Chapitre: 02
Fréquences cumulées et fonction de répartition

Chapitre 2 : Fréquences cumulées et fonction de répartition

🎯 Objectifs du chapitre

À la fin de ce chapitre, l'étudiant sera capable de :

- Construire un tableau de fréquences cumulées pour une variable ordinale, discrète ou continue.
- Définir et interpréter la fonction de répartition d'une variable statistique.
- Calculer et interpréter des fréquences cumulées dans un contexte zootechnique ou épidémiologique.
- Tracer la courbe de la fonction de répartition et l'utiliser pour répondre à des questions concrètes.
- Distinguer les cas où le cumul a un sens (variables ordinales, discrètes, continues) de ceux où il n'en a pas (variables nominales).

1. Introduction

Au chapitre précédent, nous avons appris à organiser les données sous forme de tableaux d'effectifs et de fréquences, et à les représenter graphiquement. Cependant, ces tableaux et graphiques ne répondent pas toujours aux questions concrètes que se pose l'éleveur, le technicien ou le vétérinaire. Par exemple : « Quelle proportion d'animaux a un poids inférieur à 40 kg ? », « Quel est l'âge médian au premier vêlage ? », « Quelle est la probabilité qu'un animal survive au-delà de 5 ans ? ».

Pour répondre à ce type de questions, il faut introduire la notion de **fréquences cumulées** et de **fonction de répartition**. Ces outils permettent de décrire la distribution d'une variable non plus seulement valeur par valeur, mais de manière cumulative, en indiquant la proportion d'individus qui se situent en dessous (ou au-dessus) d'un seuil donné.

Exemple introductif : Un vétérinaire souhaite savoir quelle proportion de vaches d'un troupeau a une production laitière inférieure à 20 litres par jour. Le tableau de distribution simple ne répond pas directement à cette question. En revanche, en calculant les fréquences cumulées, il peut dire : « 35 % des vaches produisent moins de 20 L/jour ». C'est l'objet de ce chapitre.

2.1. Fréquences cumulées

2.1.1. Principe et définition

La **fréquence cumulée** associée à une modalité *xi* est la proportion d'individus dont la valeur de la variable est **inférieure ou égale** à *xi*. Elle se note *Fi* et se calcule en additionnant les fréquences des modalités jusqu'à *xi* inclus :

$$F_i = f_1 + f_2 + \cdots + f_i = \sum_{j=1}^{i} f_j$$

De même, on définit l'**effectif cumulé** *Ni* comme la somme des effectifs jusqu'à *xi* :

$$N_i = n_1 + n_2 + \cdots + n_i$$

Propriété fondamentale : la dernière fréquence cumulée est toujours égale à 1 (ou 100 %), et le dernier effectif cumulé est égal à *N*.

2.1.2. Conditions d'utilisation

Le calcul des fréquences cumulées n'a de sens que si les modalités peuvent être **ordonnées**. Ainsi :

- **Variables nominales** : le cumul n'a **aucun sens**. On ne peut pas dire qu'une race est « inférieure » à une autre. Il est donc interdit de calculer des fréquences cumulées pour une variable nominale.
- **Variables ordinales** : le cumul est possible et utile, car les modalités ont un ordre naturel (par exemple, une note d'état corporel de 1 à 5).
- **Variables discrètes** : le cumul est possible, car les valeurs sont des nombres entiers ordonnés.
- **Variables continues** : le cumul est possible, mais il faut d'abord regrouper les valeurs en classes, puis cumuler les effectifs ou fréquences des classes.

2.1.3. Tableau de fréquences cumulées

Pour construire un tableau de fréquences cumulées, on ajoute deux colonnes au tableau de distribution simple : une colonne « effectifs cumulés » et une colonne « fréquences cumulées ».

Ce tableau illustre le calcul des effectifs et fréquences cumulés pour une variable ordinale, la note d'état corporel (NEC). Les colonnes « cumulés » permettent de répondre à des questions du type : « quelle proportion de vaches a une NEC inférieure ou égale à 3 ? ».

