



FIN610 – Les outils statistiques de gestion

La statistique est une branche des mathématiques qui a pour objet l'analyse et l'interprétation de données quantifiables.

On a coutume de distinguer **la statistique descriptive et l'inférence statistique**.

La **statistique descriptive**, comme son nom l'indique, sert à décrire les caractéristiques (âge, revenu...) des individus d'une population donnée à l'aide de tableaux, de graphiques, de mesures de tendance centrale (moyenne, médiane...) et de dispersion (écart-type...).

Par exemple, l'étude de la distribution des salaires dans une entreprise rentre dans ce cadre. Calculer la moyenne d'une classe ou tracer la courbe d'évolution du chômage depuis 2020 relèvent aussi de la statistique descriptive.

L'inférence statistique permet de généraliser les données tirées des échantillons à un ensemble plus vaste, la population.

L'inférence comporte deux volets : **l'estimation et le test d'hypothèse**.

L'estimation joue un grand rôle dans les sondages d'opinion et les études de marché. Dans ces deux cas, l'objectif est d'obtenir rapidement une idée précise de l'opinion ou des goûts d'une grande population.

Un recensement étant hors de question en raison de son coût et de la lenteur du processus, on recueille l'information voulue auprès d'un échantillon jugé représentatif et on généralise les résultats obtenus à l'ensemble de la population.

Par exemple lorsque j'essaie de prédire les intentions de vote à partir d'un échantillon de 1000 personnes je fais de la **statistique inférentielle**.

Les **tests d'hypothèses** servent à faire un choix entre deux hypothèses à partir des renseignements tirés d'un échantillon. Par exemple, nous faisons une affirmation (ou hypothèse) sur un paramètre de la population (un responsable pédagogique suggère que le niveau moyen du QI des étudiants qu'il a en charge est de 110), puis nous vérifions si cette hypothèse est vraie en examinant un échantillon aléatoire pris dans la population (hypothèse vérifiée si la moyenne de l'échantillon prélevé est comprise entre 109 et 111 par exemple).

Avant d'aborder sereinement les théories d'échantillonnage, champ d'étude de la statistique inférentielle, nous ferons un rappel des notions (variables aléatoires, loi normale...) qui seront souvent sollicitées.



FIN610 – Les outils statistiques de gestion



FIN610 – Les outils statistiques de gestion

1) Les variables aléatoires

a) Les variables aléatoires

Étant donné une **expérience aléatoire** (dont on ne connaît pas le résultat à l'avance¹), on appelle **variable aléatoire** toute grandeur numérique dont la valeur dépend de l'issue de l'expérience.

Exemple : Soit un jeu qui consiste à lancer un dé bien équilibré. Lancer le dé une fois constitue une **épreuve** et le résultat de cette épreuve (par exemple obtenir un 2) est un **évènement** (ou éventualité ou issue ou résultat). On a donc :

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

On appelle **univers** (ou ensemble fondamental) l'ensemble de tous les résultats (ou éventualités ou évènements) possibles au cours d'une épreuve.

- si le résultat est 2, 3, 4 ou 5, on gagne 1.50€ ;
- si le résultat est 1 ou 6, on doit verser un montant égal au résultat sur le dé. Ainsi, on paye 1€ si le résultat est 1 et 6€ si le résultat est 6.

Dans cet exemple, la variable aléatoire X représente le gain (ou la perte) du joueur. Ses valeurs possibles sont -1, -6 et 1.50. On notera $X(\Omega) = \{-1, -6, 1.50\}$. Ainsi X est une variable parce qu'elle peut prendre différentes valeurs. Et X est aléatoire parce que la valeur qu'elle prend dépend du hasard.

En associant à toutes les valeurs d'une variable aléatoire la probabilité qui lui correspond, on définit une **loi de probabilité** (ou fonction de distribution). Dans notre exemple, cette loi est donnée par le tableau ci-dessous :

x_i	-6	-1	1.50	
p_i	1/6	1/6	4/6	1

La loi de probabilité de X est l'ensemble des valeurs possibles de X (les x_i) et les probabilités de chacune d'elles (les p_i). La loi de probabilité permet donc de déterminer la probabilité de chaque modalité de la variable aléatoire.

¹ Le résultat est donc uniquement déterminé par le hasard. Par exemple lancer une pièce à deux faces ou jeter un dé à six côtés sont des expériences aléatoires.

FIN610 – Les outils statistiques de gestion

La loi de probabilité de X satisfait toujours les conditions suivantes :

$$(i) p(x_i) \geq 0 \text{ et } (ii) \sum_{i=1}^n p(x_i) = 1$$

Nous avons ici défini une **variable aléatoire discrète** car elle ne peut prendre que certaines valeurs bien précises. Ainsi la variable X « nombre de vélos défectueux dans une station de 20 vélib » est discrète car elle n'a que 21 valeurs possibles, de 0 à 20.

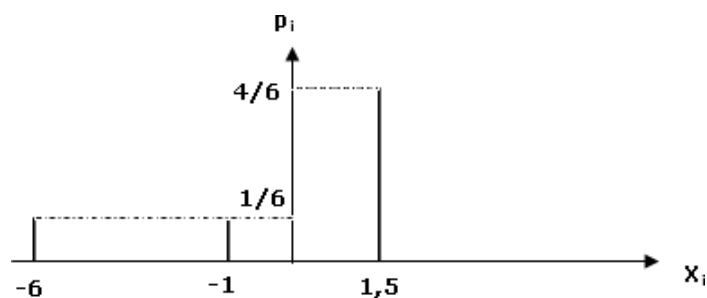
Dans notre exemple, la variable aléatoire X prend les trois valeurs -6, -1 et 1.50. Si on pose :

$$x_1 = -6, x_2 = -1 \text{ et } x_3 = 1.50$$

On constate que la variable aléatoire ne peut prendre qu'un nombre fini de valeurs. Si on note p_i la probabilité pour que X soit égal à x_i . Alors,

$$p_1 = 1/6, p_2 = 1/6 \text{ et } p_3 = 4/6$$

Les représentations graphiques d'une loi de probabilité sont obtenues en portant en abscisse les valeurs de la variable aléatoire et en ordonnée les probabilités. La **fonction de distribution** est tout simplement la représentation de la loi sous la forme d'un diagramme en bâtons, la variable aléatoire étant discrète. On observe que les segments verticaux sont proportionnels aux probabilités.



On pourrait, comme on le fait en statistique descriptive, introduire une fonction cumulative, dite **fonction de répartition F** . Elle indique la probabilité pour que la variable aléatoire X prenne une valeur *au plus* égale à une valeur donnée a :

$$F(a) = P(X \leq a)$$

Enfin la **variable aléatoire** X est dite **continue** lorsqu'elle peut prendre toutes les valeurs d'un intervalle de \mathbb{R} . Par exemple si le nombre de connexions Internet sur

FIN610 – Les outils statistiques de gestion

un site est une variable aléatoire discrète, le temps de connexion est une variable aléatoire continue².

Dans le cas d'une variable continue, il est impossible de présenter toutes les valeurs possibles du caractère étudié (théoriquement, il y en a un nombre infini). On est obligé de regrouper les données par **classes** (par exemple des tranches de revenu ou d'âge). On obtient alors une distribution qui peut être représentée par un **histogramme**. C'est un graphique où chaque classe est représentée par un rectangle dont la largeur est proportionnelle à l'amplitude de la classe et dont la hauteur est proportionnelle à l'effectif de la classe.

Ex : Considérons la répartition de 100 ménages selon le revenu du chef de famille donnée par le tableau suivant :

Revenus	Nombre de ménages	Effectifs corrigés
[3000 ; 5000[30	30/2 = 15
[5000 ; 6000[40	40
[6000 ; 7000[20	20
[7000 ; 9000[10	10/2 = 5

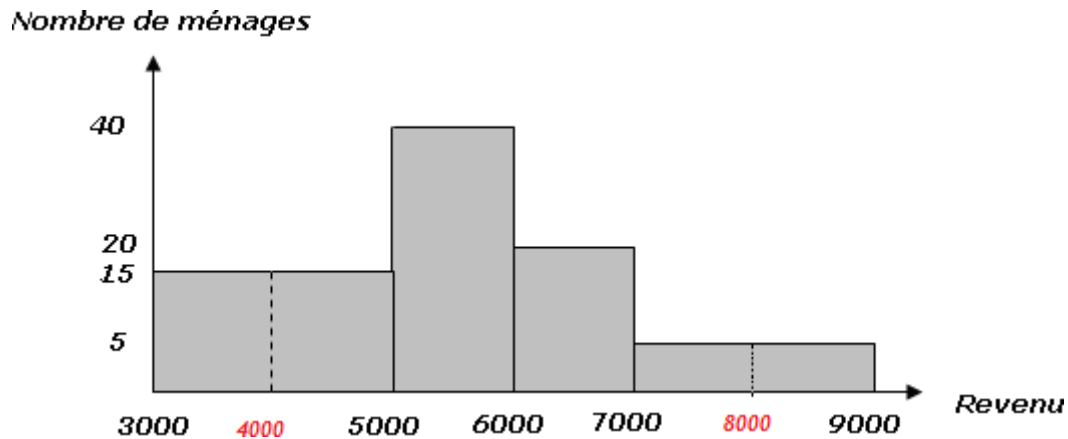
La probabilité que l'on « tire » un individu dont le revenu serait exactement de 5 320.26 euros est pratiquement nulle. Pour éviter ce problème, on a donc préféré considérer des intervalles de revenu, et non des revenus « ponctuels ». Ainsi on cherche la probabilité que le revenu de l'individu tiré soit compris dans l'intervalle [6000 ; 7000[, plutôt que celle qu'il prenne une valeur précise. Ces revenus sont ici regroupés dans des classes d'amplitude 1000 euros. Les classes étant ici inégales, le principe est le suivant : on prend une longueur de classe de référence, en général la plus petite (1000 dans le cas précédent) ou la plus répandue, et dans chaque classe dont la longueur est égale à k fois cette longueur de référence, on divise les effectifs par k^3 .

² On notera que si X est une variable aléatoire continue, alors $p(X = x_i) = 0$ (et ce, quel que soit x). Prenons l'exemple de paquets de farine dont le poids en farine est une variable aléatoire "Y" prenant ses valeurs dans l'intervalle [950 ; 1100]. Lorsque l'on prend un paquet au hasard, la probabilité d'obtenir un poids de farine rigoureusement égal à 978.25 g par exemple est nulle. En effet, c'est un paquet parmi une infinité, et $1/\text{infini} = 0$ (nombre cas favorables/nombre cas possibles). Donc l'événement poids = 978.2 g est dit "quasi impossible".

