

RECOMMANDATIONS
POUR LA REALISATION
ET LA PRESENTATION
DE TABLEAUX ET FIGURES
EN EPIDEMIOLOGIE ANIMALE

1^{ère} édition

Jean-Jacques Bénét, Julie Rivière et Bernard Toma

AUTEURS

Jean-Jacques BÉNET *Professeur émérite*
Ecole nationale vétérinaire d'Alfort

Julie RIVIÈRE *Enseignant-chercheur*
Ecole nationale vétérinaire d'Alfort

Bernard TOMA *Professeur honoraire*
Ecole nationale vétérinaire d'Alfort

© AEEMA, 2015

Association pour l'étude de l'épidémiologie des maladies animales

7 avenue du Général de Gaulle, 94704 Maisons-Alfort, France

Site internet : <http://www//aeema.vet-alfort.fr>

Reproduction ou traduction interdite sauf autorisation du premier auteur

Toute proposition d'amélioration et/ou de complémentation peut être adressée à jjbenet@vet-alfort.fr

SOMMAIRE

Remerciements	6
A L'INTENTION DU LECTEUR PRESSE	7
PRESENTATION	9

LIVRE PREMIER : ELEMENTS DE CONNAISSANCE..... 13

CHAPITRE I - MEMENTO	15
I - Tableau synthétique	15
II - Choix d'un diagramme	17
CHAPITRE II - LES VARIABLES	20
I - Les variables statistiques	20
II - Les variables quantitatives	20
B - Les variables qualitatives	22
C - Les variables temporelles	22
III - Les variables texte.....	23
CHAPITRE III - LES TABLEAUX	24
CHAPITRE IV - LES FIGURES	26
I - Terminologie	26
A - Graphiques et diagrammes.....	26
B - Schémas.....	26
C - Cartes.....	26
II - Diagrammes	27
A - Diagramme à points (en anglais : <i>dot plot</i>)	28
B - Diagramme à lignes (en anglais : <i>line graph</i>)	30
C - Diagramme A Bâtons (en anglais : <i>bar chart</i>)	32
D - Diagramme à colonnes (en anglais : <i>bar chart</i>).....	34
E - Diagramme à barres (en anglais : <i>bar chart</i>).....	36
F - Diagramme Circulaire (en anglais : <i>pie chart</i>)	38
G - Quelques autres diagrammes	40
III - Schémas	50
IV - Cartes	53

CHAPITRE V - Recommandations pour la presentation des Tableaux et des figures	57
I - Objectifs.....	57
II - Recommandations pour la présentation de toute forme d'illustration	57
III - Recommandations spécifiques pour la présentation d'un tableau	59
IV - Recommandations spécifiques pour la présentation d'une figure ou d'une carte	60

LIVRE DEUXIEME : POUR ALLER PLUS LOIN.....	61
---	-----------

CHAPITRE I - EN SAVOIR PLUS.....	63
I - Comment distinguer variables quantitatives et variables qualitatives ?.....	63
II - En savoir plus sur Les variables texte	65
III - Comment réaliser un tableau adapté aux objectifs ?.....	67
IV - En savoir plus sur les diagrammes à points : Effet attracteur de points isolés	73
V - Exemple d'utilisation inadéquate d'un diagramme à points pour un nombre limité de points	74
VI - Exemple d'utilisation d'un diagramme à lignes pour des variables discrètes regroupées en classes	75
VII - Comparaison des diagrammes à lignes et à paliers	76
VIII - Précisions sur la dénomination « diagramme à colonnes ».....	77
IX - Choix entre diagramme à colonnes et histogramme pour une variable discrète ..	77
X - Plus d'information sur la distinction entre diagrammes à colonnes et diagrammes à bâtons	78
XI - Plus de précisions sur l'utilisation des diagrammes à colonnes	79
XII - Extension de l'application des diagrammes à colonnes aux variables quantitatives discrètes	83
XIII - En savoir plus sur les histogrammes	87
XIV - En savoir plus sur le choix du nombre ou du pourcentage de cas pour une représentation cartographique	91
CHAPITRE II - COMMENT FAIRE AVEC EXCEL®	94
I - Faire des calculs avec une variable numérique en évitant le format texte.....	94
II - Placer les bâtons ou les colonnes d'un diagramme exactement sur les graduations (et non dans les intervalles, ce que fait Excel® par défaut)	95
III - Placer les colonnes d'un diagramme entre les graduations	96
IV - Réaliser un diagramme à paliers avec Excel®	96
V - Utiliser les fonctions de traitement de chaînes de caractère pour facilement réaliser la légende de l'axe des abscisses pour un histogramme.....	96
VI - Dans une figure, séparer le signe « % » des valeurs chiffrées correspondantes...	97

LIVRE TROISIEME : EXERCICES et CORRIGES 99

CHAPITRE I - ENONCES	101
I - Tableaux	101
A - Objectif.....	101
B - Enonces	101
II - Figures	104
A - Objectif 1 :	104
B - Objectif 2 :	108
 CHAPITRE II - CORRIGES.....	 121
I - Tableaux	121
A - Objectif.....	121
B - Corrigés	121
II - Figures	125
A - Objectif 1.....	125
B - Objectif 2 :	130
 BIBLIOGRAPHIE	 147
 INDEX	 149

Remerciements

Nos très vifs remerciements vont aux personnes qui nous ont aidés pour la réalisation de cette première édition, en particulier :

➤ Pour la relecture du manuscrit :

➤ Pour les conseils informatiques ayant permis la mise en ligne de ce document :

- Gilles Hattenberger
- Benoît Durand (Anses)

A L'INTENTION DU LECTEUR PRESSE

Vous êtes pressé(e) ? Vous n'avez pas les quelques heures nécessaires pour parcourir l'intégralité de ce document ? Vous devez impérativement savoir très rapidement quelle figure conviendrait à votre besoin ?

Utilisez le tableau synthétique qui présente [la liste des diagrammes les plus utilisés en épidémiologie](#) (représentés en miniature), et comportant les liens hypertextes permettant de s'y reporter immédiatement ; consultez également le résumé présentant [les éléments de choix d'un diagramme](#) situé juste après.

Plan

Cet ouvrage comporte trois parties : la première est consacrée à la présentation des éléments théoriques de connaissance, la deuxième à des compléments qui permettent d'aller plus loin, la troisième à des exercices d'application permettant leur assimilation.

La première partie comporte sept chapitres.

- Le premier chapitre est un *memento* présentant de façon très rapide les diverses formes de diagrammes possibles selon le type de variable, avec quelques indications pour orienter le choix correspondant.
- Le deuxième chapitre explicite les caractéristiques des variables, de façon à permettre de mieux comprendre les raisons du choix des diagrammes.
- Le troisième chapitre est consacré aux tableaux.
- Le quatrième chapitre porte sur les différents types de figures : les diagrammes, les schémas et les cartes.
- Le cinquième chapitre donne les recommandations pour la présentation des tableaux et figures.

La deuxième partie comporte deux chapitres compilant les divers renvois par liens hypertexte.

- Le premier permet d'aller plus loin sous le titre général de « Pour en savoir plus ».
- Le deuxième donne des indications complémentaires sur la manière de s'y prendre avec Excel pour parvenir à différents résultats présentés.

La troisième partie comporte deux chapitres.

- Le premier est consacré à la présentation des exercices.
- Le deuxième fournit leurs corrigés.

Version de travail. Ne pas diffuser

PRESENTATION

La statistique traite les données collectées de façon à les rendre plus facilement accessibles à notre compréhension grâce à ce qu'on appelle la « *réduction des données* » : celle-ci consiste à réaliser soit des résumés statistiques à l'aide des paramètres de position et de dispersion (moyenne, médiane, écart-type) à partir desquels il est possible de reconstituer l'allure générale de l'ensemble des données, soit des tableaux regroupant les données de façon agrégée, soit des représentations graphiques ou, d'une façon générale, des figures. Cet accès à une plus grande clarté, une meilleure lisibilité se fait toutefois aux dépens d'une perte d'information : il faut donc trouver le bon compromis entre ces deux aspects, fidélité de l'information et compréhensibilité pour le lecteur.

L'utilisation des outils tableaux et figures peut générer des risques de déformation, voire de faute de communication, d'où l'intérêt de respecter quelques règles simples de conception et de présentation des uns et des autres afin d'en limiter la survenue éventuelle.

Ces recommandations sont destinées à un public débutant dans l'étude de l'épidémiologie ; c'est pourquoi l'exposé vise à l'essentiel plutôt qu'à l'exhaustivité. Les **buts** sont les suivants :

- éviter les erreurs grossières dans la réalisation et la présentation des tableaux et figures destinés à un compte-rendu en épidémiologie ;
- aider à mieux comprendre comment faire sans difficulté majeure, tout en visant un niveau de base (« ...*Pour les nuls* »).

Le public ainsi visé étant concerné principalement par la rédaction de rapports, de mémoires (y compris thèse de doctorat vétérinaire) dont la vocation est d'être imprimés et éventuellement reproduits par photocopie, le choix a été fait d'une conception en noir et blanc de ces illustrations. L'adaptation des règles pour la couleur fera l'objet d'un document spécifique également utile pour la conception de diaporamas.

Le seul prérequis est d'être capable d'utiliser un tableur de type Excel[®] pour constituer un fichier de données et réaliser des tableaux et des figures. De nombreuses aides en ligne sont disponibles sur internet pour dépasser le niveau débutant requis.

Objectifs

Etre capable de :

1. reconnaître le type de variable ;
2. concevoir l'organisation d'un tableau de données en fonction de l'objectif poursuivi ;
3. choisir le type de figure appropriée au type de variable et à l'objectif souhaité ;
4. justifier ces différents choix ;
5. formuler le titre d'un tableau ou d'une figure, respecter les conditions de leur mise en forme conformément aux normes rédactionnelles en usage.

Limites

La priorité a été donnée à la simplification pour rendre ces éléments facilement utilisables¹. Visiblement, il n'existe pas encore de consensus sur les « bons usages » des représentations graphiques, qui, de plus, connaissent un développement considérable du fait des possibilités

¹ « *Ce qui est simple est toujours faux. Ce qui ne l'est pas est inutilisable* » (Paul Valéry, 1960, p. 864).

des logiciels. Raison de plus pour nous limiter à l'essentiel. Ce travail de synthèse a nécessité un effort de réflexion pour pallier l'absence de développement théorique consensuel ; étant conscients de son inachèvement, nous serons très attentifs aux remarques qui pourront nous être formulées.

Le choix a été fait de recourir à une suite bureautique classique (Excel[®], mais Open Office donne les mêmes résultats), mais qui ne respecte pas toujours l'orthodoxie des représentations graphiques : nous signalerons ces écarts. En revanche, les logiciels (gratuits) R et Epi Info[™], respectent les règles d'usage. A vous de choisir, en fonction du degré d'exigence attendue et des contraintes : pour un rapport informel, Excel[®] peut suffire, pour un rapport de stage de M2 (voire un doctorat) d'épidémiologie, R est fortement recommandé.

Ceci n'est pas un cours de statistique. Si des explications statistiques sont fournies, c'est uniquement pour servir les buts et objectifs définis : il n'est donc pas surprenant de constater des manques importants sur des développements usuels du point de vue de la méthodologie statistique.

Par conséquent, ce document est un compromis entre la rigueur, le souci d'une information suffisamment détaillée, et la facilité d'utilisation destinée à l'épidémiologiste non chevronné. C'est pourquoi, les auteurs apprécieront que toute remarque utile pour l'amélioration de ce document leur soit communiquée par courriel (jjbenet@vet-alfort.fr).

Conventions

Nous utiliserons des exemples le plus souvent publiés, en en citant les références, parfois des jeux de données en notre possession, mais non publiés : dans ce cas, nous situerons ces observations dans un pays fictif, la Mondalie, pays imaginaire, ressemblant au nôtre, utilisé dans des publications de l'OMS à vocation de formation, ce qui évitera de froisser quelque susceptibilité que ce soit.

La numérotation des tableaux et des figures est continue pour chacune des trois parties dénommées "Livre" suivi d'un numéro d'ordre. De même, la numérotation a été recommencée pour chacun des chapitres du livre second, afin de faciliter la comparaison entre l'énoncé de l'exercice et son corrigé.

Pour rester un guide pratique, tout en apportant toutes les explications indispensables à la compréhension sans alourdir le document, les développements explicatifs complémentaires ont été regroupés dans le livre deuxième ("*Pour aller plus loin*") et rendus accessibles par des liens hypertextes apparaissant en **bleu et souligné** permettant de s'y reporter rapidement. Certains de ces liens renvoient directement dans le document et non à cette partie. Dans les deux cas, pour permettre au lecteur de revenir à son point de départ pour poursuivre sa lecture, des liens permettant ce retour peuvent apparaître dans le document sous la forme suivante :

[retourner à l'article (*suivi d'un titre*)]

Si vous n'avez pas utilisé ce lien, ne vous préoccupez pas de cette indication et poursuivez votre lecture !

Les **mots et expressions** apparaissant sur fond gris et en bleu non souligné permettent (*seulement avec les versions Word, pas en format pdf*), lorsque le curseur est positionné dessus, de faire s'afficher un **court commentaire** ou une **définition** évitant de quitter l'emplacement de lecture pour y accéder (il faut juste être un tout petit peu patient, l'affichage demandant une latence d'environ une seconde : essayez sur ce qui est affiché en bleu dans les lignes qui précèdent, vous verrez).

Ce qui veut dire que ce document est conçu pour être consulté préférentiellement à l'écran afin de pouvoir disposer de l'intégralité de ces liens hypertextes. Ce document a été réalisé sous Word 2010. Si vous ne disposez pas de ce logiciel, voici quelques recommandations :

- si vous disposez de Word 2003, vous pouvez télécharger le programme vous permettant de lire cette version 2010 ;
- si vous avez une version de Word plus ancienne ou un logiciel Open Source et que vous ne parvenez pas à lire la version 2010, vous devez prendre la version 2003, compatible avec n'importe quel ordinateur ; cette version est toutefois notablement plus lourde que la version 2010 (11 Mo au lieu d'un peu plus de 8) ;
- il existe également une version au format pdf qui est beaucoup moins lourde (4 Mo), qui respecte rigoureusement la mise en page et même les liens hypertextes, mais qui ne permet pas l'affichage des bulles de commentaires au passage du pointeur.

Version de travail. Ne pas diffuser

Version de travail. Ne pas diffuser

LIVRE PREMIER : ELEMENTS DE CONNAISSANCE

Version de travail. Ne pas diffuser

Version de travail. Ne pas diffuser

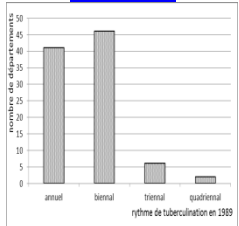
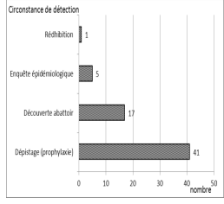
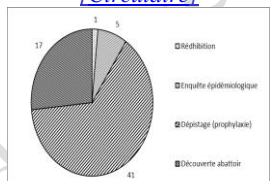
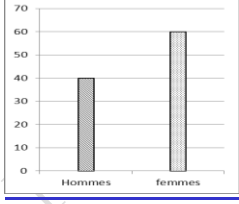
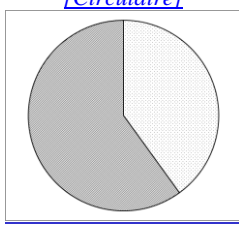
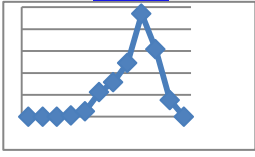
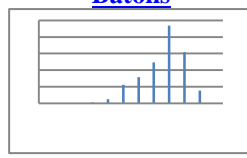
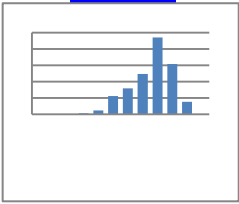
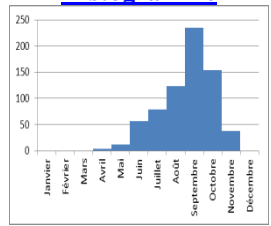
CHAPITRE I - MEMENTO

I - TABLEAU SYNTHÉTIQUE

L'essentiel est rassemblé dans un tableau synthétique (tableau 1) présentant l'éventail des diagrammes (= illustration graphique) le plus souvent utilisés en épidémiologie, et quelques conseils pour choisir celui correspondant aux besoins.

Tableau 1 : Présentation des diagrammes le plus souvent utilisés en épidémiologie en fonction de la nature de la variable à représenter (cliquer sur la dénomination pour suivre le lien)

Type de variable	Définition de la variable	Exemple	Diagramme indiqué	Utilisation possible <i>[mais non recommandée]</i>
Quantitative discrète	Seulement certaines valeurs possibles, le plus souvent entières (pas de valeurs intermédiaires)	Nombre d'animaux par élevage	<p>Bâtons</p>	<p>Points</p> <p>Colonnes (nombre élevé de valeurs par intervalle)</p>
Quantitative continue	Toutes valeurs intermédiaires possibles	Poids, taille	<p>Lignes</p> <p>Points</p> <p>(Idéal pour représentation X/Y)</p> <p>Histogramme</p>	<p>Histogramme (nombre élevé de valeurs par intervalle)</p> <p>Lignes</p>

Type de variable	Définition de la variable	Exemple	Diagramme indiqué	Utilisation possible <i>[mais non recommandée]</i>
<u>Qualitative ordinale</u>	Echelle de valeurs non quantitatives mais « ordonnées »	Degré de satisfaction (échelle de Likert)	<u>Colonnes</u> 	<u>[Barres]</u>
<u>Qualitative nominale</u>	Catégories distinguées par leur « nom »	Modalités de dépistage, races	<u>Barres,</u> 	<u>[Colonnes]</u> <u>[Circulaire]</u> 
<u>Qualitative binaire</u>	Deux valeurs possibles	Mâle / femelle ; Vrai / faux ; 0/1	<u>Bâtons</u> 	<u>[Circulaire]</u> 
<u>Temporelle</u>	Variable continue mais ni quantitative, ni qualitative	Incidence mensuelle de foyers de maladie	<u>Lignes</u> 	<u>[Points]</u> <u>Bâtons</u> 
			<u>Colonnes</u> 	<u>Histogramme</u> 

II - CHOIX D'UN DIAGRAMME

Le choix du diagramme le plus approprié dépend de trois aspects : du type de variable, de la compréhensibilité du diagramme et de l'objectif poursuivi.

Choix d'un diagramme en fonction du type de variable

Les indications et contre-indications des diagrammes doivent être respectées en fonction des variables, car certains ne conviennent pas du tout pour certaines représentations.

Le tableau 2 résume les éléments de choix des différents diagrammes en fonction des types de variable. On peut retenir les recommandations suivantes :

- Pour une variable **discrète**, la représentation appropriée est le diagramme à bâtons, mais on peut aussi envisager les lignes (surtout pour représenter plusieurs séries), ou les colonnes (dans le cas de regroupement en classes) ;
- Pour une variable **continue**, les indications majeures sont les diagrammes à points (diagramme X/Y). Il est possible d'utiliser l'histogramme dans le cas où il est régulier, ou les diagrammes à colonnes (alors jointives, en cas de regroupement en classes), ou encore les diagrammes à lignes (plusieurs séries) ;
- Les variables **ordinales** n'ont qu'une indication : le diagramme à colonnes ;
- De même, les variables **nominales** n'ont qu'une indication : le diagramme à barres ;
- Enfin, les variables **temporelles** peuvent être représentées aussi bien par des diagrammes à lignes qu'à colonnes ou par des histogrammes.

Choix d'un diagramme en fonction d'une représentation optimale pour la compréhension du lecteur

Il faut tenir compte de codes implicites de représentation présents à l'esprit de tout usager non expert, afin de ne pas induire d'erreur d'interprétation.

Par exemple, un axe des abscisses orienté par une flèche implique une notion de relation mathématique avec la variable reportée en ordonnées. C'est pourquoi il importe de respecter les codes de communication correspondant à chaque type de diagramme :

- pour un **diagramme à bâtons** ou à **points**, les valeurs sont placées à la verticale des graduations, matérialisant ainsi la nature mathématique de l'axe ;
- pour un **diagramme à colonnes**, utilisé pour les variables ordinales, les valeurs doivent être placées entre les graduations, retirant de ce fait la signification totalement mathématique de l'axe, pour n'en conserver qu'une approximation semi-quantitative toujours suggérée par l'orientation de l'axe ;
- pour les **variables nominales**, il convient de ne pas utiliser le même axe horizontal pour la dispersion des valeurs, puisque leur dispersion sur l'axe n'implique aucune relation mathématique ; c'est pourquoi l'indication du **diagramme à barres** s'impose tout naturellement, en plus de l'avantage de pouvoir écrire facilement les libellés.

Choix d'un diagramme en fonction de l'objectif poursuivi

Ce n'est pas le graphique qui va spontanément, comme par miracle, révéler une réalité cachée. Il faut d'abord être guidé par l'idée préalable que l'on veut transmettre comme information au

lecteur, ce qui conditionne la sélection d'une représentation graphique particulière : c'est un principe fondamental de travail scientifique !

Il faut donc d'abord « savoir ce que l'on veut ». En absence d'idée préconçue, il suffit de s'en tenir au principe du respect d'une description objective, avant de passer à une interprétation au plan descriptif, et ensuite envisager une réflexion explicative. Le choix des outils correspondants en découle alors.

Ensuite, il faut garder à l'esprit qu'il s'agit d'une manière de communiquer, de la façon respectant à la fois l'objectivité du travail scientifique et l'indépendance de pensée du destinataire de cette communication : celui-ci doit être amené à la liberté de partager les mêmes interprétations, ou au contraire à pouvoir les critiquer, et non à être inexorablement conduit à penser exactement comme l'auteur des représentations. Cette communication doit aussi être d'une compréhension adaptée à l'auditoire : **trop de complexité nuit à l'efficacité de la communication.**

Au total, la compréhensibilité de la représentation doit rester la priorité, ce qui peut conduire à prendre quelques degrés de liberté par rapport à une rigueur absolue, dont on retiendra :

<p style="text-align: center;">Il y a ce qui est indiqué ou souhaitable, ce qui peut être admis, et ce qui doit être évité.</p>
--

Tableau 2 : Éléments du choix d'un diagramme en fonction de la nature de la variable à représenter

Variable	Définition de la variable	Points	Bâtons	Lignes	Colonnes	Barres	Histogramme	Secteurs
<u>Quantitative</u>								
discrète	Seulement certaines valeurs possibles, le plus souvent entières (pas de valeurs intermédiaires)	Oui si deux variables (X/Y)	<u>Oui</u>	Oui = polygone de fréquence	Oui Si regroupement en classes	Non	Possible Si très grand nombre de valeurs établissant une quasi continuité	Non
continue	Toutes valeurs intermédiaires possibles	<u>Oui</u> si deux variables (X/Y)	Non	Oui après regroupement en classes, plusieurs variables	Oui après regroupement en classes, colonnes jointives	Non	<u>Oui</u>	Non
<u>Qualitative</u>								
ordinaire	Echelle de valeurs non quantitatives mais « ordonnée » (ex : échelle de Likert)	Non	Non	Non	<u>Oui</u>	Possible	Non	Non
nominaire	Catégories distinguées par leur « nom »	Non	Non	Non	Possible	<u>Oui</u>	Non	Possible
<u>Temporelle</u>	Variable continue mais non quantitative, ni qualitative	Possible	Possible	<u>Oui</u>	<u>Oui</u>	Non	Possible	Non

Les indications majeures sont soulignées.

CHAPITRE II - LES VARIABLES

Les variables sont des éléments pouvant prendre des valeurs différentes pour chaque individu. La nature des variables et leurs caractéristiques conditionnent les modalités de leur traitement, puis de leur exploitation pour produire tableaux et / ou figures. Nous opposerons les variables statistiques et les variables de type texte.

I - LES VARIABLES STATISTIQUES

Au plan statistique, on distingue les variables quantitatives et les variables qualitatives (Tableau 3) : si on peut additionner deux observations et que le résultat a un sens, la variable est quantitative, sinon, elle est qualitative (ou dite aussi catégorielle).

- Pour une variable **quantitative**, si les valeurs possibles sont infinies, la variable est **continue** ; si seulement certaines valeurs sont possibles, sans intermédiaire, elle est **discrète**.
- Pour une variable **qualitative**, si l'on peut déterminer un ordre, la variable est **ordinale** ; à défaut, elle est **nominale** (on « nomme » les différentes modalités possibles), à moins qu'elle ne comporte que deux modalités, et elle est alors **binaire**.

Tableau 3 : Différents types de variables statistiques

Type de variable	Propriétés	Exemples	Valeurs possibles
Variables quantitatives			
	<i>Numériques ; tous calculs possibles</i>		
Variable discrète	Valeurs déterminées, sans intermédiaires possibles	Nombre porcelets par truie Nombre d'animaux par élevage	10, 11, 12... 253 ; 48 ; 2 800
Variable continue	Toutes valeurs intermédiaires possibles	Poids	25,4 kg
Variables qualitatives			
	<i>Catégorielles (résultats de calculs éventuels dénués de sens)</i>		
Variable ordinale	Ordonnée Si numérique, n'autorise que le classement, pas l'addition.	Niveau d'éducation Degré de satisfaction	Primaire, secondaire, supérieur 1, 2, 3, 4, 5.
Variable nominale	Distinguées par nomination des modalités de la catégorie	Race de vache	Prim Holstein (codé 66), Charolaise (codé 38), Limousine (codé 34), etc.
Variable binaire	Deux états possibles (ou pris en compte)	Sexe État de santé	Mâle, femelle Atteint, non atteint

II - LES VARIABLES QUANTITATIVES

Une variable est quantitative lorsqu'elle exprime une grandeur, que les nombres utilisés pour la représenter permettent de faire tous types de calculs mathématiques et que les résultats obtenus ont un sens.

On distingue deux types de variables quantitatives, selon la continuité de leurs valeurs.

1. Variable quantitative discrète

1.1. Définition

Une variable quantitative est **discrète** lorsqu'elle ne peut être exprimée que par des **valeurs entières** (le plus souvent), sans valeur intermédiaire possible.

Le nombre de valeurs peut être faible ou élevé.

1.2. Nombre limité de valeurs

Considérons le nombre de porcelets nés vivants par truie à chaque portée. Il s'agit bien d'un nombre exprimant une quantité, par unités entières, et donc d'une variable discrète. On peut faire des calculs, pouvant aboutir à des nombres qui ne sont plus entiers : par exemple, en moyenne on pourra obtenir 10,43 porcelets nés vivants par truie et par portée. Le caractère entier caractérise par conséquent la donnée brute collectée, pas son résultat des calculs.

La représentation graphique d'une variable discrète doit refléter ce caractère entier de la donnée, chaque valeur étant positionnée exactement sur la graduation correspondante de l'axe des abscisses : c'est ce que font les diagrammes à **bâtons**.

1.3. Nombre élevé de valeurs

Considérons maintenant le nombre de bovins dans un élevage : il s'agit encore d'une variable discrète, puisqu'on ne peut dénombrer que des valeurs entières. Mais le nombre de valeurs possibles par élevage est beaucoup plus élevé que précédemment, d'une seule unité jusqu'à plusieurs centaines, voire plus. Il est donc préférable de procéder à des regroupements des effectifs **en classes**, par exemple de 50 en 50.

On peut donc considérer qu'il existe une certaine « continuité » des valeurs entre les bornes des intervalles de classes, mais par unités entières, sans que ce soit une continuité de même nature que pour une variable quantitative continue.

On a le choix de la représentation : soit diagramme à **colonnes**, en donnant une largeur suffisante aux colonnes pour symboliser l'étendue des valeurs possibles, sans toutefois assurer la continuité avec les classes voisines de façon à respecter la nature entière des valeurs ; soit un **histogramme** (avec les difficultés qui sont propres à ce type de figure). [\[En savoir plus\]](#)

Dans ce cas, les valeurs doivent être placées dans l'intervalle entre deux graduations de l'axe des abscisses, de façon à correctement représenter le fait que les données sont regroupées par classe.

2. Variable quantitative continue

1.4. Définition

Une variable quantitative est **continue** lorsqu'elle peut prendre **toutes les valeurs possibles**.

1.5. Exemple

Le poids, par exemple, est une variable continue.

Mais, il est possible que des variables discrètes puissent, dans certains cas, être aussi traitées comme des variables véritablement continues. Reprenons l'exemple des porcelets nés vivants par truie, mais en considérant non plus le nombre par truie (qui est un nombre entier), mais le nombre moyen par élevage (qui est un nombre décimal). La variable n'est plus discrète, mais devient **continue**.

Le graphique doit rendre compte de la continuité de la variable : c'est ce que fait un **histogramme**, ou bien un diagramme à colonnes à la condition que celles-ci soient jointives, et que les intervalles de classes soient égaux.

