

Chapitre 1 : Les lois usuelles en statistique



1. Objectifs

A la fin de ce chapitre, l'apprenant sera capable :

- D'identifier les différentes lois de distribution les plus connues en statistiques ;
- De connaître les caractéristiques de loi normale ;
- De maîtriser le calcul de probabilité et savoir utiliser la table de la loi normale centrée réduite.

2. Pré-test

Ce test permet de vérifier si l'étudiant maîtrise les notions de base en statistique.

2.1. Exercice

[solution n°4 p.21]

Qu'est ce que la loi normale centrée réduite ?

2.2. Exercice

[solution n°5 p.21]

Quels sont les paramètres descriptifs d'une variable suivant la loi normale ?

2.3. Exercice

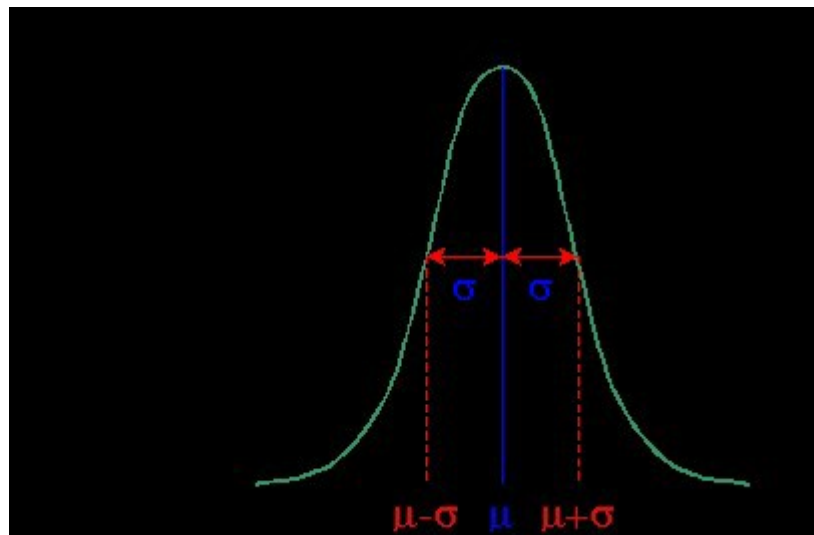
[solution n°6 p.21]

La loi normale est une distribution

- ☐ symétrique
- ☐ Asymétrique

3. La loi normale

Du point de vue historique, la nature et l'importance exceptionnelle de cette loi furent pressenties en 1773 par Abraham de Moivre lorsqu'il considéra la forme limite de la loi binomiale. En 1772, Simon Laplace l'étudia dans sa théorie des erreurs. Mais, c'est seulement en 1809 pour Carl Friedrich Gauss et en 1812 pour Simon Laplace qu'elle prit sa forme définitive. C'est ainsi qu'on l'appelle tantôt loi de Laplace, tantôt loi de Gauss, tantôt loi de Laplace-Gauss. On trouve aussi l'expression, consacrée par une longue tradition, de loi normale (ce qui ne signifie pas pour autant que les autres lois soient anormales). Elle jouit d'une importance fondamentale car un grand nombre de méthodes statistiques reposent sur elle. Ceci est lié au fait qu'elle intervient comme loi limite dans des conditions très générales. Dès que nous serons dans une situation où la variable aléatoire dépend de causes * En grand nombre et indépendantes, * Dont les effets s'additionnent, * Dont aucune n'est prépondérante, Alors nous serons en présence d'une variable aléatoire normale. C'est le cas, par exemple: · En métrologie, pour la variable des erreurs d'observation. · En météorologie, pour la distribution de phénomènes aléatoires tels que la température et la pression. · En biologie, pour les variables biométriques comme la taille ou le poids d'individus appartenant à une population homogène. En technologie, pour les dimensions des pièces usinées. · En économie, pour les fluctuations accidentelles d'une grandeur économique (production, ventes, ...) autour de sa tendance, etc. Donc on parle de la loi normale lorsqu'on a à faire à une variable aléatoire continue dépendante d'un grand nombre de causes indépendantes dont les effets s'additionnent et dont aucune n'est prépondérante. La fonction de densité de la loi normale possède la forme d'un cloche.(1)