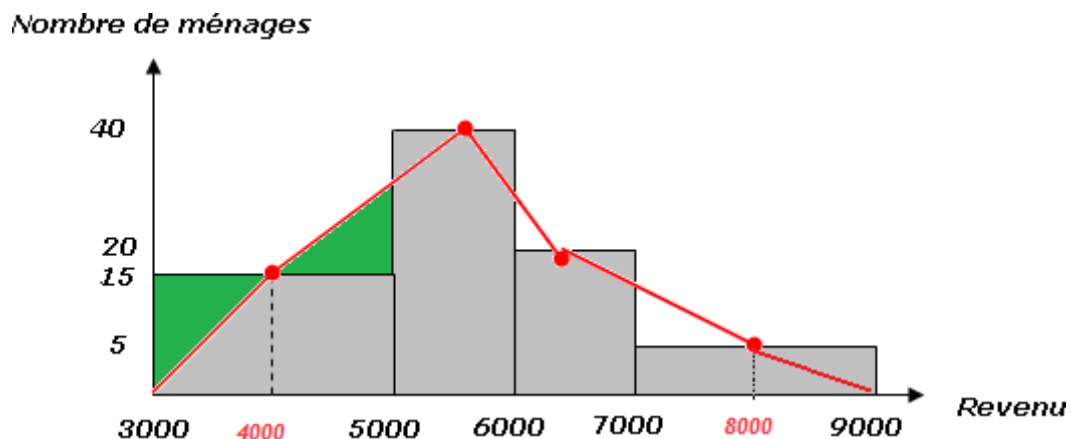
³ Comme la première et la dernière sont le double des autres, il est normal de diviser leurs effectifs par deux, de façon à rendre comparables les situations (ce qui revient, par exemple, à répartir de façon égale les 30 ménages de la première classe, allant de 3000 à 5000, en deux classes : 15 ménages dans la première, 3000-4000, 15 dans la seconde, 4000-5000 (la hauteur du tuyau correspondant égale à : $30/2 = 15$). Même chose, pour la dernière classe.

FIN610 – Les outils statistiques de gestion

Dans notre exemple, le plus petit intervalle étant de 1 000, il est pris comme unité. Un intervalle de 2 000 se trouve être égal à deux unités. La valeur du premier intervalle mesurée avec la nouvelle unité est donc égale à 2 ([3 000 ; 5 000]). La surface du rectangle correspondant est bien égale $2 \times 15 = 30$ soit l'effectif de la classe. D'où la représentation :



Enfin après rectification, on constate que la surface de l'histogramme est bien égale à l'effectif de 100 ($2 \times 15 + 1 \times 40 + 1 \times 20 + 2 \times 5$). Si nous raisonnons en fréquences, la surface de l'histogramme serait égale à la somme des fréquences, donc à l'unité.



Par ailleurs, le caractère continu de la variable peut être matérialisé par la ligne polygonale rouge des effectifs (ou des fréquences) qui joint les sommets des rectangles au niveau des **centres de classe**. Par construction, l'aire sous le **polygone des effectifs** (ou des fréquences) correspond à l'aire totale de l'histogramme. À chaque fois, on perd et on gagne simultanément deux triangles égaux, ce que montre clairement la figure précédente.

Si l'effectif de la population est important, lorsque l'amplitude de la classe tend vers zéro, c.-à-d. que le nombre de classes devient très important, la ligne

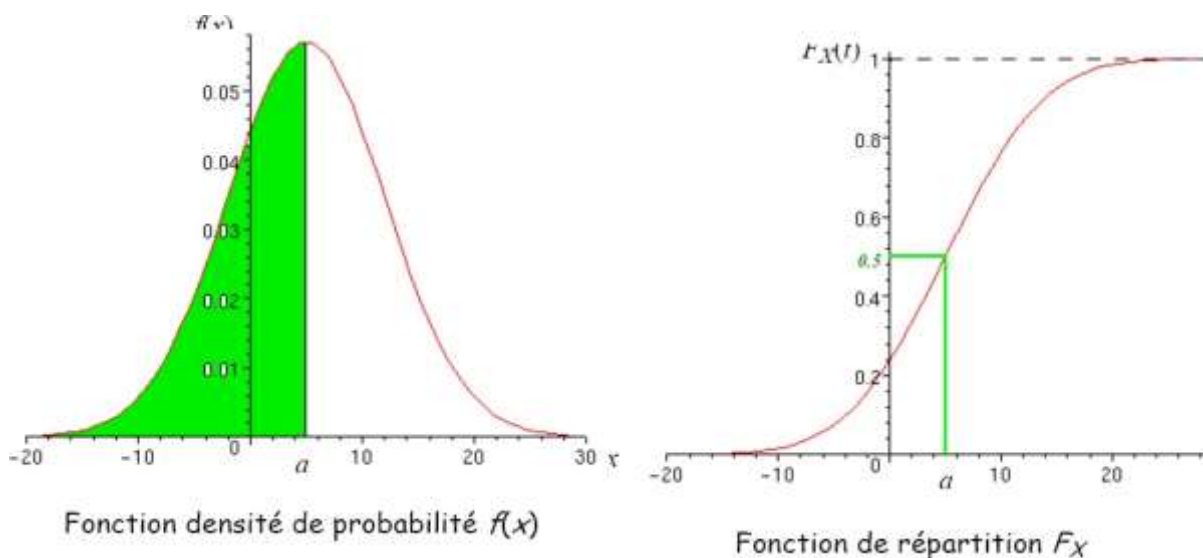
FIN610 – Les outils statistiques de gestion

polygonale du dessus tend vers une courbe de plus en plus lisse appelée **courbe des effectifs** (ou courbe des fréquences). C'est la courbe d'une fonction continue f représentative de la distribution de fréquences. Cette fonction f prend le nom de **densité de probabilité**. Dans le cas d'une variable continue, la fonction de distribution ou loi de probabilité s'appelle donc densité de probabilité $f(x)$. Par analogie avec l'histogramme, l'aire qui est sous la courbe de densité de probabilité est égale à 1 (somme des probabilités).

Enfin la **fonction de répartition** correspondrait aux probabilités cumulées associées à la variable aléatoire continue. La probabilité $P(X \leq x_i)$ est représentée par l'aire située sous la courbe à gauche de l'abscisse x_i (centre de classe)⁴.

$$P(X < x_i) = \int_{-\infty}^{x_i} f(x)dx$$

Par exemple dans le schéma suivant, l'aire hachurée en vert sous la courbe de la fonction densité de probabilité correspond à la probabilité $P(X < a)$ et vaut 0.5 car ceci correspond à la moitié de l'aire totale sous la courbe. Cette probabilité correspond à la valeur de la fonction de répartition au point d'inflexion de la courbe.



b) Espérance mathématique, moyenne et variance de variables discrètes

⁴ Puisque dans le cas d'une variable aléatoire continue, la densité de probabilité en un point est nulle (c.-à-d. que l'intervalle, base du rectangle, est sans dimension, l'aire est donc nulle), on a alors $P(X \leq x_i) = P(X < x_i)$.



FIN610 – Les outils statistiques de gestion

Soit X une variable aléatoire qui possède une loi de probabilité. **L'espérance mathématique** est la somme des valeurs prises par X , pondérée par les probabilités qui leur sont associées :

$$E(X) = x_1 p(x_1) + x_2 p(x_2) + \dots + x_n p(x_n) = \sum_{i=1}^n x_i p(x_i)$$

Ex : Pour poursuivre l'exemple de l'introduction, l'espérance mathématique de gain est,

$$E(X) = (-1)(1/6) + (1.50)(4/6) + (-6)(1/6) = -1/6 \text{ euros}$$

On peut donc s'attendre à perdre en moyenne 1/6€ par coup, ce qui signifie que le jeu est défavorable au joueur.

La variance est définie comme la différence entre la moyenne (l'espérance) des carrés et le carré de la moyenne. Ainsi :

$$V(X) = E(X^2) - [E(X)]^2$$

$$V(X) = \sum_{i=1}^n x_i^2 p(x_i) - \left[\sum_{i=1}^n x_i p(x_i) \right]^2$$

Ex : Pour la variance, nous allons calculer la moyenne des carrés moins le carré de la moyenne. Calculons d'abord la moyenne des carrés :

$$E(X^2) = 6^2 \times 1/6 + (-1)^2 \times 1/6 + 1.50^2 \times 4/6 = 7.67$$

Variance et écart-type :

$$V(X) = 7.67 - (-1/6)^2 = 7.68$$

L'**écart-type** de X , que l'on écrit σ , est la racine carrée (positive ou nulle) de $V(X)$:

$$\sigma_X = \sqrt{\text{Var}(X)}$$

Ex : D'après l'exemple précédent, $\sigma_X = \sqrt{7.68} = 2.77\text{€}$

FIN610 – Les outils statistiques de gestion

Notons que si l'espérance indique la valeur moyenne espérée d'une série de valeurs, **l'écart-type permet de mesurer la dispersion des valeurs autour de cette moyenne** espérée. S'il y a une grande dispersion (c.-à-d. si l'écart entre les valeurs et la moyenne est important), alors l'écart-type est grand ; si au contraire la dispersion est faible, alors l'écart-type est petit.