[\[retourner à l'article "diagramme à lignes"\]](#)

B - LES VARIABLES QUALITATIVES

1. Définition

Une variable est qualitative si ses valeurs ne peuvent être que **nommées** (variable qualitative *nominale*), ou **ordonnées** (variable qualitative *ordinaire*), sans que des nombres éventuellement utilisés pour ce faire permettent de faire des calculs arithmétiques.

Les variables qualitatives sont de format texte (bien qu'elles puissent être codées avec des nombres) mais ne comportent qu'un nombre limité de valeurs (*cf. infra* variables texte).

2. Exemples

Les variables qualitatives conduisent à des partitions spontanées des données, qui sont facilement classables dans chacune des catégories de la variable, ce qui n'est pas le cas d'une variable quantitative qui nécessite la définition de classes en établissant les limites pour chacune d'elles.

- Certaines variables qualitatives comportent malgré tout une notion de grandeurs que l'on peut classer, « ordonner » : ce sont des variables « **ordinales** », comme les échelles de notation ou de satisfaction (échelle de Likert). Il ne faut pas confondre ce type de classement qui résulte de la nature du phénomène étudié, avec le classement alphabétique des modalités d'une variable nominale.
- Les classes d'une variable **nominale** sont **nommées** (*cf.* tableau 3), mais elles peuvent être codées pour en simplifier le traitement, comme pour les races de bovins (Interbev, 2008) par exemple : 66 pour Prim'Holstein, 38 pour Charolaise, 34 pour Limousine, *etc.* Ces codes numériques ne permettent aucun calcul arithmétique (la Prim'Holstein, de code 66, n'est pas deux fois plus quelque chose que la Lourdaise, de code 33).
- Celles qui ne peuvent prendre que deux valeurs sont dites **binaires**. Elles sont de trois types : dichotomique, très largement utilisées en épidémiologie (atteint, non atteint) ; booléenne (vrai, faux) ; de Bernouilli (0,1).

[\[retourner à l'article "diagramme à colonnes"\]](#)

[\[retourner à l'article "diagramme à lignes"\]](#)

C - LES VARIABLES TEMPORELLES

Le temps est une variable particulière, **non quantitative** (l'an 2000 ne veut pas dire deux fois plus de temps qu'en l'an 1000, puisque tout dépend de l'origine choisie) **mais continue**. Il ne faut pas la confondre avec une durée, qui elle est une variable quantitative continue.

Les variables temporelles peuvent être exprimées selon diverses unités : années, mois, trimestres, semaines, jours....

Elles peuvent être représentées par plusieurs types de diagrammes. Du fait que le temps est une variable continue, on pourrait la représenter par un histogramme. Pour éviter la confusion avec une variable quantitative continue, on peut préférer le diagramme à colonnes. On peut aussi utiliser des diagrammes à lignes, soit pour mieux distinguer plusieurs phénomènes que l'on veut représenter, soit pour mieux observer une tendance éventuelle.

III - LES VARIABLES TEXTE

Comme leur nom l'indique, les variables texte sont destinées à contenir une chaîne de caractères, le plus souvent de type alphabétique, comme un mot par exemple. Si le nombre de possibilités est limité, il s'agit d'une variable qualitative, vue précédemment ; si le nombre de possibilités n'a pas été limité par la définition de catégories, il s'agit d'une variable texte, sans autre précision. Du texte libre peut aussi être saisi en réponse à une question ouverte, en vue d'un traitement approprié par la suite. [\[En savoir plus\]](#)

Il faut être prévenu de la distinction informatique entre variables numériques et variables texte, car un tableur ne prend en compte pour ses calculs que les valeurs numériques : si, pour une obscure raison, une valeur censée être numérique est codée comme du texte, elle ne sera pas incluse dans les calculs, ce qui sera responsable d'erreurs. [\[En savoir plus\]](#)

Version de travail. Ne pas diffuser

CHAPITRE III - LES TABLEAUX

Définition

Un tableau est un outil de communication visuelle présentant des **informations de façon structurée** pour permettre le classement de données et/ou la comparaison par le **croisement de lignes et de colonnes**.

Remarques

La question du choix entre tableau et figure se pose. Le tableau est beaucoup plus précis, mais beaucoup moins facile à lire qu'une figure, qui, à l'inverse, a l'avantage de la simplicité d'accès ; dans certains cas favorables, il est possible de compléter une figure avec des indications chiffrées, pouvant, de ce fait, rendre le tableau inutile, aux dépens toutefois de la clarté de la figure.

Il faut préciser quelques termes, pour lever des ambiguïtés pouvant résulter d'un usage parfois inapproprié de certains d'entre eux. Les données collectées sont archivées dans des *fichiers*, dont le format peut être varié. On peut être amené à concevoir un système de *base de données*, qui assure une meilleure sécurité des données et une plus grande efficacité des traitements. On peut en consulter le contenu en éditant une *liste*, le plus souvent organisée sous forme d'une *table* : les individus figurent en ligne, et les informations relatives aux diverses variables étudiées pour chacun des individus sont en colonnes. Il faut un œil exercé pour exploiter cette forme de présentation : repérer les manques, les données aberrantes, l'étendue des données, les données les plus fréquentes, *etc.* Pour rendre plus facile cette consultation, le premier traitement consiste à agréger les données de mêmes caractéristiques, ce qui génère une succession de *tableaux*.

Description

Un tableau comporte des **têtes de colonnes** (nom donné à la première ligne du tableau explicitant les contenus de chaque colonne) et une première colonne dénommée **colonne type** (explicitant les contenus de chaque ligne).

Différents types de tableaux

On peut distinguer différents types de tableaux selon leur forme et leur contenu.

➤ Selon la **forme**, on distingue :

- des tableaux à *simple entrée* : les informations sont classées selon un seul critère, organisées soit en colonnes (les libellés des rubriques occupant la colonne type, cf. Tableau 4a), soit en lignes (les libellés des rubriques figurant en tête de colonnes, cf. Tableau 4b).

Pour un grand nombre de libellés, la disposition verticale s'impose.

La disposition horizontale, en rappelant la distribution graphique selon l'axe des abscisses, permet sans doute de plus facilement faire le lien avec la figure correspondante.

Tableau 4a : Répartition des élevages laitiers selon leur taille (Mondalie, 2010).
(Source : Services vétérinaires de Mondalie)

Nombre d'animaux / élevage	Nombre d'élevages
<50	112
50-99	85
100-149	78
150-199	74
200-249	21
250-299	8

Tableau 4b : Répartition des élevages laitiers selon leur taille (Mondalie, 2010).
(Source : Services vétérinaires de Mondalie)

Nombre d'animaux / élevage	<50	50-99	100-149	150-199	200-249	250-299
Nombre d'élevages	112	85	78	74	21	8

- des tableaux à *double entrée* : les informations sont classées selon deux critères, d'où une lecture par croisement des lignes et des colonnes (cf. Tableau 5).

Tableau 5 : Répartition des départements selon le nombre d'élevages bovins par département et selon le rythme de dépistage de la tuberculose (France, 1989). (Données : DGAI)

Rythme dépistage	< 3000 élevages	[3000-6000 élevages]	> 6000 élevages	Total
Annuel	21	13	7	41
Biennal	9	21	16	46
Triennal	1	1	4	6
Quadriennal	0	0	2	2
Total	31	35	29	95

- des tableaux à *entrées multiples* : les lignes et/ou les colonnes peuvent être dédoublées pour introduire un niveau supplémentaire (non illustré ici). La lecture de ces tableaux est moins aisée.
- Selon le **contenu**, on distingue :
- Des tableaux *statistiques*, contenant des données numériques (cf. Tableaux 4 et 5) ;
 - Des tableaux *récapitulatifs*, contenant des données texte (cf. Tableau 6).

Tableau 6 : Dictionnaire des variables. Etude sur la mensuration des plis de peau
(extrait, d'après Lambert O., 2014)

VARIABLE	CONTENU	COMMENTAIRE
dI	Différence entre I3 et I0.	Epaississement dû à la tuberculine bovine (méthode de référence IDS).
dT	Différence entre T3 et T0.	Epaississement dû à la tuberculine aviaire (pour l'IDC).
dS	Différence entre S3 et S0.	Variabilité intra-site au point symétrique.
dI0T0	Différence entre I0 et T0.	Variabilité inter-sites n°1 (à J0).
dI0S0	Différence entre I0 et S0.	Variabilité inter-sites n°1 (à J0).
dT0S0	Différence entre T0 et S0.	Variabilité inter-sites n°1 (à J0).

- Des tableaux *comparatifs*, qui peuvent être statistiques ou récapitulatifs. Par exemple, la juxtaposition de deux tableaux à simple entrée pour la même variable, mais pour deux populations différentes.

Les [recommandations](#) pour la présentation des tableaux sont présentées dans le chapitre qui leur est consacré.

CHAPITRE IV - LES FIGURES

I - TERMINOLOGIE

Le terme de « figure » désigne un grand nombre d'illustrations diverses : des graphiques, des schémas, des cartes, des photographies, *etc.*. Quelle que soit la catégorie, la seule dénomination générale valable dans un article reste celle de « figure » et il faut par conséquent proscrire tout recours à d'autres dénominations diverses (graphiques, diagrammes, schémas, cartes, photographies, *etc.*).

Nous n'aborderons, pour nos besoins de définitions, que les trois premières catégories citées : graphiques / diagrammes, schémas et cartes.

A - GRAPHIQUES ET DIAGRAMMES

En statistique, un **graphique** est une image représentant un ensemble de valeurs numériques, pour en faciliter la perception de façon globale, synthétique. Les plus simples illustrent les rapports entre deux variables, l'une sur un axe horizontal, l'abscisse (ou axe des abscisses), l'autre sur un axe vertical, l'ordonnée (ou axe des ordonnées), la première variable constituant celle dont on étudie les variations et la seconde, les répercussions de cette variation.

Le mot **diagramme** sert à désigner une forme particulière : « diagramme de Tukey » (pour les « boîtes à moustaches »), « diagramme de Gantt » (planification opérationnelle de tâches), « diagramme d'Ishikawa » (arborescence de causes et d'effets), et, pour ce qui nous concerne, diagrammes à points, à lignes, à barres, *etc.*

Par conséquent, le mot graphique sert à désigner toute forme d'illustration graphique représentant un jeu de données, quelle qu'en soit la forme de diagramme, tandis que le mot diagramme sert à désigner une forme générique de représentation (indépendamment de tout jeu de données) et non la représentation elle-même d'un jeu de données.

Il est relativement courant de voir des graphiques montrant trois variables (illustrations en 3D), voire davantage, ou encore des figures complexes. Compte tenu de nos objectifs, nous nous limiterons aux graphiques à deux dimensions, et nous ne ferons qu'évoquer brièvement les quelques autres graphiques en usage en épidémiologie, sans aucun souci d'exhaustivité.

B - SCHÉMAS

Un schéma est une représentation simplifiée d'un objet, d'un processus, d'une organisation ou de tout ensemble d'éléments comportant des relations. En épidémiologie, les cycles épidémiologiques sont représentés par des schémas.

C - CARTES

Les cartes permettent la représentation spatiale de données sur un territoire.

Elles ont un intérêt tout particulier en épidémiologie, d'où leur développement distinct des schémas : l'épidémiologie descriptive s'intéresse en effet à trois dimensions, population, temps, espace, cette dernière faisant appel aux représentations cartographiques. Des outils informatiques (Systèmes d'information géographique ou SIG) ont considérablement augmenté les possibilités d'application.

Les [recommandations](#) pour la présentation des figures sont présentées dans le chapitre qui leur est consacré.

II - DIAGRAMMES

Nous retiendrons les diagrammes selon leur plus grande fréquence d'usage en épidémiologie et non en statistique d'une façon générale : diagrammes à points, à lignes, à bâtons, à colonnes, à barres, circulaires ainsi que d'autres diagrammes, comme les « boîtes à moustaches », les pyramides des âges, les histogrammes, les diagrammes radar, les diagrammes à paliers.

La terminologie n'étant pas standardisée, il est nécessaire de préciser les choix que nous avons été amenés à faire.

Diagramme « de » points est une forme relâchée, par ellipse de « nuage de points », à éviter.

Pour la plupart des diagrammes, les formules « à », ou « en » sont utilisées, avec une prédominance de l'une ou de l'autre, différente selon les diagrammes, sans qu'il soit possible de trouver de raison logique autre que celle de l'habitude.

Le préfixe « en », précédant le type de diagramme, suggère une forme générale (ellipse de l'expression « en forme de »), tandis que « à » désigne plutôt les motifs, les éléments que comporte le diagramme pour la représentation : après mûre réflexion, notre préférence incline plutôt à privilégier le motif que la forme. Aussi, compte tenu de l'hétérogénéité des usages constatée, et du besoin de cohérence dans ce document, nous avons fait le choix d'utiliser « à », pour standardiser les différentes dénominations. **Notre choix ne constitue donc nullement une recommandation, autre que celle de n'utiliser, dans un même document, qu'une seule forme d'expression, soit l'une, soit l'autre, mais pas tantôt l'une tantôt l'autre.**

[Retourner au chapitre « Diagramme à colonnes »](#)

[Retourner au chapitre "Précisions sur la dénomination Diagramme à colonnes"](#)

Le tableau 1, placé au début de ce document, résume les principales possibilités de choix de diagramme en fonction du type de variable, tout en permettant l'accès direct à chacun d'entre eux grâce aux liens hypertextes.

Chaque type de diagramme est présenté sur deux pages, les développements complémentaires éventuels étant accessibles par lien hypertexte.

A - DIAGRAMME À POINTS (en anglais : *dot plot*)

Synonymes : diagramme de dispersion, ou diagramme en nuage de points ;

Définition

Diagramme dans lequel chaque valeur est représentée par un point.

Description

Ce mode de représentation graphique est le plus élémentaire.

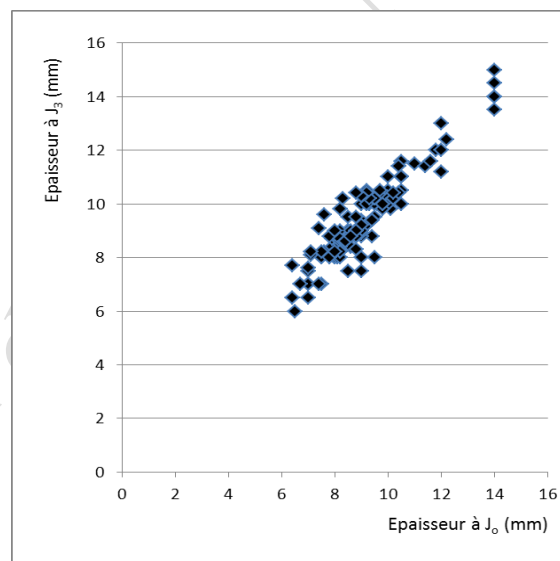
Le graphique représente une série de points dispersés. Ceux-ci peuvent ne suggérer aucune forme particulière, ou bien une orientation selon un axe rectiligne, ou encore selon une forme plus complexe (« U », ou bien U inversé « \cap »).

Exemple

Distribution des épaisseurs de pli de peau de bovins

La peau des bovins a une épaisseur variable selon l'emplacement (pour un même individu), selon les individus et selon les races. On s'en aperçoit en mesurant l'épaisseur du pli de peau lors du test de dépistage de la tuberculose (par injection intradermique de tuberculine) d'une part avant l'injection de tuberculine (J_0), d'autre part 3 jours plus tard, pour la lecture du résultat du test (J_3). Une mesure a été effectuée sur un échantillon de 144 bovins charolais (Figure 1).

Figure 1 : Relation entre l'épaisseur du pli de peau à J_0 (moment de l'injection de tuberculine) et J_3 (moment de la lecture de la réaction) ; N = 144. (Source : Gane, J., 2011)



Ce **diagramme à points** rend bien compte de la relation attendue entre l'épaisseur du pli de peau à J_0 et J_3 , sous la forme d'un alignement suivant grossièrement une bissectrice, mais dont certains individus s'écartent, soit du fait de la réaction à la tuberculine, soit d'une erreur (de lecture, de report des résultats ou pour toute autre raison !).

Avantages

Cette représentation offre une perception de la globalité des données et permet ainsi de mettre en évidence une relation entre les deux variables étudiées, si elle existe.

Inconvénients

Plusieurs individus ayant les mêmes valeurs sont représentés par un seul point et ne peuvent par conséquent être perçus : il faut alors que le logiciel puisse rendre compte de cette accumulation en indiquant le nombre d'individus en chiffres au-dessus du point ou en affichant les points selon une taille proportionnelle au nombre de points, par exemple.

Un seul point est très mal perçu par l'œil humain, induisant les deux types d'erreur, par défaut et par excès. Si le nombre de points est limité, la perception de la globalité est mauvaise, car l'œil perçoit mieux les surfaces. Mais inversement, si la perception d'un ensemble de points suffisamment nombreux est bonne, un (ou quelques) point(s) relativement écarté(s) peu(ven)t être traité(s) par l'œil en lui accordant une valeur excessive par rapport à son(leur) effet statistique véritable voir un exemple d'un tel effet [\[voir un exemple d'un tel effet attracteur\]](#).

Indications et contre-indications

Ce type de graphique convient pour la représentation de nuages de points résultant de la combinaison de deux variables (type X / Y), dont la dispersion facilite une perception globale, en particulier d'une tendance.

Lorsqu'abscisse et ordonnée ont la même unité, il est souhaitable d'utiliser la même échelle (Figure 1), sauf cas particulier [\[pour en savoir plus\]](#).

Ce mode de représentation ne convient pas lorsqu'il ne peut y avoir qu'une seule valeur possible d'ordonnée pour des valeurs d'abscisse en nombre limité [\[voir un exemple\]](#) : en effet, la perception de quelques points du graphique sera moins bonne qu'avec une représentation de type diagramme à colonnes, qui utilise la surface. Il faut alors choisir un autre mode de représentation, de façon à faciliter la perception [\[voir un exemple de diagramme plus approprié\]](#).

[Retour au tableau présentant les diagrammes](#)

B - DIAGRAMME À LIGNES (en anglais : *line graph*)

Synonymes

Diagramme à lignes brisées (ou aussi *en* lignes brisées). Diagramme linéaire.

Remarque : l'expression « diagramme en ligne(s) » est ambiguë, car elle peut à la fois signifier le mode de représentation (avec des lignes), la position (alignement des diagrammes) ou le mode d'accès (« en ligne » = accessible par internet).

Définition

Diagramme obtenu en reliant chacun des points (un seul point en ordonnée par abscisse) par des segments de droite.

Appliqué à des fréquences, il prend le nom de **polygone de fréquence** (en anglais : *frequency polygon*), qui permet un lissage des données, utile pour la perception d'ensemble, particulièrement d'une fréquence cumulée.

Description

Chaque point correspond à une seule abscisse. Les points sont reliés entre eux par une ligne brisée. Les points sont positionnés soit sur les graduations des abscisses, comme pour les diagrammes en bâtons, soit entre les graduations, comme pour les diagrammes à colonnes.

Il ne faut pas confondre ce diagramme avec une représentation d'une droite ou une courbe d'ajustement statistique d'une série de données dont le tracé résulte de calculs statistiques, tandis que pour le diagramme à lignes, il est obtenu en joignant les différents points par des segments de droite. Il faut également distinguer ce diagramme du [diagramme à paliers](#).

Exemples

La figure 2 représente la proportion d'élevages selon le nombre de bovins tuberculeux.

La figure 3 reprend les mêmes données, mais en cumulant les pourcentages en fonction de la progression des valeurs des abscisses.

Figure 2 : Distribution du pourcentage d'élevages selon le nombre de bovins tuberculeux en Mondalie de 2002 à 2010
(Source : services vétérinaires de Mondalie)
(Abscisse 10 cumule les valeurs 10 animaux et plus).

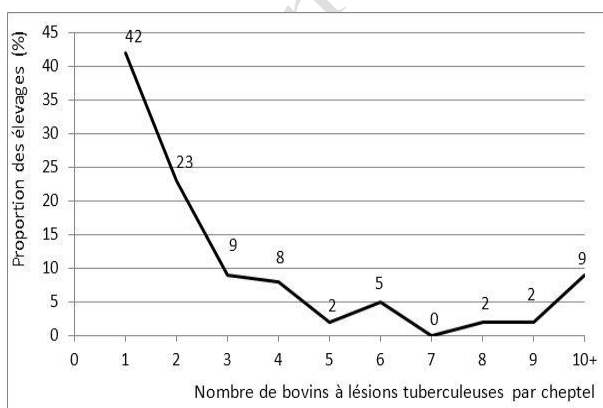
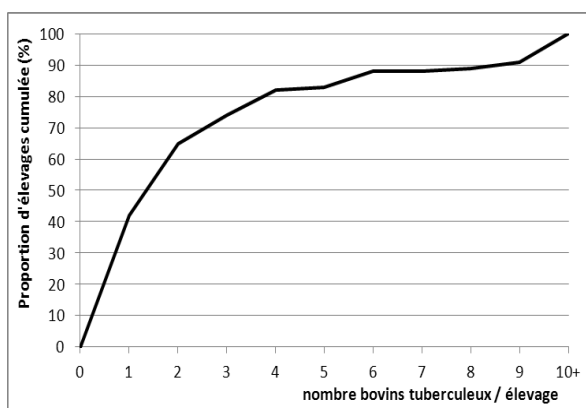


Figure 3 : Distribution du pourcentage cumulé d'élevages selon le nombre de bovins tuberculeux en Mondalie de 2002 à 2010
(Source : services vétérinaires de Mondalie)
(Abscisse 10 cumule les valeurs 10 animaux et plus).



La représentation en lignes brisées de la figure 2 a tendance à lisser les données : la perception des valeurs pour chaque abscisse est sans doute moins bonne que par un [diagramme à bâtons](#).

En revanche, la représentation de la figure 3 illustre bien l'évolution de la fréquence relative cumulée : en particulier, on perçoit plus facilement l'augmentation rapide pour les premières

abscisses, puis une relative stagnation au-delà. Ce type de courbe permet de facilement visualiser des pourcentages clés, comme par exemple, 20 %, 50 %, ou 80 %, pour des fréquences cumulées indicatrices. Dans l'exemple, les nombres de bovins tuberculeux par élevage correspondent à des valeurs différentes de fréquence cumulée : 1 bovin pour 40 %, pratiquement 2 pour plus de 60 %, et 4 pour 80 %.

Avantages

Il est possible de représenter plusieurs distributions de fréquence sur un même graphique, dont la perception peut être plus facile par rapport à d'autres modes de représentation [\[cf. Figure A.3\]](#).

Inconvénients

Le fait de joindre deux points par un segment de droite crée de l'information, en présumant qu'entre les deux points l'évolution (si elle pouvait être observée) adopterait la forme ainsi suggérée : ce type de représentation privilégie une perception de ce qui est représenté *entre les points*, et qui ne correspond à aucune réalité observée, puisque du fait de la nature de la variable, il n'y a justement pas de valeur observable entre les points reportés ; c'est donc une représentation qui privilégie déjà une interprétation, plutôt que de tenter de rendre compte du phénomène de façon la plus objective possible.

Indications et contre-indications

Ce mode de représentation convient idéalement pour la représentation de **fréquences cumulées** ; il peut être aussi utilisé pour la **représentation simultanée de plusieurs fréquences** (absolues ou relatives). Pour la représentation d'un seul jeu de fréquences, il vaut mieux utiliser le [diagramme à bâtons](#).

Il peut également convenir à la représentation de distribution d'une [variable quantitative discrète](#), ainsi que pour une [variable continue](#) mise en classes, pour répondre à des contraintes particulières de représentation (plusieurs variables) qui se prêteraient moins bien à d'autres types de représentation.

Le besoin de représenter simultanément plusieurs distributions peut conduire à utiliser le diagramme à lignes malgré ses inconvénients [\[cf. Figure A.3\]](#).

Enfin, pour la représentation de données dans le temps, qui vise à mettre en évidence une évolution, une tendance digne d'être remarquée, le diagramme à lignes est souvent choisi pour mettre en évidence ce type de tendance.

Le diagramme à [paliers](#) satisfait toutefois aux mêmes objectifs, sans dénaturer les données observées.

[Retourner au tableau de présentation des diagrammes](#)

C - DIAGRAMME A BÂTONS (en anglais : *bar chart*)

Définition

Diagramme constitué de segments verticaux en fonction d'abscisses ne pouvant prendre que des valeurs entières et dont la longueur est proportionnelle à l'effectif observé pour cette valeur, figurant en ordonnées.

Description

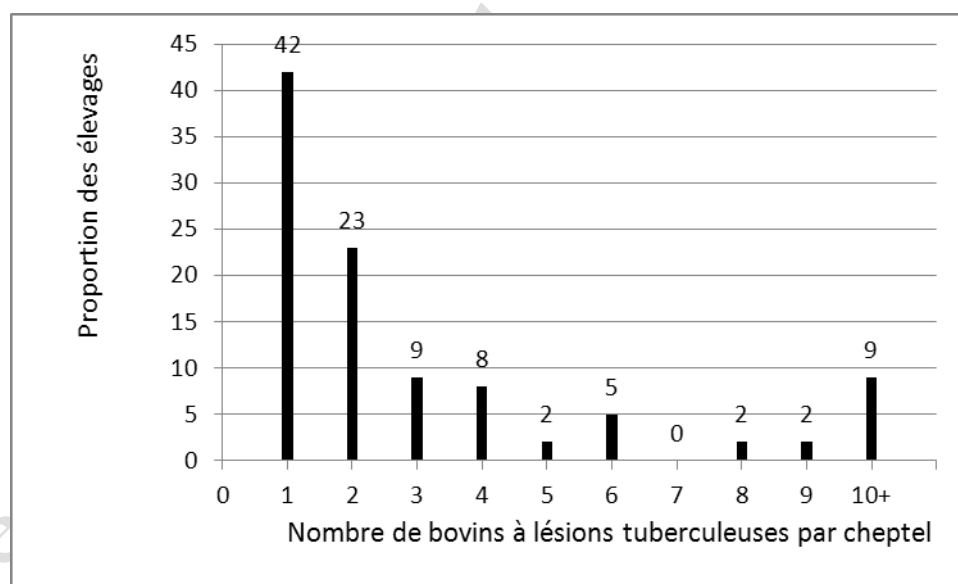
Les segments verticaux sont relativement fins, d'où le terme de bâtons. Ils sont positionnés sur la graduation de l'abscisse (*pas entre* deux graduations, comme dans un diagramme à colonnes ou dans un histogramme), de façon à rendre compte du fait que la variable en abscisse est discrète. [\[En savoir plus\]](#)

Du fait qu'il est possible d'épaissir les bâtons, certains auteurs (par exemple, les usagers de langue anglaise) ne distinguent pas les diagrammes à bâtons des diagrammes à colonnes (dénommés dans les deux cas « *bar chart* ») : en épidémiologie française, la différence repose sur le fait que les bâtons sont positionnés sur les graduations des abscisses et les colonnes entre ces graduations.

Exemple

Dans une étude portant sur le nombre de bovins tuberculeux par élevage (Figure 4), les données en abscisses sont des données entières, et en ordonnées les pourcentages d'élevages.

Figure 4 : Distribution du pourcentage d'élevages selon le nombre de bovins tuberculeux en Mondalie de 2002 à 2010 (Source : Services vétérinaires de Mondalie) (Abscisse 10 cumule les valeurs 10 animaux et plus).



Cette représentation met en évidence un premier phénomène de fréquence décroissante, pour les valeurs de 1 à 3, correspondant à pratiquement trois quarts des élevages (74 % en cumulant les pourcentages correspondants) et un deuxième phénomène, d'amplitude plus étendue, qui s'étend de 4 à 10 (et plus). Le regroupement de plusieurs valeurs d'abscisse bien plus grandes est responsable du pic observé sur la valeur 10.

[\[Retourner à la fin du chapitre « Diagramme à points »\]](#)

[\[Retourner à la fin du chapitre « Diagramme à lignes »\]](#)

Avantages

La représentation est fidèle, car en relation exacte avec la variable en abscisse, à condition d'avoir fait le nécessaire d'un point de vue informatique pour que les valeurs soient correctement placées sur les graduations. [\[En savoir plus\]](#).

Inconvénients

La perception de simples bâtons peut ne pas être optimale : on peut être amené à donner [suffisamment d'épaisseur](#) à chaque bâton pour la rendre plus satisfaisante.

Indications et contre-indications

Ce type de diagramme est classiquement utilisé pour représenter les distributions de fréquence ou de probabilités.

Il est également indiqué pour la représentation de fréquences absolues (effectifs) ou relatives (fréquences) d'une distribution d'individus pour une variable ne pouvant prendre que des valeurs entières (variable discrète), en nombre limité (au maximum une dizaine).

Si le nombre de valeurs possibles est plus élevé, on peut être amené à choisir un autre mode de représentation en raison des regroupements par classes que l'on peut être amené à réaliser : préférer dans ce cas le [diagramme à colonnes](#).

Ce type de représentation convient moins bien si l'on veut représenter simultanément plusieurs séries, qu'il est alors difficile de distinguer les unes des autres : on peut alors préférer le [diagramme à lignes](#).

