CHAPITRE 1 : INFERENCE SYSTEME STATISTIQUES DESCRIPTIVES TERMINOLOGIE ET REPRESENTATION GRAPHIQUE

1.1 TERMINOLOGIE

NOTION DE STATISTIQUES

La statistique est méthode scientifique dont l’objet est de recueillir, d’organiser, de résumer et d’analyser les données d’une enquête, d’une étude ou d’une expérience.

Elle permet de tirer des conclusions logiques et de prendre les décisions qui s’imposent à partir des analyses effectuées. Une étude statistique porte toujours sur une population. Une population statistique est l’ensemble des individus sur lesquelles porte l’étude statistique.

Si l’étude porte sur une partie de la population alors le référentiel sera appelé échantillon

Le nombre d’individus qui constituent la population représentent la taille de la population

Une population statistique est étudiée au vu d’un ou de plusieurs caractères.

Un caractère est une caractéristique commune à l’ensemble des individus constituant la population

Un caractère peut se trouver dans différentes situations appelées modalités du caractère

Un caractère est dit qualitatif lorsque ses modalités ne peuvent pas être exprimées par des valeurs numériques

Un caractère est quantitatif lorsque ses modalités peuvent être exprimées par des valeurs numériques

Un caractère quantitatif est dit discontinue ou discret si ses modalités prennent des valeurs numériques finies ou dénombrables

Le plus souvent ces valeurs sont entières

Un caractère quantitatif est dit continu si ses différentes modalités peuvent appartenir à un intervalle fixé à priori. Ces modalités sont souvent exprimées par des intervalles ou classes

1.2. Les tableaux statistiques

1.2.1 Cas d’une seule variable : TABLEAU STATISTIQUE A SIMPLE ENTREE

Le tableau brut se présente sous la forme suivante :

|  |  |
| --- | --- |
| Individus | Ci |
| 1 | C1 |
| 2 | C2 |
| . | . |
| . | . |
| . | . |
| N | Cn |
| Caractères | Effectif |
| c1 | n1 |
| c2 | n2 |
| . | . |
| . | . |
| Cn | Nk |
| TOTAL | N |

N=

|  |  |  |  |
| --- | --- | --- | --- |
| NBenfant | Eff nj | Fj% | Ni(eff cumilé crroissant) |
| 0 | 6 | 12 | 6 |
| 1 | 12 | 24 | 18 |
| 2 | 20 | 40 | 38 |
| 3 | 8 | 16 | 46 |
| 4 | 2 | 4 | 18 |
| 5 | 2 | 4 | 50 |
| T | 50 | 10 |  |

Fj= Fj(%)=\*100

=1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Revenue | Ni | Di | Ui | Nir | ECC | ECD |
| 100-200 | 15 | 0.15 | 1 | 15 | 15 | 100 |
| 200-300 | 10 | 0.10 | 1 | 10 | 25 | 85 |
| 300-400 | 30 | 0.30 | 1 | 30 | 55 | 75 |
| 400-600 | 40 | 0.20 | 1 | 40 | 95 | 45 |
| 600-700 | 3 | 0.03 | 1 | 3 | 98 | 5 |
| 700-900 | 1 | 0.005 | 2 | 0.5 | 99 | 2 |
| 900-1000 | 1 | 0.01 | 1 | 1 | 100 | 1 |
| T | 100 |  |  |  |  |  |

Ai (amplitude)= Ei+1 - Ei

Densité(Di) =

Ui=Nombre d’unité d’amplitude

1Ui=100 Am(amplitude)

Nir= (effectif rectifié)

Ci =(centre de la classe)

CL i (centre de la classe)

1.2.2 CAS DE DEUX VARIABLES

Ci-dessous le tableau de distribution conditionnelle :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X\Y | Y1 | Y2 | Yj | Yl | TOTAL |
| X1 | N11 | N12 | N1j | N1l | N1. |
| X2 | n 21 | N22 | N2j | N2l | N2. |
| Xi | Ni1 | Ni2 | Nij | Nil | Ni. |
| Xk | Nk1 | Nk2 | Nkj | Nkl | Nk. |
| TOTAL | n.1 | n.2 | n.j | n.l | N.. |

Ci-dessous les distributions marginales :

|  |  |
| --- | --- |
| X | EFF |
| X1 | N1. |
| . |  |
| XK | Nk. |
| T | N.. |

|  |  |
| --- | --- |
| Y | EFF |
| Y1 | N.1 |
| . |  |
| YL | N.L |
| T | N.. |

1.3 LES REPRESENTATIONS GRAPHIQUES

TAF

1. Déterminez les fréquences relatives

2. Trouvez la modalité la plus représentées

3. Calculez les ECC et donnez leur signification

4. Calculez les FCD et donnez leur signification

5. Trouver n (X>= 350)

N (700>=X>=400)