Tableau 2.1 – Fréquences cumulées pour une variable ordinale (NEC de 80 vaches)

NEC	Effectif *ni*	Fréquence *fi*	Effectif cumulé *Ni*	Fréquence cumulée *Fi*
1	4	0,05	4	0,05

2	12	0,15	16	0,20
3	28	0,35	44	0,55
4	24	0,30	68	0,85
5	12	0,15	80	1,00
Total	**80**	**1,00**		

Interprétation : La fréquence cumulée pour NEC = 3 est de 0,55, soit 55 %. Cela signifie que 55 % des vaches ont une NEC inférieure ou égale à 3 (c'est-à-dire un état corporel moyen ou maigre). Seules 15 % des vaches ont une NEC de 5 (très bon état).

Ce tableau montre le calcul des fréquences cumulées pour une variable discrète, le nombre de porcelets par portée. Il permet de répondre à des questions comme : « quelle proportion de portées a 9 porcelets ou moins ? ».

Tableau 2.2 – Fréquences cumulées pour une variable discrète (nombre de porcelets)

Nombre de porcelets	**Effectif *ni***	**Fréquence *fi***	**Effectif cumulé *Ni***	**Fréquence cumulée *Fi***
6	3	0,03	3	0,03
7	8	0,08	11	0,11
8	15	0,15	26	0,26
9	22	0,22	48	0,48
10	28	0,28	76	0,76
11	17	0,17	93	0,93
12	7	0,07	100	1,00
Total	**100**	**1,00**		

Interprétation : La fréquence cumulée pour 9 porcelets est de 0,48, soit 48 %. Cela signifie que près de la moitié des portées ont 9 porcelets ou moins. La fréquence cumulée pour 10 porcelets est de 0,76, soit 76 % : trois quarts des portées ont 10 porcelets ou moins.

Pour une variable continue, les données sont regroupées en classes. Le cumul se fait classe par classe. Ce tableau permet de répondre à des questions comme : « quelle proportion de veaux pèse moins de 45 kg ? ».

Tableau 2.3 – Fréquences cumulées pour une variable continue (poids de 150 veaux)

Classe de poids (kg)	**Effectif *ni***	**Fréquence *fi***	**Effectif cumulé *Ni***	**Fréquence cumulée *Fi***
[30 – 35[	9	0,06	9	0,06
[35 – 40[	27	0,18	36	0,24
[40 – 45[	36	0,24	72	0,48
[45 – 50[	45	0,30	117	0,78
[50 – 55[	27	0,18	144	0,96
[55 – 60[	6	0,04	150	1,00
Total	**150**	**1,00**		

Interprétation : La fréquence cumulée pour la classe [40 – 45[est de 0,48, soit 48 %. Cela signifie que 48 % des veaux pèsent moins de 45 kg. La fréquence cumulée pour la classe [45 – 50[est de 0,78 : 78 % des veaux pèsent moins de 50 kg.

Remarque importante : Pour une variable continue, la fréquence cumulée calculée à la borne supérieure d'une classe indique la proportion d'individus dont la valeur est **strictement inférieure** à cette borne. Par exemple, $F(45) = 0{,}48$ signifie que 48 % des veaux pèsent moins de 45 kg.

2.2. Fonction de répartition

2.2.1. Définition

La **fonction de répartition** d'une variable statistique X, notée $F(x)$, est la fonction qui, à toute valeur réelle x, associe la proportion d'individus de l'échantillon dont la valeur de X est **inférieure ou égale** à x :

$$F(x) = proportion\ d'individus\ tels\ que\ X \leq x$$

D'où, de manière équivalente :

$$F(x) = \frac{nombre\ d'individus\ avec\ X \leq x}{N}$$

Remarque terminologique : Dans certains manuels (notamment anglo-saxons), la fonction de répartition est définie avec une inégalité stricte ($X < x$). Les deux définitions sont acceptables, mais il faut être cohérent dans leur utilisation. Dans ce cours, nous adoptons la définition $F(x) = P(X \leq x)$.

