Graphes de distribution et distribution cumulée

Introduction

- Comment **visualiser graphiquement** la répartition des individus sur les modalités d'une variable ?

- À chaque type de variables correspond un graphe particulier.

Exemple : Activité préférée ?

U_X = {lecture, sport, peinture, musique}

Cas d'une variable qualitative nominale

- La distribution de la variable *X* est fournie par le tableau :

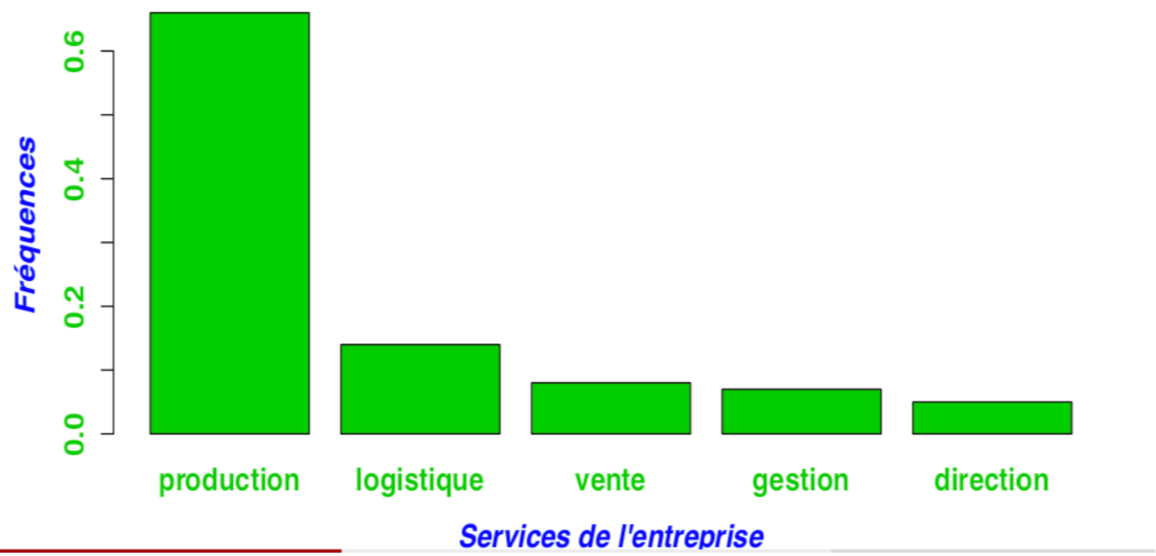
Variable	X				
Modalités m_k	m_1	m_2	$\dots$	m_C	Total
Effectifs n_k	n_1	n_2	$\dots$	n_C	n
Fréquences f_k	f_1	f_2	$\dots$	f_C	1

- Distribution - **diagramme en barres séparées**

Cas d’une variable qualitative nominale

- Exemple : répartition dans les services de l’entreprise

Service	production	logistique	vente	gestion	direction	
Effectifs	66	14	8	7	5	100
Fréquences	0.66	0.14	0.08	0.07	0.05	1



Exemple : UX = {Non satisfaisant, Peu satisfaisant, Satisfaisant, Très satisfaisant}

Cas d'une variable qualitative ordinale

- La distribution de la variable *X* est fournie par le tableau :

Variable	X				
Modalités m_k	m_1	m_2	$\ldots$	m_C	Total
Effectifs n_k	n_1	n_2	$\ldots$	n_C	n
Fréquences f_k	f_1	f_2	$\ldots$	f_C	1