Attention aussi aux confusions, courantes chez les étudiants. La première concerne la moyenne, le plus souvent notée \bar{x} , et l'espérance mathématique notée $E(X)$. **La moyenne est empirique, l'espérance est théorique.**

Ex : Imaginons un jeu de pile ou face où vous gagnez 1 euro s'il y a face et 0 euros s'il y a pile. La loi de probabilité est donc la suivante :

x_i	1	0	
p_i	1/2	1/2	1

Quel est mon espérance de gain ?

$$E(X) = \sum_{i=1}^2 x_i p(x_i) = (1)(1/2) + (0)(1/2) = 0.5\text{€}$$

C'est ce que vous pouvez espérer gagner **avant** de jouer. L'espérance de gain ne dépend pas de votre expérience : elle vaut toujours 0.50 euros. Mais supposons que vous jouez effectivement en faisant 4 lancers, et que vous avez 1 fois face, alors votre gain moyen (ou moyenne des gains) est de 0.25 euros.

$$\bar{x} = \frac{\sum_{i=1}^4 x_i}{N} = \frac{1+0+0+0}{4} = 0.25\text{€}$$

C'est surtout la formule précédente, celle avec \bar{x} , qui sera utilisée en inférence statistique. Rappelons en effet que l'inférence statistique est un procédé de généralisation des résultats échantillonnaires à l'ensemble de la population. Ainsi un fabricant d'ampoules électriques qui veut estimer la durée de vie moyenne des ampoules qu'il produit prélèvera un échantillon de sa production et estimera la vie moyenne de l'ensemble des ampoules d'après la durée moyenne observée dans l'échantillon. **On fait de l'inférence statistique lorsqu'on estime la moyenne d'une population à partir de la moyenne d'un échantillon tiré de la population.**



FIN610 – Les outils statistiques de gestion

La seconde confusion concerne les indicateurs de dispersion. Lorsqu'on se projette dans le futur, qu'on raisonne en termes probabilistes, la variance se calcule de la façon suivante :

$$V(X) = E(X^2) - [E(X)]^2$$

Lorsqu'on raisonne sur des données passées, on sollicite les formules suivantes :

$$\sigma^2 = \frac{\sum x_i^2}{N} - (\bar{x})^2 \text{ ou } \sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

Ex : Calculons l'écart type de la série observée : 4, 7, 11, 12. La moyenne vaut :

$$\bar{x} = \frac{4 + 7 + 11 + 12}{4} = 8.5$$

Soit le tableau suivant :

x_i	4	7	11	12	Total
$(x_i)^2$	16	49	121	144	330

$$V(X) = \frac{330}{4} - (8.5)^2 = 10.25 \Rightarrow \sigma(X) = \sqrt{10.25} = 3.2$$

FIN610 – Les outils statistiques de gestion

I. Notions d'analyse combinatoire

a) Les permutations et les arrangements

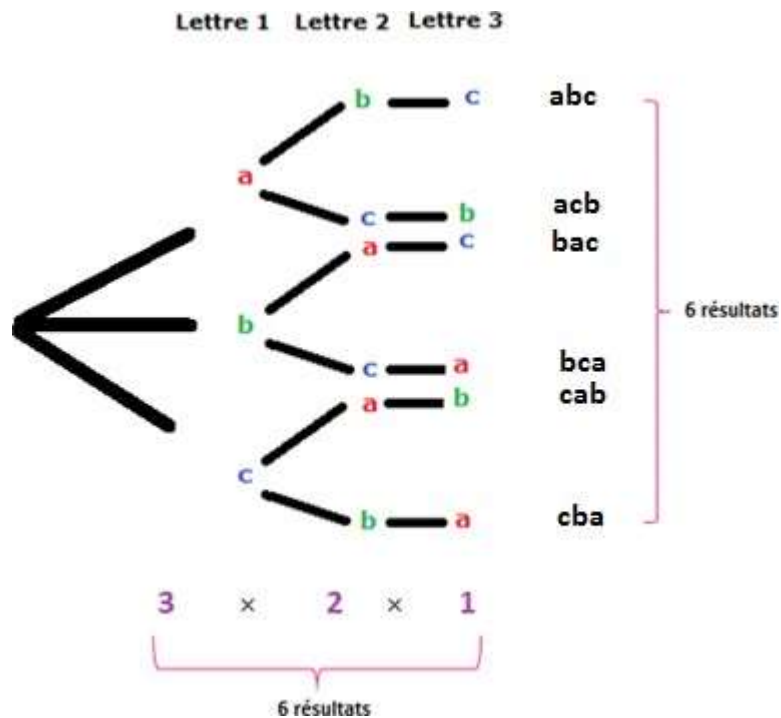
Disposant d'un ensemble de n objets *distincts*, une **permutation** est un rangement **ordonné sans répétition** de ces n objets.

Le nombre total de permutations distinctes, noté $n!$, est égal à :

$$P_n = n! = n (n - 1) (n - 2) (n - 3) \times \dots \times 2 \times 1$$

Par convention, $0! = 1$.

Ex : Le nombre de permutations de 3 éléments d'un ensemble $E = \{a,b,c\}$ est égal à $3! = 3 \times 2 \times 1 = 6$ comme le montre l'arbre ci-dessous.



Dans l'exemple, l'ensemble des permutations à trois éléments est $\{(a,b,c);(b,a,c);(b,c,a);(a,c,b);(c,a,b);(c,b,a)\}$.

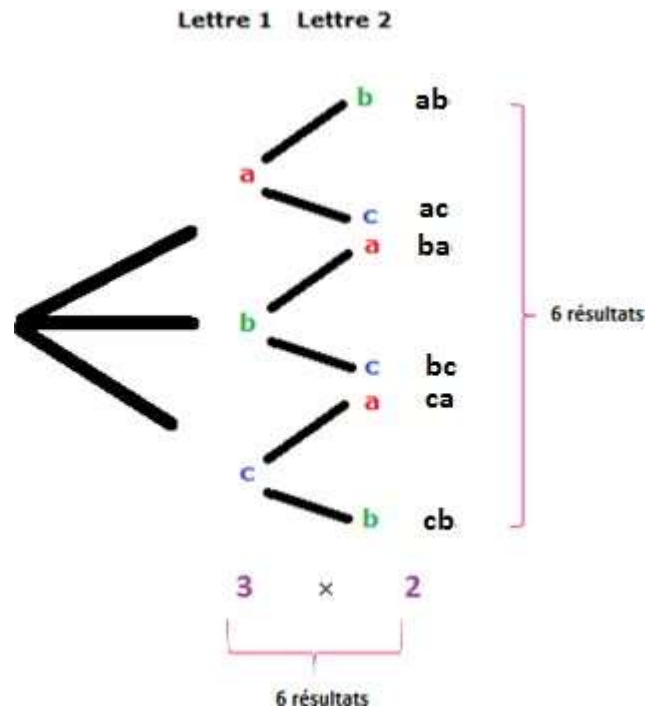
Dans un ensemble de n objets *distincts*, un **arrangement** est une disposition ordonnée de r objets choisis parmi ces n objets.

La formule pour le nombre de permutations (sans répétition) de n objets pris r à la fois est la suivante :

FIN610 – Les outils statistiques de gestion

$$A_n^r = \frac{n!}{(n-r)!}$$

Ex : La question « Combien d'arrangements de deux lettres peut-on obtenir avec les lettres a, b, c ? » implique que ab est différent de ba . Les deux mêmes lettres sont présentes, mais l'ordre d'apparition est différent.



Dans l'exemple, le nombre d'arrangements serait ab, ac, ba, bc, ca, cb . Il y a donc 3×2 arrangements (trois choix pour la première lettre et deux choix pour la seconde). Avec la formule :

$$A_3^2 = \frac{3!}{(3-2)!} = 3 \times 2 = 6$$

Pour les arrangements, la formule avec remise est : **Nombre d'arrangements** = n^r , n est le nombre total d'éléments et r est le nombre d'éléments sélectionnés.

Ex : On choisit au hasard et avec remise deux lettres parmi l'ensemble $E = \{a,b,c\}$ et on s'intéresse au nombre d'arrangements possibles.

$$\text{Nombre d'arrangements} = 3 \times 3 = 3^2$$

On reprend les six résultats précédents et on ajoute aa, bb et cc .

FIN610 – Les outils statistiques de gestion

b) Les combinaisons

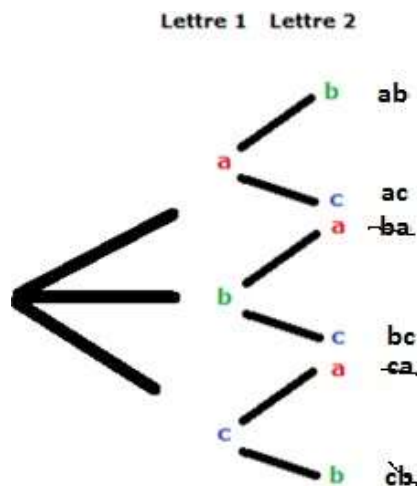
Dans un ensemble de n objets *distincts*, une **combinaison** est une disposition non ordonnée de r objets choisis parmi les n . Contrairement à l'arrangement, l'ordre des objets au sein d'un groupement n'est pas retenu dans les combinaisons. Tout comme dans l'arrangement, les répétitions ne sont pas permises. Une combinaison sera donc le résultat d'un **tirage sans ordre et sans répétition** (on dit aussi **sans remise**).

La formule générale donnant le nombre de combinaisons de n objets pris r à la fois C_n^r est donc donnée par la formule suivante :

$$C_n^r = \frac{n!}{r!(n-r)!}$$

Le nombre de combinaisons de r individus parmi n est égal au nombre d'arrangements de r individus parmi n divisé par $r!$.

Ex : La question « Combien de combinaisons de deux lettres peut-on obtenir avec les lettres a, b, c ? » implique que ab est identique à ba . Autrement dit, ce sont deux arrangements (ou permutations) mais une même combinaison puisqu'on ne tient pas compte de l'ordre.



Dans l'exemple, le nombre de combinaisons serait ab , ac , et bc .

$$C_3^2 = \frac{3!}{2!(3-2)!} = \frac{6}{2} = 3$$

On sait que a, b donnent deux permutations (ab, ba) mais une combinaison ab . Ainsi deux lettres dans l'ensemble $\{a, b, c\}$ donnent deux permutations, mais



FIN610 – Les outils statistiques de gestion

seulement une combinaison. Ici il y a au total 6 permutations, donc pour trouver le nombre de combinaisons, nous devons diviser A_3^2 par P_2 (c.-à-d. 2 !).



FIN610 – Les outils statistiques de gestion

2) La loi normale

a) Principes généraux

L'étude de certains phénomènes fait apparaître une « permanence statistique ». **Cette reproductibilité suggère l'existence de « lois statistiques »**, c.-à-d. de modèles théoriques que suivent de façon plus ou moins proche les phénomènes étudiés.

Ces lois qui font appel à la notion de probabilité sont nombreuses ; cependant, un **petit nombre d'entre elles permet de rendre compte de la majeure partie des phénomènes statistiques.**

C'est le cas en particulier de la loi normale (ou loi de Laplace-Gauss).

La loi normale est très intéressante parce que la très grande majorité des phénomènes naturels tendent vers cette distribution quand on prend un grand nombre de mesures.

