

Cours de probabilités et statistique

Table des matières



Objectifs	3
Introduction	4
I - Notions de base et vocabulaire statistique	5
1. Concepts de base de la statistique	6
1.1. Population et individu	6
2. Etude des series statistiques	8
2.1. Representation des series statistiques	8
2.2. Tableaux statistiques	9
3. Exercice	12
II - Maîtriser les notions de cours	13
1. Exercice	13
2. Exercice	13
3. Exercice	14
4. Exercice	14
Ressources annexes	15
Solutions des exercices	16

Objectifs

Ce module est destiné aux étudiants de première année. tout ce qui est requis pour aborder ce module est les mathématiques de base. Le but final visé par ce module sera une liste d'objectifs à atteindre, à savoir :

-Introduire les notions fondamentales en probabilités et en séries statistiques à une et à deux variables.

-Compréhension des concepts de base : Les étudiants doivent acquérir une compréhension approfondie des concepts fondamentaux de probabilité et de statistique, tels que les probabilités, les variables aléatoires, les distributions, les mesures de tendance centrale, les mesures de dispersion, etc.

-Application des concepts : Les étudiants doivent être en mesure d'appliquer les concepts de probabilité et de statistique à des problèmes concrets. Cela peut inclure l'analyse de données, l'estimation des paramètres, les tests d'hypothèses, la modélisation probabiliste, etc.

-Résolution de problèmes : Les étudiants doivent développer des compétences en résolution de problèmes liés à la probabilité et à la statistique. Cela implique la capacité de formuler des problèmes, de sélectionner les méthodes et les techniques appropriées, d'interpréter les résultats et de communiquer efficacement les conclusions.

-Analyse et interprétation des données :

Les étudiants doivent être capables d'analyser des ensembles de données, de les organiser de manière appropriée, d'identifier les tendances, les schémas et les relations, et d'interpréter les résultats de manière significative.

Introduction



La plupart des décisions sont souvent prises en se basant sur l'analyse d'éléments incertains, tels que :

Quelle est la possibilité que les ventes diminuent si les prix augmentent ?

Quel est le degré de probabilité qu'une nouvelle méthode économique stimule la croissance ?

Quelle est la chance que le projet soit efficace ?

Quelles sont les perspectives qu'un nouvel investissement soit rentable ?

La probabilité est une mesure numérique de la probabilité d'un événement. Ainsi, les probabilités peuvent être utilisées pour mesurer le degré d'incertitude associé aux quatre événements mentionnés ci-dessus.

La valeur de la probabilité oscille toujours entre 0 et 1. Une probabilité proche de zéro implique qu'un événement a peu de chances de se produire. Si elle est égale à zéro, l'événement est considéré comme incertain. Une probabilité proche de 1 signifie qu'un événement est très susceptible de se produire. Si elle est égale à 1, l'événement est considéré comme certain. Si la probabilité est égale à 0,5, on considère que l'événement a autant de chances de se produire que de ne pas se produire (dans ce cas, nous sommes indécis).

Le but de cette partie du cours est de présenter aux étudiants une introduction conceptuelle aux probabilités et à leurs applications. Cette partie est composée de quatre chapitres qui permettent aux étudiants d'apprendre, de comprendre et de s'exercer à des problèmes qui correspondent à leur domaine d'études. Ces quatre chapitres traitent de l'expérience et des événements, de la théorie des probabilités, du calcul des probabilités et du théorème de Bayes.

Notions de base et vocabulaire statistique



La statistique est une discipline qui permet de collecter, organiser, analyser et interpréter des données afin de tirer des conclusions significatives. Elle est largement utilisée dans de nombreux domaines tels que la recherche scientifique, l'économie, la sociologie et bien d'autres. Dans ce chapitre, nous explorerons les notions de base de la statistique, y compris les concepts de population et d'individu, ainsi que les différentes formes de représentation des données, tant pour les variables qualitatives que quantitatives.

1. Concepts de base de la statistique

1.1. Population et individu

🔑 Définition : Population

En statistique, la population fait référence à l'ensemble complet d'individus, d'objets ou d'événements que l'on souhaite étudier. Autrement dit, une population, notée Ω , est un ensemble ni de personnes ou d'objets sur lequel se fera l'étude. On note par $n = \text{card}(\Omega)$ la taille de la population.

🔑 Définition : Individu

Un individu, quant à lui, est un élément spécifique de la population.

