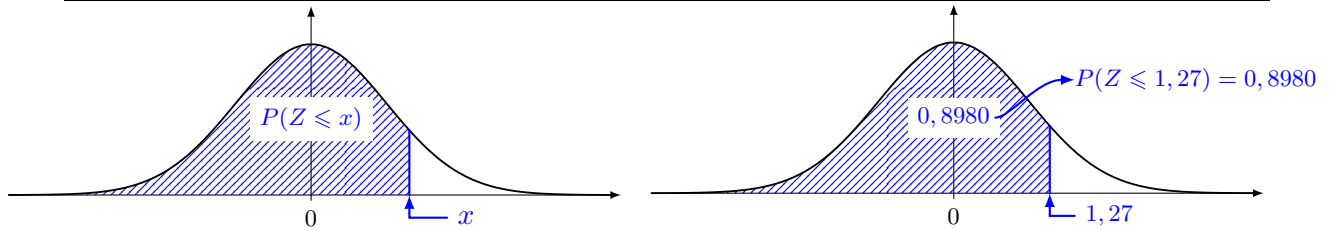
Propriété

- La fonction de fiabilité est définie pour tout $t \geq 0$ par $R(t) = e^{-\lambda t}$.
- La fonction de défaillance est définie pour tout $t \geq 0$ par $F(t) = 1 - e^{-\lambda t}$.
- La densité de probabilité de la variable aléatoire T est définie pour tout $t \geq 0$ par $f(t) = \lambda e^{-\lambda t}$.



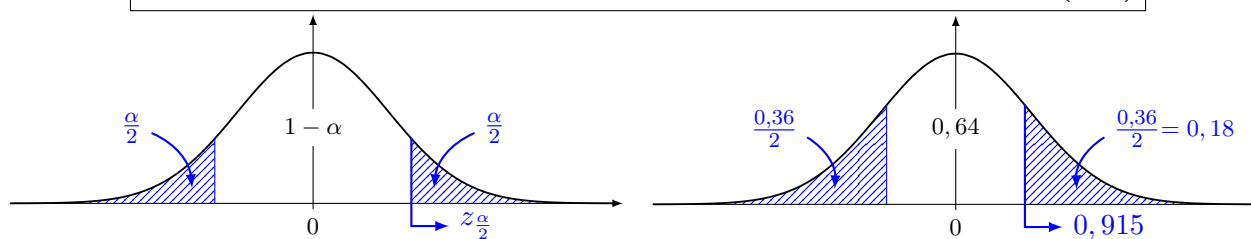
 **Propriété**

- Le temps moyen de bon fonctionnement est $E(T) = \frac{1}{\lambda}$
- L'écart-type de la variable aléatoire T est $\sigma_T = \frac{1}{\lambda}$

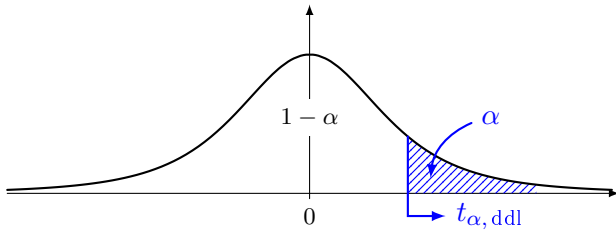
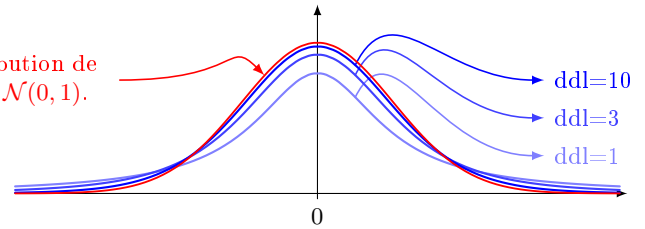
Table de la fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0, 1)$ 