2.2.2. Propriétés de la fonction de répartition

La fonction de répartition possède plusieurs propriétés fondamentales :

1. **Elle est croissante** : si $x_1 < x_2$, alors $F(x_1) \leq F(x_2)$.
2. **Elle est comprise entre 0 et 1** : pour tout x, $0 \leq F(x) \leq 1$.
3. **Elle tend vers 0 à gauche** : $\lim_{x \to -\infty} F(x) = 0$.
4. **Elle tend vers 1 à droite** : $\lim_{x \to +\infty} F(x) = 1$.
5. **Elle est constante entre deux valeurs observées** (pour une variable discrète) ou **linéaire par morceaux** (pour une variable continue regroupée en classes).

2.2.3. Construction de la fonction de répartition pour une variable discrète

Pour une variable discrète, la fonction de répartition est une **fonction en escalier**. Elle fait un saut à chaque valeur observée, et la hauteur du saut est égale à la fréquence de cette valeur.

Cette figure représente la fonction de répartition du nombre de porcelets par portée (tableau 2.2). Elle se présente sous forme d'escalier : à chaque valeur de la variable, la fonction fait un saut proportionnel à la fréquence de cette valeur.

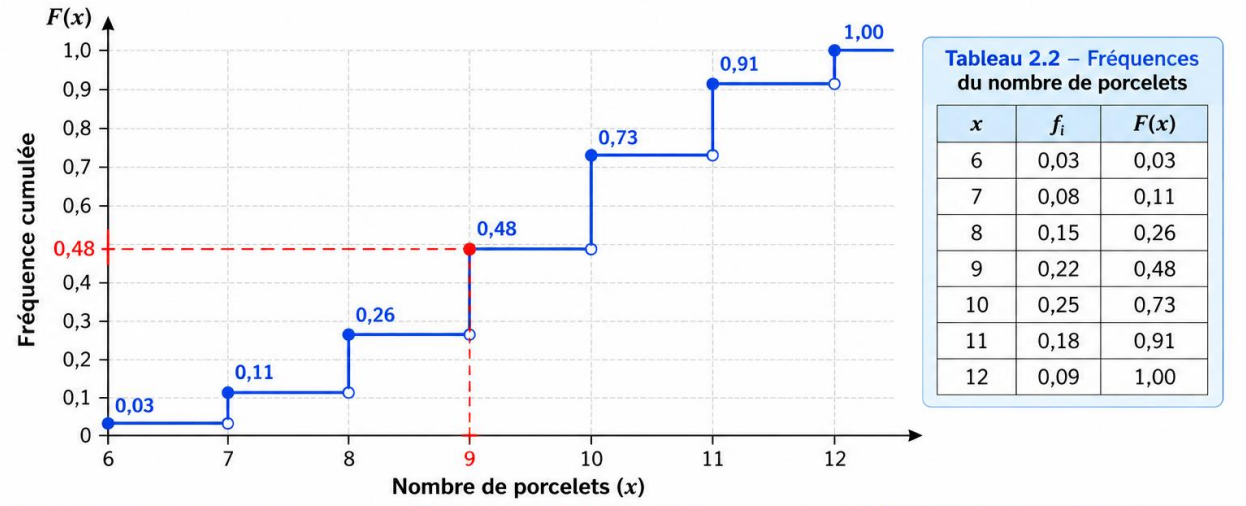


Tableau 2.2 – Fréquences du nombre de porcelets

x	f_i	$F(x)$
6	0,03	0,03
7	0,08	0,11
8	0,15	0,26
9	0,22	0,48
10	0,25	0,73
11	0,18	0,91
12	0,09	1,00

Présentation de la figure 2.1 :
Cette figure représente la fonction de répartition du nombre de porcelets par portée (tableau 2.2). Elle se présente sous forme d'escalier : à chaque valeur de la variable, la fonction fait un saut proportionnel à la fréquence de cette valeur.

Description de la figure 2.1 :
- **Axe horizontal** : nombre de porcelets (6, 7, 8, 9, 10, 11, 12).
- **Axe vertical** : fréquence cumulée $F(x)$, de 0 à 1.
- La courbe est en escalier : elle reste constante entre deux valeurs, puis fait un saut à chaque valeur.
- Les sauts correspondent aux fréquences : 0,03 pour 6 ; 0,08 pour 7 ; 0,15 pour 8 ; etc.