1. Déterminons les fréquences relatives

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Revenue (um) | Effectif(ni) | ECC | FCD | ECD | Densité |
| 100-200 | 50 | 0-50 | 90-100 | 450-500 | 0.5 |
| 200-300 | 80 | 50-130 | 74-16 | 370-80 | 0.8 |
| 300-500 | 150 | 130-280 | 44-30 | 220-150 | 0.75 |
| 500-600 | 120 | 280-400 | 20-24 | 100-120 | 1.2 |
| 600-700 | 70 | 400-470 | 6-14 | 30-70 | 0.7 |
| 700-1000 | 30 | 470-500 | 0-6 | 0-30 | 0.1 |
| T | 500 |  |  |  |  |

Fi1= (50\*100)/500= 10%

FI2= (80\*100)/500 = 16%

FI3= (150\*100)500 = 30%

FI4= (120\*100)/500 = 24%

FI5= (70\*100)/500 =14%

FI6= (30\*100)/500 =6%

2. La modalité la plus représentée

La classe la plus représentée est la classe 500-600 car elle a la plus grande densité

Ou en calculant l’effectif rectifié

3. Calculons les ECC et donnez les significations : (voir schéma)

130=nombre d’individus ayant un revenu inférieur ou égal à 300

4. Calculons les FCD et donnons les significations : (voir schéma)

44= pourcentage du nombre d’individus ayant un revenu supérieur ou égale à 500

5.

N (x>=350) = n (x>=500) +n (500>x>=350)

= 220+ 150\*0.75 🡪(di\*ai)

=333

N (700>=x>=400)= n (500>=x>=400)+n (600>=x>=500)+n (700>=x>=600)

=100\*0.75+100\*1.2+100\*0.7

=265

N (700>=x>=400)= n (x>700) - n (x<400)

= 470 – (130+ 100\*0.75)

=265

|  |  |  |  |
| --- | --- | --- | --- |
| Revenue | Ni | ECC | ECD |
| 100-200 | 50 | 50 | 500 |
| 200-300 | 80 | 130 | 450 |
| 300-500 | 150 | 280 | 370 |
| 500-600 | 120 | 400 | 220 |
| 600-700 | 70 | 470 | 100 |
| 700-1000 | 30 | 500 | 30 |
| TOTAL | 500 |  |  |

1. n (700>x>=200)

N (700>x>=200)=n(x>=200) – n(x>=700)

=450-30=420

1.3.1. Représentation graphique d’un caractère qualitatif

Dans le tableau statistique donnant la répartition du caractère qualitatif, chaque modalité xi est associée à un effectif ni .La seule représentation qui nous intéresse est celle des effectifs

Suivant la variable observée, de nombreuses représentations plus ou moins informatives peuvent être utilisées

Cependant les deux les plus classiques sont :

* Les diagrammes en tuyau d’orge ou diagramme en barre ou diagramme à bande

Les modalités de la variables sont placés sur une droite horizontale (attention ne pas orienter cette doit car les modalités ne sont pas mesurables et il n’y a donc pas de relations d’ordre entre elles)

Les effectifs ou les fréquences sont placés sur un axe vertical. La hauteur du tuyau est proportionnelle à l’effectif

La répartition des candidats retenus pour passer un test d’admissibilité pour une formation en management est organisée comme suit :

|  |  |
| --- | --- |
| Série | Effectif |
| S.ECO | 250 |
| S.MATH | 200 |
| S.EXP | 400 |
| TGA | 50 |
| TGC | 100 |
| TOTAL | 1000 |

S.EXP

400

S.MATH

200

S.ECO

250

TGC

100

TGA

50

Les diagrammes à secteur ou camembert

L’effectif total est représenté par un disque, chaque modalité est représentée par un secteur circulaire Dont la surface ou l’angle au centre est proportionnel à l’effectif correspondant

DIAGRAMME CIRCULAIRE

1.3.2. Représentation graphique d’un caractère quantitatif

Caractère q continu🡪diagramme différentiel🡪discontinu🡪Diagramme en bâton

Caractère q continu🡪diagramme différentiel🡪continu🡪Histogramme

Caractère q continu🡪diagramme intégral🡪discontinu🡪Diagramme en escalier

Caractère q continu🡪diagramme intégral🡪continu🡪Diagramme des fréquences cumulées

EXEMPLE :

Nous avons une population de 100 ménages qui est étudiée au vu du caractère nombre d’enfants

|  |  |
| --- | --- |
| Nombre d’enfants | Nombre de naissances |
| 0 | 10 |
| 1 | 25 |
| 2 | 40 |
| 3 | 10 |
| 4 | 8 |
| 5 | 7 |
| TOTAL | 100 |

40

30

10

20

0 1 2 3 4 5

Diagramme en bâton ou diagramme différentiel (nb enfants)

(diagramme en escalier, le remplir plus tard avec les ECC)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Salaire (UM) | Ni | Ui | Nir | ECC |  |
| 100-200 | 50 | 1 | 25 | 0-25 |  |
| 200-300 | 100 | 1 | 100 | 25-125 |  |
| 300-500 | 180 | 2 | 90 | 125-305 |  |
| 500-600 | 120 | 1 | 120 | 305-425 |  |
| 600-800 | 50 | 2 | 25 | 425-475 |  |
| 800-900 | 15 | 1 | 15 | 475-490 |  |
| 900-1000 | 10 | 1 | 10 | 490-500 |  |
| TOTAL | 500 |  |  |  |  |

1UA=100 UM

**CHAPITRE II : ANALYSE NUMERIQUE DES SERIES STATISTIQUES**

# I. Les paramètres de tendance centrale et de position

# 1. Le mode

C’est la valeur observée d’effectif maximum.