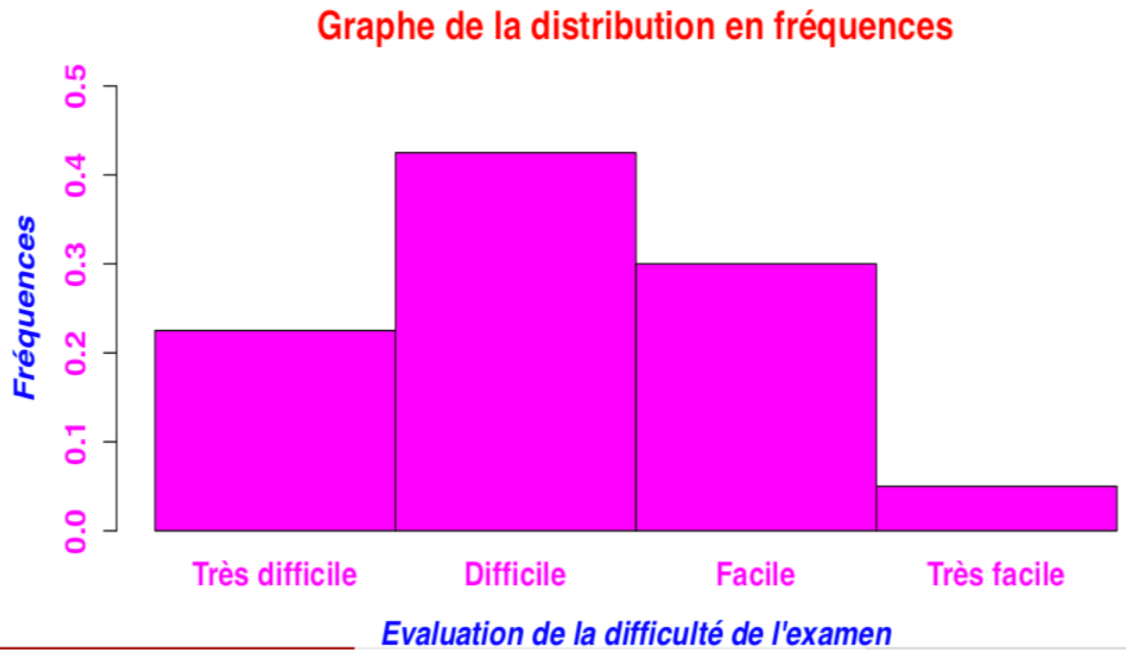
- Distribution - **diagramme en barres juxtaposées**

car nous envisageons ici l'échelle continue sous-jacente à la variable ordinale.

Cas d’une variable qualitative ordinale

Exemple : Évaluation de la difficulté d’un examen

Difficulté exam.	Très difficile	Difficile	Facile	Très facile	Total
Effectifs	9	17	12	2	40
Fréquences	0.225	0.425	0.300	0.050	1



Exemple : lancé de dé $U_X = \{1, 2, 3, 4, 5, 6\}$

Cas d'une variable quantitative discrète

- La distribution de la variable X est fournie par le tableau :

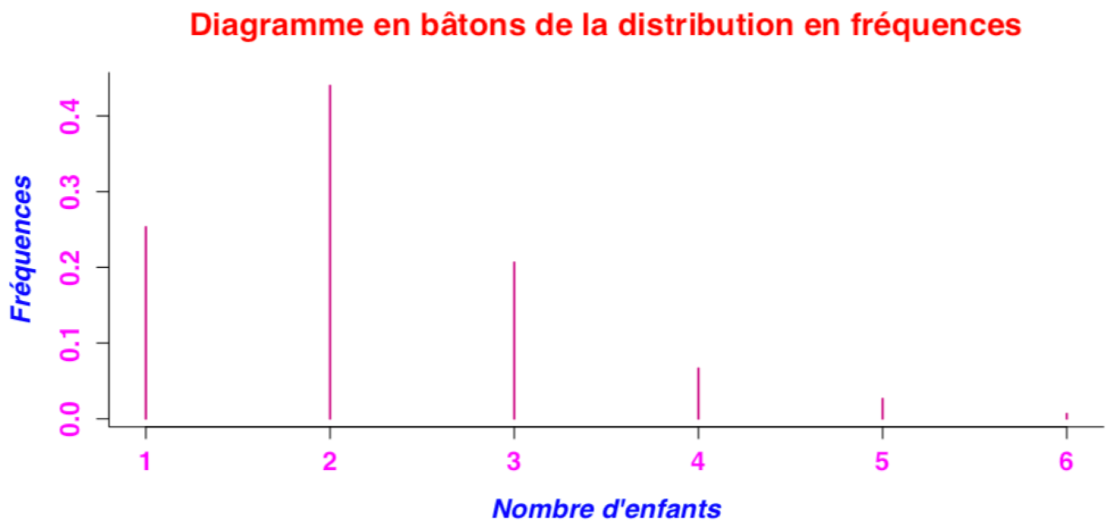
Variable	X				
Valeurs v_k	v_1	v_2	$\ldots$	v_C	Total
Effectifs n_k	n_1	n_2	$\ldots$	n_C	n
Fréquences f_k	f_1	f_2	$\ldots$	f_C	1

- Pour une variable quantitative discrète, les valeurs sont isolées les unes des autres.
- Représentation graphique de la distribution - **diagramme en bâtons**

Cas d’une variable quantitative discrète

Exemple : Nombre d’enfants dans les familles

Variable	Nombre d’enfants						
Valeurs v_k	1	2	3	4	5	6	Total
Effectifs n_k	38	66	31	10	4	1	150
Fréquences f_k	0.253	0.440	0.207	0.067	0.027	0.007	1