🔑 Définition : Variable (ou caractère)

Une variable est une caractéristique ou un attribut spécifique qui peut varier d'un individu à un autre au sein d'une population.

On appelle caractère (variable) toute application notée X définie de Ω

dans

un ensemble E telle que $\Omega = \{\omega_1, \omega_2, \dots, \omega_i\}$.

$$X : \Omega \rightarrow E$$

$$\omega_i \mapsto X(\omega_i) = x_i$$

ou $x_1, x_2, x_3, \dots, x_p$ sont les p modalités.

🔑 Définition : types de caractères

Les variables peuvent être qualitatives (catégorielles) ou quantitatives (numérique). Et ces deux types se divisent eux-mêmes en d'autres types.

🔑 Définition : Caractères qualitatifs

Les caractères qualitatifs sont souvent représentés sous forme de variables discrètes dans les analyses statistiques, et différentes méthodes peuvent être utilisées pour analyser et interpréter ces variables, telles que les tableaux de contingence, les diagrammes en barres, les diagrammes circulaires, les tests d'association, etc.

Ces caractères peuvent être subdivisés en deux sous-types principaux :

Caractère qualitatif nominal :

Il s'agit d'un type de caractère qui représente des catégories ou des groupes sans ordre ou hiérarchie particulière. Les valeurs de ce type de caractère ne peuvent pas être ordonnées ou classées de manière séquentielle. Par exemple, la couleur des yeux (bleu, vert, marron), le genre

(masculin, féminin), le statut marital (célibataire, marié, divorcé), etc.

Caractère qualitatif ordinal :

Il s'agit d'un type de caractère qui représente des catégories ou des groupes avec un ordre ou une hiérarchie spécifique. Les

valeurs de ce type de caractère peuvent être ordonnées ou classées selon une échelle ou un niveau. Par exemple, l'échelle de satisfaction (très insatisfait, insatisfait, neutre, satisfait, très satisfait), le niveau d'éducation (primaire, secondaire, supérieur), le rang dans une compétition (premier, deuxième, troisième), etc.

Définition : les caractères quantitatifs

sont des variables qui peuvent prendre des valeurs numériques et sont mesurables. Ils fournissent des informations sur une quantité ou une mesure spécifique dans une population ou un échantillon. Les caractères quantitatifs sont classés en deux sous-catégories : les caractères quantitatifs discrets et les caractères quantitatifs continus.

Caractères quantitatifs discrets : Ce sont des variables qui prennent des valeurs numériques isolées ou disjointes. Les valeurs possibles sont souvent des nombres entiers ou des nombres dénombrables. Par exemple, le nombre d'enfants dans une famille, le nombre de voitures dans un parking, le nombre de livres dans une bibliothèque, etc. Ces caractères peuvent être comptés ou dénombrés.

Caractères quantitatifs continus : Ce sont des variables qui peuvent prendre n'importe quelle valeur dans un intervalle ou une plage continue. Les valeurs possibles sont généralement des nombres réels et peuvent inclure des décimales ou des fractions. Par exemple, la taille d'une population, la température, le poids, le temps de réaction, le revenu, etc. Ces caractères peuvent être mesurés avec précision à l'aide d'instruments de mesure appropriés.

Remarque

Les caractères quantitatifs peuvent être analysés de différentes manières en statistique descriptive, en utilisant des mesures de tendance centrale (moyenne, médiane, mode), des mesures de dispersion (écart-type, variance, étendue), des représentations graphiques (histogrammes, diagrammes en boîte, diagrammes linéaires), des tests d'hypothèses, des corrélations, etc. Ces analyses permettent de résumer, de comparer et d'interpréter les données quantitatives dans un ensemble de données.

Définition : effectif

On appelle effectif de la modalité x_i le nombre d'individus ayant la modalité x_i , noté n_i .

Définition : fréquence

On appelle fréquence de la modalité x_i le rapport $f_i = \frac{n_i}{n}$.

2. Etude des séries statistiques

2.1. Représentation des séries statistiques

Pour étudier un caractère ou une variable statistique, la première opération consiste à recueillir toutes les informations voulues. Elles doivent être ordonnées dans des tableaux statistiques ou bien on donne des visualisations graphiques, qui donnent un résumé plus clair et facilite l'interprétation des données.