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974

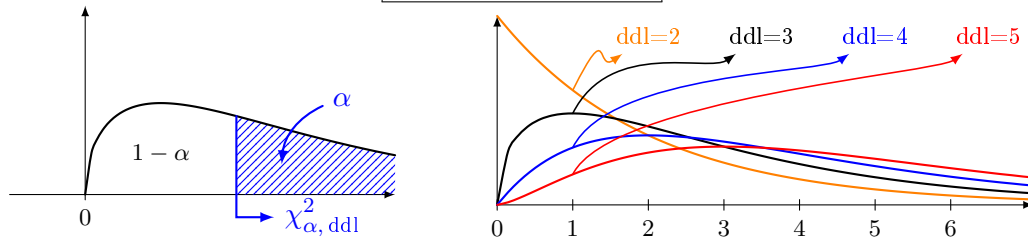
x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Table de l'écart réduit de la loi normale centrée réduite $\mathcal{N}(0,1)$ 

α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

Table de la loi de Student $\mathcal{T}(\text{ddl})$ Distribution de
la loi $\mathcal{N}(0, 1)$.

ddl \backslash \alpha	0,005	0,01	0,025	0,05	0,10	0,20	0,25	0,30	0,40
1	63,657	31,821	12,706	6,314	3,078	1,376	1,000	0,727	0,325
2	9,925	6,965	4,303	2,920	1,886	1,061	0,816	0,617	0,289
3	5,841	4,541	3,182	2,353	1,638	0,978	0,765	0,584	0,277
4	4,604	3,747	2,776	2,132	1,533	0,941	0,741	0,569	0,271
5	4,032	3,365	2,571	2,015	1,476	0,920	0,727	0,559	0,267
6	3,707	3,143	2,447	1,943	1,440	0,906	0,718	0,553	0,265
7	3,499	2,998	2,365	1,895	1,415	0,896	0,711	0,549	0,263
8	3,355	2,896	2,306	1,860	1,397	0,889	0,706	0,546	0,262
9	3,250	2,821	2,262	1,833	1,383	0,883	0,703	0,543	0,261
10	3,169	2,764	2,228	1,812	1,372	0,879	0,700	0,542	0,260
11	3,106	2,718	2,201	1,796	1,363	0,876	0,697	0,540	0,260
12	3,055	2,681	2,179	1,782	1,356	0,873	0,695	0,539	0,259
13	3,012	2,650	2,160	1,771	1,350	0,870	0,694	0,538	0,259
14	2,977	2,624	2,145	1,761	1,345	0,868	0,692	0,537	0,258
15	2,947	2,602	2,131	1,753	1,341	0,866	0,691	0,536	0,258
16	2,921	2,583	2,120	1,746	1,337	0,865	0,690	0,535	0,258
17	2,898	2,567	2,110	1,740	1,333	0,863	0,689	0,534	0,257
18	2,878	2,552	2,101	1,734	1,330	0,862	0,688	0,534	0,257
19	2,861	2,539	2,093	1,729	1,328	0,861	0,688	0,533	0,257
20	2,845	2,528	2,086	1,725	1,325	0,860	0,687	0,533	0,257
21	2,831	2,518	2,080	1,721	1,323	0,859	0,686	0,532	0,257
22	2,819	2,508	2,074	1,717	1,321	0,858	0,686	0,532	0,256
23	2,807	2,500	2,069	1,714	1,319	0,858	0,685	0,532	0,256
24	2,797	2,492	2,064	1,711	1,318	0,857	0,685	0,531	0,256
25	2,787	2,485	2,060	1,708	1,316	0,856	0,684	0,531	0,256
26	2,779	2,479	2,056	1,706	1,315	0,856	0,684	0,531	0,256
27	2,771	2,473	2,052	1,703	1,314	0,855	0,684	0,531	0,256
28	2,763	2,467	2,048	1,701	1,313	0,855	0,683	0,530	0,256
29	2,756	2,462	2,045	1,699	1,311	0,854	0,683	0,530	0,256