[Retourner au tableau de présentation des diagrammes](#)

[Retourner au chapitre "Diagramme à lignes"](#)

D - DIAGRAMME À COLONNES (en anglais : *bar chart*)

Synonymes : diagramme en colonnes, ou diagramme colonnes.

[\[En savoir plus sur la justification de cette dénomination\]](#)

Définition

Diagramme constitué de colonnes **non contiguës**, comportant en abscisses les classes d'une **variable qualitative ordinale** et en ordonnées les valeurs de la distribution (effectifs ou proportion).

Description

Pour chaque classe, la hauteur de la colonne est proportionnelle à l'effectif de la classe.

Les colonnes sont placées entre deux graduations. Elles peuvent avoir une épaisseur variable, selon le besoin pour une perception satisfaisante. [\[En savoir plus sur la distinction entre diagrammes à colonnes et diagrammes à bâtons\]](#)

L'axe horizontal suggère une augmentation progressive de l'intensité de la variable représentée, même si la relation n'est pas strictement mathématique : implicitement, tout ce qui est à droite est perçu comme « plus », par rapport à ce qui est à gauche.

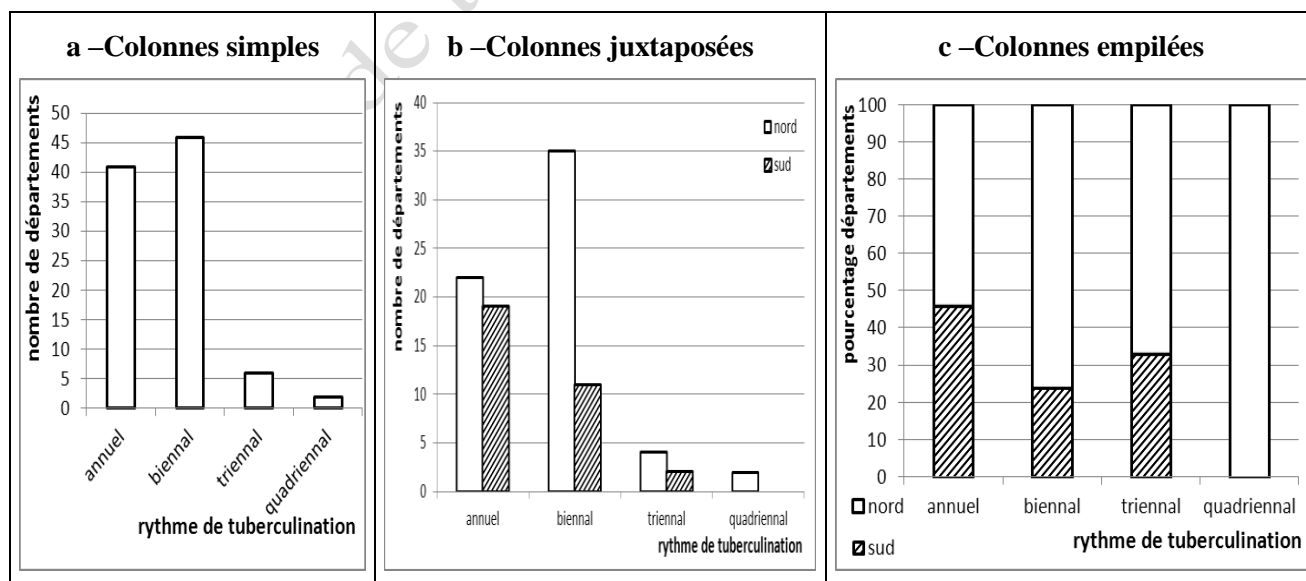
Une étiquette avec les effectifs peut être placée au-dessus de chaque colonne.

Pour représenter plus d'une série de données, deux modalités de représentation des séries de données sont possibles : juxtaposées, empilées (voir les exemples suivants).

Exemples

La figure 5 illustre les modalités de diagramme à colonnes (simples, juxtaposées, empilées) par la représentation de la distribution des départements selon leur rythme de dépistage de la tuberculose bovine en France en 1989.

Figure 5 : Distribution des départements selon le rythme de dépistage de la tuberculose bovine en France en 1989 : (a) toute la France ; (b) et (c) : en fonction de la situation géographique des départements, nord ou sud, par rapport à la ligne Bordeaux-Annecy (Source : Données DGA).



Le total de 100 % est obtenu, pour les colonnes juxtaposées par sommation des valeurs d'une série (ici, nord ou sud), et pour les colonnes empilées, par sommation des valeurs d'une modalité (ici, annuel, biennal, triennal, quadriennal).

Les différentes séries de données peuvent être représentées de diverses façons selon les objectifs poursuivis. [\[En savoir plus\]](#)

Avantages

La perception est meilleure qu'avec un diagramme en bâtons, du fait de l'épaisseur de la colonne.

Inconvénients

La perception peut être difficile pour la représentation simultanée de plus de deux séries de données.

Si l'on représente plusieurs séries afin de les comparer, il est indispensable d'utiliser des pourcentages afin de les rendre comparables malgré leur différence éventuelle d'effectifs : dans le cas des colonnes juxtaposées, l'utilisation des données brutes peut donner une figure d'aspect différent de celle obtenue avec les pourcentages effectués sur le total d'une série ; il faut aussi faire attention à la base sur laquelle sont constitués les pourcentages servant à la représentation, car le tableur peut faire des choix par défaut ne correspondant pas aux objectifs poursuivis (soit sur le total de la série, soit par catégorie selon la configuration des données). [\[En savoir plus\]](#)

Indications et contre-indications

Le diagramme à colonnes est classiquement utilisé pour la représentation des variables **qualitatives ordinales**.

Il est aussi utilisé pour les **variables quantitatives discrètes**, lorsque le nombre de valeurs possibles est trop élevé pour un diagramme à bâtons et nécessite un **regroupement en classes** [\[En savoir plus\]](#) ou lorsque le traitement (mise en classes) altère le caractère quantitatif de la variable en abscisses [\[En savoir plus\]](#)

Le **regroupement de données d'une variable discrète** ne doit surtout pas être effectué avec des classes d'amplitude inégale, car cette opération dénature la relation implicite de l'axe des abscisses. Si l'on souhaite conserver le caractère quantitatif de l'axe des abscisses tout en procédant à ces regroupements d'effectifs, le diagramme à colonnes ne convient pas [\[En savoir plus\]](#). Si, compte tenu de ces regroupements, on veut conserver malgré tout la forme de diagramme à colonnes, il faut parvenir à transformer la variable de façon à ce qu'elle soit perçue comme ordinale (et non comme quantitative) ([voir un exemple](#)), sinon il faudra trouver une autre solution (histogramme par exemple).

Le diagramme à colonnes est aussi utilisé pour la **description dans le temps** de phénomènes de santé.

En revanche, on ne doit pas utiliser les diagrammes à colonnes pour des **variables qualitatives nominales**, car l'axe horizontal suggère trop fortement une idée d'ordre, de classement, de degré d'intensité que ne comporte pas une variable nominale et qui peut même nuire à une perception correcte, voire heurter la sensibilité de certains. Pour les variables nominales, on préfère utiliser les diagrammes à barres.

Par exemple, une représentation par un diagramme à colonnes de la distribution d'une variable relative à des catégories selon le sexe (hommes / femmes), selon la nationalité ou l'ethnie susciterait inévitablement des réactions de la part de personnes pouvant s'estimer atteintes par le classement implicite suggéré par ce type de diagramme.

[\[Retourner au tableau de présentation des diagrammes\]](#)

[\[Retourner à la fin du chapitre "Diagrammes à lignes"\]](#)

[\[Retourner à la fin du chapitre "Diagrammes à bâtons"\]](#)

E - DIAGRAMME À BARRES (en anglais : *bar chart*)

Synonymes : diagramme à bandes, à **barres horizontales**.

Définition

Diagramme constitué d'éléments horizontaux (barres) **non contigus**, comportant en ordonnées les états d'une **variable qualitative nominale** et en abscisses les effectifs de la distribution.

Description

Le diagramme comporte sur l'axe vertical les intitulés des différentes classes de la variable servant au dénombrement. Ces intitulés sont écrits horizontalement, par conséquent, selon le sens naturel de l'écriture.

Il comporte, selon l'axe horizontal, des barres dont la longueur est proportionnelle à l'effectif de la classe. Toutes les barres ayant la même largeur, la surface est également proportionnelle à l'effectif de la classe.

On peut représenter soit des effectifs, soit des pourcentages.

Une étiquette avec la valeur peut être placée auprès de chaque barre.

Plusieurs choix sont possibles pour l'ordre des classes de la variable nominale, dont le choix peut constituer une interprétation. L'ordre alphabétique est la solution la plus simple, en l'absence d'autre recours. L'ordre peut être dicté par une organisation sous-jacente (par exemple, catégories socio-professionnelles qui font l'objet d'une codification numérique permettant le classement).

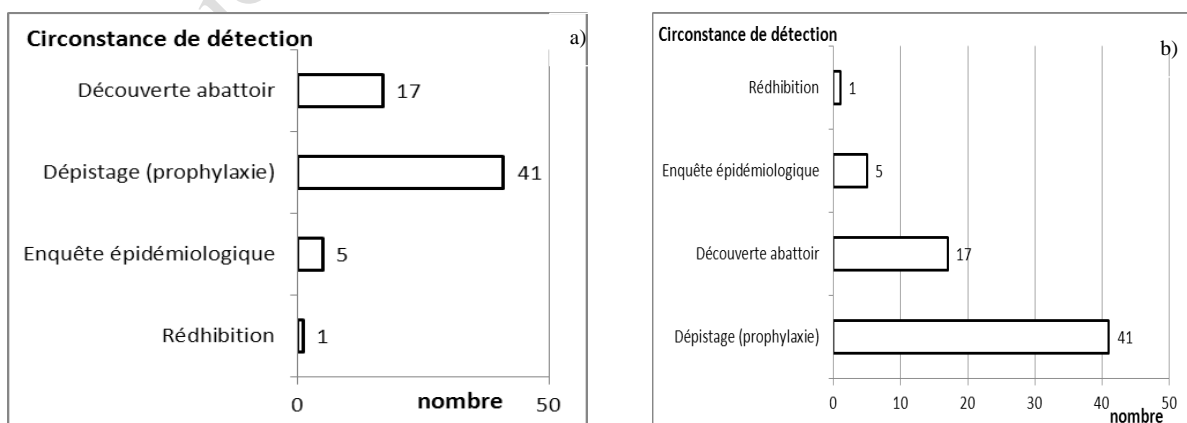
La représentation selon un mode ordonné en fonction des effectifs (croissants ou décroissants) peut permettre de mettre en valeur, le cas échéant, une relation intéressante connue sous le nom de « *distribution de Pareto* » ou tout autre rapport de proportions reflétant un déséquilibre suggestif.

Pour représenter plus d'une série de données, deux modalités de représentation des séries de données sont possibles : juxtaposées, empilées horizontalement (voir exemples, figure 7).

Exemple 1

La Figure 6 montre un exemple de diagramme à barres, dont les classes de la variable sont ordonnées de deux façons : selon un ordre alphabétique (a) et selon l'ordre croissant des effectifs (b).

Figure 6 : Distribution du nombre d'élevages tuberculeux (foyers cumulés de 2001 à 2010) en Mondalie selon les circonstances de détection de 64 foyers (Source : services vétérinaires de Mondalie).

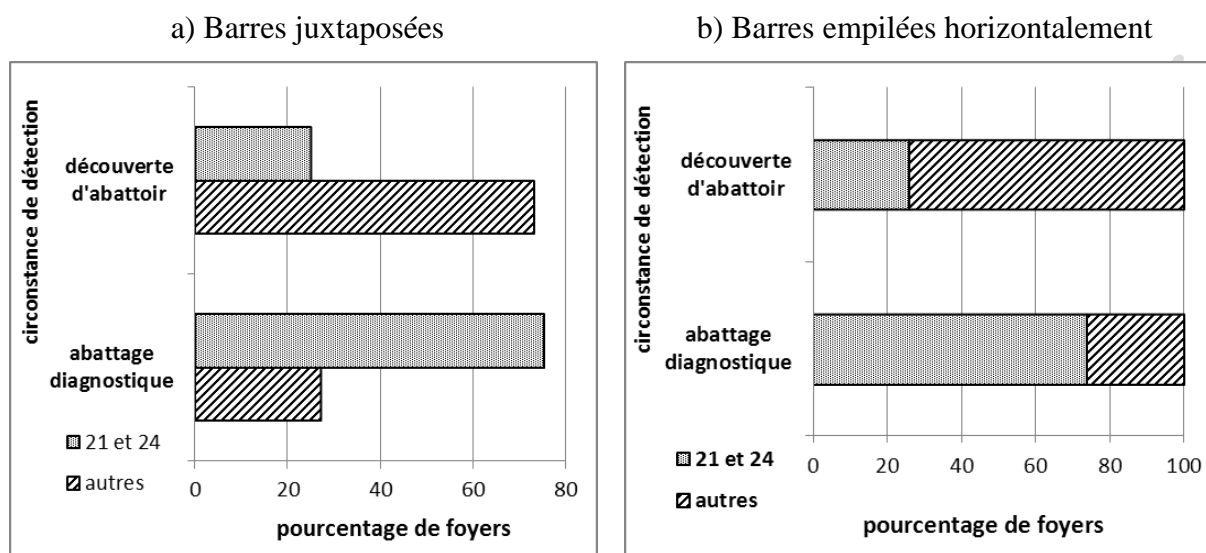


Le choix d'un classement selon les valeurs de fréquence observée (figure b) rend plus facile la perception qu'un classement selon l'ordre alphabétique des modalités (figure a).

Exemple 2

La figure 7 compare les données de deux groupes de départements (Côte-d'Or et Dordogne, d'une part et les autres départements, d'autre part), selon le mode de découverte de la tuberculose.

Figure 7 : Distribution du nombre d'élevages tuberculeux (foyers cumulés de 2005 à 2007) en France (21 : Côte-d'Or, n = 83 ; 24 : Dordogne, n = 83) selon le mode de détection des foyers (Du Breil, 2009).



Pour faciliter la comparaison, les données ont été transformées en pourcentages. La figure a représente la distribution pour chaque série géographique ; les données brutes pouvant être d'effectifs différents pour chaque série, les pourcentages ont été utilisés pour la représentation et ont été calculés sur le total de chaque série pour faciliter la comparaison. La figure b vise à comparer la contribution *relative* de chaque série géographique : les pourcentages ont été effectués pour chaque modalité de détection des foyers.

Avantages

Ce type de diagramme convient idéalement pour les graphiques dont les intitulés des différentes classes de la variable sont **relativement longs** et peuvent ainsi être écrits sans difficulté, ce qui facilite leur lecture.

Inconvénients

Comme pour le diagramme à colonnes, la superposition des valeurs de plusieurs séries peut ne pas être facilement lisible si les distributions sont relativement voisines.

Il faut faire attention à la base sur laquelle sont calculés les pourcentages servant à la représentation dans le cas des barres empilées horizontalement, le tableur pouvant faire des choix par défaut ne correspondant pas aux objectifs poursuivis. Les remarques formulées à ce sujet à propos des diagrammes à colonnes sont également valides. [\[Voir le développement correspondant au sujet des diagrammes colonnes\]](#)

Indications et contre-indications

Ce mode de représentation est indiqué pour des **variables qualitatives nominales**, tout particulièrement lorsque leur libellé est long.

[Retourner au tableau de présentation des diagrammes](#)

F - DIAGRAMME CIRCULAIRE (en anglais : *pie chart*)

Synonymes : diagramme en camembert ou diagramme à secteurs.

Définition

Diagramme constitué d'un cercle, divisé en secteurs représentant les classes d'une **variable qualitative nominale** et dont la surface est proportionnelle à la fréquence de l'effectif des classes.

Description

Le diagramme est composé d'une surface circulaire, divisée en autant de secteurs que de classes que comporte la variable nominale étudiée. Les libellés sont rédigés dans une légende séparée.

Les secteurs sont proportionnels aux effectifs, que ce soit par la longueur de l'arc ou par la surface du secteur.

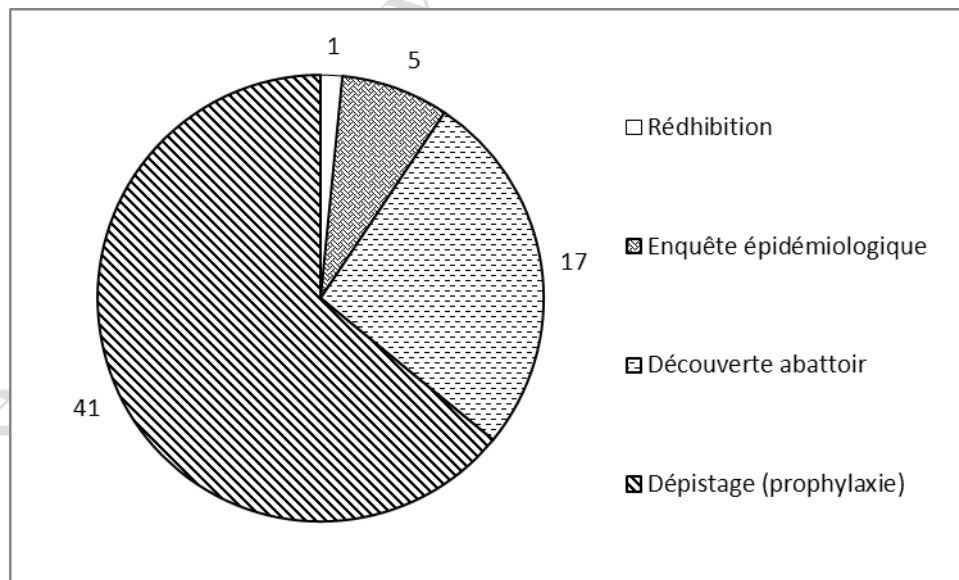
Des étiquettes peuvent être ajoutées pour indiquer les effectifs (ou les pourcentages, mais il n'y a aucune différence de représentation par rapport à l'utilisation des effectifs).

Pour faciliter la lecture d'un tel diagramme, il est recommandé de disposer la légende des secteurs du haut vers le bas en commençant à l'endroit correspondant à midi et en tournant dans le sens des aiguilles d'une montre.

Exemple

La figure 8 illustre une représentation par un diagramme circulaire. On pourra comparer avec la figure 6 (diagramme à barres) qui correspond au même jeu de données.

Figure 8 : Distribution du nombre d'élevages tuberculeux (foyers cumulés de 2005 à 2007) en Mondalie selon les circonstances de détection de 64 foyers (Source : services vétérinaires de Mondalie).



Les données de la figure 8 sont les mêmes que celles de la figure 6 (diagramme à barres). Même en couleurs, la représentation ne permettrait pas de distinguer aussi bien que sur le diagramme à barres les rapports d'échelle entre les différentes valeurs. Comme dans ce cas, il n'existe aucune raison impérieuse de ne pas respecter un ordre des variables plutôt qu'un autre, on n'a donc aucune raison de choisir ce mode de représentation et on choisira plutôt le diagramme à barres.

Avantages

Un cercle, en théorie, ne comporte pas de début, pas de fin, donc pas d'ordre possible, donc pas de suggestion possible d'une échelle de valeurs. Toutefois, en pratique la lecture s'effectue en partant de midi, dans le sens des aiguilles d'une montre, ce qui fait que l'avantage est tout à fait virtuel.

Inconvénients

Le premier inconvénient majeur de ce type de représentation est que les libellés des classes figurent dans une légende séparée qui procure moins d'espace que dans le diagramme à barres, alors que les variables nominales peuvent nécessiter un espace important pour leur rédaction, au point qu'on peut en être amené à ne placer que des abréviations figurant dans la légende placée dans la figure, qui doivent être explicitées dans une légende placée sous la figure. De ce fait, la correspondance entre le libellé et sa représentation est moins immédiate qu'avec un diagramme à barres.

Un autre inconvénient de ce type de représentation est que l'œil humain perçoit difficilement les différences de taille des surfaces angulaires, tandis qu'il perçoit bien mieux les différences de rapports entre des longueurs. Par conséquent, d'une façon générale, il vaudra mieux utiliser un diagramme à barres, sauf dans les cas où le choix de l'ordre des catégories pourrait suggérer implicitement une échelle de valeurs risquant d'affecter l'objectivité de la représentation.

Indications et contre-indications

Le diagramme circulaire peut être utilisé pour les **variables qualitatives nominales**, à la place d'un diagramme à barres, à la condition que le nombre de classes soit limité (au maximum cinq ou six), et pour satisfaire à un souci de **ne pas suggérer une idée de hiérarchie de valeur** entre les catégories (par exemple, selon le sexe).

En dehors de cette indication particulière, mieux vaut se dispenser d'utiliser le diagramme circulaire et préférer le **diagramme à barres**, qui comporte bien plus d'avantages.

Le diagramme circulaire ne convient pas pour une **variable ordinale**, pour laquelle, justement, la notion de classement est indispensable.

[Retourner au tableau de présentation des diagrammes](#)

G - QUELQUES AUTRES DIAGRAMMES

1. DIAGRAMME A BOITES (en anglais : *box plot*)

Synonyme : diagramme en **moustaches** (box and **whiskers** plot), diagramme de **Tukey**.

Définition

Diagramme résumant diverses caractéristiques d'une variable quantitative : médiane, quartiles, minimum, maximum.

Description

Un rectangle (« boîte ») représente chacun des deuxième et troisième quartiles d'une variable ; ils sont réunis par la médiane de la distribution. De chaque côté de ces rectangles sont figurés des segments qui s'étendent jusqu'aux valeurs extrêmes (minimale et maximale).

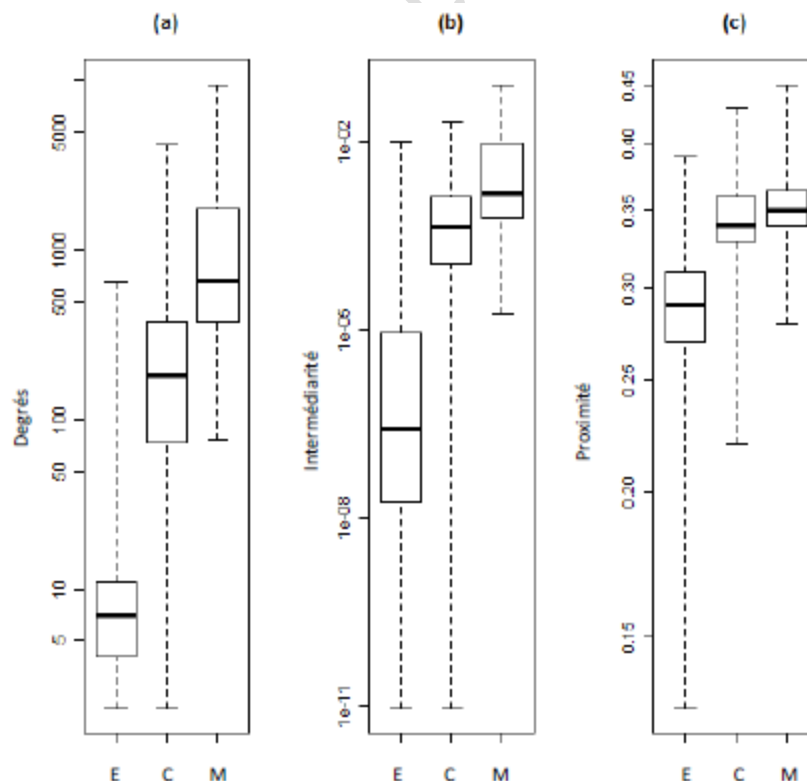
La représentation originelle était horizontale, ce qui a justifié sa dénomination (moustaches, les deux segments extérieurs figurant les « moustaches »). Mais l'usage actuel l'a consacré avec une orientation verticale beaucoup plus fréquente, bien que les deux orientations soient tout à fait possibles.

Exemple

L'usage des diagrammes à boîtes est illustré par la figure 9.

Figure 9 : Mesures de centralité des établissements du réseau d'échange bovin (Rautureau S, 2012).

Mesures de centralité (a : degré, b : intermédianité, c : proximité) pour les élevages (E), les centres de rassemblement (C) et les marchés (M).



Avantages

Cette représentation, utilisant la médiane plutôt que la moyenne, est la description la plus objective d'une distribution : la moyenne n'a en effet de validité que sous condition que la distribution suive une loi normale et ne reflète pas suffisamment les caractéristiques de distribution, en particulier les valeurs extrêmes.

Inconvénients

Il est indispensable de systématiquement rappeler que le trait qui apparaît à l'intérieur de la boîte est la médiane, et non la moyenne, comme pourrait le penser spontanément un usager insuffisamment familiarisé avec cet outil : la moyenne est en effet une notion statistique plus présente dans l'esprit des profanes que la médiane.

Indications et contre-indications

Ce type de graphique est tout indiqué pour représenter de façon graphique les valeurs de paramètres dits de position d'une distribution d'une variable **quantitative**.

Version de travail. Ne pas diffuser

2. PYRAMIDE DES ÂGES (en anglais : *age pyramid* ou *population pyramid*)

Définition

Diagramme représentant la structure par âge d'une population à un instant donné, selon un critère dichotomique (sexe par exemple).

Description

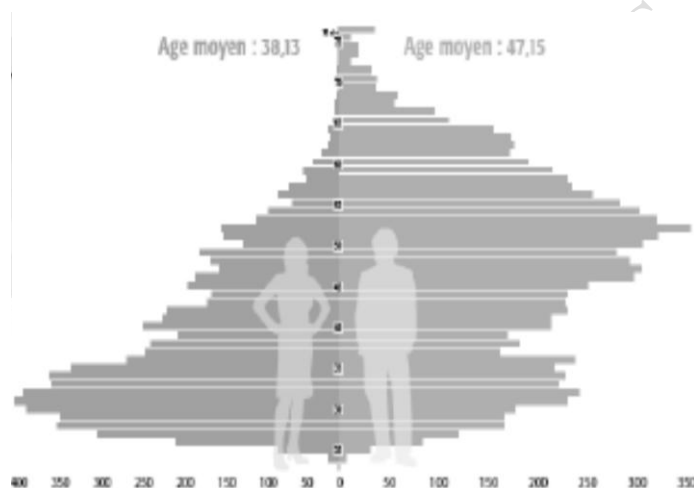
La pyramide des âges comporte en abscisses les fréquences (ou les effectifs), disposées selon deux axes opposés en miroir, portant chacun une modalité de la répartition dichotomique de la population selon une variable d'intérêt (le sexe en médecine humaine).

Les données sont représentées par deux histogrammes juxtaposés selon un axe vertical où sont reportés les âges.

Exemples

La figure 10 représente la pyramide des âges de la profession vétérinaire en 2012.

Figure 10 : Distribution des effectifs de vétérinaires inscrits au Conseil de l'Ordre des vétérinaires en France en 2012 (source : rapport annuel 2012 du CSO).



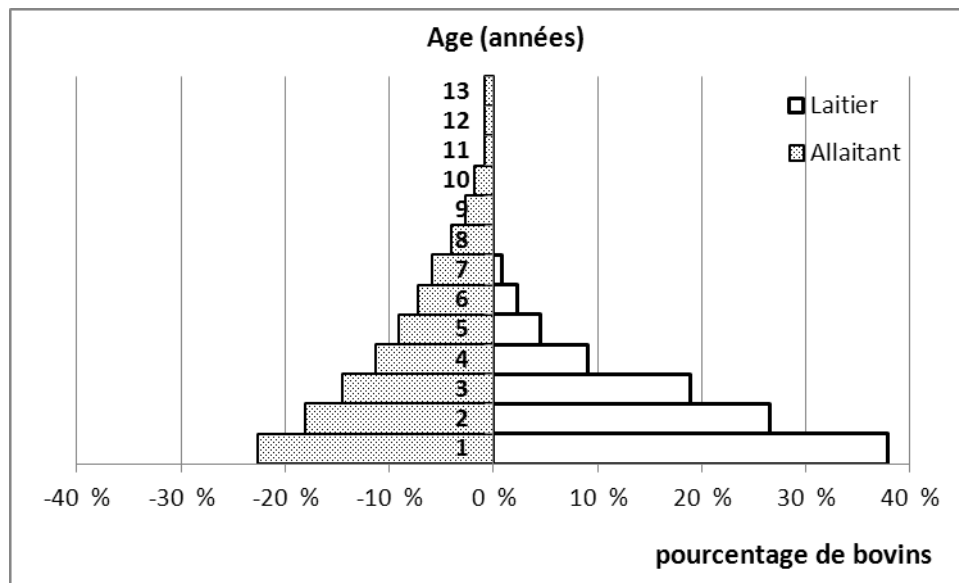
Cette pyramide montre bien l'évolution démographique de la population des vétérinaires français, où la prédominance passée de vétérinaires de sexe masculin, illustrée par leur plus grande fréquence chez les vétérinaires âgés, fait place à celle des femmes, plus fréquentes chez les jeunes diplômés.