Fonction de densité de la loi normale

3.1. Paramètre descriptifs de la loi normale

Une variable aléatoire X continue suit une loi normale si l'expression de sa fonction de densité de probabilité est de la forme suivante :

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

La loi dépend des deux réels μ et σ appelés paramètres de la loi normale. On la note $N(\mu, \sigma)$.

- L'espérance de la loi normale est : $E(X) = \mu$.
- La variance de la loi normale est : $V(X) = \sigma^2$, et l'écart-type $\sigma(X) = \sigma$.

3.2. Propriétés de la loi normale

3.2.1. Somme de deux variables normales

$$X_1 \rightarrow N(m_1, \sigma_1) \text{ et } X_2 \rightarrow N(m_2, \sigma_2), \text{ alors, } X_1 + X_2 \rightarrow N(m_1 + m_2, \sigma_1 + \sigma_2)$$

c'est à dire, la somme des deux variable suit la loi normale caractérisée par les deux paramètres : la somme des moyennes et la somme des variances.

3.2.2. Loi normale centrée réduite ou loi normale standardisée

A toute variable aléatoire X , on peut associer une variable dite standardisée d'espérance nulle et de variance unité (ceci résultait des propriétés de translation et de changement d'échelle). On montre assez facilement que si on effectue cette transformation sur une variable suivant une loi normale, la variable standardisée suit encore une loi normale mais cette fois-ci de paramètres 0 et 1. La loi standardisée est appelée loi normale centrée réduite, et notée $N(0, 1)$. Donc si :

$$X \rightarrow N(m, \sigma), \text{ on pose } T = \frac{X - m}{\sigma} \rightarrow N(0, 1)$$

Il faut garder à l'esprit que concrètement T est le nombre d'écarts-type entre la valeur de X et la moyenne. La loi $N(0,1)$ est tabulée à l'aide de la fonction de répartition des valeurs positives. Cette table permet de déterminer les probabilités de n'importe quelle distribution normale.

3.2.3. Calcul de probabilités

Si une variable aléatoire X suit une loi normale centrée réduite alors pour tous réels a et b tel que inférieur ou égale à b , on a :

$$P[X > a] = 1 - P[X \leq a] = 1 - \Phi(a)$$

$$P[X \leq a] = \Phi(a)$$

$$P[a < X < b] = \Phi(b) - \Phi(a)$$

$$P[X \leq -a] = \Phi(-a) = 1 - \Phi(a)$$

Exemple : Comment utiliser la table de la loi normale standardisée ?

La première colonne de la table indique les unités et les dixièmes des valeurs de t alors que les centièmes des valeurs de t se lisent sur la ligne supérieure de la table. La valeur trouvée à l'intersection de la ligne et de la colonne adéquates donne la valeur cherchée $\Phi(t)$. 1. Je cherche la valeur de $\Phi(0.5)$ A l'intersection de la ligne "0.5" et de la colonne "0.00", je lis 0.1915. 2. Je cherche la valeur de $P(-0.5 < T < 0)$. J'utilise la symétrie de la courbe par rapport à l'axe des ordonnées et j'en conclus que $P(-0.5 < T < 0) = P(0 < T < 0.5) = 0.1915$. 3. Je cherche la valeur de $P(-2.24 < T < 1.12)$. L'aire cherchée correspond à la somme suivante :

$P(-2.24 < T < 1.12) = P(-2.24 < T < 0) + P(0 < T < 1.12) = 0.4875 + 0.3686 = 0.8561$. 4. Je cherche la valeur de $P(1 < T < 2)$. L'aire cherchée correspond à la différence suivante $P(1 < T < 2) = P(0 < T < 2) - P(0 < T < 1) = 0.4772 - 0.3413 = 0.1359$. 5. Je cherche la valeur t de T telle que $P(0 < T < t) = 0.4750$. C'est le problème inverse de celui des exemples précédents. Il s'agit de localiser dans la table l'aire donnée et de déterminer la valeur de T correspondante. Je trouve $t = 1.96$.