Table du χ^2 (ddl)

ddl \ α	0,005	0,010	0,025	0,050	0,100	0,250	0,750	0,900	0,950	0,990	0,995
1	7,879	6,635	5,024	3,841	2,706	1,323	0,1015	0,01579	0,003932	0,000157	0,000039
2	10,60	9,210	7,378	5,991	4,605	2,773	0,5754	0,2107	0,1026	0,02010	0,01003
3	12,84	11,34	9,348	7,815	6,251	4,108	1,213	0,5844	0,3518	0,1148	0,07172
4	14,86	13,28	11,14	9,488	7,779	5,385	1,923	1,064	0,7107	0,2971	0,2070
5	16,75	15,09	12,83	11,07	9,236	6,626	2,675	1,610	1,145	0,5543	0,4117
6	18,55	16,81	14,45	12,59	10,64	7,841	3,455	2,204	1,635	0,8721	0,6757
7	20,28	18,48	16,01	14,07	12,02	9,037	4,255	2,833	2,167	1,239	0,9893
8	21,95	20,09	17,53	15,51	13,36	10,22	5,071	3,490	2,733	1,646	1,344
9	23,59	21,67	19,02	16,92	14,68	11,39	5,899	4,168	3,325	2,088	1,735
10	25,19	23,21	20,48	18,31	15,99	12,55	6,737	4,865	3,940	2,558	2,156
11	26,76	24,72	21,92	19,68	17,28	13,70	7,584	5,578	4,575	3,053	2,603
12	28,30	26,22	23,34	21,03	18,55	14,85	8,438	6,304	5,226	3,571	3,074
13	29,82	27,69	24,74	22,36	19,81	15,98	9,299	7,042	5,892	4,107	3,565
14	31,32	29,14	26,12	23,68	21,06	17,12	10,17	7,790	6,571	4,660	4,075
15	32,80	30,58	27,49	25,00	22,31	18,25	11,04	8,547	7,261	5,229	4,601
16	34,27	32,00	28,85	26,30	23,54	19,37	11,91	9,312	7,962	5,812	5,142
17	35,72	33,41	30,19	27,59	24,77	20,49	12,79	10,09	8,672	6,408	5,697
18	37,16	34,81	31,53	28,87	25,99	21,60	13,68	10,86	9,390	7,015	6,265
19	38,58	36,19	32,85	30,14	27,20	22,72	14,56	11,65	10,12	7,633	6,844
20	40,00	37,57	34,17	31,41	28,41	23,83	15,45	12,44	10,85	8,260	7,434
21	41,40	38,93	35,48	32,67	29,62	24,93	16,34	13,24	11,59	8,897	8,034
22	42,80	40,29	36,78	33,92	30,81	26,04	17,24	14,04	12,34	9,542	8,643
23	44,18	41,64	38,08	35,17	32,01	27,14	18,14	14,85	13,09	10,20	9,260
24	45,56	42,98	39,36	36,42	33,20	28,24	19,04	15,66	13,85	10,86	9,886
25	46,93	44,31	40,65	37,65	34,38	29,34	19,94	16,47	14,61	11,52	10,52
26	48,29	45,64	41,92	38,89	35,56	30,43	20,84	17,29	15,38	12,20	11,16
27	49,64	46,96	43,19	40,11	36,74	31,53	21,75	18,11	16,15	12,88	11,81
28	50,99	48,28	44,46	41,34	37,92	32,62	22,66	18,94	16,93	13,56	12,46
29	52,34	49,59	45,72	42,56	39,09	33,71	23,57	19,77	17,71	14,26	13,12
30	53,67	50,89	46,98	43,77	40,26	34,80	24,48	20,60	18,49	14,95	13,79
31	55,00	52,19	48,23	44,99	41,42	35,89	25,39	21,43	19,28	15,66	14,46
32	56,33	53,49	49,48	46,19	42,58	36,97	26,30	22,27	20,07	16,36	15,13
33	57,65	54,78	50,73	47,40	43,75	38,06	27,22	23,11	20,87	17,07	15,82
34	58,96	56,06	51,97	48,60	44,90	39,14	28,14	23,95	21,66	17,79	16,50
35	60,27	57,34	53,20	49,80	46,06	40,22	29,05	24,80	22,47	18,51	17,19
36	61,58	58,62	54,44	51,00	47,21	41,30	29,97	25,64	23,27	19,23	17,89
37	62,88	59,89	55,67	52,19	48,36	42,38	30,89	26,49	24,07	19,96	18,59
38	64,18	61,16	56,90	53,38	49,51	43,46	31,81	27,34	24,88	20,69	19,29