La figure 11 représentant la pyramide des âges d'une population de bovins selon le type de production (laitier, allaitant), illustre la différence démographique de ces deux sous-populations.

La représentation de cette figure a été obtenue avec Excel, au prix de manipulations laborieuses, qui conduisent à ne pas pouvoir éviter le signe négatif pour les pourcentages de l'axe des abscisses situés à gauche de la valeur zéro.

Les effectifs de bovins laitiers et allaitants étant notablement différents (447 pour les bovins allaitants et 221 pour les bovins laitiers), la représentation en fréquence était indispensable pour rendre la comparaison possible. Elle montre la différence, les bovins laitiers étant plus souvent d'âge relativement jeune, les bovins allaitants pouvant être beaucoup plus âgés.

Figure 11 : Distribution de fréquence des bovins, allaitants ou laitiers, selon leur âge
(source : services vétérinaires de Mondalie).



Avantages

Ce diagramme facilite la perception de la différence de la structure de deux groupes d'une population, par une figure qui combine la représentation des effectifs ou de leur fréquence par une barre, l'opposition en miroir facilitant la comparaison.

Inconvénients

Ce type de diagramme n'est vraiment pas facile à réaliser avec Excel® ! Il faut utiliser pour cela les aides en ligne qui sont heureusement suffisamment explicites, mais nécessitent quand même pas mal de persévérance. Sinon, un logiciel spécialisé fait très facilement ce type de graphique !

Indications et contre-indications

Ce diagramme convient pour la représentation de la structure d'âge d'une population divisée en deux classes (sexe, ou autre). Selon les buts poursuivis, on peut utiliser soit des effectifs, soit des fréquences propres à chaque classe.

Ne convient pas si la variable étudiée comporte plus de deux classes.

3. HISTOGRAMME (en anglais : *histogram*)

Définition

Diagramme représentant la distribution de fréquence d'une **variable quantitative continue** par des rectangles verticaux **contigus** de **surface proportionnelle** aux effectifs.

Remarques :

Etymologiquement, le terme histogramme renvoie à la notion de tissu ou encore de voile (de bateau), soulignant ainsi la caractéristique fondamentale de ce type de figure, la *surface*.

Un histogramme **n'est pas** un diagramme à colonnes dont les colonnes sont contiguës. Cette confusion est pourtant très courante : le tableur Excel regroupe d'ailleurs dans son menu, sous l'étiquette de la catégorie « Colonne », diverses formes (histogrammes, cylindres, coniques, pyramidales, *etc.*), démontrant bien l'assimilation erronée faite par ce logiciel du diagramme à colonnes et de l'histogramme.

Son usage en épidémiologie étant beaucoup moins fréquent que les diagrammes à colonnes et sa réalisation nécessitant une maîtrise nettement supérieure à celle du public visé, nous avons choisi de lui donner la place secondaire actuelle.

Description

Les colonnes sont jointives, reflétant la continuité de distribution des valeurs.

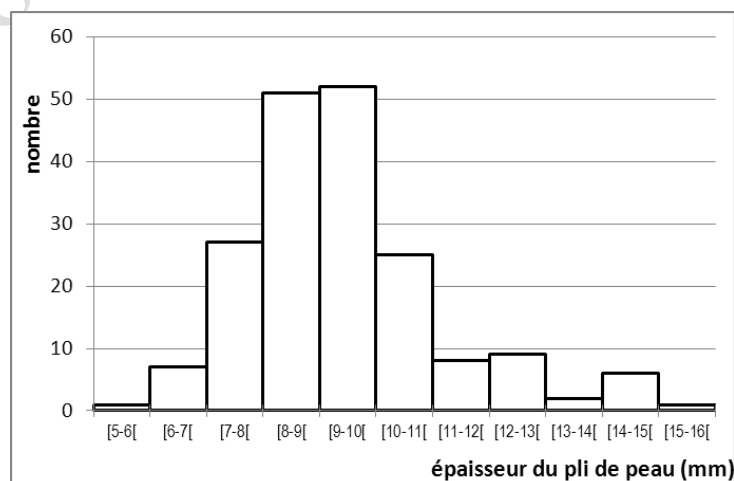
Les valeurs en abscisses sont délimitées par les bornes des intervalles de classes. Toutes les valeurs de la variable sont possibles (« **variable continue** »).

En ordonnées, la hauteur de la colonne est proportionnelle aux effectifs uniquement dans le cas particulier où les intervalles sont de même amplitude (« **histogramme régulier** »). Si certains intervalles sont d'amplitude différente, la représentation doit être adaptée pour en tenir compte et respecter le principe de proportionnalité vis-à-vis de la surface en transformant les effectifs en fréquence ($f_i = n_i/n$). [\[En savoir plus\]](#)

Le choix du nombre de classes, du fait qu'il détermine les valeurs en abscisses, peut avoir une répercussion importante sur l'allure finale du graphique : un nombre de classes trop élevé ou trop faible ne permet pas de révéler une structure intéressante. [\[En savoir plus\]](#)

Exemple

Figure 12 : Distribution du nombre de bovins charolais selon l'épaisseur de leur pli de peau (N = 189) (GANE, 2011).



[\[En savoir plus sur la manière de faire un histogramme conforme avec Excel\]](#)

Avantages

La perception visuelle d'un histogramme est fondée sur la surface délimitée, à comparer avec un diagramme à lignes. [\[Comparer Histogramme et diagramme à lignes\]](#)

La contiguïté des bandes illustre la continuité de la variable, tandis qu'un diagramme à colonnes souligne le caractère discret des valeurs.

Si les amplitudes des classes sont égales, la figure peut être traitée comme un diagramme à colonnes (donc facilement réalisable avec un tableur comme d'Excel[®]), dont on élargit les colonnes pour qu'elles soient jointives, ce qui permet de combiner la perception à la fois de la hauteur de la colonne (dans ce cas particulier les effectifs correspondent à la fréquence) et de la surface.

Inconvénients

Dans un histogramme véritable, la surface de chaque rectangle doit être proportionnelle à la fréquence des effectifs : si les amplitudes de classe ne sont pas égales, il faut alors transformer l'ordonnée en fonction de ce principe pour ne pas déformer la représentation. Seuls des logiciels adaptés peuvent réaliser facilement cette opération.

L'histogramme est un type de diagramme de conception fondamentalement différente des autres types qui montrent la distribution des effectifs, base même de la statistique appliquée à l'épidémiologie descriptive qui procède par dénombrements et leur représentation directe par une longueur. L'histogramme véritable est plus difficile à réaliser, mais aussi moins facilement accessible à la compréhension.

Indications et contre-indications

Un histogramme est adapté *a priori* pour représenter la distribution des fréquences d'une **variable quantitative continue** : les valeurs possibles étant infinies, il faut au préalable les regrouper en classes, ce qui permet de déterminer les effectifs correspondants.

Par extension, il peut aussi être utilisé pour une **variable quantitative discrète** lorsque le nombre de valeurs est relativement élevé, ce qui conduit à devoir également les regrouper en classes. [\[En savoir plus sur les indications de l'histogramme »\]](#)

[Retourner au tableau de présentation des diagrammes](#)

4. DIAGRAMME EN RADAR (en anglais : *radar chart*)

Synonymes : graphique de Kiviat, en toile d'araignée, en étoile.

Définition

Diagramme permettant de représenter simultanément la distribution de fréquence selon **plusieurs variables ordinales** grâce à différents axes partant d'un point central.

Description

La figure comporte autant d'axes que de variables étudiées, tous partant d'un même point central. Les unités des axes doivent être standardisées de façon à rendre possible la comparaison.

Une seule série peut être représentée (diagramme simple), ou plusieurs (diagramme complexe).

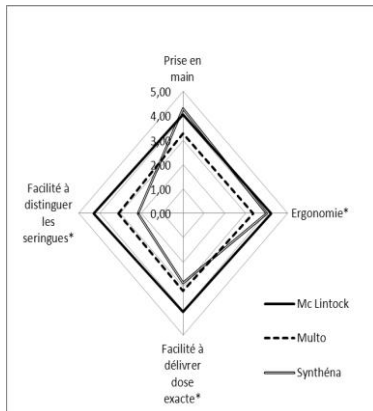
Exemples

Le diagramme représentant l'intensité des vents à un lieu géographique donné pendant une certaine période est l'exemple le plus connu (« rose des vents »).

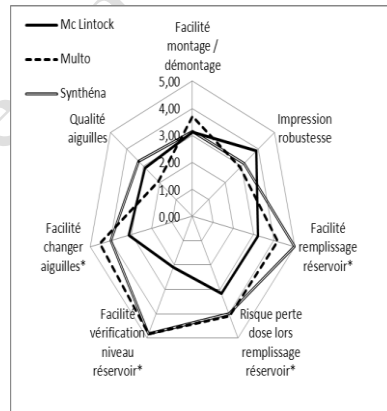
La figure 13 montre trois exemples de diagrammes en radar, comportant 4, 7 et 3 axes. Trente-huit vétérinaires ont donné leur avis sur la comparaison de trois seringues à tuberculiner, dans le cadre du dépistage de la tuberculose bovine : quatre questions portaient sur la commodité d'usage du matériel (a), sept sur la qualité du matériel (b) et trois sur la qualité des injections (c). Les vétérinaires devaient donner leur avis selon une échelle de Likert comportant cinq valeurs.

Figure 13 : Comparaison de trois modèles de seringues à tuberculiner [Bénet et Dufour, 2014].

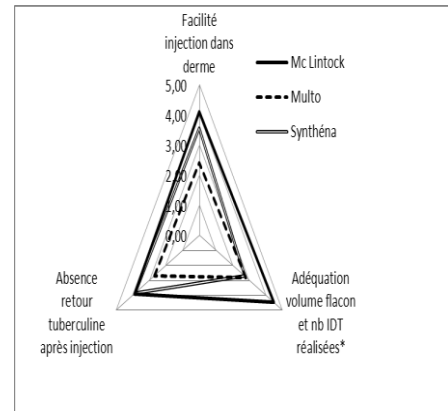
a) Commodité d'usage du matériel



b) Qualité du matériel



c) Qualité des injections



Moyenne des évaluations (échelle de 1 à 5) pour chaque critère. * = écart significatif (seuil de 0,05).

Pour la commodité d'usage du matériel, qui comporte quatre axes, on voit nettement que la seringue Mc Lintock a été mieux appréciée que les deux autres. Pour la qualité du matériel, la Synthéna et la Multo ont des résultats assez voisins, supérieurs à la Mc Lintock. Enfin, la qualité des injections montre à nouveau une supériorité des avis donnés par les vétérinaires enquêtés pour la seringue Mc Lintock.

Avantages

Ce type de diagramme a le double avantage de solliciter la perception par la ligne, qui délimite le contour, et par la surface, du fait que la forme tracée par la ligne est fermée. Il est donc possible de voir d'un seul coup d'œil si une des séries comporte plus (ou moins) de surface que les autres et quelles sont les variables pour lesquelles il peut y avoir des différences par rapport à cette perception d'ensemble.

La réalisation de ce type de diagramme ne pose aucune difficulté avec les tableurs.

Il est possible de représenter plusieurs séries à comparer.

Inconvénients

Les différents axes doivent comporter des unités « comparables », c'est-à-dire de même étendue.

La perception de la représentation de la valeur d'une variable est dépendante de celle des deux variables voisines.

Si la perception des diagrammes radars simples est facile, elle peut être plus difficile pour les diagrammes radars complexes (représentant plusieurs séries).

Indications et contre-indications

Ce type de diagramme peut convenir pour les **variables qualitatives ordinales** de séries que l'on souhaite comparer d'un seul coup d'œil.

Il peut également convenir pour des **variables quantitatives continues ou discrètes**.

La condition toutefois est que les différentes variables doivent être d'une **amplitude comparable**, ou avoir été **standardisées** pour ce faire.

Il ne convient pas si les variables ne respectent pas cette contrainte, ou si leur transformation ne permet pas de la satisfaire.

Il peut ne pas convenir non plus si les relations entre les différentes séries sont complexes, ne permettant pas de dégager facilement une opinion sur chacune d'elles.

5. DIAGRAMME À PALIERS (en anglais : *stairs diagram*)

Synonymes : Diagramme de survie, diagramme en escalier.

Définition

Diagramme comportant des segments horizontaux (« paliers ») pour représenter chaque point d'une série situé entre les graduations des abscisses, joints entre eux par des segments verticaux, à chaque changement de palier, pour assurer la continuité du tracé.

Remarque

L'expression de diagramme en escalier ne convient que pour les séries de données monotones (ascendantes ou descendantes, ce qui est toujours le cas pour des probabilités cumulées), car un escalier ne peut à la fois monter et descendre dans la même volée de marches sans passer par un palier : si la série évolue en alternant montée et descente, le terme de palier convient donc mieux, car il ne préjuge aucunement du sens.

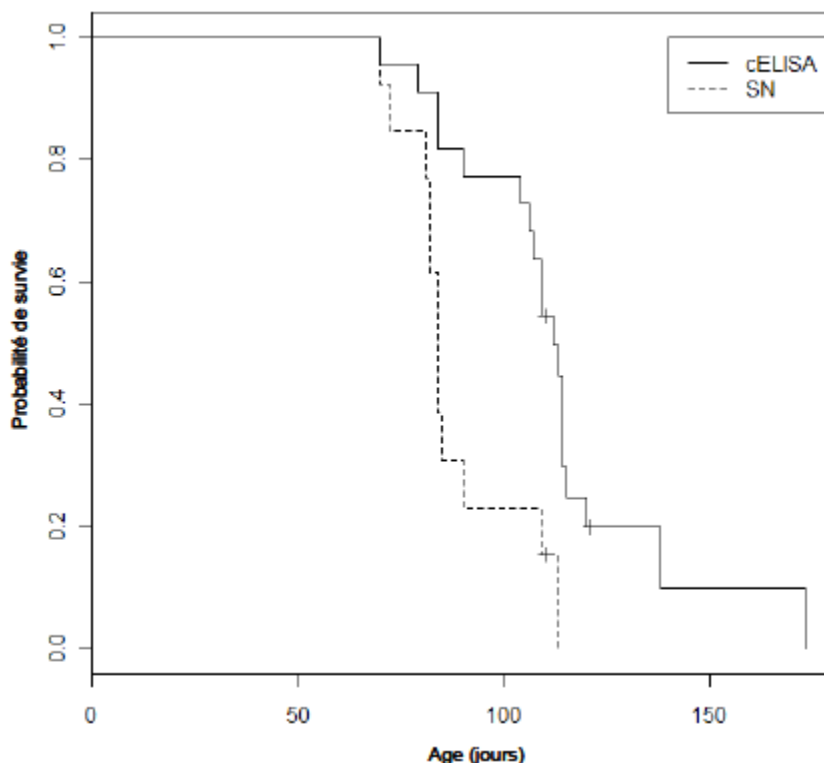
Description

Ce tracé apparaît sous la forme du contour extérieur d'un diagramme à colonnes.

Exemples

Les courbes de survie constituent l'exemple actuellement le plus connu (cf. figure 14).

Figure 14 : Courbes de survie de Kaplan-Meier estimant le délai requis pour obtenir une disparitions des anticorps maternels spécifiques du BTV parmi 22 veaux avec les tests cELISA ou SN (Vitour *et al.*, 2010)



Avantages

Le trait horizontal matérialisant les coordonnées du point a l'avantage de le représenter de la façon la plus objective en n'introduisant aucune interprétation comme le fait le diagramme à lignes, qui crée de l'information en joignant les points entre eux.

Inconvénients

La réalisation de ce diagramme, facile avec des logiciels statistiques, demande quelques efforts mais il est [réalisable avec Excel](#)[®] : cette servitude se résout aisément en enregistrant le premier graphique ainsi réalisé en tant que modèle.

Indications et contre-indications

L'indication majeure de ce type de diagramme est la représentation des probabilités cumulées, en usage pour les courbes de survie par exemple.

De fait, cette ligne brisée, mais continue, suggère une évolution.

Pour ces deux raisons, le diagramme à paliers pourrait aussi être utilisé pour la description de données d'évolution dans le temps d'un épisode épidémique, en raison de la meilleure objectivité procurée ([à comparer avec les diagrammes à lignes et à colonnes](#)).

Pour cette indication, il permet de représenter plusieurs séries et le résultat serait à comparer avec des diagrammes à colonnes juxtaposés pour choisir le plus lisible.

Le diagramme à paliers n'est pas à recommander pour une représentation de distribution de fréquence de population ; le diagramme à colonnes convient mieux.

[Retour au diagramme à lignes](#)

[Retour au tableau de choix des diagrammes](#)

III - SCHÉMAS

Définition

Un **schéma** est une représentation graphique, simplifiée, rationnelle et organisée d'une réalité plus ou moins complexe, destinée à faire comprendre sa conformation et/ou son fonctionnement.

« *Un schéma vaut mieux qu'un long discours* »². Certes, sous certaines conditions toutefois !

Objectifs

Les objectifs découlent de la définition :

- Un schéma vise à une simplification ;
- Un schéma sert à organiser des informations ;
- Un schéma doit faciliter la compréhension ;
- Un schéma est un complément de l'expression écrite.

Types de schémas et exemples

Quelques types sont présentés en les distinguant selon leur forme propre et selon leur degré de conceptualisation (figure 15).

• **Selon leur forme propre**

- Schéma linéaire : mettant en relation selon un axe (horizontal ou vertical) des éléments reliés entre eux par des traits ou des flèches (figure 15a) ;
- Organigramme : mettant en relation différents éléments selon une logique d'organisation fonctionnelle (figure 15b) ;
- Arborescence (figure 15c) ;
- Schéma circulaire : la roue de Deming, un cycle épidémiologique (figure 15d)...

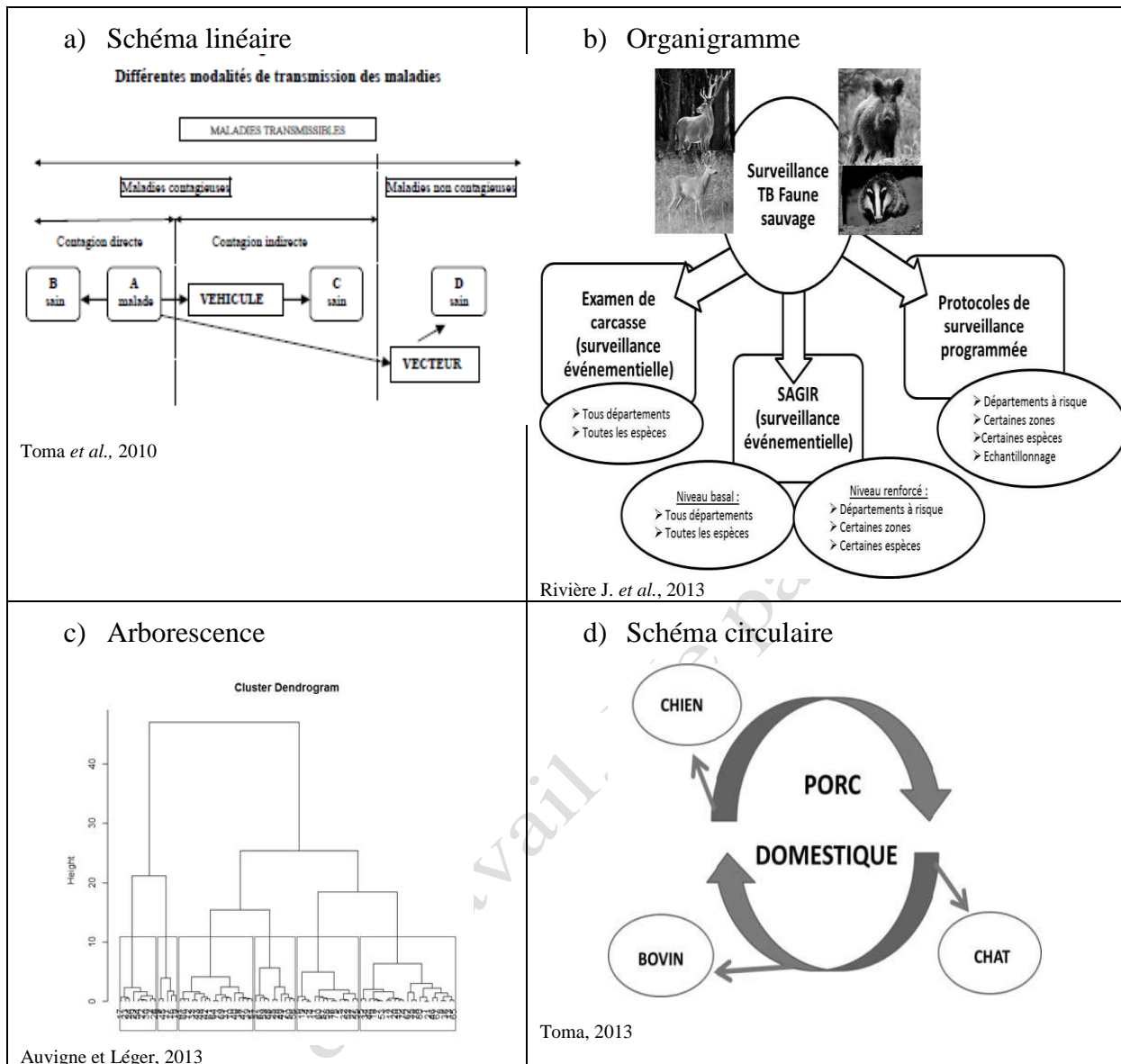
• **Selon leur degré de conceptualisation**

- Dessin : tracé représentant de manière simplifiée un objet réel, (ex : planche anatomique), ou un objet abstrait (comme un protocole)(figure 15e) ;
- Carte : Représentation schématique d'un espace, utilisant des codes permettant d'en illustrer la structure, la dynamique, afin de décrire et de comprendre cet espace (figure 15f) ;
- Chronogramme : représentation graphique de l'évolution temporelle des états d'un élément, ou de succession d'étapes d'un processus (figure 15g) ;
- Schéma systémique : représentation graphique symbolisant les éléments constitutifs d'une réalité pouvant être complexe et leurs relations fonctionnelles (ex : cycle épidémiologique)(figure 15h).

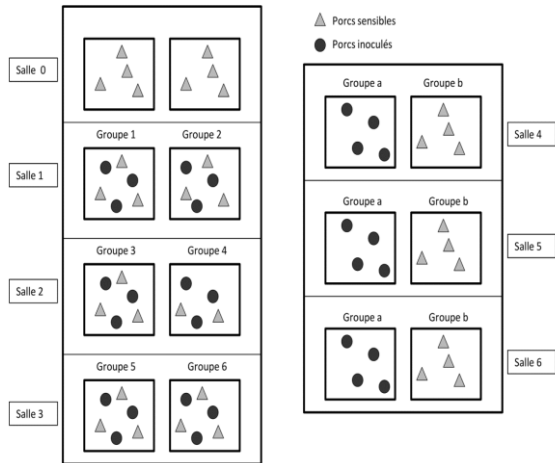
Ces quelques types sont loin d'être exhaustifs, car il est aussi possible d'assembler de diverses manières ces éléments dans un graphique composite, pour condenser l'information tout en la rendant accessible.

² Plus exactement : « *Un bon croquis vaut mieux qu'un long discours* » (Napoléon Bonaparte).

Figure 15 : Types de schémas en usage en épidémiologie

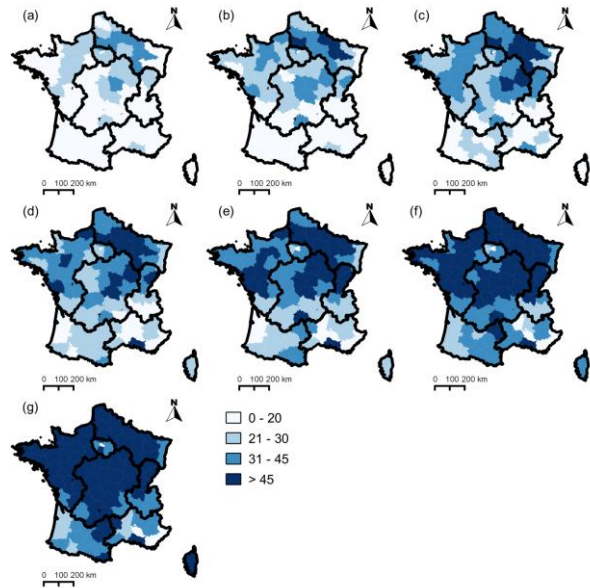


e) Dessin



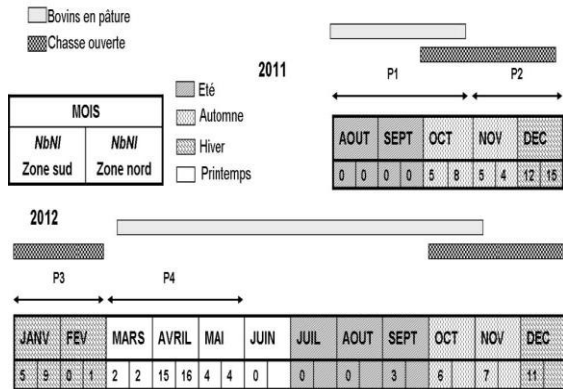
Andraud M *et al*, 2013

f) carte



Bekara M *et al*, 2014

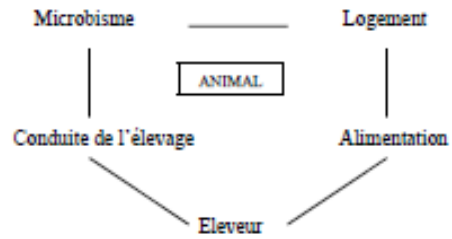
g) Chronogramme



Payne A. *et al*, 2013

h) Schéma systémique

Représentation schématique des catégories de facteurs de l'environnement pouvant jouer un rôle sur la réceptivité (et la sensibilité) des animaux à un agent pathogène



Toma *et al*, 2010

IV - CARTES

Définition

Une carte permet la représentation spatiale de données sur un fond représentant un territoire.

Les cartes sont des catégories de schémas qui ont un intérêt majeur en épidémiologie, ce qui justifie un développement particulier.

Description

Le territoire peut être suggéré de façon minimale, par ses seuls contours géographiques ou administratifs, ou bien représenté de façon plus sophistiquée en intégrant des éléments comme le relief, les cours d'eau, les villes ou habitations, la végétation, les cultures, voire des informations élaborées (comme des flux d'activités économiques, des informations climatiques, des ressources disponibles pour la lutte contre les maladies) dans le cas des « Systèmes d'information géographique » (S.I.G.), sous forme de couches (« calques ») superposables en fonction des besoins. De nombreux logiciels permettent de réaliser ces cartes (par exemple, payant, ArcView ou gratuit, Quantum GIS).

Objectifs

En épidémiologie, une carte est un outil de description, ainsi que de prospection de relation avec des facteurs éventuels.

Elle vise en premier à visualiser la localisation de la maladie (**où ?**) : les points représentant chacun des cas ou des foyers constituent le mode de représentation le plus simple.

Elle permet de connaître aussi l'importance de la maladie (**combien ?**). On peut utiliser diverses modalités (taille des points, densité de couleurs ou de motifs) pour apporter cette information supplémentaire qui peut être plus explicite que la seule accumulation de points. La représentation du nombre des cas ou des foyers doit impérativement être complétée par celle de la proportion par rapport à la population susceptible : à défaut, l'interprétation peut être erronée, imputant à la maladie la responsabilité de la répartition observée, alors qu'en fait elle n'est que le reflet de celle de la population susceptible. [\[En savoir plus sur l'importance du choix entre nombre et pourcentage de cas\]](#)

On peut aussi chercher à en savoir plus sur la manière dont est répartie la maladie (**comment ?**) : dispersée, par agrégats, de manière régulière ou non.

La relation avec le temps peut aussi être envisagée (**quand ?**) grâce à une série de cartes successives ou à une représentation différente des cas ou foyers en fonction du moment.

Enfin, la mise en relation avec des substrats, des facteurs sous-jacents permet de chercher les relations potentielles à prospector (**pourquoi ?**).

Quelques exemples et types de cartes

Nous nous contenterons de quelques exemples, pour illustrer l'intérêt de l'utilisation des cartes en épidémiologie et fournir les éléments nécessaires à la formulation des recommandations.

1. Selon les modalités de représentation des cas ou foyers

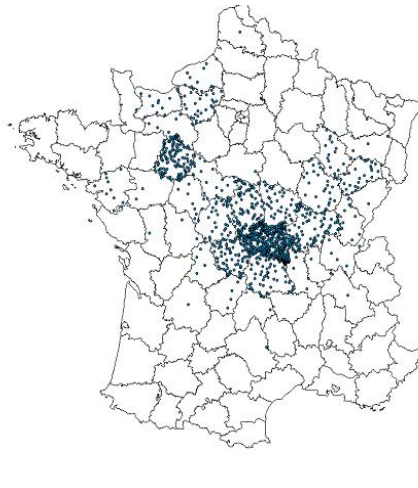
Les cas ou foyers peuvent être représentés par des points, des codes de couleur ou de motifs ou par une ligne symbolisant une limite d'extension.