Les contextes d'utilisation sont très nombreux, par exemple :

- La distribution de la durée de vie des ampoules électriques,
- Les distributions du poids des baleines
- Les notes d'un examen, La température

La plupart des caractéristiques physiques mesurables respectent la loi normale. **On observe la distribution normale partout dans la nature.**

De façon générale, lorsque l'on a affaire à une variable aléatoire continue dont la valeur dépend d'un grand nombre de causes indépendantes dont les effets s'additionnent et dont aucune n'est prépondérante (conditions de Borel), on montre que la distribution des valeurs de cette variable aléatoire suit une loi de Laplace-Gauss.

Ainsi la taille corporelle d'un animal dépend des facteurs environnementaux (disponibilité pour la nourriture, climat, prédation, etc.) et génétiques. Dans la mesure où ces facteurs sont indépendants et qu'aucun n'est prépondérant, on peut supposer que la taille corporelle suit une loi normale.

La loi normale est entièrement déterminée par deux paramètres :

- sa moyenne " μ "
- et son écart-type " σ ".

FIN610 – Les outils statistiques de gestion

Si une variable aléatoire continue "X" obéit à une loi normale, alors "X" admet la **fonction de densité** suivante :

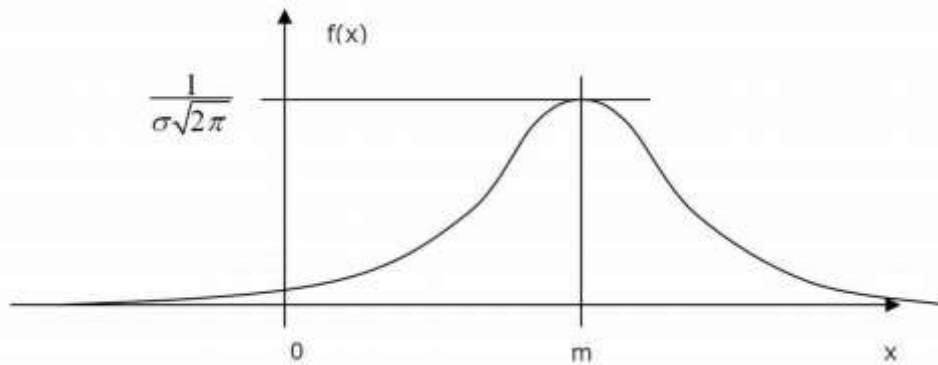
$$f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} e^{-\frac{1}{2} \left[\frac{x-m}{\sigma} \right]^2}$$

Avec, $-\infty < x < +\infty$ et $\sigma > 0$

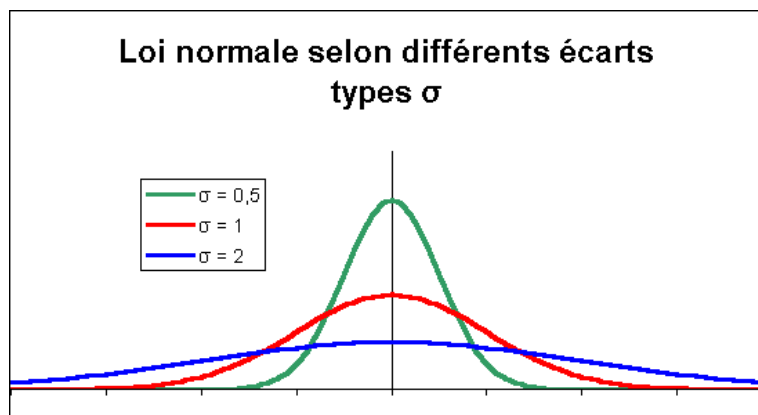
$$\text{Et : } \int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} e^{-\frac{1}{2} \left[\frac{x-m}{\sigma} \right]^2} dx = 1$$

FIN610 – Les outils statistiques de gestion

Ainsi l'expression, $f(x)$ est la courbe de fréquence ou **densité de probabilité**⁵, et la représentation graphique de cette fonction a la forme d'une « courbe en cloche » comme le montre le graphique suivant.



Dans une distribution normale, **la moyenne arithmétique, la médiane et le mode sont donc confondus**⁶. On a donc une courbe symétrique, aplatie aux deux extrémités et telle que la fréquence des données décroît à mesure qu'on s'éloigne du centre, que ce soit vers la gauche ou vers la droite. C'est là l'idée que l'on a généralement de la normalité : beaucoup de données se situent autour de la moyenne et le nombre de données diminue à mesure qu'on s'éloigne de celle-ci.



Bien que toutes les courbes soient symétriques, leur évasement correspond à l'écart-type. Plus il est élevé, plus les mesures sont dispersées autour de la valeur moyenne et plus la courbe s'aplatit. Plus il est faible (les valeurs sont rassemblées autour du centre), plus la courbe est pointue en sa moyenne.

⁵ Rappelons qu'une fonction f est une densité de probabilité sur un intervalle si elle est continue (on trace la fonction d'un seul trait), positive (sa courbe est au-dessus de l'axe des abscisses), et si l'aire sous la courbe est égale à 1.

⁶ Dans une distribution, le **mode** (M_o) correspond à la valeur ou la catégorie qui possède l'effectif le plus élevé. La **médiane** (M_d) correspond à la valeur qui se trouve au centre d'une distribution de données qui sont ordonnées en ordre croissant ou décroissant.



FIN610 – Les outils statistiques de gestion

La distribution normale tient compte de deux paramètres la moyenne m et l'écart-type σ . Et puisqu'il existe un nombre infini de combinaisons moyenne et écart-type, il existe un nombre infini de lois normales.

Comme pour toute variable aléatoire continue, c'est la fonction de répartition $F(x)$ qui sert à calculer les probabilités. Ainsi la probabilité que la variable aléatoire X ait une valeur inférieure à "x" est donnée par la **fonction de répartition** suivante :

$$F(x) = p(X < x) = \int_{-\infty}^x \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} e^{-\frac{1}{2} \left[\frac{t-m}{\sigma} \right]^2} dx$$

Mais compte tenu de la complexité de la formule ci-dessus, le calcul des valeurs à partir de la fonction de répartition de la variable normale X qui suit $N(m, \sigma)$ est quasi impossible. C'est la raison pour laquelle on effectue un changement de variable pour passer à une loi normale particulière appelée **loi normale centrée réduite**.

Soit "T" la variable aléatoire obtenue à partir de X :

$$T = \frac{X-m}{\sigma}$$

L'espérance et l'écart-type d'une variable suivant une loi normale centrée réduite sont :

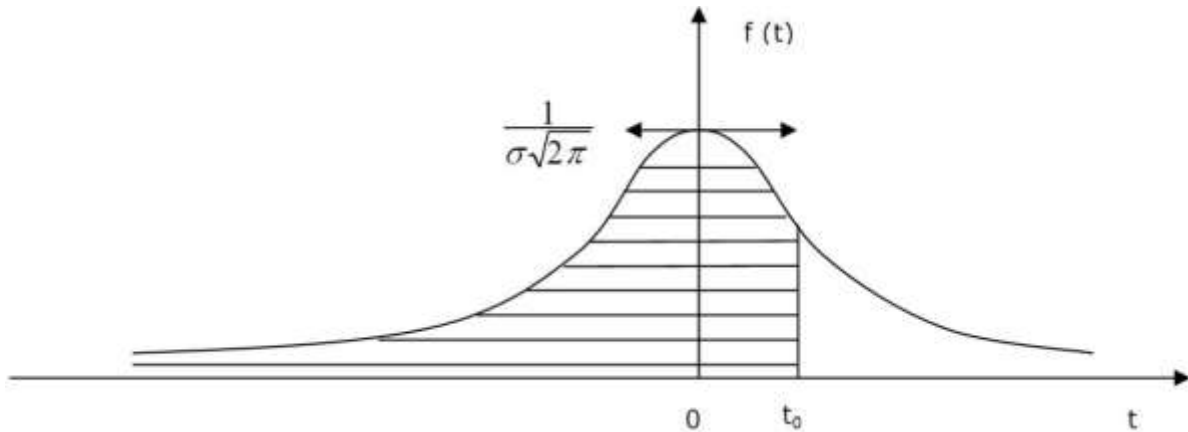
- $E(T) = 0$;
- $\sigma(T) = 1$.

La loi de T , notée $N(0 ; 1)$ admet une fonction de densité définie par :

$$f(t) = \frac{1}{\sqrt{2 \cdot \pi}} e^{-\frac{1}{2} t^2}$$

La représentation graphique de cette fonction a la forme de la courbe ci-dessous, dite fonction de densité de la loi normale centrée réduite.

FIN610 – Les outils statistiques de gestion



Evidemment puisque la moyenne est nulle, il en est de même pour le mode et la médiane.

Enfin, la surface sous une courbe de distribution normale représente la probabilité qu'un phénomène se produise dans cet intervalle. Ainsi la probabilité qu'une variable prenne une valeur inférieure à t_0 est égale à l'aire sous la courbe entre $-\infty$ et la valeur t_0 . Ainsi l'aire totale comprise entre la courbe de densité de probabilité et l'axe des abscisses est égale à 1. La fonction de répartition est notée :

$$\Pi(t) = F(T) = P(T \leq t) = \int_{-\infty}^t f(t) dt = \int_{-\infty}^t \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2} dt$$

Nous verrons que $\Pi(t) = F(T) = P(T \leq t)$ indique que cette fonction est tabulée pour des valeurs positives de t .

a) Calculs de probabilités avec la loi normale

Comment calculer la probabilité pour que X soit < à 25 000 ?