Remarque

Si la valeur de l'aire ne peut se lire directement dans les valeurs de la table, on pourra toujours effectuer une interpolation linéaire entre deux valeurs adjacentes ou prendre la valeur la plus proche.

4. D'autres lois statistiques plus usuelles

4.1. La loi de Pearson (χ^2) et la loi de Student

4.1.1. La loi de Pearson (χ^2)

Elle a été découverte en 1905 par le mathématicien britannique Karl Pearson (1857-1936) qui travailla également sur les problèmes de régression avec le généticien Sir Francis Galton. Cette distribution (qui se prononce khi-deux) est très importante pour tester l'ajustement d'une loi théorique à une distribution expérimentale (test du χ^2) et pour déterminer la loi de la variance d'un échantillon. Ses paramètres descriptifs sont : $E(X) = n$, $\text{Var}(X) = 2n$. A mesure que n augmente, la loi du χ^2 tend vers la loi normale. En pratique, on peut considérer que pour $n \geq 30$, on peut remplacer la loi du χ^2 à n degrés de liberté par la loi normale :

$$N(n, \sqrt{2n})$$

a) Utilisation de la table de Pearson

Pour des raisons de commodité, au lieu de donner la table des fonctions de répartition des variables aléatoires

$$X_n^2$$

pour les différentes valeurs de n , on donne, en fonction de n (nombre de degrés de liberté) et d'une probabilité α que l'on peut choisir, la valeur

$$X_{\alpha, n}^2$$

définie par :

$$P[X^2 > X_{\alpha, n}^2] = \alpha$$

α est un seuil et a en fait une signification particulière dans les problèmes d'estimation et de tests. Il sera défini ultérieurement.

4.1.2. La loi de Student

Student est le pseudonyme de V.S Gosset, 1908. Soient X et Y deux variables aléatoires indépendantes, la première étant distribuée selon une loi normale centrée réduite $N(0, 1)$ et la deuxième selon une loi de khi-deux à n degrés de liberté. La quantité :

$$T = \frac{X\sqrt{n}}{\sqrt{Y}}$$

est une variable aléatoire qui suit une loi de Student à n degré de liberté. On écrit :

$$T = T_n$$

$$E(T_n) = 0$$

Si

$$n > 1$$

et

$$V(T_n) = \frac{n}{n-2}$$

si

$$n > 2$$

a) Approximation par la loi normale et utilisation de la table de Student

A mesure que n augmente, la distribution de Student à n degrés de liberté se rapproche de plus en plus de celle de la loi normale centrée réduite. En pratique si :

$$T \Rightarrow T_n \text{ pour } n \geq 30$$

On pourra écrire :

$$T \Rightarrow N(0,1)$$

Les valeurs tabulées de la variable T dépendent d'un seuil α que l'on peut choisir et du nombre de degré de liberté n . La table donne la valeur :

$$t_{\alpha, n}$$

définie par :

$$P(|T| > t_{\alpha, n}) = \alpha$$

5. Post -test

Le test vise à vérifier si l'étudiant a assimilé les connaissances données dans ce chapitre.

5.1. Exercice : Exercice sur la loi normale

Une enquête a été exécutée auprès des familles afin de connaître leur achat de lait en un mois. Sur l'ensemble des personnes interrogées, la consommation a une distribution normale avec une moyenne de 20L et un écart-type de 6L. En vue d'une campagne publicitaire, on souhaite connaître :

1. Le pourcentage des grands consommateurs (plus de 30 L par mois).
2. Le pourcentage des personnes qui consomment entre 15L et 25L par moi