1.1. Points

C'est la modalité la plus simple de représentation : un point représente un cas ou un foyer (figure 16).

Figure 16 : Répartition géographique des foyers de fièvre catarrhale ovine de sérotype 8 par département au 25 juillet 2008 (source : DGAI/Afssa)

Chaque point correspond à un foyer



Dans certaines circonstances, il peut être nécessaire de se poser la question de savoir à *quoi correspond la localisation géographique du cas ou du foyer*. En effet, s'agit-il du lieu de déclaration (l'instance administrative qui l'a enregistré), de sa survenue (l'endroit où était localisé le sujet qui a présenté les premières manifestations de la maladie ayant permis son diagnostic), du lieu de résidence de l'individu concerné (qui correspond à l'adresse postale, pas forcément au lieu de vie le plus fréquent), ou bien du lieu où vraisemblablement il a pu contracter la maladie.... Ces quelques questions permettent de constater qu'il est très important de se poser au préalable la question de ce que l'on veut représenter avec cette carte.

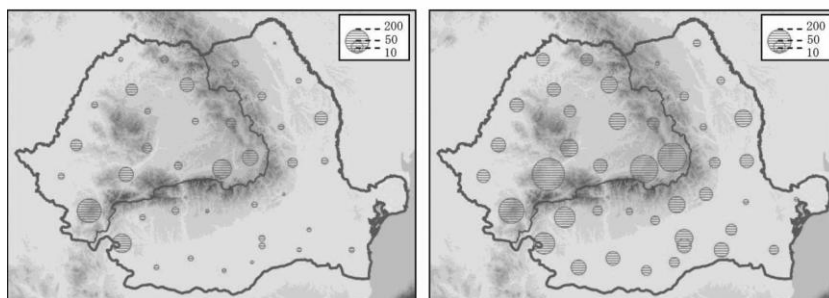
L'accumulation des points suggère l'intensité du phénomène étudié. Mais cette accumulation peut conduire à en masquer un certain nombre lorsqu'ils sont trop nombreux. C'est pourquoi il peut être utile d'utiliser une taille des points qui soit proportionnelle au nombre des cas (figure 17).

Figure 17 : Taux d'incidence cumulé de la trichinellose en Roumanie

[source : *Am. J. Trop. Med. Hyg.*, 2007, **76**, 983]
(en nombre de cas pour 10 millions de personnes par an)

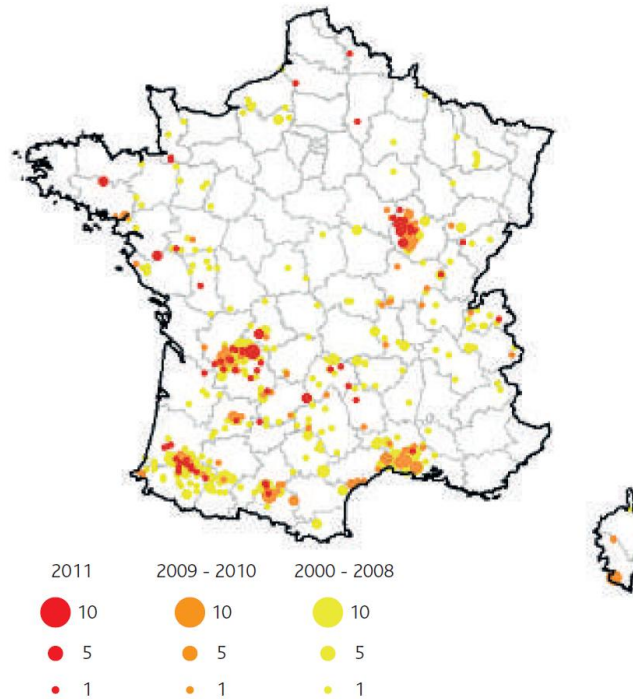
1980-1989

1990-2004



Si l'on cumule des données sur une période suffisamment longue, il peut être intéressant de distinguer différentes périodes : on peut les distinguer en utilisant des formes (ronds, carrés, triangles, étoiles, *etc.*) ou des couleurs différentes (figure 18).

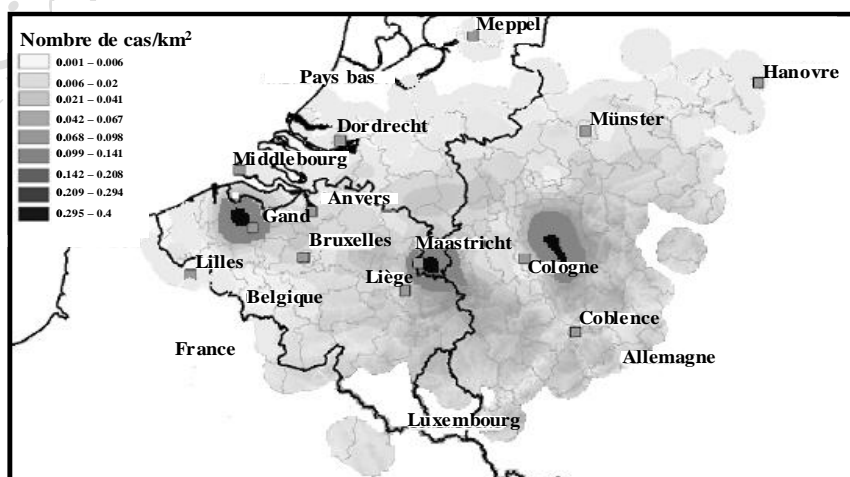
Figure 18 : Distribution des foyers incidents de tuberculose bovine en France métropolitaine de 2000 à 2011 (Fediaevsky *et al.*, 2012)



1.2. Densité de couleur ou de symboles

Les nombres de cas ou de foyers par unité géographique peuvent alors être distribués en classes représentées par une palette de couleurs ou de motifs, choisis selon une échelle croissante d'intensité (figure 19).

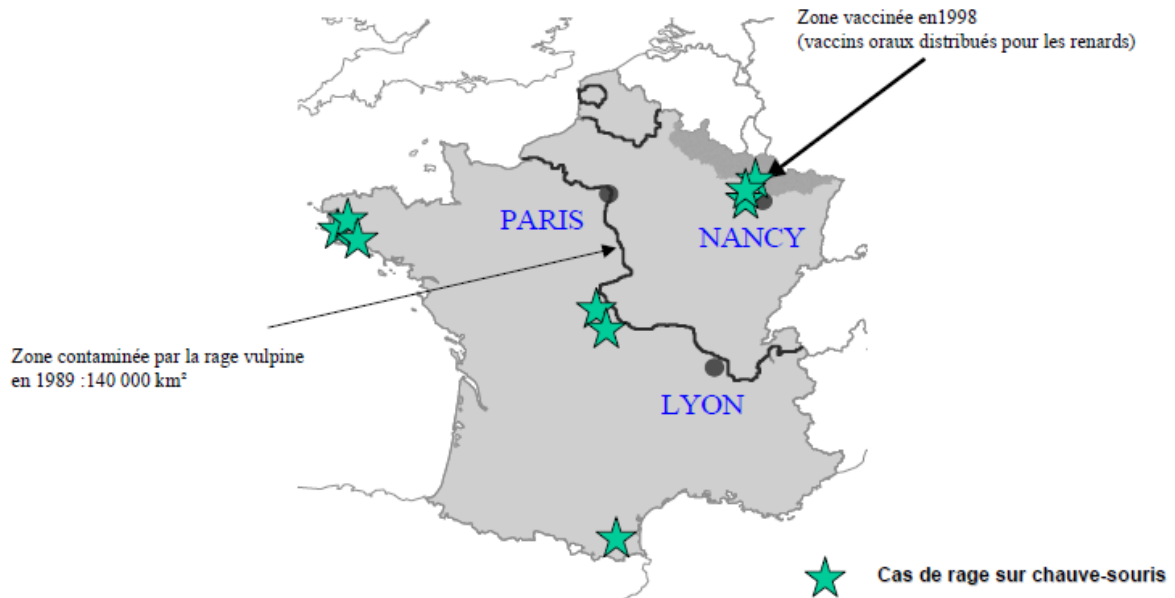
Figure 19 : Densité de foyers de fièvre catarrhale ovine de sérotype 8 par kilomètre carré dans le nord de l'Europe au 1^{er} février 2007 (source : AESA)



1.3. Ligne de limite d'extension

Pour les maladies infectieuses qui progressent dans l'espace, on peut représenter la ligne symbolisant la limite du front de progression (figure 20).

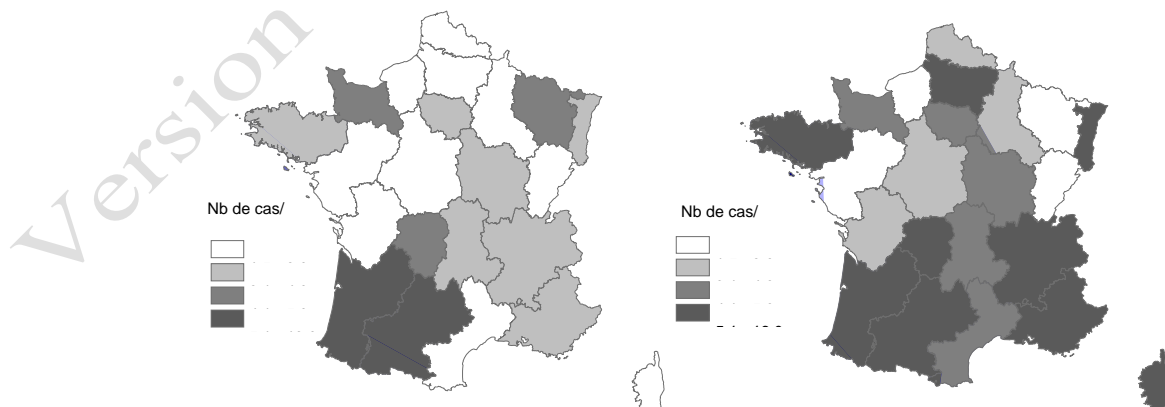
Figure 20 : Extension maximale de la rage vulpine en France et répartition des cas de rage sur chauves-souris : deux phénomènes indépendants (Moutou *et al.*, 2000).



2. Selon un substrat sous-jacent de cartographie

- Le substrat sous-jacent peut être réduit au strict minimum : une carte muette ne comportant que les limites administratives ayant vocation au repérage géographique (cf. figure 18), ou bien ces limites administratives servent à délimiter des zones utiles pour la comparaison (figure 21).

Figure 21 : Evolution de l'incidence régionale de la listériose humaine en France de 2001-2005 à 2006-2007 [source : BEH, 2008, n° 30-31]



- Des logiciels (systèmes d'information géographique ou SIG) permettent de représenter simultanément les cas et des informations complémentaires, telles que les zones géographiques naturelles, les voies de communication, la végétation, les cultures ou des densités de ressources sous forme de « calques » superposables.

CHAPITRE V - RECOMMANDATIONS POUR LA PRESENTATION DES TABLEAUX ET DES FIGURES

I - OBJECTIFS

Pour posséder les qualités requises, une illustration (un tableau ou une figure) doit atteindre deux types d'objectifs, les uns centrés sur l'auteur du tableau et son travail, les autres sur son lecteur.

➤ Objectifs centrés sur l'auteur de l'illustration

- L'illustration doit être **pertinente** : elle doit refléter au mieux la pensée de l'auteur, être adaptée à la question posée, aux résultats qu'il désire communiquer.
- Elle doit avoir le **degré de complétude** adapté : suffisamment d'information (pour répondre à toutes les questions), mais aucune information inutile (pour éviter d'égarer l'attention).

➤ Objectifs centrés sur le lecteur de l'illustration

- L'illustration doit être **informative** : pas besoin d'un tableau pour exprimer un simple pourcentage.
- Elle doit être **compréhensible** : c'est sa qualité première ! En quelque sorte, c'est une déclinaison de l'aphorisme de Boileau : « Ce qui se conçoit bien s'énonce clairement. Et les mots pour le dire arrivent aisément »³. Autrement dit, une difficulté de compréhension peut être révélatrice d'un défaut de conception.
- Elle doit être **facilement lisible** : elle doit permettre un accès rapide à l'information par la vision synoptique qu'elle procure ; elle doit permettre également une analyse rapide de cette information par le classement effectué selon un ou plusieurs critères et en facilitant les comparaisons.
- Elle doit être **esthétique**, car une mise en forme esthétique contribue à faciliter la compréhension, tandis qu'une surcharge inutilement lourde peut entraver la compréhension.
- Elle doit être **autonome**, c'est-à-dire qu'elle doit être compréhensible sans le besoin du texte.

Voyons comment procéder en pratique pour atteindre ces objectifs.

II - RECOMMANDATIONS POUR LA PRÉSENTATION DE TOUTE FORME D'ILLUSTRATION

L'objectif des recommandations portant sur la présentation est la **mise en forme** d'une illustration pour en faciliter la compréhension. Cette mise en forme doit être centrée sur le lecteur, afin de s'assurer qu'il va pouvoir comprendre le message, de la façon la plus efficace possible. Les règles correspondantes portent sur la numérotation, sur la mise en forme (centrée sur le fond), et sur l'esthétique.

• Recommandation générale de numérotation

Par convention, toutes les illustrations sont **numérotées**, les tableaux de façon distincte des figures, dans l'ordre d'apparition dans le texte, et sont systématiquement « **appelées** » dans le

³ Nicolas Boileau, L'art poétique, 1674

texte dans l'ordre, par une simple indication entre parenthèse, en évitant toute allusion à son emplacement, comme « *tableau suivant* », moins explicite que son numéro.

Pour l'appel des illustrations, éviter des formules lourdes : « *comme le montre le tableau x* », « *Si on regarde la figure y* », « *On voit dans le tableau x* ». Préférer un style direct, consacré à ce que l'on veut dire : « *les élevages ne comportant qu'un faible nombre de bovins tuberculeux sont les plus fréquents (tableau x)* ».

- **Recommandations de conception visant à faciliter la compréhension**

Il faut appliquer diverses **règles de mise en forme** visant toutes à une meilleure compréhension pour le lecteur.

- La **conception** doit être **logique**, en respectant l'organisation naturelle dans notre culture occidentale : lecture de gauche à droite, de haut en bas. L'organisation doit aller du plus simple au plus complexe, du normal au pathologique, du descriptif à l'explicatif.
- L'illustration doit apporter toutes les informations nécessaires à sa compréhension, de façon **autonome**, sans besoin de se reporter au texte pour la comprendre : toutes les informations nécessaires doivent par conséquent y figurer, d'une façon ou d'une autre, que ce soit dans le titre, dans les libellés des têtes de colonne ou de ligne, ou encore dans une légende.
- De ce point de vue, le **titre** joue un rôle crucial et sa rédaction doit par conséquent retenir toute l'attention... et il faut un certain entraînement pour en maîtriser les conditions de qualité :
 - Il doit pouvoir être compris sans l'aide du texte de fond. Pour apporter les précisions indispensables sans nuire à la clarté du titre, il faut utiliser une légende où les abréviations éventuelles doivent être explicitées ;
 - Son caractère autonome est souligné par le fait qu'il commence par une majuscule.
 - Il doit donner aussi les informations permettant de situer le contexte. On peut s'aider des questions classiques : quoi, quand, où, comment, *etc* ;
 - Il doit être explicite, non ambigu : il doit porter sur ce qui est représenté, non sur ce que qu'on souhaiterait que le lecteur comprenne.

Toutefois, deux écoles s'opposent en ce qui concerne la conception du titre.

- Les usages de la communication scientifique ne portent que sur ce qui est représenté, sans aller jusqu'à l'interprétation ;
 - Les usages de la vulgarisation et de la communication d'entreprise (management, marketing) mettent au contraire l'accent sur l'interprétation.
- **Recommandations centrées sur des règles esthétiques pour faciliter la compréhension**

La présentation doit aussi respecter différentes **règles esthétiques** pour faciliter la compréhension, en évitant les lourdeurs excessives.

Un bon principe général est celui de l'**économie d'encre**, pas véritablement pour des préoccupations écologiques (quoique...), mais surtout pour éviter les surcharges inutiles. Ce qui conduit à éviter les grisés, les **fonds** de couleur, voire à choisir tel type de représentation plutôt que telle autre, qui serait pourtant théoriquement mieux indiquée ([Voir un exemple](#)).

III - RECOMMANDATIONS SPÉCIFIQUES POUR LA PRÉSENTATION D'UN TABLEAU

Pour que la présentation d'un tableau concoure aux objectifs définis précédemment, il faut que sa **conception** soit elle-même adaptée aux objectifs.

[\[En savoir plus sur les recommandations relatives à la conception d'un tableau\].](#)

Pour cela, il faut savoir atteindre le [besoin d'information](#) préalablement défini.

Les recommandations esthétiques suivantes sont spécifiques aux tableaux.

- **Éviter les lignes verticales** et **limiter les lignes horizontales** au strict minimum : pour souligner la ligne d'en têtes de colonnes, et pour souligner les marges, haute et basse, du tableau.
- Limiter l'utilisation du **gras** aux **titres de colonnes et de lignes** : le corps du tableau doit être dans une police normale.
- Le corps du tableau ne doit comporter **que des nombres**, jamais d'unité de mesure. Pour les pourcentages, on a le choix, soit de deux colonnes, l'une dédiée aux effectifs bruts, l'autre aux pourcentages, soit d'une seule colonne donnant les effectifs, les pourcentages figurant entre parenthèses. Dans tous les cas, le signe % ne doit pas apparaître. Les indications doivent permettre de comprendre facilement cette convention (*cf.* tableau 7).

Tableau 7 : Exemple de mise en forme (vue partielle)

	Zone 1	Zone 2	Zone 3
Pourcentage de prévalence	51	74	79
Intervalle de confiance 95 %	[44 ; 58]	[68 ; 80]	[74 ; 84]

- Ensuite, il faut respecter les bons **alignements**.
 - Le titre du tableau est centré.
 - Les nombres sont centrés sur le séparateur décimal. Dans tout le tableau, tous les nombres comportent le même nombre de chiffres après la virgule. Le signe du pourcentage (%) doit être séparé des chiffres par une espace.
- Enfin, **tout doit être explicite**.
 - Les données manquantes sont signalées : « - », *, ND (non déterminé), DM (donnée manquante), avec légende, jamais laissées en blanc.
 - Les notes de bas de tableau pour comprendre les abréviations ne doivent jamais comporter de chiffre, mais soit des lettres (a), soit un symbole, dans l'ordre : *, †, ‡, §, ||, ¶ (que l'on peut trouver dans la police de caractères Vrinda).

[\[En savoir plus, par des exemples, sur la mise en forme centrée sur le fond et sur la forme\]](#)

IV - RECOMMANDATIONS SPÉCIFIQUES POUR LA PRÉSENTATION D'UNE FIGURE OU D'UNE CARTE

Comme signalé en introduction, il faut savoir si le choix de la couleur est pertinent, en fonction de la méthode qui sera choisie pour la reproduction du document : les couleurs initiales peuvent ne plus donner une palette de gris procurant un rendu tout aussi satisfaisant qu'avec les couleurs. Si c'est le cas, il faudra définir une palette adaptée aux gris.

Il faut apporter un soin tout particulier à la légende des figures, puisqu'elles constituent une ressource capitale pour leur compréhension : il convient de leur appliquer les recommandations générales déjà mentionnées.

Les cartes doivent toujours comporter :

- L'indication de l'orientation (nord cadastral),
- Une indication de l'échelle,
- Et une légende explicitant les codes de couleurs ou de représentation utilisés.

La mise en forme des diagrammes doit respecter les conventions typographiques, en particulier pour le signe du pourcentage (%) qui doit être séparé des chiffres par une espace.

[Pour savoir comment faire avec Excel](#)

LIVRE DEUXIEME : POUR ALLER PLUS LOIN

Version de travail. Ne pas diffuser

Version de travail. Ne pas diffuser

CHAPITRE I - EN SAVOIR PLUS

I - COMMENT DISTINGUER VARIABLES QUANTITATIVES ET VARIABLES QUALITATIVES ?

C'est donc si difficile de les distinguer ?

Apparemment, c'est simple : une variable quantitative s'exprimant par des nombres et permettant des calculs, c'est à de tels nombres qu'on peut la reconnaître. Malheureusement, ce raisonnement est faux, car on peut utiliser des nombres pour coder un type de race : il est évident qu'un calcul effectué sur ce type de code n'aura aucun sens. De même, un codage du type 0/1 pour distinguer deux catégories d'individus (mâles, femelles, par exemple) n'implique aucunement un sens mathématique.

Les valeurs d'appréciation qualitative, comme « pas du tout d'accord », « pas d'accord », « ni en désaccord ni en accord », « d'accord », « tout à fait d'accord » (échelle de Likert), sont des exemples de valeurs ordinales : elles ne sont pas quantitatives, mais on peut quand même les classer selon une échelle croissante ou décroissante. On peut utiliser pour leur codification des valeurs chiffrées (1, 2, 3, 4, 5, par exemple dans l'appréciation du degré de saleté des vaches, ou de développement musculo-graisseux d'une carcasse) pour traduire ce degré d'intensité, sans pour autant qu'il soit possible de faire des calculs sur ces valeurs.

Comment facilement les distinguer ?

Si le résultat de l'addition des valeurs a un sens, la variable est bien quantitative ; sinon, elle est qualitative. L'addition du poids de deux individus a bien un sens, pas celle de leur degré de satisfaction. On ne peut pas dire qu'un individu qui a exprimé sa satisfaction en choisissant une réponse ("d'accord") qui sera traduite par une note de 4 est quatre fois plus content qu'une personne qui a choisi une autre réponse ("pas du tout d'accord") correspondant à la valeur 1.

Conséquences d'une confusion éventuelle au plan de l'étude d'un phénomène

Où est le problème, si on peut malgré tout calculer moyenne et écart-type ?

Pour une variable quantitative, les intervalles entre les valeurs peuvent être connus, ou modélisés. Les valeurs relevées représentent bien un phénomène *réel*, et ces valeurs ont le même sens de représentation de ce réel d'une observation à une autre.

Dans le cas d'une échelle de Likert par exemple, on ne peut même pas considérer la note traduisant le choix de réponse aux items proposés comme une forme de variable quantitative discrète : les intervalles entre les valeurs ne représentent pas toujours le même phénomène réel, car l'appréciation d'un degré de satisfaction, par exemple, est subordonnée à la subjectivité du répondant. La transformation de cette subjectivité en une valeur chiffrée donne l'illusion d'une standardisation rigoureuse : on sait très bien que les scores de douleur, par exemple, ne sont comparables que pour un même individu, à deux moments différents, ou pour deux épreuves différentes réalisées au même moment ; en revanche, la comparaison des scores entre individus est beaucoup plus difficile, car un même phénomène réel peut être traduit par des valeurs différentes selon les individus.

On peut objecter toutefois que le dénombrement des répondants pour chaque note pour chaque item permet lui de faire des calculs, puisqu'on travaille sur des populations dénombrées. Ou bien, qu'on peut additionner les notes obtenues pour chaque individu à chacun des items et obtenir un score permettant de caractériser l'individu et le classer dans la population observée.

C'est vrai mais trompeur malgré tout, car s'il est vrai que les dénombrements peuvent effectivement donner lieu à toutes formes de calculs, les phénomènes qui ont donné lieu à la formulation d'un degré d'accord par les répondants n'ont pas le caractère objectif garantissant la standardisation des observations de chaque individu.

La conséquence possible est que l'étude de phénomènes réels par l'intermédiaire de telles échelles peut s'avérer difficile, voire décevante, car les phénomènes réels transcrits selon des valeurs de scores, qui apportent les vertus objectives des nombres, sont en fait ainsi caractérisés avec beaucoup d'incertitude : la mesure donne une image floue du phénomène, ce qui altère la qualité des représentations nécessaires à leur étude.

Conséquences du point de vue de la qualité d'une représentation

Du fait de l'absence de continuité des variables qualitatives, la seule représentation qui convient est les diagrammes soit **à colonnes** pour les variables **ordinales** (l'axe des abscisses permettant de représenter l'échelle de valeur croissante), soit **à barres** pour les variables **nominales** (la dénomination relativement longue des catégories prend trop de place et l'agencement vertical des classes permet d'éviter d'introduire une notion de valeur entre les classes).

La représentation graphique d'une variable quantitative doit tenir compte de sa caractéristique fondamentale de calcul numérique et les axes doivent pouvoir être perçus selon cette logique calculatoire :

- les unités doivent être indiquées ;
- l'axe quantitatif doit être homogène, ne pas comporter de ruptures d'unités (par exemple en raison d'un regroupement de données effectué pour certaines valeurs et pas d'autres) ;
- un mélange sur un même axe d'une conception strictement numérique et d'une conception qualitative est à proscrire ([voir exemples](#) à propos des diagrammes à colonnes).

Inversement, tout observateur d'une description d'une variable numérique la considère comme telle : si le traitement appliqué lui fait perdre son caractère quantitatif, il faut que le lecteur puisse en avoir immédiatement conscience.

[Retour au chapitre "Variables statistiques"](#)

II - EN SAVOIR PLUS SUR LES VARIABLES TEXTE

Compte tenu des possibilités maintenant offertes par l'informatique, il est tentant de saisir des informations sous forme de texte sans avoir préalablement défini de catégories, « pour en tirer quelque chose par la suite ». En l'absence de formation spécifique dans ce domaine pour les étudiants vétérinaires (ce qui n'est pas le cas des étudiants par exemple en sciences humaines), il est utile de donner quelques recommandations minimales.

Savoir quels traitements seront appliqués par la suite

Il est capital de bien savoir ce que l'on va appliquer comme traitement ultérieur à tout texte saisi : ne pas y avoir pensé au préalable expose au risque d'avoir dépensé de l'énergie pour saisir des informations sous forme de texte, pour se rendre compte au final qu'on n'est pas capable de les traiter de façon pertinente. Si l'on peut imaginer que pour un fichier comportant un nombre raisonnable d'enregistrements, il soit possible de lire l'intégralité des textes saisis, pour ensuite procéder à une codification en vue d'un traitement statistique, cela ne peut être le cas pour un nombre trop important, qui découragera les meilleurs bonnes volontés (ou alors, il faudra envisager une sélection aléatoire de ces informations).

Pour cela, il ne faut pas reporter le soin de se préoccuper de cette question à plus tard : mieux vaut procéder sur un jeu d'essai, pour bien prendre la dimension des difficultés et des moyens éventuels pour les résoudre.

Prendre en compte la distorsion entre l'orthographe et le sens.

Une fois l'information saisie, « il n'y aura qu'à » lui appliquer des traitements informatiques, se dit-on ! Certes, sauf que le moindre caractère d'écart est distingué par l'ordinateur (attention aux accents, car l'ordinateur distingue « e » et « é », « è », sans compter « er », « ai », ou « aie », etc.), voire simplement la « casse » (distinction entre minuscule et majuscule), ce qui se répercute sur les opérations de tri ou de dénombrement. Si l'information est saisie soit par différents opérateurs (ce qui est souvent le cas dans des structures administratives ou des entreprises), soit directement par des personnes enquêtées (pour les enquêtes réalisées en ligne), penser non seulement aux fautes de frappe, mais aussi aux conceptions de l'orthographe parfois très éloignées de l'orthodoxie.

Mais, cette standardisation orthographique (le même mot étant écrit exactement de la même manière) ne suffit pas. Il faut aussi penser que deux mots différents peuvent avoir été utilisés dans la même intention. Par exemple, le mot animal peut avoir été utilisé au lieu et place de chien ou chat : le contexte permet alors de faire le remplacement nécessaire à un traitement informatique pertinent, car la distinction dans cet exemple n'aurait pas de sens. Mais, parfois, des mots différents peuvent être utilisés, tantôt avec le même sens que le terme générique, tantôt avec un sens spécifique : c'est le contexte qui permet de lever l'ambiguïté... Ou pas ! Dans ces conditions, le choix de substituer le terme employé par un terme générique constitue une interprétation qui peut altérer le sens véritable, et de toute façon constitue une entorse à la neutralité scientifique qui doit prévaloir dans la récolte des données avant leur traitement. Il n'y a pas d'autre solution que de définir les règles de décision pour ce type de situation soit avant, soit, au moment où la circonstance est rencontrée, de façon à standardiser le traitement appliqué et surtout en tenir compte au moment de l'interprétation.

Au total, il faut prévoir de relire l'information saisie afin de la standardiser sur les plans orthographique et éventuellement sémantique avant de pouvoir lui appliquer le traitement prévu. Mais il faudra aussi que la discussion prenne en compte ces différents traitements pour bien s'assurer qu'ils n'ont pas altéré le sens intime de l'information : la standardisation peut donner une apparence d'objectivité, mais elle peut aussi en modifier le sens.

Les types de traitement des variables texte

Le traitement des variables texte peut porter sur le mot lui-même (ou une expression), pour en détecter l'occurrence (afin de décrire le contexte de cette occurrence ou son importance) et en effectuer le dénombrement éventuel.