Cas d’un caractère quantitatif discontinue :

|  |  |
| --- | --- |
| Xi | Ni |
| 0 | 5 |
| 1 | 10 |
| 2 | 25 |
| 3 | 5 |
| 4 | 3 |
| 5 | 2 |
| T | 50 |

Cas d’un caractère quantitatif continue :

|  |  |  |
| --- | --- | --- |
| Classe | Ni | Nir |
| 100-200 | 5 | 5 |
| 200-400 | 15 | 7.5 |
| 400-500 | 20 | 20 |
| 500-600 | 12 | 12 |
| 600-800 | 25 | 12.5 |
| 800-900 | 5 | 5 |
| 900-1000 | 3 | 3 |
| T | 80 | - |

Classe modale [400,500[

**2. La médiane**

**Cadre série d’observation**

Xi : 1 3 4 10 1 2 7 8

1. Ranger les observations par valeurs croissantes ou décroissantes

Xi : 1 1 2 3 4 7 8 10

2. Regarder la parité du nombre d’observation

-N paire

N = 2p avec p=4

Intervalle médian [p ; p+1] 🡪 [3,4]

Me= 3

-Si n impair

Xi : 1 3 4 10 1 2 7 8 12

N=2p+1 🡪n=9=2\*4+1

Me 🡪 p+1 = 4

-Cadre d’un caractère quantitatif

Lorsque nous avons un tableau statistique qui décrit la distribution d’un caractère quantitatif continue, on va considérer que la médiane est une modalité du caractère qui est associée à une fréquence cumulée de 50% ou bien à un effectif cumulé égal à la moitié de l’effectif total

|  |  |  |
| --- | --- | --- |
| Classe | Ni | ECC |
| 100-200 | 5 | 0-5 |
| 200-400 | 15 | 5-20 |
| 400-500 | 20 | 20-40 |
| 500-600 | 12 | 40-52 |
| 600-800 | 25 | 52-77 |
| 800-900 | 5 | 77-82 |
| 900-1000 | 3 | 82-85 |
| T | 80 |  |

(1) 500🡪40

(2) Me🡪 85/2= 42.5

(3)600🡪52

Classe médiane [500, 600[

Interpolation linéaire :

(2) – (1) 🡪 = = 🡪 Me = 500 +100(2.5/12)

----------

(3) – (1)

3. La moyenne arithmétique

**Cas d’une série d’observation**

Supposons qu’une population de n individus étudiés au vue du caractère revenu

Xi : x1 x2 ….. xn

La moyenne arithmétique de ces observations sera égale := =

Cette moyenne arithmétique qui par définition est une moyenne arithmétique simple représente le revenu que recevrait chaque individu si la masse globale des revenus était répartie de manière égale.

Dans le cas d’un tableau statistique, les observations sont affectées d’un coefficient de pondération

Représentés par l’effectif ou la fréquence relative

Dans ce cas on va calculer une moyenne arithmétique pondérée :

=

|  |  |  |
| --- | --- | --- |
| Xi | Ni | Ni.xi |
| 0 | 5 | 0 |
| 1 | 10 | 10 |
| 2 | 25 | 50 |
| 3 | 5 | 15 |
| 4 | 3 | 12 |
| 5 | 2 | 10 |
| T | 50 | 97 |

= =1.94

|  |  |  |  |
| --- | --- | --- | --- |
| Classe | Ni | Xi | Ni.Xi |
| 100-200 | 5 | 150 | 750 |
| 200-400 | 15 | 300 | 4500 |
| 400-500 | 20 | 450 | 9000 |
| 500-600 | 12 | 550 | 6600 |
| 600-800 | 25 | 700 | 17500 |
| 800-900 | 5 | 850 | 4250 |
| 900-1000 | 3 | 950 | 2850 |
| T | 85 |  |  |

= = 534.7

## 4. Les Quantiles

Ce sont des paramètres de condition

On a une médiane qui sépare les observations en deux groupes d’effectifs égaux

Il y’a 3 quartiles qui séparent les observations en 4 groupes d’effectifs égaux

Q2=Me

Il y’a 9 déciles qui séparent les observations en dix groupes d’effectifs égaux

D5=Me

Il y’a 99 centiles c1 c2… c99

Il y’a 99 centiles qui séparent les observations en 100 groupes d’effectifs égaux

La détermination de ces caractéristiques est identique à celle de la médiane

Les quartiles sont obtenus lorsqu’on a cumulé 25, 50, 75% de la population

Les déciles sont obtenus lorsqu’on a cumulé 10,20 etc. jusqu’à 90% de la population

Les centiles sont obtenus lorsqu’on a cumulé 1, 2, 3 etc. jusqu’à 99% de la population

Les notions de décile et de centile n’ont de sens que si il y’a beaucoup d’observation et donc essentiellement si il y’a une variable classée