2.2. Tableaux statistiques

2.2.1. Cas d'une variable discrete

Considérons une population de N individus, décrite suivant une variable statistique discrète X ayant les valeurs x_i . On s'intéresse donc à

connaître, pour chaque valeur x_i , le nombre d'individus prenant cette valeur, ce nombre est noté par $n_i, i = 1, \dots, k$. Nous obtenons donc le tableau statistique suivant :

Les valeurs de x_i	n_i
x_1	n_1
x_2	n_2
.....
x_k	n_k
TOTAL	N

Tableau 1

Définition : Fréquence absolue et fréquence relative

- Le nombre d'individus n_i de la population, pour lesquels la variable X prend la valeur x_i , est appelé effectif ou fréquence absolue de la valeur x_i .
- La fréquence relative f_i de la valeur x_i d'effectif n_i est donné par la formule $\frac{n_i}{N}$, où N est l'effectif total de la population.
- Le pourcentage p_i de la valeur x_i d'effectif n_i est donnée par la formule $\frac{n_i}{N} \times 100$.

Remarque

- k est le nombre des valeurs différentes.
- $\sum_{i=1}^k n_i = N$ et $\sum_{i=1}^k f_i = 1$.
- La correspondance entre les valeurs de x_i et leurs effectifs s'appelle la distribution d'effectifs.

Exemple

Un fabricant de tissus essaie une nouvelle machine, il compte

le nombre de défauts sur 75 échantillons de 10 mètres. Il a obtenu les résultats

suivants :

Nombre de défauts x_i	Nombre d'échantillons n_i
0	38
1	15

2	11
3	6
4	3
5	2
Total	75

Tableau 2

Effectifs cumulés

Il peut être intéressant par la lecture du tableau de répondre à des questions de la forme:

- Quel est le nombre d'individus pour lesquels la variable X prend au moins X_j ?
 - Quel est le nombre d'individus pour lesquels la variable X prend au plus X_j ?
1. La réponse à la première question se fait en additionnant les effectifs à partir de la première valeur n_1 jusqu'à n_j ($1 \leq j \leq k$). Les nombres ainsi obtenus sont appelés effectifs cumulés croissants ou fréquences absolues cumulées croissantes, notés par n_{ic}^\uparrow .
 2. La réponse à la deuxième question se fait en additionnant les effectifs à partir de n_j ($1 \leq j \leq k$) jusqu'à la dernière valeur n_k . Les nombres ainsi obtenus sont appelés effectifs cumulés décroissants ou fréquences absolues cumulées décroissantes, notés par n_{ic}^\downarrow .

Par exemple, la quatrième ligne du tableau -2-, se lit aussi :

- 6 échantillons contiennent 3 défauts.
- 70 échantillons contiennent au plus 3 défauts.
- 11 échantillons contiennent au moins 3 défauts.

Le tableau -2- pourra être complété de la manière suivante :

Nombre de défauts X_i	Nombre d'échantillons n_i	n_{ic}^\uparrow	n_{ic}^\downarrow
0	38	38	75
1	15	53	37
2	11	64	22
3	6	70	11
4	3	73	5
5	2	75	2
Total	75		

 **Remarque**

De la même manière, on peut définir les fréquences relatives cumulées (croissantes et décroissantes) ainsi que les pourcentages cumulés (croissants et décroissants).

2.2.2. Cas d'une variable discrete

Dans le cas d'une variable continue, théoriquement les valeurs recueillies sont infinies et très proches l'une de l'autre. Alors, pour simplifier l'étude, on construit des classes (intervalles) en divisant l'étendue de la série statistique en plusieurs intervalles.

Définition

- L'étendue d'une série statistique est la différence entre la plus grande et la plus petite valeur dans la série.
- Les classes sont des intervalles de la forme $[a_i, a_{i+1}[$, tels que a_0 et a_k sont respectivement la plus petite et la plus grande valeur de la série.
- Dans la classe $[a_i, a_{i+1}[$, les valeurs a_i et a_{i+1} sont les bornes ou les limites de cette classe.
- Le nombre $\frac{a_i + a_{i+1}}{2}$ s'appelle le centre de la classe $[a_i, a_{i+1}[$.
- Le nombre $a_{i+1} - a_i$ s'appelle l'étendue ou l'amplitude de la classe $[a_i, a_{i+1}[$.
- L'effectif n_i de la classe $[a_i, a_{i+1}[$ correspond au nombre de valeurs appartenant à cette classe.