α ddl	0,005	0,010	0,025	0,050	0,100	0,250	0,750	0,900	0,950	0,990	0,995
39	65,48	62,43	58,12	54,57	50,66	44,54	32,74	28,20	25,70	21,43	20,00
40	66,77	63,69	59,34	55,76	51,81	45,62	33,66	29,05	26,51	22,16	20,71
41	68,05	64,95	60,56	56,94	52,95	46,69	34,58	29,91	27,33	22,91	21,42
42	69,34	66,21	61,78	58,12	54,09	47,77	35,51	30,77	28,14	23,65	22,14
43	70,62	67,46	62,99	59,30	55,23	48,84	36,44	31,63	28,96	24,40	22,86
44	71,89	68,71	64,20	60,48	56,37	49,91	37,36	32,49	29,79	25,15	23,58
45	73,17	69,96	65,41	61,66	57,51	50,98	38,29	33,35	30,61	25,90	24,31
46	74,44	71,20	66,62	62,83	58,64	52,06	39,22	34,22	31,44	26,66	25,04
47	75,70	72,44	67,82	64,00	59,77	53,13	40,15	35,08	32,27	27,42	25,77
48	76,97	73,68	69,02	65,17	60,91	54,20	41,08	35,95	33,10	28,18	26,51
49	78,23	74,92	70,22	66,34	62,04	55,27	42,01	36,82	33,93	28,94	27,25
50	79,49	76,15	71,42	67,50	63,17	56,33	42,94	37,69	34,76	29,71	27,99
51	80,75	77,39	72,62	68,67	64,30	57,40	43,87	38,56	35,60	30,48	28,73
52	82,00	78,62	73,81	69,83	65,42	58,47	44,81	39,43	36,44	31,25	29,48
53	83,25	79,84	75,00	70,99	66,55	59,53	45,74	40,31	37,28	32,02	30,23
54	84,50	81,07	76,19	72,15	67,67	60,60	46,68	41,18	38,12	32,79	30,98
55	85,75	82,29	77,38	73,31	68,80	61,66	47,61	42,06	38,96	33,57	31,73
56	86,99	83,51	78,57	74,47	69,92	62,73	48,55	42,94	39,80	34,35	32,49
57	88,24	84,73	79,75	75,62	71,04	63,79	49,48	43,82	40,65	35,13	33,25
58	89,48	85,95	80,94	76,78	72,16	64,86	50,42	44,70	41,49	35,91	34,01
59	90,72	87,17	82,12	77,93	73,28	65,92	51,36	45,58	42,34	36,70	34,77
60	91,95	88,38	83,30	79,08	74,40	66,98	52,29	46,46	43,19	37,48	35,53
61	93,19	89,59	84,48	80,23	75,51	68,04	53,23	47,34	44,04	38,27	36,30
62	94,42	90,80	85,65	81,38	76,63	69,10	54,17	48,23	44,89	39,06	37,07
63	95,65	92,01	86,83	82,53	77,75	70,16	55,11	49,11	45,74	39,86	37,84
64	96,88	93,22	88,00	83,68	78,86	71,23	56,05	50,00	46,59	40,65	38,61
65	98,11	94,42	89,18	84,82	79,97	72,28	56,99	50,88	47,45	41,44	39,38
66	99,33	95,63	90,35	85,96	81,09	73,34	57,93	51,77	48,31	42,24	40,16
67	100,6	96,83	91,52	87,11	82,20	74,40	58,87	52,66	49,16	43,04	40,94
68	101,8	98,03	92,69	88,25	83,31	75,46	59,81	53,55	50,02	43,84	41,71
69	103,0	99,23	93,86	89,39	84,42	76,52	60,76	54,44	50,88	44,64	42,49
70	104,2	100,4	95,02	90,53	85,53	77,58	61,70	55,33	51,74	45,44	43,28
71	105,4	101,6	96,19	91,67	86,64	78,63	62,64	56,22	52,60	46,25	44,06
72	106,6	102,8	97,35	92,81	87,74	79,69	63,58	57,11	53,46	47,05	44,84
73	107,9	104,0	98,52	93,95	88,85	80,75	64,53	58,01	54,33	47,86	45,63
74	109,1	105,2	99,68	95,08	89,96	81,80	65,47	58,90	55,19	48,67	46,42
75	110,3	106,4	100,8	96,22	91,06	82,86	66,42	59,79	56,05	49,48	47,21
76	111,5	107,6	102,0	97,35	92,17	83,91	67,36	60,69	56,92	50,29	48,00
77	112,7	108,8	103,2	98,48	93,27	84,97	68,31	61,59	57,79	51,10	48,79
78	113,9	110,0	104,3	99,62	94,37	86,02	69,25	62,48	58,65	51,91	49,58
79	115,1	111,1	105,5	100,7	95,48	87,08	70,20	63,38	59,52	52,72	50,38
80	116,3	112,3	106,6	101,9	96,58	88,13	71,14	64,28	60,39	53,54	51,17
81	117,5	113,5	107,8	103,0	97,68	89,18	72,09	65,18	61,26	54,36	51,97
82	118,7	114,7	108,9	104,1	98,78	90,24	73,04	66,08	62,13	55,17	52,77
83	119,9	115,9	110,1	105,3	99,88	91,29	73,99	66,98	63,00	55,99	53,57
84	121,1	117,1	111,2	106,4	101,0	92,34	74,93	67,88	63,88	56,81	54,37
85	122,3	118,2	112,4	107,5	102,1	93,39	75,88	68,78	64,75	57,63	55,17