On cherche $p(X < 25\,000)$ sachant que la variable aléatoire continue "X" suit une loi normale de paramètres " m " = 20 000 et " σ " = 8 000. On ne peut pas répondre à cette question directement puisqu'il faudrait utiliser pour cela la formule suivante :

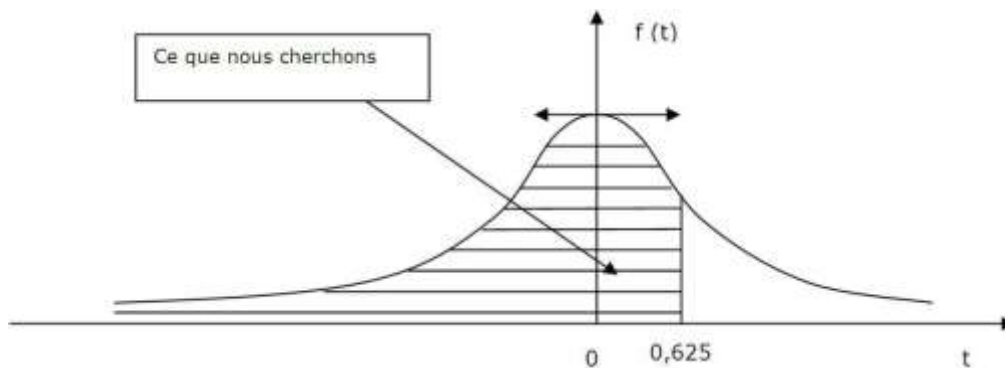
$$F(x) = p(X < x) = \int_{-\infty}^x \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} e^{-\frac{1}{2} \left[\frac{t-m}{\sigma} \right]^2} dt$$

FIN610 – Les outils statistiques de gestion

Or grâce au changement de variable de Laplace-Gauss, nous allons pouvoir *centrer réduire* la variable X et la remplacer par " T ". Il vient :

$$p(X < a) = p\left(T < \frac{a - m}{\sigma}\right) \Rightarrow p(X < 25\ 000) = p\left(T < \frac{25\ 000 - 20\ 000}{8\ 000}\right) = p(T < 0,625)$$

Ce que nous recherchons graphiquement :



Pour trouver cette probabilité nous allons donc utiliser une des tables de loi normale. Il existe en effet une table de la loi normale centrée réduite $P(T < t)$ qui nous donne ce que nous recherchons, à savoir l'aire grisée de la courbe précédente. Comme son nom l'indique cette table donne les probabilités : $p(T < t) = \Pi(t)$.

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500 0	0,504 0	0,508 0	0,512 0	0,516 0	0,519 9	0,523 9	0,527 9	0,531 9	0,535 9
0,1	0,539 8	0,543 8	0,547 8	0,551 7	0,555 7	0,559 6	0,563 6	0,567 5	0,571 4	0,575 3
0,2	0,579 3	0,583 2	0,587 1	0,591 0	0,594 8	0,598 7	0,602 6	0,606 4	0,610 3	0,614 1
0,3	0,617 9	0,621 7	0,625 5	0,629 3	0,633 1	0,636 8	0,640 6	0,644 3	0,648 0	0,651 7
0,4	0,655 4	0,659 1	0,662 8	0,666 4	0,670 0	0,673 6	0,677 2	0,680 8	0,684 4	0,687 9
0,5	0,691 5	0,695 0	0,698 5	0,701 9	0,705 4	0,708 8	0,712 3	0,715 7	0,719 0	0,722 4
0,6	0,725 7	0,729 1	0,732 4	0,735 7	0,738 9	0,742 2	0,745 4	0,748 6	0,751 7	0,754 9
0,7	0,758 0	0,761 1	0,764 2	0,767 3	0,770 3	0,773 4	0,776 4	0,779 3	0,782 3	0,785 2

Dans ce cas, $\Pi(0,62) = P(T < 0,62) = 0,7324$.

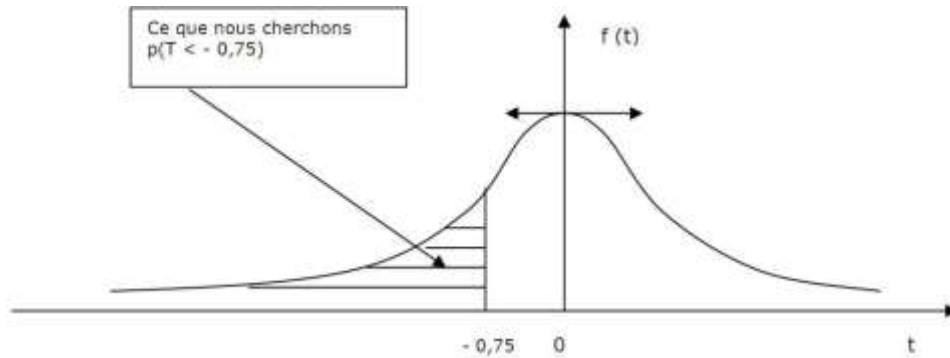
Comment calculer la probabilité pour que X soit < à 14 000 ?

On cherche $p(X < 14\ 000)$ sachant que la variable aléatoire continue " X " suit une loi normale de paramètres " m " = 20 000 et " σ " = 8 000. Grâce au changement de variable, nous allons pouvoir *centrer réduire* la variable X et la remplacer par " T ". Il vient :

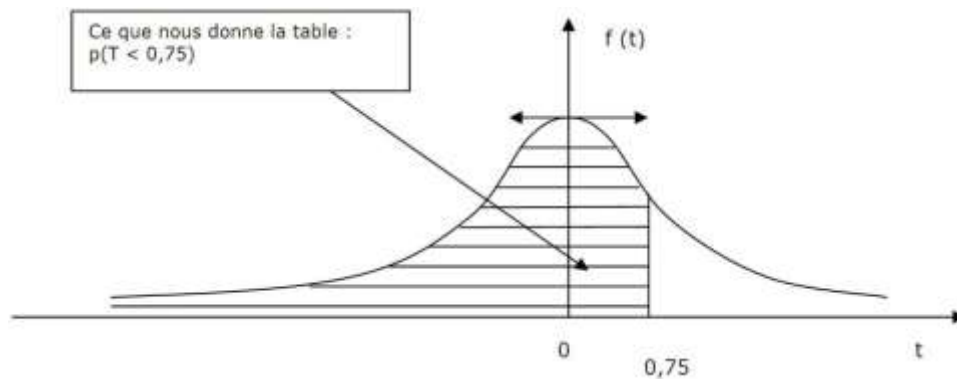
FIN610 – Les outils statistiques de gestion

$$p(X < 14\ 000) = p\left(T < \frac{14\ 000 - 20\ 000}{8\ 000}\right) = p(T < -0,75)$$

La courbe suivante nous donne schématiquement ce que nous recherchons :



La table $p(T < t)$ ne donne les probabilités que pour $t > 0$:



On note toutefois que si " t " est $< 0 \Rightarrow P(-t) = p(T < -t) = 1 - p(T < t)$.

Compte tenu de la symétrie (par rapport à $t = 0$) de la table centrée réduite on voit bien en comparant les deux schémas que :

$$p(T < -0,75) = 1 - p(T < 0,75)$$

Donc en appliquant ce principe, il vient :

$$p(T < -0,75) = p(T > 0,75) = 1 - p(T < 0,75) = 1 - 0,7734 = 0,2266$$

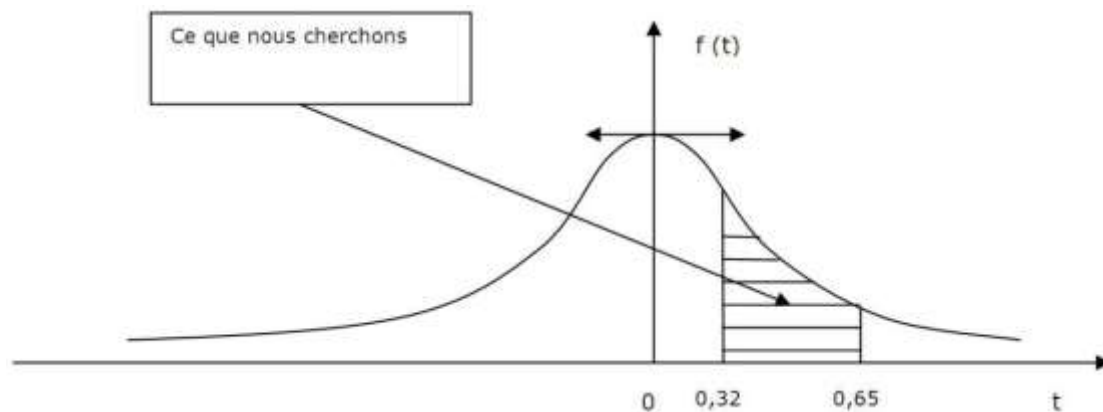
Comment calculer la probabilité pour que X soit comprise entre 22 560 et 25 200 ?

FIN610 – Les outils statistiques de gestion

Cette fois on cherche $p(22\ 560 < X < 25\ 200)$ sachant que la variable aléatoire continue "X" suit une loi normale de paramètres "m" = 20 000 et "σ" = 8 000. Grâce au changement de variable, nous allons pouvoir centrer réduire la variable X et la remplacer par "T".

$$\Rightarrow p(22\ 560 < X < 25\ 200) = p\left[\left(\frac{22\ 560 - 20\ 000}{8\ 000}\right) < T < \left(\frac{25\ 200 - 20\ 000}{8\ 000}\right)\right]$$

$$\Rightarrow p(22\ 560 < X < 25\ 200) = p(0,32 < T < 0,65)$$



En regardant le schéma, on voit bien que :

$p(0,32 < T < 0,65) = p(T < 0,65) - p(T < 0,32)$. Or ici 0,32 et 0,65 sont des valeurs positives. Nous pouvons donc trouver le résultat directement dans la table $\Pi(t) = p(T < t)$. Il vient :

$$p(T < 0,65) - p(T < 0,32) = 0,7422 - 0,6255 = 0,1167 = 11,67\%$$

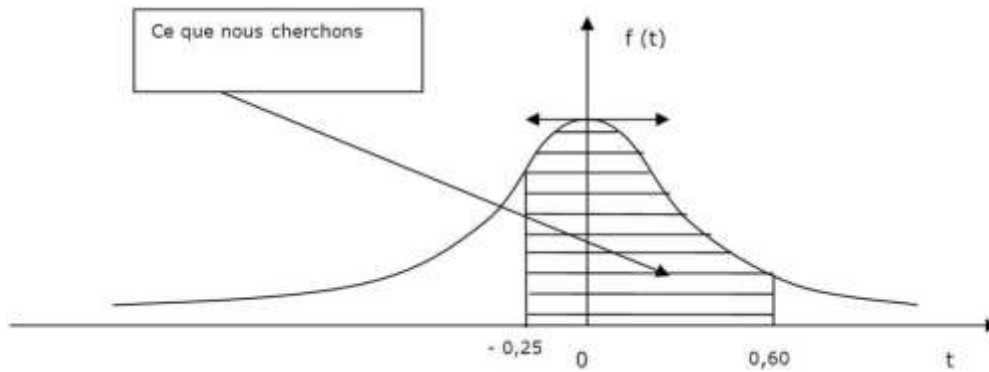
Comment calculer la probabilité pour que X soit compris entre 18 000 et 24 800 ?

