

TP n° 1 : Distributions

L'objectif de ce TP est de réaliser avec le logiciel R des représentations graphiques basiques des variables quantitatives et qualitatives, puis le test d'ajustement à une distribution connue.

I. REPRÉSENTATIONS GRAPHIQUES

I. Diagramme à bâtons

Considérons le tableau suivant :

Nombre de personnes	Effectif
1	125
2	200
3	295
4	100
5	50

Table 1: Distribution du nombre de personnes par ménage

1. Quel est le type de la variable étudiée ?
2. Rentrer les modalités puis les effectifs dans un vecteur.
3. Créer une variable n calculant l'effectif total de l'échantillon étudié.
4. Créer ensuite le vecteur des fréquences de cette distribution.
5. Tracer le diagramme en bâtons des effectifs de cette distribution. Pour y arriver, utiliser la fonction `plot` avec l'argument `type="h"`.
6. Tracer le diagramme en bâtons des fréquences de cette distribution en utilisant la fonction `barplot`.

II. Polygone des fréquences

On va maintenant comparer les trois distributions de la table 2 en superposant leur polygone des fréquences.

1. Créer les vecteurs contenant les centres des classes (on pourra utiliser la fonction `seq`) et les fréquences des trois distributions.
2. Tracer avec la fonction `plot` le polygone des fréquences de la distribution des célibataires.
3. Avec la fonction `lines`, ajouter sur le même graphique les polygones des fréquences de deux autres distributions. Contrôler le type de symbole de chaque polygone par l'argument `pch`.
4. Utiliser la fonction `legend` pour ajouter une légende au graphique.

Age	Etat civil		
	Célibataires	Veuves	Divorcées
15-19	44827	20	86
20-24	81345	175	2558
25-29	21774	391	6162
30-34	5216	394	4785
35-39	1768	425	2958
40-44	833	529	2121
45-49	521	795	1589
50-54	383	932	1125
55-59	265	1041	563
60-64	176	947	235

Table 2: Répartition des mariages au Canada selon l'âge et l'état civil de l'épouse-1976

Continent	Fréquence
Afrique	0.111
Amérique du Nord	0.082
Amérique du Sud	0.0554
Asie	0.608
Europe	0.139
Océanie	0.005

Table 3: Distribution de la population mondiale sur les continents

III. Camembert

La distribution de la population mondiale sur les continents est donnée par la table 3.

1. Quel est le type de la variable observée ?
2. Créer un vecteur contenant les fréquences puis un autre contenant les modalités.
3. Tracer le camembert de cette distribution (fonction *pie*).

II. TEST D'AJUSTEMENT

Le logiciel R permet d'effectuer le test d'ajustement à une distribution connue avec la fonction *chisq.test* et les arguments *x* et *p* (?chisq.test pour plus de détails).

On considère la distribution suivante de 300 accouchements selon les jours de la semaine :

Jour	L	M	M	J	V	S	D	Total
Effectif	50	42	47	42	44	40	35	300

Effectuer un test d'ajustement pour déterminer si les accouchements se répartissent uniformément.

ν	1	2	3	4	5	6	7	8	9	10
Point critique	3.84	5.99	7.82	9.49	11.07	12.59	14.07	15.51	16.92	18.31