Compte tenu des remarques sur la variabilité de la saisie, pour toutes les variables statistiques qualitatives dont les valeurs sont des mots ou des expressions, il est préférable d'imposer une standardisation de la saisie à l'aide de dispositif de contrôle n'autorisant l'entrée que de mots ou expressions préalablement déterminés, voire, en utilisant un code constitué de chiffres qui n'auront pas une valeur numérique puisqu'ils représenteront des mots.

Le traitement peut porter sur le sens, lorsqu'il s'agit de texte libre (question à réponse ouverte) ou de la transcription d'un entretien. La lecture permet de repérer le sens général du discours : par exemple, on peut repérer le thème abordé, et la valeur associée (par exemple positif / négatif). On peut donc créer des variables interprétatives qui faciliteront le traitement ultérieur : par exemple, dénombrement des occurrences du thème, et des valeurs associées.

Cette approche est toutefois très délicate, car deux personnes différentes confrontées aux mêmes textes peuvent ne pas avoir exactement la même interprétation pour chaque proposition ainsi analysée. Il faut donc définir une procédure qui permette d'assurer la fiabilité (répétabilité, reproductibilité) de ce traitement de décriptage... ce qui est loin d'être évident ! Une manière de contourner la difficulté est de soumettre le texte à au moins deux lecteurs, afin que tout écart d'interprétation soit discuté, et tranché après consensus (voir un exemple de ces difficultés dans Delahousse, 2014).

[Retour au chapitre des variables textes](#)

III - COMMENT RÉALISER UN TABLEAU ADAPTÉ AUX OBJECTIFS ?

On peut distinguer trois phases dans la réalisation d'un tableau :

- d'abord, la phase d'exploration des données, le plus souvent en se faisant aider de logiciels adaptés, qui aboutit à faire surgir les premiers constats ;
- puis, celle de son enrichissement, par exemple en situant ces premiers résultats dans un contexte permettant de faire surgir ou de conforter le sens escompté ;
- enfin, celle de son adaptation pour le rendre le plus propice à la communication, car le tableau exploratoire nécessite que la présentation, voire son organisation, soient travaillées ou bien parce que l'on désire le combiner avec un (ou plusieurs) autre(s) tableau(x).

1. Première phase : exploration des données et élaboration de la première esquisse du tableau

Lors de l'exploration des données, on est à la recherche d'éléments significatifs, cachés dans l'anarchie des données contenues dans un fichier.

La démarche classique de **tri** statistique a ici un sens très fort : il s'agit de trier l'information intéressante parmi une nuée de données sans signification intéressante. En statistique, on utilise le tri à plat (une seule variable) puis les tris croisés (au moins deux variables) par exemple. Tout l'art consiste à trouver la bonne partition des données, les bons seuils pour cela, tout en respectant l'objectivité consistant à éviter d'adapter la méthode de partition en fonction du résultat obtenu, par exemple en utilisant des critères indépendants des premières productions de résultats, mais au contraire en s'appuyant sur les hypothèses formulées au préalable. Ce travail est purement *descriptif* au plan des données ; il n'est que partiellement adapté à la description du phénomène étudié en raison des comparaisons multiples effectuées qui diminuent la puissance démonstrative des données : on peut être amené à observer des rapprochements, des associations dus au simple fait du hasard. Pour éviter cet inconvénient, il faut des stratégies statistiques adaptées, que nous ne développerons pas davantage.

Le but est donc de voir s'il existe ou non des points d'accumulation, et de décrire la dispersion des données.

2. Phase suivante : enrichir cette première approche

Un premier résultat doit être considéré avec le recul permettant de le situer dans son contexte :

- soit par comparaison avec au moins une autre population (sur une seule variable étudiée),
- soit en combinant deux variables (pour une seule population, voire deux).

Ces deux approches correspondent aux deux formes de tableaux de données, à simple ou à double entrée. On peut donc constater que le choix du type de tableau est subordonné à l'idée que l'on a en tête.

L'outil majeur de ces comparaisons est le **pourcentage** : la meilleure et la pire des choses, car, il est essentiel de savoir quel effectif placer au dénominateur. Or, une difficulté majeure provient de ce qu'il y a toujours **trois** façons de calculer un pourcentage avec la même donnée :

- par rapport à l'ensemble de la population, de l'échantillon du tableau,
- par rapport à un total de ligne, par exemple, conditionnellement à une catégorie,

- par rapport à un total de colonne, par exemple, conditionnellement à un état donné.

Soit la distribution de données du tableau A.1.

Tableau A.1 : Illustration de la possibilité de disposer de trois pourcentages différents pour une même donnée

	Colonne 1	Colonne 2	Total
Ligne 1	10	40	50
Ligne 2	20	30	50
Total	30	70	100

L'effectif de la première case (10) peut en effet donner trois pourcentages, selon la manière utilisée pour effectuer le calcul :

- 10 %, par rapport au total général,
- 20 %, par rapport au total de ligne,
- 33 %, par rapport au total de colonne.

Ces notions de *conditions* sont capitales pour soit décider du sens véritable que l'on veut donner au résultat et déterminer quel est le bon correspondant, soit, inversement, réfléchir au sens que peut avoir véritablement tel pourcentage.

Bien entendu, cela veut dire qu'il y a au moins trois possibilités d'illustration par des figures, selon le choix de pourcentage qui sera fait : or, les tableurs usuels font spontanément des choix par défaut, auxquels il convient d'être particulièrement vigilant pour éviter d'utiliser un mode de pourcentage qui ne correspond pas à celui que l'on souhaite...

[Voir des exemples](#) illustrant les différents types de pourcentage.

Si vous souhaitez retourner dès maintenant au chapitre « Recommandations pour la présentation d'un tableau », [c'est ici](#).
Sinon, vous pouvez poursuivre.

3. Troisième phase : clarification du besoin d'information

Si dans la première phase, l'intérêt était centré sur les données, dans cette dernière phase, il est centré sur le lecteur, afin de s'assurer qu'il va pouvoir comprendre, de la façon la plus efficace possible.

En effet, un tableau peut paraître tout à fait informatif à son auteur, mais un regard « naïf », c'est-à-dire qui n'a pas profité des heures de travail sur les données et des réflexions sur le sujet, peut ne pas parvenir à voir les mêmes choses. Il faudra donc vérifier si le tableau est utilisable tel quel pour que le lecteur parvienne lui aussi à accéder au sens qui y a été trouvé, ou, si besoin, le modifier pour atteindre cet objectif.

On pourrait croire qu'il suffit, en complément de l'illustration, de *dire* les choses, pour que le lecteur comprenne : grave erreur, car le tableau (mais aussi la figure !) impose une manière de *voir*, qui supprime tout discours. Par conséquent, il faut que le tableau (mais aussi la figure) donne à voir véritablement ce que l'auteur souhaite : c'est véritablement faire œuvre de maïeutique, c'est-à-dire de ne pas imposer *sa* vérité par son discours, mais d'aider son interlocuteur à la trouver par lui-même dans l'illustration fournie, en lui facilitant toutefois ce parcours. C'est la garantie d'une communication objective réussie.

Le fin du fin consiste même à ne donner que le tableau (ou la figure) qui doit normalement suffire à la compréhension, sans le besoin de l'aide d'un texte complémentaire : c'est d'ailleurs un des objectifs formulés (le tableau doit être **autonome**).

Compte tenu des objectifs et limites fixés au départ, nous nous plaçons dans le cadre de la communication scientifique, et n'envisageons pas celui de la vulgarisation.

L'étape de **clarification du besoin informatif** consiste à se poser la question : quelle compréhension veut-on induire chez le lecteur ? La mise par écrit de cette étape est très utile pour la suite, comme autant d'objectifs qui permettront de vérifier point par point si tous les éléments sont facilement accessibles à la compréhension du lecteur du tableau.

Exemple 1

Dans une étude épidémiologique descriptive par sondage, l'exploration descriptive de l'échantillon a porté sur certaines variables démographiques choisies pour permettre la comparaison à celle de la population cible afin d'en vérifier la représentativité.

On veut donc que le lecteur puisse d'un coup d'œil, d'une part, percevoir la *description démographique* de l'échantillon (en pourcentage des différentes modalités des variables choisies) afin, d'autre part, de pouvoir *la comparer* à celle de la population sur les différentes variables choisies afin de savoir si les écarts sont significatifs ou non.

Par conséquent, le tableau devra permettre, d'une part, ce rapprochement, en évitant au lecteur de devoir le faire par lui-même en consultant deux tableaux qui peuvent être relativement distants dans le document, et d'autre part permettre cette comparaison, car ils peuvent être organisés d'une façon légèrement différente, pouvant gêner la comparaison.

Par exemple, dans une thèse de doctorat vétérinaire, l'échantillon des propriétaires d'animaux enquêtés était décrit dans la deuxième partie, dite expérimentale, d'après la distribution selon l'âge, tandis que la description de la population des propriétaires avait été donnée en première partie (environ 30 pages auparavant !), selon des tranches d'âge qui ne correspondaient pas. Autant dire qu'il était impossible de pouvoir facilement effectuer la comparaison : il aurait fallu rappeler dans le tableau les éléments d'information du premier.

Exemple 2

Dans une étude sur l'identité de la profession vétérinaire (Delahousse, 2014), la question suivante a été posée aux individus constituant plusieurs échantillons (population générale, vétérinaires praticiens, vétérinaires non praticiens, étudiants vétérinaires) « *citez entre 5 et 10 mots ou groupes de mots, qui vous viennent spontanément à l'esprit, en association avec le mot vétérinaire* ». Après traitement sémantique approprié pour tenir compte des différences d'orthographe ou de formulation pour un sens équivalent, les termes ont été classés selon, d'une part, leur fréquence (élevée ou non, selon un seuil approprié) et, d'autre part, le rang de citation (dans les premiers ou non, selon un seuil approprié).

On veut donc que le lecteur puisse d'un coup d'œil percevoir les termes majeurs organisés selon ce *classement à double entrée*, tout en lui permettant une *comparaison* facile entre les échantillons afin d'apprécier les différences ou au contraire les similitudes éventuelles entre les échantillons.

Donc, l'idée d'un tableau à double entrée est confortée, mais son organisation devra être conçue de façon à permettre ces comparaisons multiples.

[Retour au point de départ de « besoin d'information »](#)
[Retour au chapitre "Recommandations pour la présentation d'un tableau".](#)

4. Exemples de mise en forme centrée sur le fond

Exemple 1 : formulation d'un titre

Prenons, par exemple, des données décrivant la tuberculose dans des élevages bovins (tableau A.2), dont le titre reste à concevoir.

Tableau A.2 : Données décrivant la tuberculose dans des élevages bovins

Nombre de bovins porteurs de lésions de tuberculose par élevage	Pourcentage d'élevages
1	42
2	23
3	9
4	8
5	2
6	5
7	0
8	2
9	2
10+	9

Ce tableau montre qu'un faible nombre de bovins tuberculeux par élevage est le plus fréquemment observé, alors que, s'agissant d'une maladie infectieuse et contagieuse, on s'attendrait au contraire à constater un nombre bien plus élevé par élevage.

Une première formulation du titre serait : « répartition du nombre de bovins tuberculeux par élevage ». En fait, les données portent sur des nombres d'élevages (transformés ensuite en pourcentage) pour un nombre donné de bovins tuberculeux par élevage. Par conséquent, les dénombrements correspondant aux observations portent sur des élevages (et non sur les animaux), même si le sens concerne le faible nombre de bovins tuberculeux par élevage.

Un autre titre possible selon la conception axée sur le sens, pourrait être : « Faible nombre de bovins porteurs de lésions tuberculeuses par élevage ». Mais, nous avons vu que nous ne retenons pas cette forme de communication, tout à fait adaptée en revanche à la vulgarisation.

Pour atteindre les objectifs de mise en forme fixés pour une communication scientifique, le titre doit mentionner en premier de « quoi » il s'agit, à savoir des élevages (« Distribution du pourcentage d'élevages »), en précisant « comment » les effectifs ont été obtenus (selon le nombre de bovins tuberculeux par élevage). Il faut par ailleurs donner des précisions d'espace (« où ? »), de temps (« quand ? »), de source des données, en fournissant les précisions complémentaires indispensables en légende, ce qui donne au final :

Distribution du pourcentage d'élevages selon le nombre de bovins tuberculeux par élevage en Mondalie de 2002 à 2010 (Source : Services vétérinaires de Mondalie)

En légende : (la classe 10 cumule les valeurs 10 et plus).

Exemple 2 : organisation du contenu d'un tableau

Le tableau A.3 rapporte les fréquences d'utilisation de certains items (définis par une méthode de référence et identifiés par un numéro) par des vétérinaires (V2, V6, V7, V8) dans une consultation. L'auteur, dans son commentaire, veut souligner deux faits : d'une part, certains vétérinaires utilisent certains items, pas certains autres ; d'autre part, certains items sont utilisés par tous les vétérinaires, d'autres seulement par certains vétérinaires.

Tableau A.3 : Exemple de tableau difficilement lisible, dans son état initial (Source Crépin, 2014)

V2		V6		V7		V8	
Item	Utilisation	Item	Utilisation	Item	Utilisation	Item	Utilisation
15	9 fois	13	10 fois	3	10 fois	15	10 fois
40	9 fois	15	10 fois	13	10 fois	40	9 fois
4	7 fois	26	10 fois	15	10 fois	47	7 fois
21	7 fois	27	10 fois	21	10 fois	30	6 fois
30	7 fois	28	10 fois	30	10 fois	37	6 fois
47	7 fois	30	10 fois	37	10 fois	13	4 fois
13	6 fois	40	10 fois	40	10 fois	43	4 fois
8	5 fois			47	10 fois		
53	5 fois			53	10 fois		

Le tableau ne répond pas à l'objectif de compréhension facile par le lecteur, car celui-ci ne peut pas facilement dégager les éléments synthétiques que l'auteur souligne dans son texte.

L'organisation du tableau devrait permettre cette perception : si l'individualisation des vétérinaires a été respectée, ce n'est pas le cas des items, qui sont classés par ordre de fréquence de citation pour un même vétérinaire : de ce fait, sur une même ligne, ils ne se correspondent pas selon les vétérinaires et les mêmes items apparaissent à des lignes différentes selon les vétérinaires. Dans sa phase exploratoire, l'auteur a procédé par tri pour chaque vétérinaire, dont le résultat a été simplement reporté, sans organisation préalable. Le résultat *brut* a été utilisé tel quel, sans que l'auteur ne se soit posé la question de son adaptation à l'objectif de compréhension.

Le tableau A.4 a restauré l'organisation des items, ce qui permet de facilement percevoir ce qui est dit dans le texte, quitte à laisser des cases vides.

Tableau A.4 : Données du tableau A.3 réorganisées pour être plus lisibles (stade de réalisation intermédiaire)

Item	V2	V6	V7	V8
3			10 fois	
4	7 fois			
8	5 fois			
13	6 fois	10 fois	10 fois	4 fois
15	9 fois	10 fois	10 fois	
21	7 fois		10 fois	10 fois
26		10 fois		
27		10 fois		
28		10 fois		
30	7 fois	10 fois	10 fois	6 fois
37			10 fois	6 fois
40	9 fois	10 fois	10 fois	9 fois
43				4 fois
47	7 fois		10 fois	7 fois
53	5 fois		10 fois	

Retourner au chapitre sur les recommandations sur la présentation d'un tableau, en poursuivant par les [règles centrées sur la mise en forme esthétique](#).

5. Exemple de mise en forme centrée sur des règles esthétiques

La mise en forme finale du tableau A.4 aboutit au tableau A.5.

Tableau A.5 : Occurrence des items les plus fréquents selon les vétérinaires

Item	V2	V6	V7	V8
3	-	-	10	-
4	7	-	-	-
8	5	-	-	-
13	6	10	10	4
15	9	10	10	-
21	7	-	10	10
26	-	10	-	-
27	-	10	-	-
28	-	10	-	-
30	7	10	10	6
37	-	-	10	6
40	9	10	10	9
43	-	-	-	4
47	7	-	10	7
53	5	-	10	-

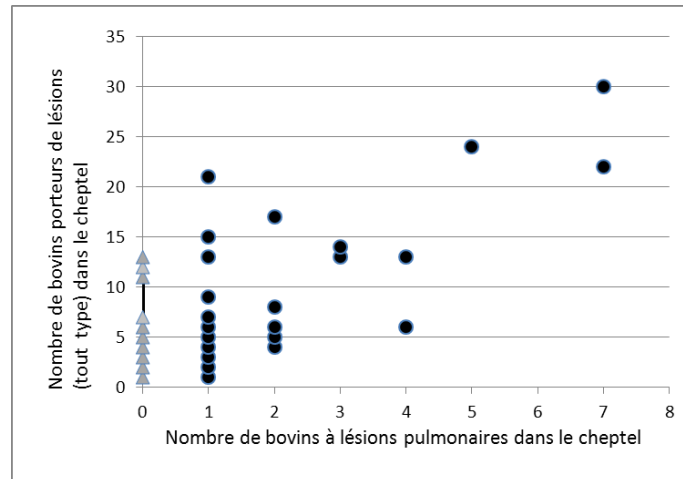
Les traits inutiles ont été supprimés. Les polices ont été standardisées (certaines indications étaient en italique, on ne sait pourquoi). Les cases vides comportent un tiret pour matérialiser l'absence de donnée d'observation (à distinguer d'un éventuel oubli de report dans le tableau). Seuls les nombres figurent : la mention « fois » a été supprimée.

[Retour à la fin du chapitre sur la mise en forme](#)

IV - EN SAVOIR PLUS SUR LES DIAGRAMMES À POINTS : EFFET ATTRACTEUR DE POINTS ISOLÉS

La figure A.1 met en relation le nombre de bovins présentant des lésions de tuberculose dans un élevage, quelle qu'en soit la localisation organique, en fonction du nombre de bovins présentant des lésions pulmonaires dans le même élevage. Sont comparés le groupe des élevages dont aucun animal ne présentait de lésion pulmonaire (représentés par des triangles gris) à celui des élevages en comportant au moins 1 (ronds noirs).

Figure A.1 Diagramme à points illustrant l'effet « attracteur » de points isolés par rapport à un nuage (Source Rambaud 2012)



Globalement, l'ensemble des points suggère une progressivité du nombre de bovins porteurs de lésions en fonction du nombre de bovins porteurs de lésions pulmonaires. Mais, pour les élevages ne comportant qu'un ou deux bovins à lésion pulmonaire, on ne voit pas de relation nette, tandis que pour un nombre plus élevé (de 4 à 7), on a « l'impression » d'une augmentation progressive du nombre de bovins à lésions toutes localisations confondues. Si cette impression suggérée par cette représentation est cohérente avec l'idée que l'on peut se faire de la contagiosité (« *il est logique que plus le nombre d'animaux excréant par voie pulmonaire est élevé, plus il risque d'y avoir d'autres animaux contaminés, et donc porteurs de lésions diverses* »), l'analyse statistique ne permet pas de la confirmer. Cette interprétation visuelle résulte en fait de la puissance suggestive de seulement trois points (abscisses 5 et 7) auxquels l'œil accorde plus volontiers du sens par rapport aux 30 points situés à gauche du graphique (abscisses 0 à 4), qui ne montrent pas de relation aussi claire.

Dans la figure A.1, la contrainte de devoir utiliser la même échelle pour l'abscisse et l'ordonnée n'a pas été respectée, bien que l'unité soit la même (nombre d'animaux dans les deux cas) : les nombres en ordonnées étant bien plus élevés, la représentation aurait manqué de lisibilité, les points étant écrasés sur le côté gauche du graphique, ce qui aurait été contraire à l'objectif visé de facilitation de la perception. Le but étant de montrer la *dispersion* des données, il était logique de choisir une échelle qui prenne en compte l'étendue de chacune des distributions.

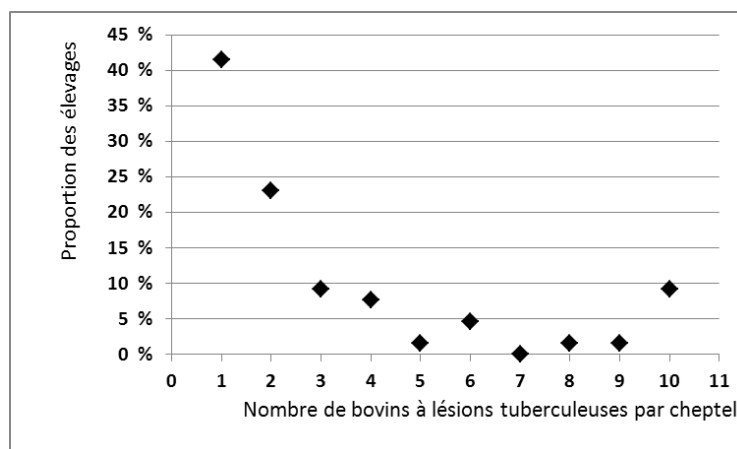
[Retour au chapitre « Diagramme à points »](#)

V - EXEMPLE D'UTILISATION INADÉQUATE D'UN DIAGRAMME À POINTS POUR UN NOMBRE LIMITÉ DE POINTS

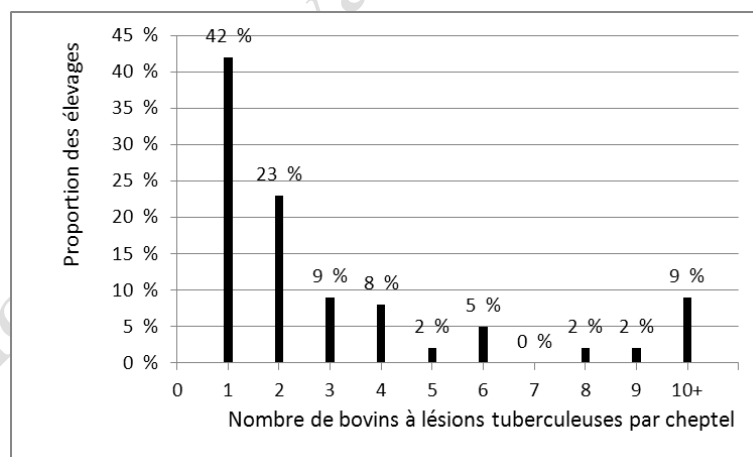
L'intérêt de la représentation par points est de constituer une image globale de nuage lors de la représentation d'un nombre suffisant de points, facilement perceptible à l'œil. Dans le cas d'un nombre limité de points, cet effet de globalisation n'est pas possible. À l'évidence, des points offrent moins de surface au regard que des colonnes : la comparaison des figures A.2 (a et b) permet de vérifier que la perception est meilleure avec un diagramme à bâtons qu'avec un diagramme à points.

Figure A.2 : Distribution du pourcentage d'élevages selon le nombre de bovins tuberculeux en Mondalie de 2002 à 2010 (Source : services vétérinaires de Mondalie) (Abscisse 10 cumule les valeurs 10 et plus)

a) Diagramme à points



b) Diagramme à bâtons



Le diagramme à points ne supporte pas l'adjonction des valeurs dans des étiquettes : nous ne l'avons pas représentée, car elle devient illisible (la valeur chiffrée brouille la perception du point). Le diagramme à bâtons, au contraire, tolère beaucoup mieux ces étiquettes de valeurs, car le bâton occupe beaucoup plus d'espace que l'indication chiffrée.

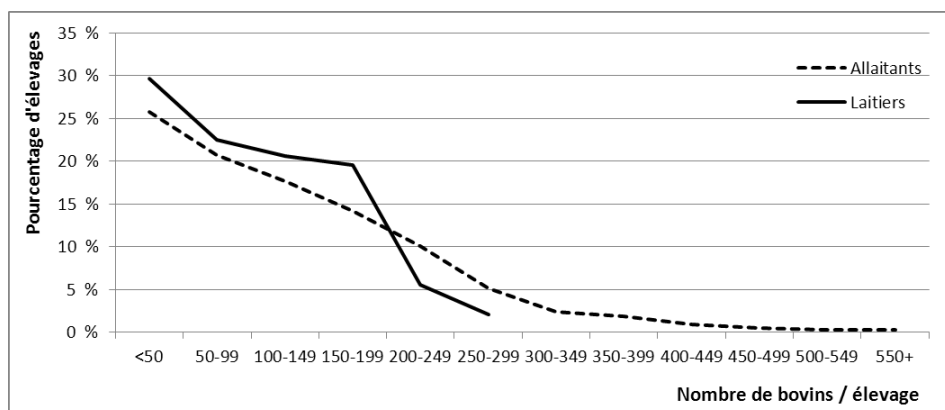
[Retour au chapitre "Diagramme à points"](#)

VI - EXEMPLE D'UTILISATION D'UN DIAGRAMME À LIGNES POUR DES VARIABLES DISCRÈTES REGROUPEES EN CLASSES

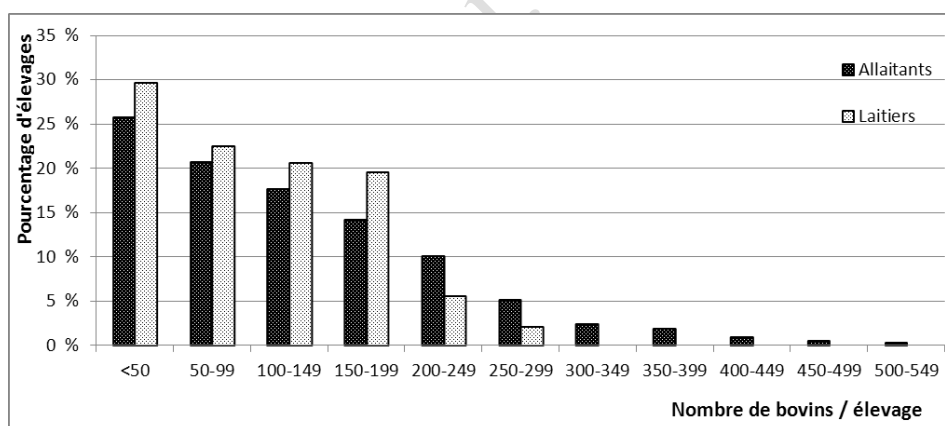
La figure A.3 vise à comparer les distributions des élevages (allaitants et laitiers) selon leur taille. Il s'agit d'une variable discrète, puisque les bovins qui constituent ces élevages sont des unités entières. Si, théoriquement, le diagramme à bâtons est indiqué pour une telle variable, l'étendue des valeurs estompée le caractère discret, au profit d'une relative continuité des valeurs : de fait, le regroupement des valeurs par classes est tout indiqué.

Figure A.3 : Distribution du pourcentage d'élevages, allaitants et laitiers, selon le nombre de bovins par élevage et le type de production (Source : services vétérinaires de Mondalie, 2010)

a) Diagramme à lignes



b) Diagramme à colonnes



Dans le diagramme à lignes (figure A.3a), les données ayant été regroupées en classes, les points doivent être positionnés *au centre* de chaque segment d'abscisse, de façon à respecter la représentation par classes.

La représentation par un diagramme à colonnes (figure A.3b) serait plus adaptée pour une telle variable regroupée en classes. Mais, il faut reconnaître que la comparaison des distributions est moins facile qu'avec le diagramme à lignes de la figure A.3a.

Cette comparaison illustre le bien-fondé du principe de *l'économie d'encre*, évoqué dans les recommandations de présentation des tableaux et des figures.

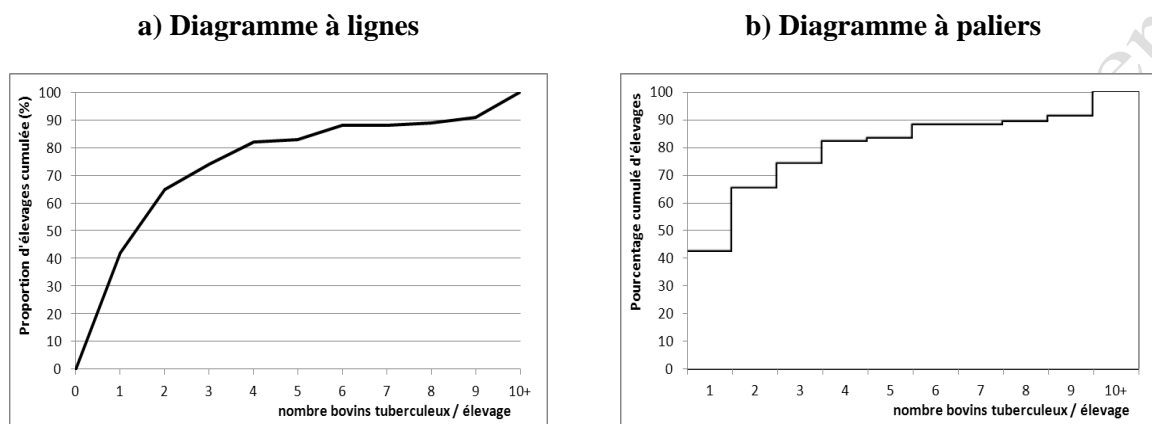
[Retour au chapitre « Diagramme à lignes »](#)

[Retour au chapitre "Recommandations pour la présentation des tableaux et figures"](#)

VII - COMPARAISON DES DIAGRAMMES À LIGNES ET À PALIERS

Les diagrammes à paliers sont utilisés en première indication pour représenter des probabilités cumulées des courbes de survie. La figure A.4 permet de comparer les représentations de ces probabilités utilisées dans un contexte de description avec un diagramme à lignes et un diagramme à paliers.