Cette fois on cherche $p(18\ 000 < X < 24\ 800)$ sachant que la variable aléatoire continue "X" suit une loi normale de paramètres "m" = 20 000 et "σ" = 8 000. Grâce au changement de variable, nous allons pouvoir centrer réduire la variable X et la remplacer par "T".

$$\Rightarrow p(18\ 000 < X < 24\ 800) = p\left[\left(\frac{18\ 000 - 20\ 000}{8\ 000}\right) < T < \left(\frac{24\ 800 - 20\ 000}{8\ 000}\right)\right]$$

$$\Rightarrow p(-0,25 < T < 0,60)$$

FIN610 – Les outils statistiques de gestion



En regardant le schéma, on voit bien que :

$$p(-0.25 < T < 0.60) = p(T < 0.60) - p(T < -0.25)$$

Or ceci revient à dire : $p(-0.25 < T < 0.60) = p(T < 0.60) - [1 - p(T < 0.25)]$.
Donc en lisant sur la table, il vient :

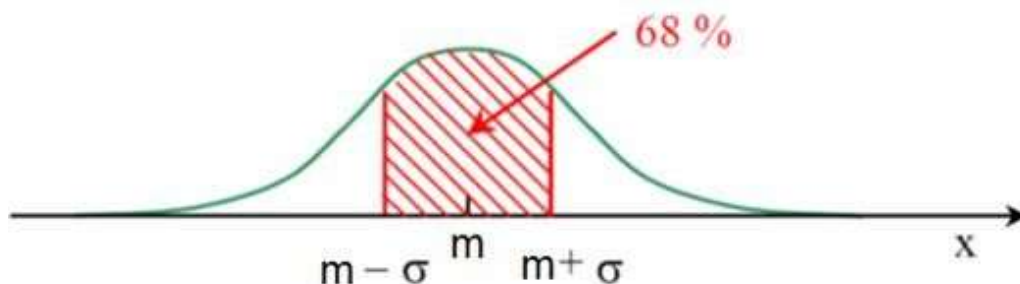
$$p(T < 0.60) - [1 - p(T < 0.25)] = 0.7257 - (1 - 0.5987) = 0.3244 = 32.44\%$$

b) Propriétés fondamentales de la loi normale

Une variable aléatoire X normale de moyenne m et d'écart type σ prend une valeur à moins d'un écart type de sa moyenne environ 68% des fois.
Mathématiquement :

$$P(m - \sigma \leq X \leq m + \sigma) = 0.6826$$

Graphiquement :

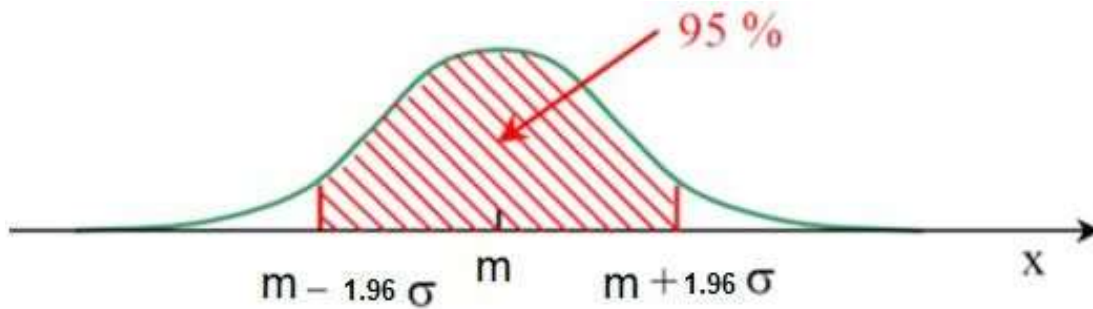


Une variable aléatoire X normale de moyenne m et d'écart type σ prend une valeur à moins de 1.96 écarts type de sa moyenne environ 95% des fois.
Mathématiquement :

FIN610 – Les outils statistiques de gestion

$$P(m - 1.96\sigma \leq X \leq m + 1.96\sigma) = 0.954$$

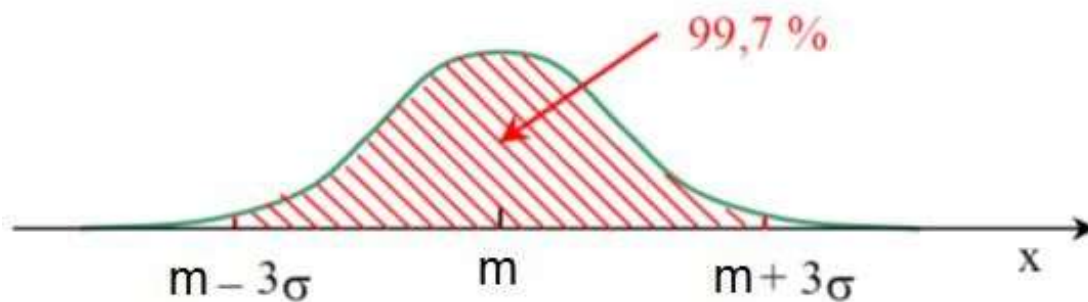
Graphiquement :



Une variable aléatoire normale de moyenne m et d'écart type σ prend une valeur à moins de trois écarts types de sa moyenne plus de 99% des fois.

$$P(m - 3\sigma \leq X \leq m + 3\sigma) = 0.997$$

Graphiquement :



Ex : Pour le Q.I. avec $m = 100$ et $\sigma = 15$, à peu près 68% des valeurs de la population se retrouvent entre 85 et 115, environ 95% des valeurs de la population se retrouvent entre 70 et 130, presque 100% des valeurs de la population se retrouvent entre 55 et 145.

Si X est une variable aléatoire normale, alors on sait que $(X - m)/\sigma$ est la variable aléatoire centrée réduite de moyenne 0 et d'écart-type 1. On aura alors les égalités suivantes :

$$P(-1 \leq T \leq +1) = 0.6826$$

$$P(-1.96 \leq T \leq +1.96) = 0.95$$

$$P(-3 \leq T \leq +3) = 0.9974$$



FIN610 – Les outils statistiques de gestion

Ex : Calculons $P(-1 \leq T \leq +1)$

$$\begin{aligned}P(-1 \leq T \leq +1) &= p(T < 1) - [1 - p(T < 1)] \\ &= 2 p(T < 1) - 1 \\ &= 2 \Pi(1) - 1\end{aligned}$$

Sur la table on peut lire $\Pi(1) = 0.84134$. D'où :

$$P(-1 \leq T \leq +1) = (2 \times 0.84134) - 1 = 0.6826$$



FIN610 – Les outils statistiques de gestion

II. Applications

• Soit l'expérience aléatoire : "On tire une carte dans un jeu de 32 cartes." On considère le jeu suivant :

- si on tire un cœur, on gagne 2€,
- si on tire un roi, on gagne 5€,
- si on tire une autre carte, on perd 1€.

On appelle X la variable aléatoire qui à une carte tirée associe un gain ou une perte. a) Déterminer la loi de probabilité de X . b) Calculer l'espérance mathématique de gain et commentez. c) Calculer l'écart type.

a) La variable aléatoire X peut prendre les valeurs 2, 5, -1 mais aussi 7. En effet, si on tire le roi de cœur, on gagne $5(\text{roi}) + 2(\text{cœur}) = 7\text{€}$.

- Si la carte tirée est un cœur (autre que le roi de cœur), $X = 2$.

$$P(X = 2) = \frac{7}{32}$$

- Si la carte tirée est un roi (autre que le roi de cœur), $X = 5$.

$$P(X = 5) = \frac{3}{32}$$

- Si la carte tirée est le roi de cœur, $X = 7$.

$$P(X = 7) = \frac{1}{32}$$

- Si la carte tirée n'est ni un cœur, ni un roi, $X = -1$.

$$P(X = -1) = \frac{21}{32}$$

La loi de probabilité de X est :

x_i	-1	2	5	7
$P(X = x_i)$	$\frac{21}{32}$	$\frac{7}{32}$	$\frac{3}{32}$	$\frac{1}{32}$



FIN610 – Les outils statistiques de gestion

On constate que la somme des probabilités est égale à 1.

$$b) E(X) = \sum_{i=1}^4 x_i p(x_i) = -1 \times \frac{21}{32} + 2 \times \frac{7}{32} + 5 \times \frac{3}{32} + 7 \times \frac{1}{32} = \frac{15}{32} \text{ €}$$

On peut donc espérer gagner en moyenne 0.5€ en jouant. En répétant un grand nombre de fois l'expérience, la loi des grands nombres nous permet d'affirmer que les fréquences se rapprochent des probabilités théoriques. La moyenne des résultats se rapproche donc de l'espérance de la loi de probabilité. **L'espérance est donc la moyenne que l'on peut espérer si l'on répète l'expérience un grand nombre de fois.**

c) Variance et écart type

$$V(X) = \sum_{i=1}^n x_i^2 p(x_i) - \left[\sum_{i=1}^n x_i p(x_i) \right]^2$$

\Leftrightarrow

$$V(X) = [(-1)^2 \times \frac{21}{32} + (2)^2 \times \frac{7}{32} + (5)^2 \times \frac{3}{32} + (7)^2 \times \frac{1}{32}] - 0.5^2 = 5.19$$

$$\sigma(X) = \sqrt{5.19} = 2.28 \text{ €}$$

L'écart-type est environ égal à 2.28€, ce qui signifie qu'avec une espérance proche de 0.50€ le risque de perdre de l'argent est important.

- Soit X la variable aléatoire qui caractérise le nombre de garçons dans une famille de quatre enfants. Calculer E(X), V(X) et $\sigma(X)$.

