

Exercice 1 : Tableau du nombre d'essais pour réussir un test pour 40 individus.

Nb. essais	5	6	7	8	9	10	11	12	13	14	15
Effectifs	1	3	1	4	3	5	7	4	5	3	4

1. S'agit-il d'une variable nominale, quantitative discrète ou quantitative continue ?

quantitative discrète

2. Quelle est la définition de la médiane ?

C'est la valeur qui vérifie que "la moitié des individus a une valeur inférieure à la médiane" mais aussi "la moitié des individus a une valeur supérieure à la médiane"

3. Comment détermine-t-on cette médiane ?

On calcule les effectifs cumulés et on s'arrête dès que l'on dépasse 20 (la moitié de l'effectif))

Nb. essais	5	6	7	8	9	10	11	12	13	14	15
Effectifs Cumulés	1	4	5	9	12	17	24				

4. Quelle est alors la médiane "Med" de cette distribution ?

la médiane est 11

5. Quelle est la fréquence des observations inférieures ou égales à "Med" ? Commentez !

Freq(≤ 11) = 24/40 : c'est bien supérieure ou égal à 0.5

6. Quelle est la fréquence des observations supérieures ou égales à "Med" ? Commentez !

Freq(≥ 11) = 23/40 : c'est bien supérieure ou égal à 0.5

Exercice 2 : Une étude sur le chiffre d'affaires d'une population de PME a permis d'obtenir les résumés suivants (en milliers d'euro)

minimum	3500
moyenne	4900
écart-type	650
écart interquartile	2100
médiane	4600
skewness	1.2
premier quartile	4100
étendu	5000

- Quelle est la variable de cette situation statistique ? Quelle est sa nature ?
variable = le chiffre d'affaires d'une population de PME ; quantitative continue
- Lesquels de ces indices sont des paramètres de localisation (ou position) ?
minimum, moyenne, médiane, premier quartile
- Lesquels de ces indices sont des paramètres de dispersion ?
écart-type , écart interquartile , étendu
- Lequel de ces indices n'est ni l'un, ni l'autre ? Que reflète-t-il ?
skewness ; c'est un indice qui mesure la symétrie. Ici il semble différent de 0 : on a donc a priori une distribution NON symétrique
- Quelle la plus grande valeur des observations ?
max = minimum + étendu = 8500
- Quel est le 3^e quartile de cette distribution ?
 $Q_3 = \text{premier quartile} + \text{écart interquartile} = 6200$
- Construire le box-plot de cette distribution. Quel est votre commentaire sur la symétrie (ou son contraire) de cette distribution ?
*Axe avec : le min = 3500 ; $Q_1 = 4100$; Med = 4600 ; $Q_3 = 6200$; max = 8500
On retrouve la NON symétrie de cette distribution*

Exercice 3 : Le tableau ci-dessous donne la distribution de l'âge des populations de l'île de la Réunion et de la Métropole, exprimée en pourcentage.

âge (X) résidence(Y)	[0 ; 15[[15 ; 30[[30 ; 50[[50 ; 70[[70 ; 97[
la Réunion	29	30	25	12	4
Métropole	20	22	28	20	10

1. Construire l'histogramme de l'âge pour la Réunion.

histogramme sur :

les classes [0 ; 15[; [15 ; 30[; [30 ; 50[; [50 ; 70[; [70 ; 97[

de hauteur proportionnelle à :

$29/15 = 1.933$, $30/15 = 2$, $25/20 = 1.25$, $12/20 = 0.6$, $4/23 = 0.1739$

par ex : $29 ; 30 ; 25 * 15/20 = 18.75$; $12 * 15/20 = 9$; $4 * 15/27 = 2.22$

2. Calculer une valeur de la médiane des observations pour la Réunion ?

âge (X)	[0 ; 15[[15 ; 30[[30 ; 50[[50 ; 70[[70 ; 97[
la Réunion	29	59	84	96	100

la médiane se trouve dans la classe [15 ; 30[:

$med = 15 + (30-15) * (50-29) / (59 - 29) = 15 + 10.5 = 25.5$

3. Calculer une valeur de la moyenne des observations pour la Métropole ?

$moyenne = (20 * 7.5 + 22 * 22.5 + 28 * 40 + 20 * 60 + 10 * (97+70)/2) / 100 = 38$

4. Calculer la dispersion totale des observations pour la Métropole autour de la moyenne au sens de la distance des carrés en supposant que la population est de taille 100.