Interprétation :
La fonction de répartition permet de lire directement la proportion de portées ayant un nombre de porcelets inférieur ou égal à une valeur donnée.
Par exemple, $F(9) = 0{,}48$: **48 %** des portées ont 9 porcelets ou moins.

Figure 2.1 – Fonction de répartition d'une variable discrète (nombre de porcelets)

Description de la figure 2.1 :

- Axe horizontal : nombre de porcelets (6, 7, 8, 9, 10, 11, 12).
- Axe vertical : fréquence cumulée $F(x)$, de 0 à 1.

- La courbe est en escalier : elle reste constante entre deux valeurs, puis fait un saut à chaque valeur.
- Les sauts correspondent aux fréquences : 0,03 pour 6 ; 0,08 pour 7 ; 0,15 pour 8 ; etc.

Interprétation : La fonction de répartition permet de lire directement la proportion de portées ayant un nombre de porcelets inférieur ou égal à une valeur donnée. Par exemple, $F(9) = 0{,}48$: 48 % des portées ont 9 porcelets ou moins.

2.2.4. Construction de la fonction de répartition pour une variable continue

Pour une variable continue regroupée en classes, la fonction de répartition est construite par **interpolation linéaire** entre les bornes des classes. On procède en plusieurs étapes :

1. **Calculer les fréquences cumulées aux bornes des classes** :
$$F(a_0) = 0, F(a_1) = f_1, F(a_2) = f_1 + f_2, \dots, \qquad F(a_k) = 1$$
2. **Relier les points par des segments de droite** dans chaque classe : on fait une interpolation linéaire entre $F(a_{i-1})$ et $F(a_i)$.
3. **Prolonger la courbe par 0 à gauche de a_0 et par 1 à droite de a_k.**

Ce tableau présente les fréquences cumulées aux bornes des classes de poids. Il sert de base à la construction de la courbe de répartition par interpolation linéaire.

Tableau 2.4 – Fonction de répartition pour les poids de 150 veaux

Classe de poids (kg)	Effectif *ni*	Fréquence *fi*	Fréquence cumulée *Fi*
[20 – 25[	9	0,06	0,06
[25 – 30[	27	0,18	0,24
[30 – 35[	36	0,24	0,48
[35 – 40[	45	0,30	0,78
[40 – 50[	27	0,18	0,96
[50 – 60[	6	0,04	1,00
Total	**150**	**1,00**	

Interprétation : La fonction de répartition $F(30)=0{,}24$ signifie que 24 % des veaux pèsent moins de 30 kg. $F(35)=0{,}48$: 48 % des veaux pèsent moins de 35 kg. $F(40)=0{,}78$: 78 % des veaux pèsent moins de 40 kg.

Cette figure représente la fonction de répartition des poids de 150 veaux. Elle est construite par interpolation linéaire entre les bornes des classes. Elle permet de lire rapidement la proportion de veaux en dessous d'un seuil donné.