Distribution de probabilité de X

x	p(x)
0	0,0625
1	0,2500
2	0,3750
3	0,2500
4	0,0625
Total	1



FIN610 – Les outils statistiques de gestion

$$E(X) = \sum_{i=1}^5 x_i p(x_i) = 0 \times 0.0625 + 1 \times 0.25 + 2 \times 0.375 + 3 \times 0.25 + 4 \times 0.0625 = 2$$

$$E(X^2) = \sum_{i=1}^5 x_i^2 p(x_i) = 0^2 \times 0.0625 + 1^2 \times 0.25 + 2^2 \times 0.375 + 3^2 \times 0.25 + 4^2 \times 0.0625 = 5$$

$$V(X) = E(X^2) - [E(X)]^2 = 5 - 2^2 = 1 \Rightarrow \sigma(X) = \sqrt{1} = 1$$

- Soit les 6 notes suivantes, 3, 5, 8, 11, 12, 15. Calculez la moyenne et l'écart type.

$$\text{Moyenne } \bar{x} = \frac{\sum_{i=1}^6 x_i}{N} = \frac{3+5+8+11+12+15}{6} = \frac{54}{6} = 9$$

Pour calculer la variance, dressons le tableau suivant :

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i)^2$
3	-6	36	9
5	-4	16	25
8	-1	1	64
11	2	4	121
12	3	9	144
15	6	36	225
54	Total	102	588

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N} = \frac{102}{6} = 17 \Rightarrow \sigma(X) = \sqrt{17} = 4.2 \text{ (dispersion plutôt élevée)}$$

On notera que l'élévation au carré permet d'éviter les compensations entre écarts positifs et négatifs (entre -6 et +6 par exemple).

Avec l'autre formule, plus commode :

$$\sigma^2 = \frac{\sum x_i^2}{N} - (\bar{x})^2 = \frac{588}{6} - 9^2 = 17 \Rightarrow \sigma(X) = \sqrt{17} = 4.2$$

En supposant que ces valeurs soient des notes obtenues lors d'un contrôle de statistique, un écart-type faible signifie que les étudiants ont dans l'ensemble une note proche de la moyenne, donc que leurs notes se ressemblent. À la limite, un écart-type nul signifie que tous les étudiants ont la même note.



FIN610 – Les outils statistiques de gestion

- On a prélevé un échantillon de 5 nouveau-nés qu'on a mesurés. La série statistique des longueurs (cm) obtenues est :

53, 47, 51, 49, 50

Calculer la variance et l'écart-type. Que constatez-vous ?

Avant de calculer la variance, il faut d'abord trouver la moyenne de cette série :

$$\bar{x} = \frac{\sum_{i=1}^5 x_i}{N} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{N} = \frac{53 + 47 + 51 + 49 + 50}{5} = 50 \text{ cm}$$

La longueur moyenne de ces nouveau-nés est de 50 cm. La variance est :

$$\begin{aligned} V(X) = \sigma^2 &= \frac{\sum_{i=1}^5 (x_i - \bar{x})^2}{N} = \frac{(x_1 - 50)^2 + (x_2 - 50)^2 + (x_3 - 50)^2 + (x_4 - 50)^2 + (x_5 - 50)^2}{5} \\ &= \frac{(53 - 50)^2 + (47 - 50)^2 + (51 - 50)^2 + (49 - 50)^2 + (50 - 50)^2}{5} \\ &= 5 \text{ cm}^2 \end{aligned}$$

La variance est donc de 5 cm².

On constate que la variance ne s'exprime pas dans les mêmes unités que la variable. Ainsi lorsque des données sont exprimées en centimètres, alors la variance est exprimée en cm². Afin d'exprimer la dispersion dans les mêmes unités que la variable, on définit l'écart type.

$$\sigma(X) = \sqrt{5 \text{ cm}^2} = 2.24 \text{ cm}$$

- On dispose des anticipations suivantes concernant l'action A (cf tableau), sachant qu'elle cote 1 000€ en date 0. On vous demande de a) Calculer l'espérance de gain et b) le risque de l'action A.

FIN610 – Les outils statistiques de gestion

Anticipation du cours à une date "d"	Dividendes possibles	Probabilité $p(R_A)$
800	0,00	20 %
1 200	25,00	60 %
1 350	50,00	20 %

a) Il faut d'abord calculer les rentabilités successives en fonction des différents cours possibles. Il vient :

Anticipation du cours à une date "d"	Dividendes possibles	Rentabilité (R_A)
800	0	$(800 - 1\,000 + 0) / 1\,000 = -0,2 = -20\%$
1 200	25	$(1\,200 - 1\,000 + 25,00) / 1\,000 = 0,225 = 22,5\%$
1 350	50	$(1\,350 - 1\,000 + 50,00) / 1\,000 = 0,40 = 40\%$

Rappelons que la formule de l'espérance mathématique d'une variable x est la suivante :

$$E(x) = \sum_{i=1}^n x_i \times p(x_i)$$

L'espérance de la rentabilité de l'action A , $E(R_A)$ se déduit de la formule précédente, la variable aléatoire étant la rentabilité R_A .

Anticipation du cours à une date "d"	Dividendes possibles	R_A	Probabilité $p(R_A)$	$R_A \cdot p(R_A)$
800	0,00	-20,00	0,20	-4,00
1 200	25,00	22,50	0,60	13,50
1 350	50,00	40,00	0,20	8,00
Total = $E(R_A) = 17,5$				

Pour des raisons pratiques, les rentabilités n'ont pas été exprimées en pourcentage (20 au lieu de 0.20...). Exprimée en pourcentage, l'espérance $E(R_A)$ est donc de 17.50%.

b) Le risque d'une action est mesuré par l'écart type des rentabilités futures. Rappelons que la formule de la variance d'une variable x est la suivante :

FIN610 – Les outils statistiques de gestion

$$V(x) = \left[\sum_{i=1}^n x_i^2 \times p(x_i) \right] - \left[\sum_{i=1}^n x_i \times p(x_i) \right]^2$$

La variance est donc la moyenne des carrés moins le carré des moyennes. L'écart type est la racine carrée de la variance. Plus l'écart type de la rentabilité est grand, plus le risque de l'action est élevé, autrement dit plus il y a de chances que l'on s'éloigne de la moyenne calculée par l'espérance.

Anticipation du cours à une date "d"	Dividendes possibles	R _A	Probabilité p(R _A)	R _A · p(R _A)	R _A ²	R _A ² · p(R _A)
800	0	- 20,00	0,20	- 4,00	400,00	80,00
1 200	25	22,50	0,60	13,50	506,25	303,75
1 350	50	40,00	0,20	8,00	1 600,00	320,00
Total = E(R_A) = 17,5						Total = 703,75

$$V(R_A) = 703.75 - (17.50)^2 = 397.50 \Rightarrow \sigma(r_A) = 19.94\%$$

Rappelons que la variance n'est jamais munie d'une unité ; on ne dit pas 397.50 euros ou 397.50 kg ou 397.50%, mais simplement 397.50. Ici, les variances ont été calculées en élevant au carré des pourcentages (les taux de rentabilité). Des pourcentages au carré n'auraient pas de signification concrète. C'est pourquoi nous ne faisons pas suivre la valeur 397.50 du symbole %. En revanche, l'écart type s'exprime toujours dans la même unité que la variable aléatoire. Ainsi la rentabilité de l'action s'exprimant en %, l'écart type s'exprimera en %.

- Imaginons une entreprise nouvellement cotée et dont les cours mensuels ont été les suivants depuis sa première cotation en Bourse :

Cours action
48,17
48,24
49,06
47,98
48,34
48,72
48,77
46,68
48,02
50,17
51,47
50,46
51,43

FIN610 – Les outils statistiques de gestion

Calculer l'espérance, la variance et l'écart type passés de la rentabilité de ce titre.

Ici on dispose bien de la population entière ; nous disposons de tous les cours depuis le début de la cotation. Si N est le nombre d'observations (nombre de mois, semestres, trimestres, années...) alors :

$$\text{Moyenne de } x = \bar{x} = \frac{1}{N} \sum_{i=1}^n x_i$$

$$\text{Variance} = \text{moyenne des carrés} - \text{carré de la moyenne} = \left[\frac{1}{N} \sum_{i=1}^n x_i^2 \right] - (\bar{x})^2$$

Pour évaluer la *rentabilité moyenne passée* et le *risque passé*, il faut donc **faire appel à des formules qui font intervenir des observations passées, non des probabilités.**

Cours action	Mois	Rentabilité action en % RA	RA ²
48,17			
48,24	1	0,1453	0,02112
49,06	2	1,6998	2,88944
47,98	3	- 2,2014	4,84610
48,34	4	0,7503	0,56297
48,72	5	0,7861	0,61795
48,77	6	0,1026	0,01053
46,68	7	- 4,2854	18,36484
48,02	8	2,8706	8,24039
50,17	9	4,4773	20,04623
51,47	10	2,5912	6,71427
50,46	11	- 1,9623	3,85065
51,43	12	1,9223	3,69529
		6,8965	69,85977

Moyenne observée (passée) de la population = 6.8965/12 = 0.575. La rentabilité moyenne passée est donc de 0.575%. Á noter qu'il y a 13 cours de sorte qu'on calcule 12 rentabilités.

Variance observée de la population = (69.86/12) - (0.575)² = 5.491

Écart-type observé = $\sqrt{5.491}$ = 2.34%



FIN610 – Les outils statistiques de gestion

- Donnez le nombre de tiercés que l'on peut former si une course comporte 12 chevaux.

Nombre de tiercés :

$$A_{12}^3 = \frac{12!}{(12-3)!} = 12 \times 11 \times 10 = 1\,320$$

- Combien y-a-t-il de numéros de téléphone à 10 chiffres qui commencent par 06 ?

Le 0 et le 6 étant pris, nous avons un choix pour le 0 et un choix pour le 6 puis 10 choix (entre 0 et 9) pour chacun des huit numéros restant. Soit :

$$1 \times 1 \times 10^8 = 100 \text{ millions de numéros à 10 chiffres commençant par 06.}$$

- Au loto, il s'agit de sélectionner les 6 numéros gagnants parmi tous les nombres de 1 à 49. Combien y a-t-il de grilles de loto possibles ?

$$C_{49}^6 = \frac{49!}{6!(49-6)!} = 13\,983\,816$$

- Quel est l'intérêt de la loi normale centrée réduite ?

Grâce à la loi normale centrée réduite, pas besoin d'effectuer des intégrales pour évaluer les probabilités. Ces dernières seront directement estimées à l'aide de tables.

- Supposons qu'on choisisse au hasard un étudiant d'une Université et qu'on mesure sa taille. La variable aléatoire X associée à chaque étudiant sa taille. Supposons que X soit une variable aléatoire normale de moyenne 180 cm et d'écart type 4 cm. Quel pourcentage d'étudiants ont une taille comprise entre 172 cm et 188 cm ?