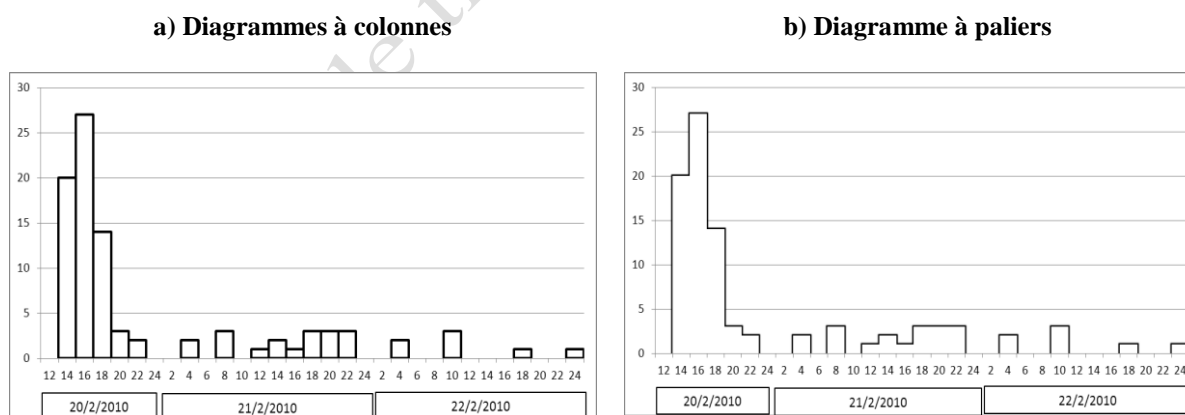
Figure A.4 : Pourcentage cumulé d'élevages selon le nombre de bovins tuberculeux par élevage, en Mondalie de 2002 à 2010 (Source : services vétérinaires de Mondalie) (Abscisse 10 cumule les valeurs 10 et plus)



Le diagramme à lignes (figure A.4a) donne indiscutablement une allure plus lisse, qui convient mieux pour repérer des valeurs clés correspondant à des pourcentages particuliers.

Le diagramme à paliers (figure A.4b), qui représente le même jeu de valeurs, donne toutefois une représentation plus objective, dont on peut juger l'effet sur un jeu de données d'une épidémie (figure A.5).

Figure A.5 : Représentation des données d'une épidémie par un diagramme à colonnes et un diagramme à paliers (d'après Viriot *et al.*, 2011)



Le diagramme à colonnes (figure A.5a) met l'accent sur la colonne, le diagramme à paliers (figure A.5b) sur le contour.

[Retour au chapitre Diagrammes à paliers](#)

VIII - PRÉCISIONS SUR LA DÉNOMINATION « DIAGRAMME À COLONNES »

Les deux expressions, diagramme en colonnes et à colonnes, semblent utilisées de façon équivalente, sans raison évidente de privilégier l'une plutôt que l'autre. En revanche, celle de « diagramme colonnes » étant une ellipse des deux précédentes serait plutôt à éviter car relevant d'une forme d'expression relâchée. Nous avons expliqué antérieurement (cf. texte sur les [diagrammes](#)) les raisons de notre choix pour « à » plutôt que « en ».

La nomenclature des différents types de diagrammes, aussi bien française qu'internationale, n'est pas fixée. L'expression française « diagrammes en barres verticales », utilisée par certains, nécessite de préciser que les barres sont verticales, pour distinguer le cas où elles sont horizontales.

Comme une barre est un « *objet long et élancé servant à barrer, c'est-à-dire à empêcher le passage* », tandis qu'une colonne est un « *élément de forme verticale et étirée* », il n'est pas besoin de dénaturer les mots par un usage inapproprié. De plus, (pour une fois !) Excel[®] fait cette distinction.

C'est pourquoi, nous avons choisi de retenir, dans le premier cas (barres horizontales), l'expression « diagramme à barres », et dans le deuxième (barres verticales), celle de « diagrammes à colonnes » : ces expressions évitent toute ambiguïté, et elles respectent le sens premier des termes.

[Retour au chapitre "Diagrammes à colonnes"](#)

IX - CHOIX ENTRE DIAGRAMME À COLONNES ET HISTOGRAMME POUR UNE VARIABLE DISCRÈTE

- Si les amplitudes de classe sont égales, la surface et les effectifs sont proportionnels. Le choix du diagramme à colonnes est en cohérence avec la nature discrète de la variable ; le choix de l'histogramme est valide si on peut véritablement l'assimiler à une variable pratiquement continue du fait du grand nombre de valeurs regroupées en classes, sans dénaturation. L'histogramme est alors dit régulier, et ne pose pas de difficulté de réalisation, l'axe des ordonnées pouvant indifféremment porter les effectifs ou leur fréquence (rapport des effectifs au total du nombre d'observations).
- Si les amplitudes de classes ne sont pas égales et si l'on veut conserver le caractère mathématique de l'axe des abscisses, on est conduit à devoir réaliser un histogramme, pour lequel il faut utiliser un logiciel spécialisé, ou des paramètres spécifiques avec un logiciel classique ; si l'on peut renoncer au caractère rigoureusement mathématique pour se contenter d'une relation d'une approximation semi-quantitative, on peut utiliser un diagramme à colonnes. Pour cela, il faut transformer la variable de façon à ce qu'elle devienne véritablement ordinale : il faut que son caractère non mathématique soit perçu d'une façon suffisamment évidente pour le lecteur afin que ne subsiste aucune ambiguïté.

Si, pour l'axe des abscisses d'un diagramme à colonnes, on conserve un mélange de relation mathématique pour une partie des données, et de représentation de type ordinal pour les autres, on introduit une confusion source d'artefact, ce qui bien évidemment est à proscrire.

[Retour au chapitre "variables discrètes"](#)

X - PLUS D'INFORMATION SUR LA DISTINCTION ENTRE DIAGRAMMES À COLONNES ET DIAGRAMMES À BÂTONS

Pourquoi faut-il maintenir cette subtilité de distinction entre des diagrammes constitués tous deux d'éléments verticaux, colonnes ou bâtons ?

Le diagramme à bâtons est utilisé pour représenter des *variables quantitatives discrètes*. Les valeurs possibles sont celles figurant sur les graduations, car on est sûr qu'il n'existe pas de valeur intermédiaire. Il est donc logique qu'elles soient représentées à l'endroit qui en matérialise la valeur sur l'axe des abscisses.

Le diagramme à colonnes est utilisé pour les *variables qualitatives ordinales* ; il est possible de les ordonner selon un degré d'intensité, comme pour un indice de satisfaction par exemple, formulé par des mots (désaccord total, désaccord partiel, neutre, accord partiel, accord total). Du fait de cet ordonnancement, leur codification par des nombres est possible (dans notre exemple de 1 à 5) afin de mieux refléter la progressivité, mais elle n'a aucune signification arithmétique : on ne peut pas dire que les personnes qui ont formulé un accord partiel codé 4 sont quatre fois plus satisfaites que celles qui ont manifesté un désaccord total codé 1. C'est pourquoi ces valeurs, représentées par des nombres, doivent être placées *entre les graduations*, comme pour tout diagramme à colonnes représentant une variable qualitative. Il ne serait donc pas correct de placer les colonnes sur les graduations d'un axe ce qui alors symboliserait des valeurs quantitatives et pourrait être source de confusion dans l'esprit du lecteur.

Le diagramme à colonnes est aussi utilisé pour la représentation de *variables quantitatives discrètes*, lorsque les valeurs sont **regroupées en classes**. Il est donc logique de placer les colonnes *entre* les graduations pour matérialiser ce regroupement. Bien plus, l'espace entre les colonnes (que l'on doit éviter d'accoler les unes aux autres) matérialise l'absence de valeur intermédiaire entre les classes du fait qu'il s'agit d'une variable discrète. Dans ce cas, les graduations de l'axe des abscisses servent donc simplement à faire la distinction entre les classes, non à symboliser des valeurs auxquelles correspondraient des observations et les valeurs de l'axe des abscisses doivent également être placées entre les graduations.

[Retour au chapitre « Diagrammes à colonnes »](#)

XI - PLUS DE PRÉCISIONS SUR L'UTILISATION DES DIAGRAMMES À COLONNES

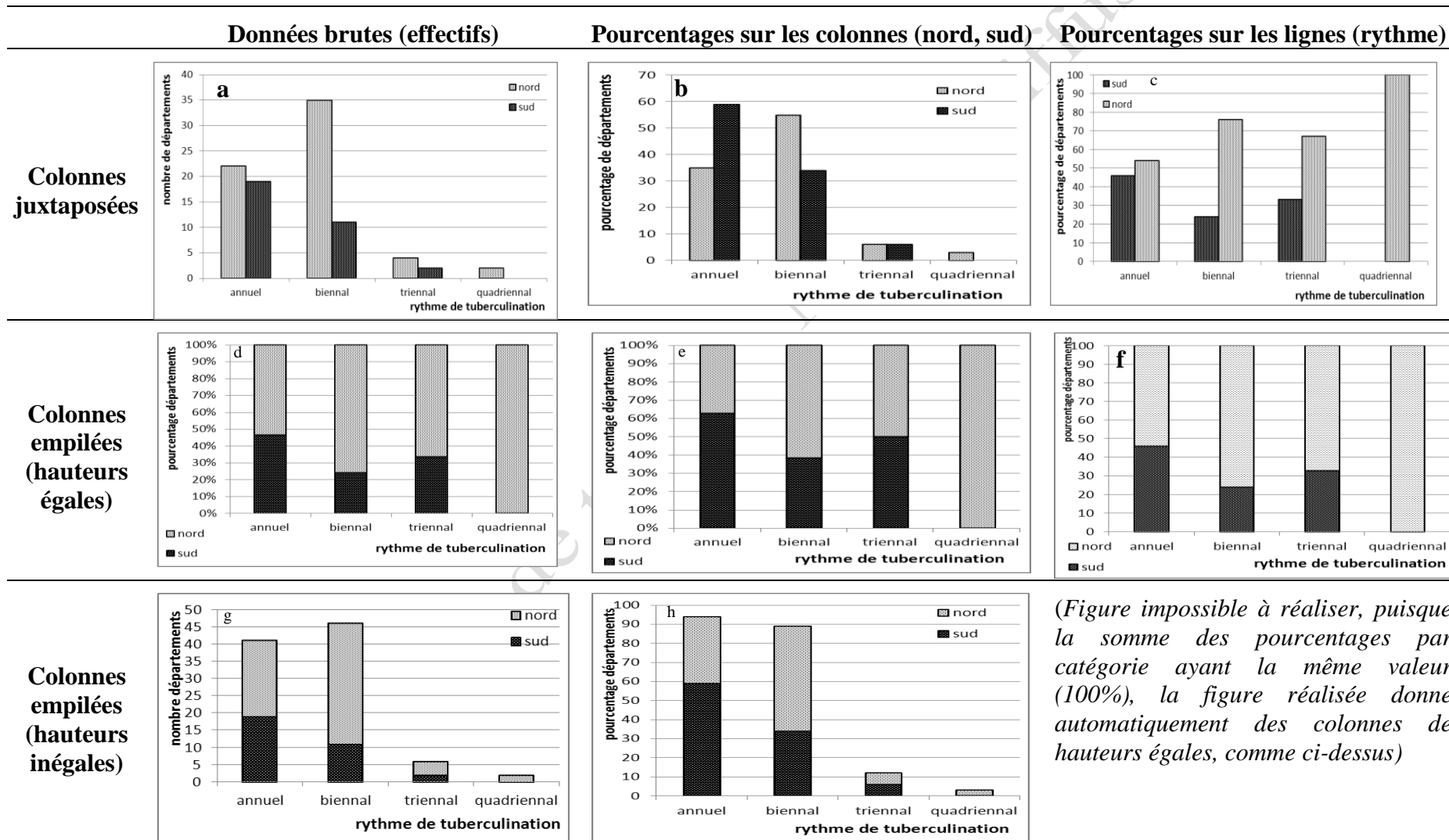
La représentation d'une seule série de données ne montre aucune différence si l'on utilise les données brutes ou les pourcentages correspondants. Dès que l'on représente plus d'une série, les représentations sont d'aspect différent selon que l'on choisit les données brutes ou les pourcentages (par série ou par catégorie) : comme Excel peut faire le choix spontanément du type de pourcentage, mieux vaut en être prévenu pour éviter d'obtenir un graphique qui ne correspond pas véritablement aux intentions. Nous allons prendre le jeu de données correspondant à la distribution des départements selon le rythme de dépistage pour illustrer ces difficultés.

La base de données brutes du tableau A.6 est illustrée par les différentes représentations de la figure A.6.

Tableau A.6 : Distribution du nombre de départements du nord et du sud de la France (par rapport à une ligne joignant Bayonne à Annecy) selon le rythme de dépistage de la tuberculose bovine par tuberculination (Données DGAI)

Rythme	nord	sud
annuel	22	19
biennal	35	11
triennal	4	2
quadriennal	2	0
total	63	32

Figure A.6 : Différentes modalités de représentation de la distribution des nombre de départements (données du tableau A.6) du nord et du sud de la France (par rapport à une ligne imaginaire joignant Bayonne à Annecy) selon leur rythme de dépistage de la tuberculose bovine par tuberculation en 1989 : avec les effectifs de chaque case du tableau ; avec les pourcentages des effectifs de chaque case calculés par rapport aux effectifs totaux de la colonne de la région correspondante (nord, sud) ; avec les pourcentages des effectifs calculés par rapport aux effectifs totaux de la ligne de la modalité correspondante de rythme (annuel, biennal, triennal, quadriennal)



Analyse des différentes représentations

La figure A.6 montre que selon les choix des données (données brutes, pourcentages en ligne ou en colonnes), ou de représentation par des colonnes juxtaposées ou empilées, on aboutit à des illustrations différentes. Certaines se ressemblent suffisamment pour qu'on puisse ne pas se rendre compte du choix effectué s'il est différent de celui escompté.

- Les colonnes juxtaposées permettent trois types de représentation : avec les données brutes (a), avec les pourcentages réalisés sur les données en colonnes (secteur géographique : b), ou réalisés sur les données en ligne (rythme : c).
- Les colonnes empilées peuvent être représentées avec des colonnes égales ou inégales ; on peut utiliser les données brutes ou les pourcentages, ceux-ci étant réalisés sur la totalité des valeurs de la colonne (région : nord ou sud), soit sur la totalité de la ligne (rythme), ce qui devrait conduire à 6 graphiques au total, mais celui des colonnes empilées pour les pourcentages, réalisées par catégorie, ne peut pas être réalisé en colonnes inégales du fait que les pourcentages par catégories ont toujours la même somme de 100 %, ce qui donne automatiquement des colonnes de hauteur égale.

Le choix de représentation par des colonnes empilées de hauteur égale à partir de données brutes (d) conduit automatiquement à la même figure que celle reposant sur les pourcentages réalisés par catégorie (f). Les deux figures utilisant des colonnes empilées de hauteur égale pour des pourcentages réalisés sur la série géographique (e) ou sur la catégorie (f) se ressemblent, mais ne sont pas identiques.

Le choix de représentation par des colonnes de hauteurs inégales conduit à des résultats différents selon qu'il s'agit de données brutes (g) ou de pourcentages effectués sur le total des effectifs de la colonne correspondant à une région (h).

Au total, on peut donc obtenir huit représentations différentes pour un seul tableau de données ! Les représentations peuvent être deux à deux fort différentes, voisines ou identiques ! Autant dire qu'il est capital de bien savoir ce que l'on veut illustrer pour choisir le (ou les) graphique(s) correspondant au propos de l'auteur.

Le meilleur choix selon l'objectif poursuivi

Il est certain que l'utilisation des **données brutes** ne facilite pas la comparaison entre les séries lorsqu'elles sont d'importance inégale. Mais d'un autre côté, cette représentation permet une comparaison globale de l'importance (en masse) respective de chacune des séries. La **figure a** est la plus neutre et peut être **recommandée** pour la représentation de données brutes. La figure d est interprétative, du fait de la standardisation par colonne : comme elle n'est pas adaptée du fait que les données brutes ne sont pas standardisées en vue d'une comparaison, elle est à éviter. La figure g est de lecture moins facile que la figure a, bien qu'elle utilise les mêmes données, car la comparaison des hauteurs de colonnes nécessite une gymnastique de l'esprit : elle est donc à éviter.

- Si le propos est d'étudier la distribution **conditionnellement à une série** (ici, les colonnes), pour les comparer entre elles, mieux vaut les rendre *comparables*, donc s'affranchir des écarts éventuels qui pourraient résulter des différences de taille de chacune des séries en utilisant les **pourcentages effectués sur l'effectif total de cette série** (illustrations centrales correspondant aux pourcentages effectués sur les totaux de colonne). La représentation par colonnes juxtaposées (b) respecte l'allure générale de la distribution de chaque série. De plus, les deux séries peuvent être

facilement comparées, du fait de la juxtaposition sur une même ligne horizontale (abscisses). Le choix de colonnes de hauteurs égales (e) dénature la comparaison des séries, car la cohérence de la série est perdue au profit du rapport pour une même catégorie après standardisation des importances respectives de chaque série d'interprétation plus difficile. Le choix de colonnes de hauteurs inégales (h) respecte les séries, mais la comparaison des séries est moins facile que dans la figure (b) : en effet, pour comparer les hauteurs, mieux vaut que les segments à comparer soit au même niveau. **Le meilleur choix serait donc celui de la figure b (colonnes juxtaposées construites à partir de pourcentages par séries)**. Les figures e et h sont à éviter.

- Si le propos est d'étudier la distribution *conditionnellement à une catégorie* (ici, le rythme), le meilleur choix est celui de la **figure d** (colonnes empilées réalisées à partir des données brutes – choix spontanément réalisé par Excel avec les colonnes empilées de hauteurs égales) ou, ce qui revient au même, la **figure f (à partir des pourcentages calculés par séries)** dont l'aspect est identique à la figure d. La figure c est à éviter.

Attention toutefois, car si l'on dispose d'un tableau des pourcentages selon la série, Excel peut en faire la représentation en colonnes empilées égales (e), laissant à croire qu'il s'agit d'une représentation du type de celle de la figure f, qui certes, pour le jeu de données utilisé, est proche, mais suffisamment différente toutefois pour bien révéler que ces représentations ne sont pas équivalentes : chaque pourcentage obtenu par calcul sur la colonne est rapporté par Excel au pourcentage correspondant de l'autre colonne pour une même ligne....

[Retourner au chapitre « Diagramme à colonnes »](#)

[Retourner au chapitre « Diagrammes à barres »](#)

[Retourner au chapitre "Comment réaliser un tableau adapté aux objectifs ?"](#)

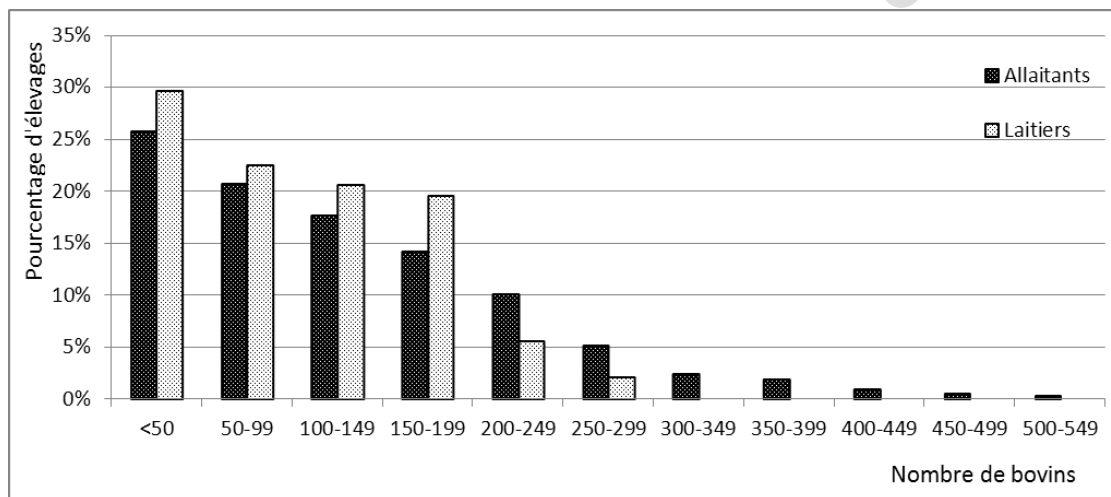
XII - EXTENSION DE L'APPLICATION DES DIAGRAMMES À COLONNES AUX VARIABLES QUANTITATIVES DISCRÈTES

Les variables qualitatives ordinales constituent l'indication majeure des diagrammes à colonnes. Mais ils peuvent aussi être utilisés pour les variables quantitatives discrètes, sous condition d'utiliser des classes d'amplitude égale afin d'éviter des erreurs pouvant fausser la représentation.

1. Regroupement en classes d'amplitude égale des valeurs d'une variable quantitative discrète

Le diagramme à colonnes est aussi utilisé pour les variables quantitatives discrètes, lorsque le nombre de valeurs possibles est trop élevé pour un diagramme à bâtons (Figure A.7), ou lorsque le traitement (mise en classes) altère le caractère quantitatif de la variable en abscisses (Figure A.8).

Figure A.7 : Distribution du nombre de bovins par élevage selon le type de production (bovins laitiers ou allaitants). (Services vétérinaires de Mondalie, 2010). (Mêmes données que celles de la figure A.3)



Bien qu'il s'agisse d'une distribution de fréquence, la représentation utilisée ici n'est pas un diagramme à bâtons, car le nombre d'abscisses aurait été beaucoup trop élevé, ce qui aurait nui considérablement à sa qualité. Les données ont été regroupées par classes de 50 individus. Il n'est donc plus possible de placer les bâtons *sur* les graduations, puisqu'elles ne correspondent pas à une seule valeur symbolisée par la graduation, mais à un ensemble de valeurs. C'est pourquoi, du fait du grand nombre de valeurs ainsi regroupées, le choix du diagramme à colonnes est judicieux, car il permet de placer les colonnes *entre* les graduations, concrétisant ainsi ce regroupement de valeurs.

L'amplitude des classes est constante, ce qui conserve à l'axe des abscisses sa valeur quantitative régulière, à défaut d'être continue (puisque toutes les valeurs ne sont pas possibles, seulement des valeurs entières).

Comparer avec la figure A.3, qui représente les mêmes données, mais avec un diagramme à lignes : le choix de l'une ou de l'autre de ces modalités est affaire personnelle, car il comporte des avantages et des inconvénients dans chaque cas.

Toutefois, il faut que les amplitudes de classes soient égales : dans le cas contraire, la représentation pourrait occasionner une déformation trompeuse (voir [paragraphe suivant](#)).

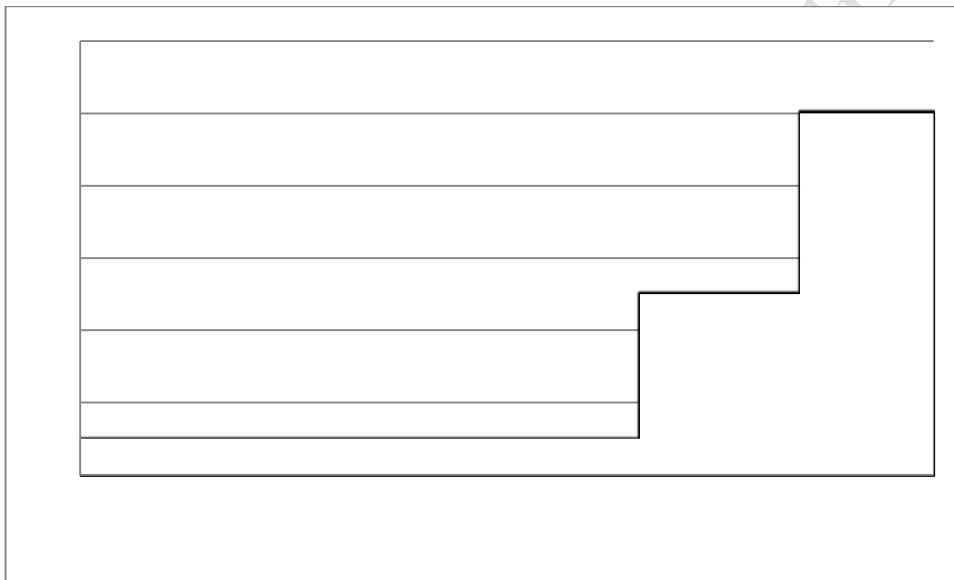
[Retour au chapitre "Diagramme à colonnes"](#)

2. Danger des classes inégales

Imaginez une distribution constante d'un nombre de cas pour chaque valeur de l'abscisse, par exemple de 0 à 100 (totalement improbable, mais c'est pour le besoin de la démonstration) : sa représentation donnerait une ligne strictement horizontale à un certain niveau.

Imaginez maintenant que pour des raisons tenant, par exemple, au fait qu'on est intéressé surtout par les valeurs de l'abscisse les plus basses (disons que c'est une partie de la population qui est l'objet de notre intérêt), nous modifions la répartition des valeurs : par exemple, nous conservons les données observées pour chacune des unités de 0 à 20, puis ensuite, nous regroupons les suivantes par paquets de 5 en 5 de 21 jusqu'à 50, puis de 10 en 10 de 51 à 100 (figure A.8).

**Figure A.8 : Distribution du nombre de bovins par élevage selon le type de production (bovins laitiers ou allaitants). (Services vétérinaires de Mondalie, 2010).
Effet d'un regroupement irrégulier de données *a priori* continues**



Vous obtiendrez trois paliers, le premier sur le niveau initial, le deuxième à un niveau 5 fois plus haut, le troisième à un niveau 10 fois plus élevé.

Ces paliers progressifs vont suggérer une progressivité du nombre de cas en fonction de la valeur de l'abscisse, ce qui est faux, nous le savons, compte tenu des conditions de départ. La progressivité du nombre de cas résulte simplement de l'opération de regroupement effectuée : il s'agit donc d'un *artefact*, c'est-à-dire d'un fait artificiellement créé en raison du traitement appliqué, et qui a déformé la nature réelle du phénomène qui, ici, était constant.

Cette progressivité suggérée résulte du fait qu'un axe des abscisses portant les valeurs d'une variable quantitative matérialise une échelle croissante, sur laquelle des opérations mathématiques sont possibles : une abscisse de 20 signifie une valeur 2 fois plus grande qu'à une abscisse de 10, ou 20 fois plus grande qu'à une abscisse de 1.

L'opération de regroupement des valeurs d'abscisses en a altéré la continuité mathématique, et l'a transformée en représentation de type ordinal. Sauf que le début de la série conserve son caractère mathématique ! Une telle représentation serait donc contradictoire.

Si nous devons procéder à des regroupements d'amplitude inégale, parce que le sens le commande, tout en conservant l'idée d'un diagramme à colonnes, nous devons alors respecter le principe d'une variable ordinale ([voir paragraphe suivant](#)).

[Retour au chapitre "Diagrammes à colonnes"](#)

[Retour à la fin du chapitre "Distinguer variables quantitatives et qualitatives"](#)

3. Transformation d'une variable discrète en variable ordinale

Si le regroupement en classes d'amplitude égale n'est pas possible, une solution possible est de transformer la variable discrète en variable ordinale.

Le tableau A.7, produit à partir des données d'une enquête réalisée sur les foyers de tuberculose apparus en France de 2005 à 2007, fournit les dénombrements des foyers observés et des départements correspondants en fonction du nombre de foyers par département.

Le nombre de foyers par département est une variable discrète, très irrégulière : de nombreuses valeurs de nombres de foyers par département ne sont pas représentées dans le tableau car elles ne correspondent à aucune observation. Pour éviter que le graphique ne comporte des vides inutiles, il est indispensable de regrouper les données. Mais nous avons vu au paragraphe précédent que les regroupements sur une base quantitative devaient être évités.