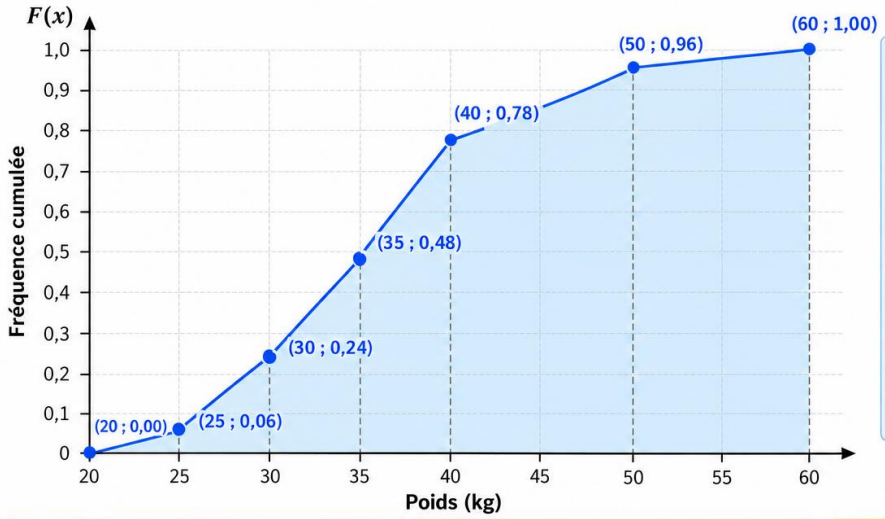


Tableau (exemple) – Distribution des poids de 150 veaux

Classe (kg)	Effectif n_i	Fréquence f_i	Fréq. cumulée $F(x)$
[20 – 25[	9	0,06	0,06
[25 – 30[	27	0,18	0,24
[30 – 35[	36	0,24	0,48
[35 – 40[	45	0,30	0,78
[40 – 45[	18	0,12	0,90
[45 – 50[	9	0,06	0,96
[50 – 60]	6	0,04	1,00
Total	**150**	**1,00**	–

Présentation de la figure 2.2 :
Cette figure représente la fonction de répartition des poids de 150 veaux. Elle est construite par interpolation linéaire entre les bornes des classes. Elle permet de lire rapidement la proportion de veaux en dessous d'un seuil donné.

Description de la figure 2.2 :
- **Axe horizontal :** poids (kg), de 20 à 60.
- **Axe vertical :** fréquence cumulée $F(x)$, de 0 à 1.
- La courbe part de 0 à gauche de 20 kg, monte linéairement dans chaque classe, et atteint 1 à droite de 60 kg.
- Points remarquables : $F(25) = 0,06$; $F(30) = 0,24$; $F(35) = 0,48$; $F(40) = 0,78$; $F(50) = 0,96$; $F(60) = 1$.

Interprétation :
La courbe est croissante et en forme de « S » légèrement asymétrique. Elle montre que la majorité des veaux pèsent entre 30 et 45 kg. La pente est maximale dans la classe [35 – 40[, ce qui correspond à la classe modale.

Par exemple :
- $F(35) = 0,48$: 48 % des veaux pèsent 35 kg ou moins.
- $F(50) = 0,96$: 96 % des veaux pèsent 50 kg ou moins.

Figure 2.2 – Fonction de répartition des poids de 150 veaux

Présentation de la figure 2.2 : **Description de la figure 2.2** :

- Axe horizontal : poids (kg), de 20 à 60.
- Axe vertical : fréquence cumulée $F(x)$, de 0 à 1.
- La courbe part de 0 à gauche de 20 kg, monte linéairement dans chaque classe, et atteint 1 à droite de 60 kg.
- Les points remarquables : $F(25)=0,06$; $F(30)=0,24$; $F(35)=0,48$; $F(40)=0,78$; $F(50)=0,96$; $F(60)=1$.

Interprétation : La courbe est croissante et en forme de « S » légèrement asymétrique. Elle montre que la majorité des veaux pèsent entre 30 et 45 kg. La pente est maximale dans la classe [35 – 40[, ce qui correspond à la classe modale.

2.2.5. Utilisation de la fonction de répartition pour répondre à des questions concrètes

La fonction de répartition est un outil puissant pour répondre à des questions du type :

- « Quelle proportion d'animaux a un poids inférieur à x ? » → lire $F(x)$.
- « Quelle proportion d'animaux a un poids supérieur à x ? » → calculer $1-F(x)$.
- « Quelle est la valeur x telle que 50 % des animaux ont un poids inférieur ? » → c'est la médiane, qui sera étudiée au chapitre 3.

Exemple d'application :

Dans le tableau 2.4, on veut savoir quelle proportion de veaux pèse moins de 38 kg.