FIN610 – Les outils statistiques de gestion

On sait que l'écart type de la variable X est $\sigma = 4$ cm. Il est facile de constater que 172 cm = 180 cm - 2 (4 cm) et que 188 cm = 180 cm + 2 (4 cm). De là, on peut écrire :

$$P(172 \leq X \leq 188) = 0.95$$

On peut en conclure que 95% des étudiants ont une taille comprise entre 172 cm et 188 cm.

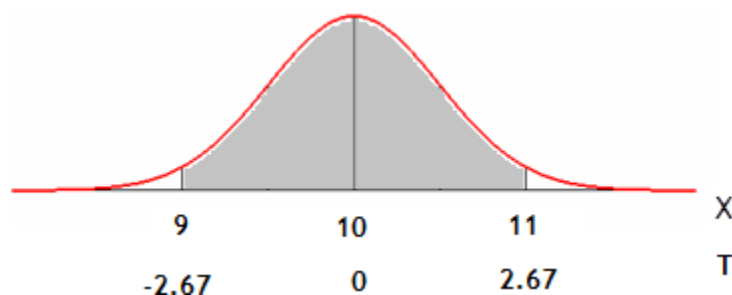
• Supposons qu'on choisisse au hasard un étudiant d'une Université et qu'on enregistre sa note en statistiques. La variable aléatoire X associée à chaque étudiant sa note en statistiques. Supposons que X soit une variable aléatoire normale de moyenne 10 et d'écart type 0.375. Quel pourcentage d'étudiants ont une note comprise entre 9 et 11 ? b) Que représente T ?

a) On cherche $p(9 < X < 11)$, problème familier où il faut trouver l'aire sous une courbe normale. On commence par transformer la variable aléatoire X en variable aléatoire normale centrée réduite T de moyenne 0 et d'écart type 1. Voici les équations :

$$T = \frac{X - m}{\sigma}$$

Quand $X = 11$, $T = \frac{11 - 10}{0.375} = 2.67$, et quand $X = 9$, $T = \frac{9 - 10}{0.375} = -2.67$

Représentation de la distribution des notes :



Autrement dit :

$$p(9 < X < 11) = p(-2.67 < T < 2.67) = \Pi(2.67) - \Pi(-2.67)$$



FIN610 – Les outils statistiques de gestion

$$\begin{aligned} &= 2\Phi(2.67) - 1 \\ &= 0.99 \end{aligned}$$

b) Que représente T

T est d'abord la variable aléatoire normale centrée réduite de moyenne 0 et d'écart type 1. C'est précisément le nombre d'écart types à partir de la moyenne. Ainsi dans l'exercice, la valeur 11 se situe à 2.67 écart types à droite de la moyenne de 10, et la valeur 9 se situe à 2.67 écart types à gauche de la moyenne de 10. En général, la valeur de la cote T est comprise entre -3 et +3, de sorte que la plupart des données se trouvent à moins de 3 écart types de la moyenne.

• Le délai de livraison d'un matériel, compté en jours, est une variable aléatoire X qui suit la loi normale N (30 ; 5). Quelle est la probabilité que le délai soit compris entre 22 et 38 jours ?

On pose $T = \frac{X-30}{5}$. Alors T suit une loi normale centrée réduite N (0 ; 1). On

pose :

$$22 \leq X \leq 38 \Leftrightarrow -8 \leq X - 30 \leq 8 \Leftrightarrow -1.6 \leq \frac{X-30}{5} \leq 1.6. \text{ Soit :}$$

$$P(22 \leq X \leq 38) = P(-1.6 \leq T \leq 1.6) = 2 P(T < 1.6) - 1 = 2(0.9452) - 1 = 0.8904$$

Il y a 89,04 % de chances que le délai de livraison soit compris entre 22 et 38 jours.

• La durée de vie (X) d'un lecteur MP4 est une variable qui obéit à une loi normale. La durée de vie moyenne de ces appareils est de 7 ans et l'écart type est de 1.2 ans. a) Compléter $X \sim N(;)$, b) Quelle proportion de lecteurs MP4 ont une durée de vie comprise entre 8 et 9 ans ? c) Quelle proportion de lecteurs MP4 ont une durée de vie supérieure à 7.5 ans ?

a) On a $X \sim N(7 ; 1.2)$



FIN610 – Les outils statistiques de gestion

$$b) P(8 < X < 9) = P\left(\frac{8-7}{1.2} < T < \frac{9-7}{1.2}\right) = P(0.83 < T < 1.67)$$

$$P(8 < X < 9) = \Pi(1.67) - \Pi(0.83) = 0.9525 - 0.7967 = 0.1558$$

De sorte qu'environ 15.58% des lecteurs MP4 ont une durée de vie comprise entre 8 et 9 ans.

$$c) P(X > 7.5) = P\left(T > \frac{7.5-7}{1.2}\right) = P(T > 0.42) = 1 - P(T < 0.42) = 1 - 0.6628 = 0.3372$$

De sorte qu'environ 33.72% des lecteurs MP4 ont une durée de vie supérieure à 7.5 ans.

• On suppose que la distribution des achats effectués par les clients dans une pâtisserie peut être modélisée par une loi normale de moyenne $m = 20\text{€}$ et d'écart type $\sigma = 5\text{€}$.

a) Calculer la proportion de clients qui dépensent :

- Plus de 30€,
- Moins de 15€,
- Entre 15 et 30€.

b) 50% des clients dépensent moins d'un montant x_0 . Calculer ce montant.

$$a) \text{ On a } X \rightarrow N(20; 5); T = \frac{X-20}{5}; T \rightarrow N(0; 1)$$

$$P(X > 30) = P\left(T > \frac{30-20}{5}\right) = P(T > 2) = 1 - \Pi(2) = 1 - 0.9772 = 0.023$$

$$P(X < 15) = P\left(T < \frac{15-20}{5}\right) = P(T < -1) = 1 - \Pi(1) = 1 - 0.8413 = 0.159$$

$$\begin{aligned} P(15 < X < 30) &= P\left(\frac{15-20}{5} < T < \frac{30-20}{5}\right) \\ &= P(-1 < T < 2) \\ &= \Pi(2) - \Pi(-1) \\ &= 0.9772 - 0.159 = 0.818 \end{aligned}$$

$$b) \text{ On cherche } x_0 \text{ tel que : } P(X < x_0) = P\left(T < \frac{x_0 - 20}{5}\right) = 0.50$$



FIN610 – Les outils statistiques de gestion

Par lecture de la table, on a $\frac{x_0 - 20}{5} = 0$ d'où $x_0 = 20$.

On trouve un résultat évident puisque, dans une distribution normale, la valeur médiane est aussi moyenne de la série.

• Imaginons que l'on étudie chez une PME le compte de trésorerie dont le montant suit une loi normale de moyenne $m = 7\,000\text{€}$ et d'écart type $3\,000\text{€}$. En arrondissant à 10^{-3} près, répondez aux questions suivantes :

- Quelle est la probabilité pour que l'entreprise ait une trésorerie supérieure à $10\,500\text{€}$?
- Quelle est la probabilité d'une rupture de trésorerie ?
- Quelle est la probabilité d'une situation saine ?

a) On cherche $P(X > 10\,500)$

Après le changement de variable, on cherche donc $P(T > 1.16) = 1 - P(T < 1.16) = 0.123$. Il y a donc 12.3% de chances que le montant de la trésorerie dépasse $10\,500\text{€}$.

b) On cherche $P(X < 0)$ soit, après changement de variable, $P(T < -2.33)$. Par symétrie, $P(T < -2.33) = P(T > 2.33) = 1 - P(T < 2.33) = 0.01$.

c) On cherche $P(X > 0) = 1 - P(X < 0) = 0.99$

• Si $X \sim N(m ; \sigma)$, vérifiez que $P(m - 1.96\sigma < X < m + 1.96\sigma) = 0.95$ et représentez cette information dans un graphique.

On sait que :

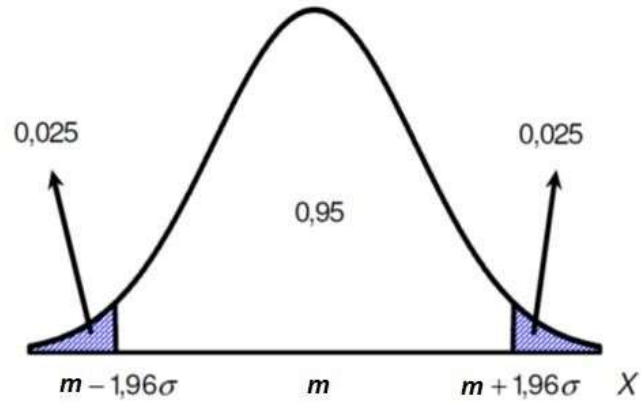
$$\begin{aligned} P(m - 1.96\sigma < X < m + 1.96\sigma) &= P\left(\frac{m - 1.96\sigma - m}{\sigma} < \frac{X - m}{\sigma} < \frac{m + 1.96\sigma - m}{\sigma}\right) \\ &= P(-1.96 < T < 1.96) \\ &= p(T < 1.96) - [1 - p(T < 1.96)] \\ &= 2p(T < 1.96) - 1 \\ &= 2\Phi(1.96) - 1 \end{aligned}$$

FIN610 – Les outils statistiques de gestion

Sur la table on peut lire $\Pi(1.96) = 0.9750$. D'où :

$$P(-1.96 \leq T \leq +1.96) = (2 \times 0.9750) - 1 = 0.95$$

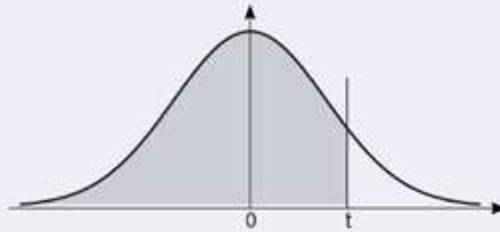
D'où la courbe en cloche suivante :



FIN610 – Les outils statistiques de gestion

ANNEXE

Table donnant $P(Z < t)$ pour une variable aléatoire suivant $N(0,1)$



	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
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	0,8413	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	0,9772	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
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

Table pour les grandes valeurs

t	3	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4	4,5
$\phi(t)$	0,9987	0,9990	0,9993	0,9995	0,9997	0,9998	0,9998	0,999928	0,999968	0,999997



FIN610 – Les outils statistiques de gestion



FIN610 – Les outils statistiques de gestion