Fondamental

Le nombre de classes k ne doit pas être trop petit (perte d'information) ni trop grand (le regroupement en classe est alors inutile). Le nombre de classes qu'on peut construire est donné par la formule $k = \sqrt{N}$.

Exemple

Une étude concernant le poids de 80 nouveaux-nés dans une maternité a donné les résultats suivants :

Les classes	Les centres de classes x_i	Les effectifs n_i	n_{ic}^{\uparrow}	n_{ic}^{\downarrow}
[2.2, 2.5[2.35	2	2	80
[2.5, 2.8[2.65	5	7	78
[2.8, 3.1[2.95	20	27	73
[3.1, 3.4[3.25	19	46	53
[3.4, 3.7[3.55	20	66	34
[3.7, 4.0[3.85	8	74	14
[4.0, 4.3[4.15	4	78	6
[4.3, 4.6[4.45	2	80	2
Total		80		

3. Exercice

[solution n°1 p.16]

On s'intéresse à l'étude de la taille de 100 étudiants de l'université de Bejaia, la variable étudiée est de type

- Quantitatif discret
- quantitatif continue
- qualitatif ordinal

Maîtriser les notions de cours

II

1. Exercice

[solution n°2 p.16]

On dispose de dix jetons numérotés de 1 à 10 et on en extrait simultanément trois pour former un « paquet ». Combien de « paquets » contenant au moins un jeton ayant un numéro pair peut-on ainsi former ?

Veillez choisir une réponse.

- 330
- 110
- 180
- 900

2. Exercice

[solution n°3 p.16]

Dans une série statistique de N données qui prennent p valeurs différentes, l'écart-type :

Veillez choisir une réponse.

- comporte deux solutions opposées
- est toujours inférieur à la variance
- peut être négatif

3. Exercice

[solution n°4 p.17]

Un quartier résidentiel comprend 99 unités d'habitation ayant une valeur locative moyenne de 10000 Da. Deux nouvelles unités d'habitation sont construites dans le quartier : l'une a une valeur locative de 7000 Da et l'autre, une villa luxueuse, a une valeur locative de 114000 Da.

– Quelle est la nouvelle moyenne de valeur locative pour le quartier ?

4. Exercice

[solution n°5 p.17]

Dans une série statistique de N données qui prennent p valeurs différentes, l'écart-type mesure :

Veillez choisir une réponse.

- la dispersion de la série de p valeurs autour de la moyenne
- la distribution des p valeurs
- l'intervalle entre la valeur la plus basse et celle la plus élevée

5

•

 *Conseil : Lien utile*

* (cf. p.15) (cf. p.) <https://www.youtube.com/watch?v=FNunfXkPq14&list=PLJOPLcr85yakekhpCsq32BbC1KGSAAtjOD&index=2> [mp4]

Ressources annexes



>

>



Solutions des exercices



> Solution n°1

Exercice p. 12

On s'intéresse à l'étude de la taille de 100 étudiants de l'université de Bejaia, la variable étudiée est de type

- Quantitatif discret
- quantitatif continue
- qualitatif ordinal

> Solution n°2

Exercice p. 13

On dispose de dix jetons numérotés de 1 à 10 et on en extrait simultanément trois pour former un « paquet ». Combien de « paquets » contenant au moins un jeton ayant un numéro pair peut-on ainsi former ?

Veillez choisir une réponse.

- 330
- 110
- 180
- 900

> Solution n°3

Exercice p. 13

Dans une série statistique de N données qui prennent p valeurs différentes, l'écart-type :

Veillez choisir une réponse.

- comporte deux solutions opposées
- est toujours inférieur à la variance

- peut être négatif

> **Solution n°4**

Exercice p. 14

Un quartier résidentiel comprend 99 unités d'habitation ayant une valeur locative moyenne de 10000 Da. Deux nouvelles unités d'habitation sont construites dans le quartier : l'une a une valeur locative de 7000 Da et l'autre, une villa luxueuse, a une valeur locative de 114000 Da.

– Quelle est la nouvelle moyenne de valeur locative pour le quartier ?

110000

> **Solution n°5**

Exercice p. 14

Dans une série statistique de N données qui prennent p valeurs différentes, l'écart-type mesure :

Veillez choisir une réponse.

- la dispersion de la série de p valeurs autour de la moyenne
- la distribution des p valeurs
- l'intervalle entre la valeur la plus basse et celle la plus élevée