- 38 kg se trouve dans la classe [35 – 40[.
- On sait que $F(35)=0{,}48$ et $F(40)=0{,}78$.
- Par interpolation linéaire :

$$F(38) = F(35) + \frac{38-35}{40-35} \times \big(F(40) - F(35)\big)$$

$$F(38) = 0{,}48 + \frac{3}{5} \times (0{,}78 - 0{,}48) = 0{,}48 + 0{,}6 \times 0{,}30 = 0{,}48 + 0{,}18 = 0{,}66$$

- Donc 66 % des veaux pèsent moins de 38 kg.

2.3. Application à la zootechnie et à l'épidémiologie

2.3.1. En zootechnie

Les fréquences cumulées et la fonction de répartition sont utilisées pour :

- **Évaluer la croissance** : quelle proportion d'animaux atteint un poids cible à un âge donné ?
- **Comparer des lots** : deux lots d'engraissement peuvent avoir des distributions de poids différentes ; la fonction de répartition permet de comparer leur homogénéité.
- **Définir des seuils de sélection** : par exemple, sélectionner les 20 % d'animaux les plus lourds à un âge donné.

2.3.2. En épidémiologie

Les fréquences cumulées sont également précieuses pour :

- **Estimer la prévalence cumulée** d'une maladie : proportion d'animaux ayant été malades au moins une fois jusqu'à un âge donné.
- **Étudier la survie** : la fonction de répartition de l'âge au décès permet de calculer la proportion d'animaux morts avant un âge donné.
- **Définir des seuils de risque** : par exemple, quelle proportion d'animaux a une température rectale supérieure à 40 °C ?

2.4. Synthèse du chapitre

- Les **fréquences cumulées** indiquent la proportion d'individus dont la valeur est inférieure ou égale à une modalité donnée.
- Le cumul n'a de sens que pour les variables **ordinales, discrètes ou continues**. Il est interdit pour les variables nominales.

- La **fonction de répartition** $F(x)$ associe à toute valeur x la proportion d'individus tels que $X \leq x$.
- Pour une variable discrète, la fonction de répartition est une **fonction en escalier**.
- Pour une variable continue, la fonction de répartition est construite par **interpolation linéaire** entre les bornes des classes.
- La fonction de répartition permet de répondre à des questions concrètes : proportion d'animaux en dessous d'un seuil, proportion au-dessus, etc.

2.5. Conclusion

Ce chapitre a introduit deux outils fondamentaux pour l'analyse descriptive : les fréquences cumulées et la fonction de répartition. Ces outils complètent les tableaux et graphiques du chapitre 1 en permettant de répondre à des questions cumulatives, essentielles en zootechnie et en épidémiologie. Ils constituent également la base de la notion de médiane et de quantiles, qui seront étudiées au chapitre 3.

La fonction de répartition est un objet central en statistique : elle résume toute l'information contenue dans une distribution et permet de calculer des proportions, des seuils et des quantiles. Sa maîtrise est donc indispensable avant d'aborder les caractéristiques de tendance centrale et de dispersion.

Maintenant que nous savons calculer des fréquences cumulées et construire une fonction de répartition, nous pouvons aborder les **caractéristiques de tendance centrale et de dispersion**. La médiane, en particulier, est directement liée à la fonction de répartition : c'est la valeur x telle que $F(x)=0{,}5$. Le chapitre III introduira également la moyenne, le mode, les quartiles, la variance et l'écart-type, qui permettent de résumer une distribution par quelques indicateurs clés.

? Questions de révision

1. Définissez la fréquence cumulée. Pour quels types de variables peut-on la calculer ?
2. Pourquoi ne peut-on pas calculer de fréquences cumulées pour une variable nominale ?
3. Construisez le tableau des fréquences cumulées pour la variable « nombre de traitements antibiotiques » (données : 0 : 5 ; 1 : 15 ; 2 : 20 ; 3 : 15 ; 4 : 5).
4. Définissez la fonction de répartition. Quelles sont ses propriétés ?
5. Pour une variable discrète, quelle est la forme de la fonction de répartition ?

6. Pour une variable continue, comment construit-on la fonction de répartition ?
7. À partir du tableau 2.4, calculez la proportion de veaux pesant moins de 42 kg.
8. À partir du tableau 2.1, interprétez la fréquence cumulée pour NEC = 4.