$dispersion = 20 * (7.5 - 38)^2 + 22 * (22.5 - 38)^2 + 28 * (40 - 38)^2 + 20 * (60 - 38)^2 + 10 * ((97 + 70)/2 - 38)^2 = 54385$

5. Donner alors la valeur de l'écart-type de l'âge de la population de la Métropole.

$écart-type = \sqrt{54385/100} = 23.32059$

6. Ces deux distributions sont-elles des distributions conditionnelles ? Si oui, en quel sens ?

ce sont 2 distributions conditionnelles de l'âge (conditionnellement à chacune des modalités de la variable résidence)

Exercice 4 : Les 36087 permis de conduire délivrés en Languedoc-Roussillon en 2004 par département et par nature de permis (P. O. pour Pyrénées-Orientales) :

1. Quels sont les individus et les variables ? Pourquoi ce tableau est-il un tableau de contingence ?

Individus : "Les 36087 permis de conduire délivrés en Languedoc-Roussillon en 2004"

2 variables qualitatives : "département" et "nature de permis" à resp. 4 et 5 modalités.

On a bien dans chaque croisement de "département" et "nature de permis" l'effectif du nombre de permis associé.

2. Donner la signification et la valeur des expressions suivantes : n ; n_{24} ; f_{42} ; $n_{2.}$.

$$n = 36087 ; n_{24} = 58 ; f_{42} = 1880/36087 = 5.21\% ; n_{2.} = 256 + 477 + 628 + 58 + 315 = 1734$$

3. Calculer et reporter sur le tableau les 2 distributions marginales en effectifs et en fréquences.

Permis	Aude	Gard	Hérault	Lozère	P. O.		
Auto	3 644	7 149	11 709	1 098	4 700	28300	0.7842
PL	256	477	628	58	315	1734	0.0480
TC	19	68	165	24	43	319	0.0088
Moto	656	1 880	2 227	203	768	5734	0.1589
	4575	9574	14729	1383	5826		
	0.1268	0.2653	0.4082	0.0383	0.1614		

4. Dans quel département les permis motos sont-ils relativement les plus nombreux ?

la fréquence de permis moto par département est :

$$0.1433880 \quad 0.1963652 \quad 0.1511983 \quad 0.1467824 \quad 0.1318229$$

C'est donc dans le Gard (avec 19.63 %) qu'ils sont relativement les plus nombreux.

5. On demande de compléter les cases vides des 2 tableaux des distributions conditionnelles en fréquence. Donner un titre explicite à chacun.

Titre - tableau 1 : Département conditionnellement à chaque nature du permis

Nature du permis	Aude	Gard	Hérault	Lozère	P. O.
Tourisme	0.1288	0.2526	0.4137	0.0388	0.1661
Poids lourds	XXX	XXX	0.3622	XXX	XXX
Transports en commun	XXX	XXX	0.5172	XXX	XXX
Motocyclette	XXX	XXX	0.3884	XXX	XXX

Titre - tableau 2 : Nature du permis conditionnellement à chaque département

Nature du permis	Aude	Gard	Hérault	Lozère	P. O.
Tourisme	0.7965	0.7467	0.7950	0.7939	0.8067
Poids lourds	XXX	XXX	0.04264	XXX	XXX
Transports en commun	XXX	XXX	0.01120	XXX	XXX
Motocyclette XXX	XXX	XXX	0.1512	XXX	XXX