Exemple : le nombre de femmes qui ont voté lors des dernières élections

Cas d'une variable quantitative continue

- Comme pour toutes les variables, l'information sur les valeurs observées sont :
  - soit des données brutes
  - soit un tableau de distribution
- **Mais** à la différence de toutes les autres variables, **il ne peut pas en théorie y avoir de valeurs répétées** dans l'observation d'une variable quantitative continue.
  - pour regrouper les valeurs, on construit des **classes**
  - on **perd alors de l'information** sur les valeurs réelles observées
- **Attention :** Ce regroupement en classes est nécessaire pour construire une représentation graphique de la distribution

Cas d’une variable quantitative continue

- La distribution de la variable X est fournie par le tableau :

Variable	X				
Classes	$[b_0; b_1[$	$[b_1; b_2[$	$\dots$	$[b_{C-1}; b_C[$	Total
Effectifs n_k	n_1	n_2	$\dots$	n_C	n
Fréquences f_k	f_1	f_2	$\dots$	f_C	1

- **Attention :** Pour la représentation graphique de cette distribution, il faut tenir compte des **amplitudes** des classes (ou largeurs des classes).
  - Affirmer que 10 individus mesurent entre 155 cm et 165 cm ne signifie pas la même chose que d’affirmer que 10 individus mesurent entre 160 et 162 cm.
  - Les effectifs sont identiques (les fréquences aussi) mais dans le second cas, il y a une **concentration** beaucoup plus forte d’observations (l’intervalle est plus petit).
- Représentation graphique de la distribution – **histogramme**

Cas d’une variable quantitative continue

- Nous allons donc calculer pour chaque classe :
  - sa **densité de fréquence** : dk = fk / ak
  - équivalente (proportionnelle) à sa **densité d’effectif :** dk = nk / ak

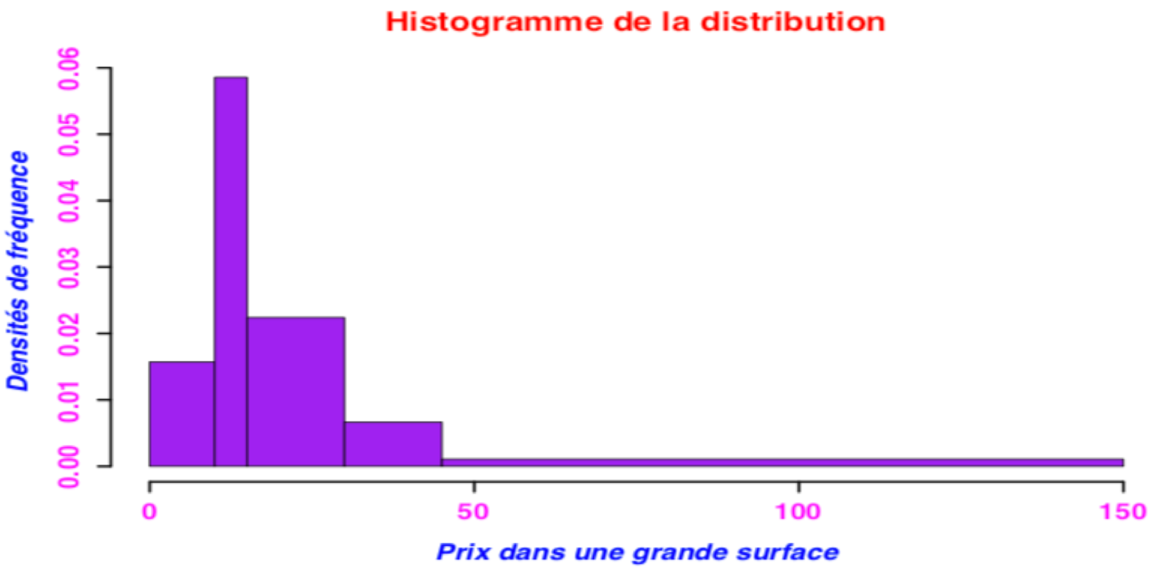
- On complète alors le tableau :

Variable	X				
Variable X	$[b_0 ; b_1[$	$[b_1 ; b_2[$	$\dots$	$[b_{C-1} ; b_C[$	Total
Effectifs n_k	n_1	n_2	$\dots$	n_C	n
Fréquences f_k	f_1	f_2	$\dots$	f_C	1
Amplitudes a_k	a_1	a_2	$\dots$	a_C	
Densités d_k	d_1	d_2	$\dots$	d_C	