ddl \ α	0,005	0,010	0,025	0,050	0,100	0,250	0,750	0,900	0,950	0,990	0,995
86	123,5	119,4	113,5	108,6	103,2	94,45	76,83	69,68	65,62	58,46	55,97
87	124,7	120,6	114,7	109,8	104,3	95,50	77,78	70,58	66,50	59,28	56,78
88	125,9	121,8	115,8	110,9	105,4	96,55	78,73	71,48	67,37	60,10	57,58
89	127,1	122,9	117,0	112,0	106,5	97,60	79,68	72,39	68,25	60,93	58,39
90	128,3	124,1	118,1	113,1	107,6	98,65	80,62	73,29	69,13	61,75	59,20
91	129,5	125,3	119,3	114,3	108,7	99,70	81,57	74,20	70,00	62,58	60,00
92	130,7	126,5	120,4	115,4	109,8	100,8	82,52	75,10	70,88	63,41	60,81
93	131,9	127,6	121,6	116,5	110,9	101,8	83,47	76,01	71,76	64,24	61,63
94	133,1	128,8	122,7	117,6	111,9	102,8	84,42	76,91	72,64	65,07	62,44
95	134,2	130,0	123,9	118,8	113,0	103,9	85,38	77,82	73,52	65,90	63,25
96	135,4	131,1	125,0	119,9	114,1	104,9	86,33	78,73	74,40	66,73	64,06
97	136,6	132,3	126,1	121,0	115,2	106,0	87,28	79,63	75,28	67,56	64,88
98	137,8	133,5	127,3	122,1	116,3	107,0	88,23	80,54	76,16	68,40	65,69
99	139,0	134,6	128,4	123,2	117,4	108,1	89,18	81,45	77,05	69,23	66,51