Cas d'une variable quantitative continue

- Exemple : Prix des produits d'une grande surface

Variable	Prix					
Classes	]0 ; 10[	[10 ; 15[	[15 ; 30[	[30 ; 45[	[45 ; 150]	Total
n_k	110	205	235	70	80	700
f_k (%)	15.7	29.3	33.6	10.0	11.4	100
a_k	10	5	15	15	105	/
d_k	1.57	5.86	2.24	0.67	0.109	/



Distribution cumulée d'une variable discrète

Distribution cumulée d'une variable discrète

- Les valeurs étant numériques, elles peuvent être bien sûr rangées :

$$v_1 < v_2 < \ldots < v_k < \ldots < v_C$$

- Le tableau contenant la **distribution** et la **distribution cumulée** :

Variable	X				
Valeurs v_k	v_1	v_2	$\ldots$	v_C	Total
Effectifs n_k	n_1	n_2	$\ldots$	n_C	n
Fréquences f_k	f_1	f_2	$\ldots$	f_C	1
Eff. cum. N_k	$N_1 = n_1$	$N_2 = n_1 + n_2$	$\ldots$	$N_C = n$	
Fréq. cum. F_k	$F_1 = f_1$	$F_2 = f_1 + f_2$	$\ldots$	$F_C = 1$	

- **Attention :** les cumuls se font au niveau des valeurs puisque les individus se situent exactement sur chaque valeur.

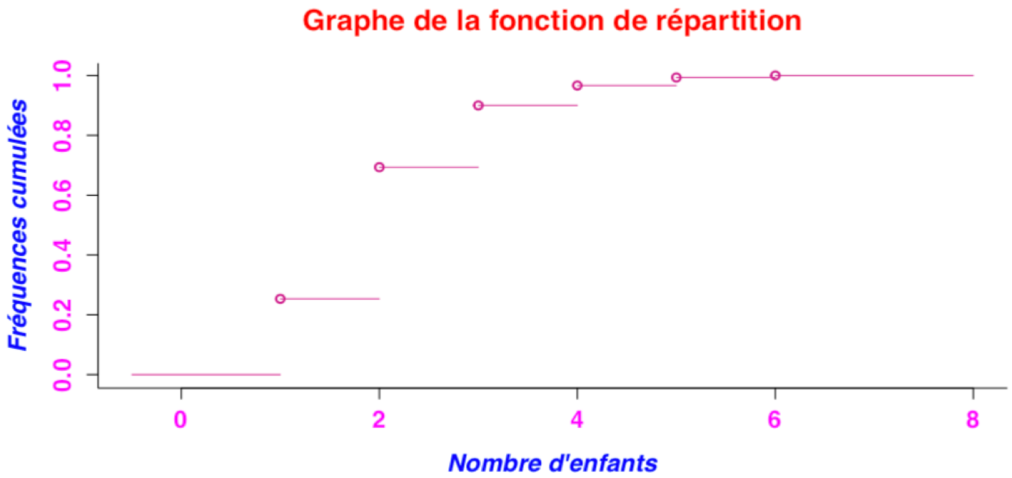
Distribution cumulée d'une variable discrète

Exemple : Nombre d'enfants dans les familles

Variable	Nombre d'enfants					
Valeurs v_k	1	2	3	4	5	6
Eff. cum. N_k	38	104	135	145	149	150
Fréq. cum. F_k	0.253	0.693	0.900	0.967	0.993	1

Distribution cumulée :

Distribution cumulée – fonction en escalier



Distribution cumulée d'une variable continue

Distribution cumulée d’une variable continue

- On rajoute des effectifs cumulés et des fréquences cumulées au niveau des bornes de classe.

Variable	X				
Variable X	$[b_0; b_1[$	$[b_1; b_2[$	$\dots$	$[b_{C-1}; b_C[$	Total
Effectifs n_k	n_1	n_2	$\dots$	n_C	n
Fréquences f_k	f_1	f_2	$\dots$	f_C	1

Bornes	b_0	b_1	b_2	$\dots$	b_{C-1}	b_C
Eff. cum. N_k	0	N_1	N_2	$\dots$	N_{C-1}	n
Fréq. cum. F_k	0	F_1	F_2	$\dots$	F_{C-1}	1

- on trace alors la fonction de répartition empirique par un **graphe linéaire par morceaux.**

Distribution cumulée d’une variable continue

- Exemple : Prix des produits d’une grande surface

Variable	Prix					
Classes	]0 ; 10[	[10 ; 15[	[15 ; 30[	[30 ; 45[	[45 ; 150]	Tot.
n_k	110	205	235	70	80	700
f_k (%)	15.7	29.3	33.6	10.0	11.4	100

Bornes	0	10	15	30	45	150
N_k	0	110	315	550	620	700
F_k (%)	0	15.7	45	78.6	88.6	100