Tableau A.7 : Dénombrement des départements (n = 37) et du total des foyers (n = 165) en fonction du nombre de foyers par département. Enquête nationale sur la tuberculose bovine en France de 2005 à 2007, Camargue exclue (Du Breil, 2009)

Nombres de foyers par département	Nombre de départements	Nombre total de foyers pour cette fréquence
1	19	19
2	10	20
3	2	6
4	1	4
5	2	10
6	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	1	16
17	0	0
24	1	24
66	1	66

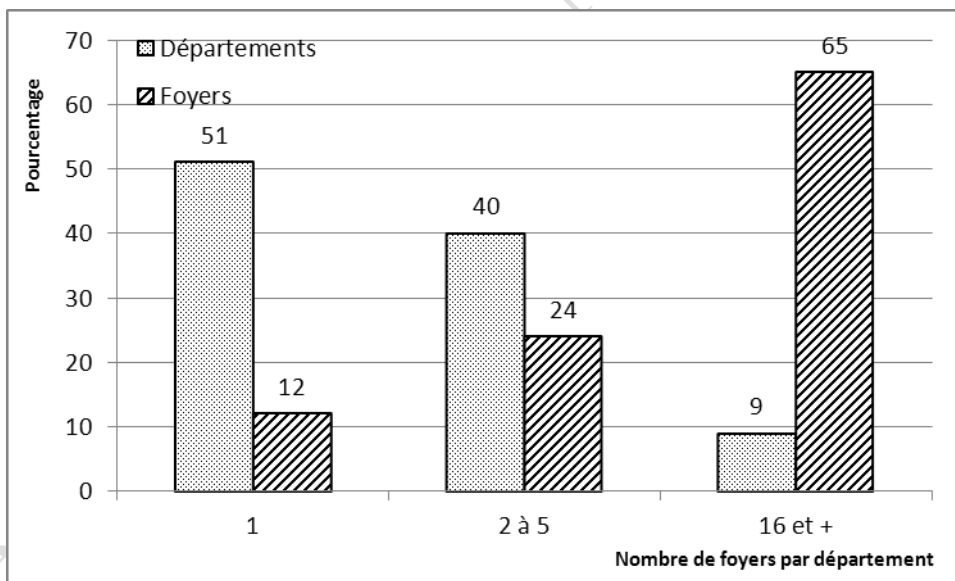
Quel commentaire suggère ce tableau ? Manifestement, les effectifs les plus élevés sont répartis de façon opposée pour le nombre de département et le nombre total de foyers en fonction du nombre de foyers par département. En fait, c'est cette notion que le graphique devrait illustrer.

Si, pour éviter de déformer les faits, nous regroupons les données non pas de manière quantitative, mais qualitative, nous pouvons opposer trois classes ordinales : « *un seul foyer* », « *quelques-uns* » (2 à 5), « *beaucoup* » (16 et plus, puisqu'entre 6 et 15, il n'y a aucune observation).

Pour permettre la comparaison entre les deux variables étudiées d'effectifs inégaux (37 et 165), nous en prenons les pourcentages, ce qui donne la figure A.9.

Cette représentation permet de facilement visualiser l'opposition des distributions : le pourcentage le plus élevé de foyers correspond au pourcentage le plus faible de départements, et réciproquement. Ce type d'opposition entre les distributions de deux variables comparées est appelée « *distribution de Pareto*⁴ » : 65 % des foyers correspondent à 9 % des départements, tandis qu'à l'inverse, 51 % des départements ne cumulent que 12 % des foyers ; en d'autres termes, la majorité des foyers sont regroupés dans un nombre limité de départements et la majorité des départements ne comprennent qu'un nombre limité de foyers.

Figure A.9 : Distributions comparées du pourcentage de foyers de tuberculose bovine en France selon le nombre de foyers par département et du pourcentage de départements correspondant à ce nombre de foyers par département (Du Breil, 2009 ; N = 165 foyers). Le total peut être supérieur à 100 à cause des arrondis.



[Retour au chapitre "Diagrammes à colonnes"](#)

[Retour à la fin du chapitre "Distinguer variables quantitatives et qualitatives"](#)

⁴ Du nom du sociologue et économiste Vilfredo Pareto, qui avait constaté que 80 % des richesses en Italie étaient détenues par 20 % de la population.

XIII - EN SAVOIR PLUS SUR LES HISTOGRAMMES

1. Pourquoi un histogramme est-il intrinsèquement différent d'un diagramme à colonnes ?

Les tableurs de bureautique ne savent pas faire d'histogramme (tout au moins pas facilement, sans recours à des astuces de programmation ou à des modules complémentaires ou « *add-ins* » adaptés) : ils se contentent de réaliser des diagrammes à colonnes dont la largeur est étendue de façon à assurer leur contiguïté.

Or, un histogramme n'est pas une forme particulière de diagramme à colonnes, qui résulterait de la contiguïté des colonnes, mais il s'agit d'une figure bien plus complexe et ne peut leur être assimilé. En effet, pour tous les autres diagrammes, la relation est simple entre les valeurs qui permettent la partition de l'ensemble des données et les effectifs (ou leurs pourcentages) correspondants : la *longueur* du bâton, de la colonne ou de la barre est proportionnelle aux effectifs. Pour un histogramme, c'est la *surface* représentée qui est proportionnelle aux effectifs, ce qui complique singulièrement les choses, sauf dans le cas d'un histogramme *régulier* !

Un histogramme est dit « **régulier** » lorsque les classes ont la **même amplitude**, car, dans ce cas particulier, la surface et la hauteur de la bande sont proportionnelles aux effectifs. Pour un histogramme dont les classes n'ont pas la même amplitude, ce sont les *surfaces* des rectangles qui doivent être proportionnelles aux effectifs, et par conséquent l'ordonnée devrait porter une *fonction de densité*, d'où l'utilisation de la fréquence ($f_i = n_i/n$) plutôt que les effectifs. La lecture d'articles spécialisés (Régnier, 1998) illustre la richesse mathématique de ce concept qui dépasse largement le cadre de ce document tout en démontrant l'usage le plus souvent abusif et inapproprié des histogrammes.

Notre besoin n'est pas celui de l'orthodoxie mathématique (qui est toutefois nécessaire dans les travaux de modélisation qui imposeront donc de réviser sérieusement les conceptions sommaires exposées ici), mais d'une représentation facilement perceptible et compréhensible, mettant simplement en relation les effectifs avec la variable étudiée, comme le font tous les autres diagrammes, ce qui induit automatiquement ce type de représentation dans l'esprit des usagers et rend plus difficile la perception exacte de la nature de l'histogramme. Nous avons donc fait le choix de ne traiter que le cas particulier où *toutes les classes ont la même amplitude* : de ce fait, la longueur de la bande est proportionnelle à la fois à la surface et à l'effectif. Pour toute situation nécessitant impérieusement le respect des principes de la construction d'un véritable histogramme, il faudra se reporter à des outils spécifiquement adaptés (modules complémentaires pour XL, feuilles XL programmées, Epi-info, R).

Par conséquent, dans ce document, tout doit être compris comme ne s'appliquant qu'au cas particulier mentionné permettant de répondre à nos besoins spécifiques en épidémiologie et dans le cadre du public défini en introduction et non une présentation générale de la notion d'histogramme.

[Retour au chapitre « Histogramme »](#)

2. Comment déterminer le nombre de classes d'un histogramme ?

Il faut éviter d'ajuster le nombre de classes selon l'allure du graphique finalement observée, ce qui serait une entorse grave au principe de description *objective*, c'est-à-dire indépendante de la subjectivité de l'observateur alimentée par la perception du graphique.

Le choix du nombre de classes doit, par conséquent, résulter de l'application d'une règle indépendante de la subjectivité de l'observateur.

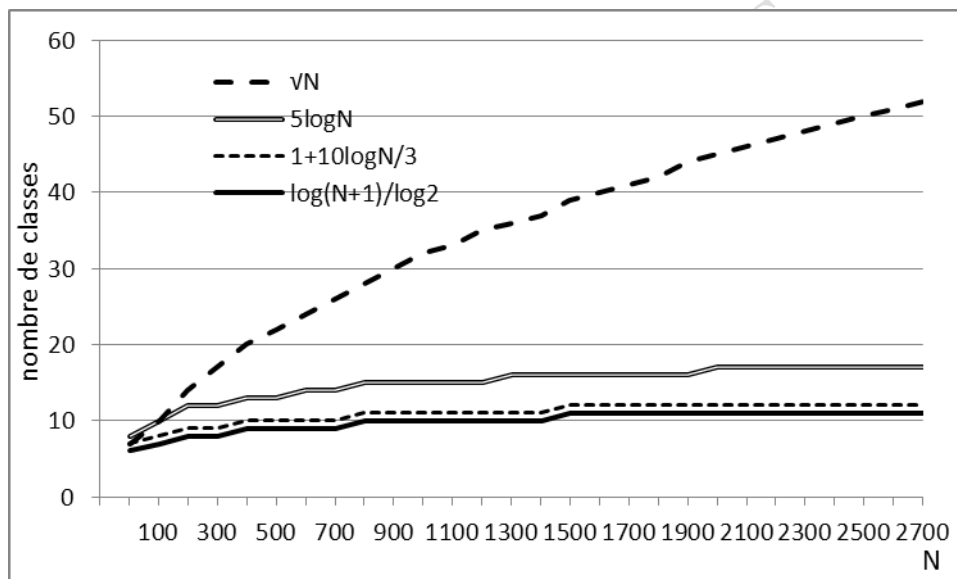
On utilise une des formules proposées par divers auteurs :

k (le nombre de classes) =

- \sqrt{N}
- $5 * \log_{10}(N)$ (Brooks-Carruthers)
- $1 + (10 \log_{10}(N))/3$ (Huntsberger)
- $\log_{10}(N+1)/\log_{10}(2)$ (Sturges)

La Figure A.10 montre les différences de performance de ces méthodes de calcul pour des échantillons comportant entre 30 et 300 individus. La formule $5 * \log N$ constitue une solution intermédiaire, qui peut convenir pour une première approche. Selon les besoins, on pourra être amené à choisir une autre solution.

Figure A.10 : Comparaison des nombres de classes obtenues avec quatre formules courantes



[Retour au chapitre « Histogramme »](#)

3. Comment réaliser en pratique un histogramme avec Excel® ?

Même si le terme « histogramme » est utilisé par ce logiciel, il s'agit en fait de diagramme à colonnes, qu'il faut donc adapter pour que la représentation soit conforme à un histogramme. Nous n'évoquerons que le cas de l'**histogramme régulier**, pour lequel toutes les classes sont égales.

Le Tableau A.8 donne la distribution des valeurs, organisées par classes d'un millimètre, des épaisseurs de pli de peau mesurées lors de la réalisation d'une tuberculination de 189 bovins charolais. Remarquez que, par convention, la borne supérieure d'une classe est exclue.

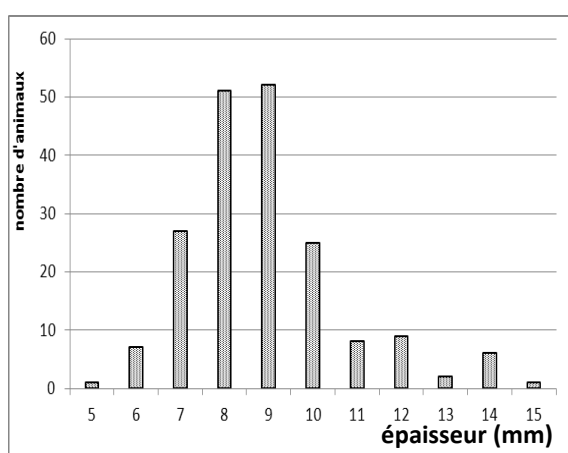
Tableau A.8 : Distribution du nombre de bovins charolais selon l'épaisseur de leur pli de peau (N = 189)

Borne inférieure	Nombre
5	1
6	7
7	27
8	51
9	52
10	25
11	8
12	9
13	2
14	6
15	1

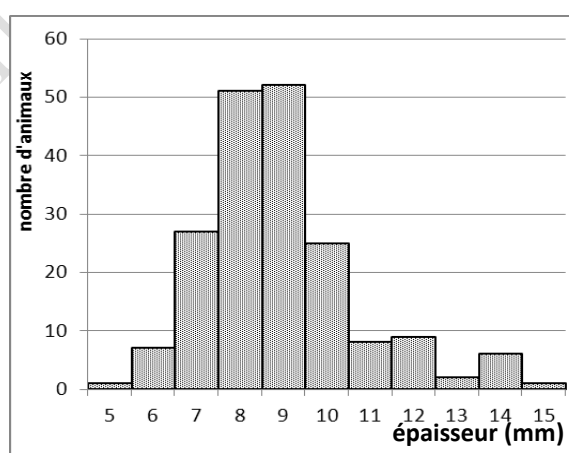
La Figure A.11 montre le graphique obtenu par défaut (on peut le dire !) avec Excel®, c'est-à-dire sans avoir procédé à un paramétrage particulier. Nous avons juste complété les légendes.

Figure A.11 : Distribution des épaisseurs de pli de peau de bovins (N = 189) (d'après Gane, 2011)

a) Diagramme par défaut avec Excel



b) Après élargissement des colonnes



La représentation (a) de la figure A.11 a été obtenue avec Excel® en sélectionnant dans le menu d'insertion d'un graphique le choix « colonnes », puis « histogramme », qui a fourni un diagramme en colonnes. Pour rendre les colonnes jointives (comme en b de la figure A.11) : *clic droit* sur l'une des colonnes, *mettre en forme une série de données*, *options des séries*, *largeur de l'intervalle* (mettre à zéro).

La représentation (a) fournie par Excel® est inexacte, car les valeurs des intervalles de classes sont placées entre les graduations ; elle introduit une ambiguïté, du fait qu'elle laisse à penser que la valeur indiquée dans l'intervalle de classe est la valeur centrale de la classe, les bornes étant par conséquent situées à $-0,5$ et $+0,5$ autour de cette valeur, alors qu'en fait l'intervalle est délimité par des valeurs entières, la valeur du milieu de classe devant être estimée en ajoutant $0,5$ à la valeur ainsi affichée.

Pour positionner les valeurs des abscisses *entre les graduations*, une solution rapide consiste à déplacer les valeurs de l'axe (*clic* sur l'échelle des abscisses), en les justifiant à gauche (*format paragraphe*). Le résultat est imparfait, car les valeurs ne sont pas *exactement* sur la graduation, mais légèrement à droite. De plus, il manque la valeur supérieure, ce que l'on peut facilement compenser en ajoutant une ligne au tableau avec la valeur immédiatement supérieure pour l'abscisse (16), et zéro pour les données, puis en étendant la sélection des données. Le résultat final (non reproduit) n'est quand même pas satisfaisant.

Pour corriger ce défaut, il suffit de porter en abscisses les valeurs des intervalles qui seront positionnées *entre* les graduations, comme reproduit dans la figure A.12a : pour cela, créer une colonne avec les valeurs des intervalles, et les choisir comme valeurs à porter en abscisses.

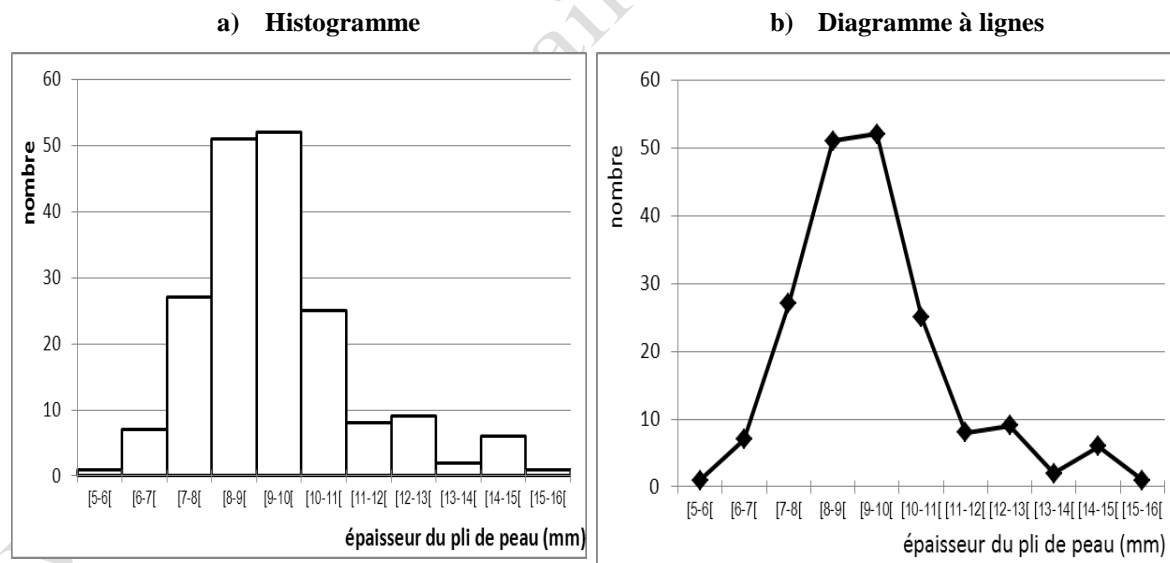
Pour éviter l'effort fastidieux d'entrer à la main les valeurs des bornes des intervalles, on peut facilement utiliser les possibilités de traitement de chaînes de caractères d'Excel® [\[En savoir plus\]](#)

[Retour au chapitre « Histogramme »](#)

4. Comparer un histogramme et un diagramme à lignes

La figure A.12 comporte bien en abscisses les indications des intervalles. En plus, pour respecter le principe d'économie d'encre, les grisés ont été supprimés (figure A.12 a).

Figure A.12 : Distribution du nombre de bovins charolais selon l'épaisseur de leur pli de peau (N = 189) (représentation corrigée, portant les valeurs des intervalles en abscisses) (d'après Gane, 2011)



L'histogramme (ou un diagramme à colonnes ; figure A.12a) est plus objectif que le diagramme à lignes (figure A.12b) : les segments de droite reliant les points constituent indiscutablement des interprétations critiquables.

[Retourner au chapitre "Histogramme"](#)

5. Indications, contre-indications d'un histogramme : en savoir plus

Une telle représentation graphique permet de juger de l'opportunité de transformer une variable, afin d'en « normaliser » la distribution et ainsi permettre l'utilisation de la

statistique paramétrique sur les valeurs transformées. D'une façon plus générale, il s'agit d'un outil permettant d'ajuster mathématiquement les données à une loi de probabilité.

Rappelons la mise en garde contre tous usages abusifs de prétendus histogrammes : pour représenter des effectifs, mieux vaut utiliser des diagrammes à colonnes (quitte à ce que les colonnes soient contigües pour souligner le caractère continu de la variable) et ne pas les dénommer « histogrammes », en prenant soin que les amplitudes de classes soient identiques. Si les amplitudes de classes ne sont pas identiques, il faut utiliser un histogramme, en respectant alors la condition impérative de correspondance de la surface avec les fréquences représentées, ce qui avec un tableur comme Excel nécessite quelques calculs complémentaires (et laborieux) pour y parvenir.

[Retourner au chapitre "Histogramme"](#)

XIV - EN SAVOIR PLUS SUR LE CHOIX DU NOMBRE OU DU POURCENTAGE DE CAS POUR UNE REPRÉSENTATION CARTOGRAPHIQUE

Supposons que nous disposions des données épidémiologiques à l'échelle d'une zone géographique donnée, en ayant toute l'information nécessaire dans un fichier : la première réaction consisterait à pointer les cas sur une carte « pour voir une accumulation éventuelle de cas ». Une telle accumulation pourrait peut-être apparaître, mais on ne saurait pas finalement si la concentration des cas ne résulterait pas tout simplement d'une plus grande densité des individus.

Cette façon de procéder consiste à n'utiliser que le numérateur (*les cas*) et à oublier le dénominateur (*la population*). Mais on peut *rappor*ter ces cas à des unités géographiques pour lesquelles on dispose des effectifs de population nécessaires au calcul de l'indicateur approprié. L'unité dépend des préoccupations de l'observateur : le canton, la commune, le département, ou une zone définie selon un critère d'intérêt (*zone climatique, zone d'influence d'un prestataire de service, etc.*).

La représentation graphique correspondante consiste alors à déterminer des seuils permettant de transformer les valeurs de rapports en classes. Pour chaque classe de valeurs, on utilise un code de densité de couleur ou de symboles permettant de traduire l'intensité du phénomène.

Exemple :

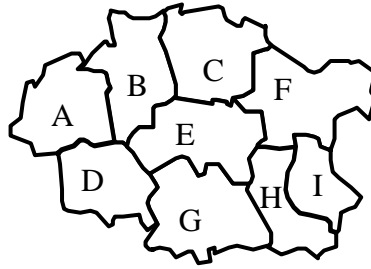
Le tableau A.9 fournit des données fictives pour 9 zones (A à I) correspondant à des nombres de cas ($n+$), la population totale de la zone considérée (N) et les pourcentages correspondants.

Tableau A.9 : Distribution géographique du nombre de cas selon la zone géographique (données fictives)

Zone	n+	N	%
A	20	100	20%
B	40	200	20%
C	60	1200	5%
D	60	300	20%
E	60	1000	6%
F	140	3000	5%
G	80	4200	2%
H	160	3200	5%
I	180	7200	3%

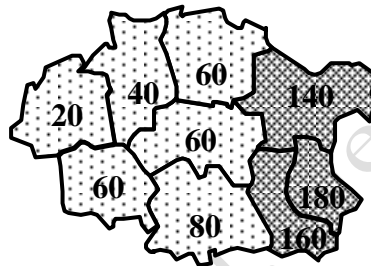
La représentation géographique des zones est donnée dans la figure A.13.

Figure A.13 : Représentation des zones géographiques du tableau A.9 (données fictives)



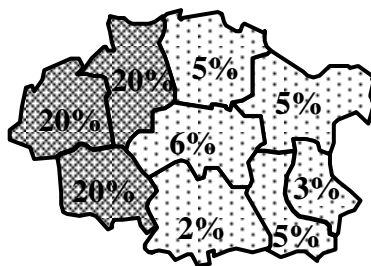
Les données du tableau ont été reportées sur la carte figurant ci-dessous, avec une mise en valeur en densité de gris par rapport au seuil moyen de 90 cas par zone. Cette représentation suggère une accumulation des cas en zones est.

Figure A.14 : Représentation de la distribution géographique des cas du tableau A.9 (données fictives ; effectifs absolus)



Dans la représentation de la figure A.15, ce sont les valeurs de pourcentages de cas par zone qui ont été reportées, en utilisant comme seuil le pourcentage moyen de 9 % pour la mise en valeur selon la densité de gris. Cette fois, la représentation suggère une accumulation de la répartition plutôt du côté ouest de la région.

Figure A.15 : Représentation de la distribution géographique des cas du tableau A.9 (données fictives ; pourcentages)



Cet exemple simpliste permet de souligner plusieurs difficultés pour procéder à une représentation satisfaisante dans l'espace.

Il faut tout d'abord déterminer différentes zones géographiques, comme autant d'unités dans lesquelles il sera procédé à la représentation appropriée. Les découpages administratifs ne sont pas forcément les plus pertinents ; on peut choisir des limitations selon d'autres caractéristiques, d'ordre géographique (*cultures, climat, sol, etc.*). En l'absence d'hypothèse, on peut aussi déterminer des zones géographiques en constituant des cellules de même taille (*carrés, hexagones*).

Ensuite, il faut disposer des données nécessaires pour les numérateurs et les dénominateurs de chaque cellule, en disposant de la même qualité d'observation : pour chaque individu, il faut avoir la même fiabilité du classement en atteint ou non atteint et pas se contenter du dénombrement des cas connus spontanément.

Il faut choisir la durée la plus appropriée pour la description, en fonction de l'incidence : durée courte (*la semaine par exemple*) pour une maladie épizootique, durée longue (*mois, trimestre, voire année*) pour une maladie rare.

Puis, il faut déterminer les seuils de classes qui permettront la représentation. Le nombre de classes doit être nécessairement limité (une demi-douzaine au maximum) pour la lisibilité de la représentation.

Le code visuel utilisé pour la représentation des classes d'incidence ou de prévalence doit être adapté à la progression des valeurs de ces classes : pour une représentation en noir et blanc, d'une trame la plus légère (*faible incidence*) à la trame la plus chargée (*forte incidence*). Il convient de rester prudent pour des représentations en couleurs variées qui, en photocopie, peuvent être illisibles (*cf. annexe Présentation des tableaux et des figures*).

Enfin, la représentation graphique pourra s'accompagner d'un jugement statistique pour déterminer si la perception visuelle d'une accumulation est bien légitime.

[\[Pour retourner au chapitre sur les cartes\]](#)

I - FAIRE DES CALCULS AVEC UNE VARIABLE NUMÉRIQUE EN ÉVITANT LE FORMAT TEXTE

Cela peut paraître absurde d'indiquer comment faire des calculs sur une variable numérique avec Excel, dont c'est justement l'indication première. Pourtant, la méconnaissance des subtilités de ce logiciel est la première cause d'erreurs dont Excel peut être responsable. En effet, dans certaines circonstances, les chiffres d'une cellule peuvent très bien ne pas être utilisés pour un calcul, car ils ne sont pas reconnus comme tels, mais comme du *texte* : l'opérateur croit que toutes les données d'une colonne sont utilisées pour le calcul, car, sous son regard, ce sont bien des nombres qui sont alignés, mais, pour une raison qu'il ignore sur l'instant, certaines cellules peuvent être sautées par le logiciel, qui considère qu'elles contiennent du texte, impropre au calcul.

C'est pourquoi, il faut être attentif au *format* de la cellule : pour que les calculs puissent être réalisés, il faut que la cellule soit réputée contenir des nombres. On peut le vérifier facilement, car, à moins que l'opérateur ait modifié les conditions d'alignement des contenus des cellules, les contenus numériques et reconnus comme tels sont alignés à *droite*, tandis que les contenus reconnus comme du texte sont alignés à *gauche*.

Comment est-il possible que dans une colonne supposée contenir des valeurs numériques, certaines cellules puissent être au format texte ? Cet incident ne peut survenir avec une feuille vierge dans laquelle on effectue des saisies pour la première fois. Mais, dans une feuille comportant déjà des valeurs texte et des valeurs numériques, cet incident peut survenir, comme vous allez pouvoir le vérifier avec l'exercice suivant.

Dans une feuille Excel, créez une colonne de texte en insérant une succession de lettres : Excel les reconnaît comme du texte et elles seront alignées à gauche. D'ailleurs, sur le ruban « Accueil », consacré au format, vous pouvez constater que dans le pavé du ruban appelé « Nombre », le mot « texte » s'affiche dans le menu déroulant, ce qui confirme le format de la cellule.

Dans la colonne située immédiatement à droite, insérez une succession de nombres : ils seront automatiquement reconnus comme des valeurs numériques, et alignés à droite ; le format affiché est « Standard », qui permet les calculs.

Maintenant, supposons que vous vouliez introduire une colonne de chiffres entre la colonne de texte et la colonne de chiffres que vous venez de saisir : vous vous placez *sur* la colonne de chiffres, et insérez une colonne (« Ctrl + »), ce qui décale la colonne de chiffres d'une colonne à droite. Saisissez une succession de chiffres, de la même façon que précédemment, c'est-à-dire directement (comme on le fait spontanément sur Excel) : vous constatez que vos chiffres sont alignés cette fois-ci à gauche, et que la cellule comporte un petit triangle vert dans son coin gauche ; si vous cliquez sur la cellule, une bulle contenant un point d'exclamation attire votre attention ; si vous cliquez sur cette bulle, une information est affichée : « *Nombre stocké sous forme de texte* », ce qui empêche tout calcul.

Sur une colonne comportant ce petit triangle vert, l'attention de l'opérateur est attirée, mais si vous avez inséré une cellule, ou procédé à un copier / coller, votre attention ne sera pas forcément attirée sur le fait que dans une longue colonne de chiffres, une ou plusieurs cellules ont le format texte, alors que les données affichées sont bien des nombres. Si l'opérateur ne comprend pas que l'alignement à gauche est un signal d'alerte

et qu'il corrige ce défaut d'alignement, il aura masqué cette erreur : seul le petit triangle vert persistera encore.

Pour rectifier, on pense qu'il suffit d'assigner le bon format (Partie du ruban « Nombre »/ menu déroulant « nombre ») : rien n'y fait, le petit triangle persiste et la mention texte demeure de façon incompréhensible ! Pour résoudre cette difficulté, il faut se positionner dans la cellule de saisie, à gauche du nombre et supprimer une espace constituée d'une apostrophe non visible : après cette manœuvre, la cellule est correctement formatée.

Pour effectuer cette manœuvre de manière automatique sur toute une série de données, il faut utiliser la fonction de traitement de chaîne de caractères « CNUM() » (sans doute pour « copie numérique ») qui copie le contenu d'une cellule au format texte dans une autre cellule en la transformant en valeur numérique.

Ex : la cellule B3 contient le chiffre 5 malheureusement formaté en texte. Pour en reporter la valeur numérique dans la cellule C3, il suffit de se placer dans cette cellule C3 et de taper :

=CNUM(B3).

La cellule de réception (C3) contiendra bien la valeur 5 au format numérique.

[Retour au chapitre des variables texte](#)

II - PLACER LES BÂTONS OU LES COLONNES D'UN DIAGRAMME EXACTEMENT SUR LES GRADUATIONS (ET NON DANS LES INTERVALLES, CE QUE FAIT EXCEL® PAR DÉFAUT)

Pour réaliser un diagramme à bâtons avec Excel®, il faut choisir le type « diagramme à colonnes », puis « histogramme » et en donnant la largeur minimale (option des séries « *intervalle large* »).

On voit ici une manifestation de la confusion fréquente qui règne dans la dénomination des graphiques, y compris dans des outils complexes comme ce tableau, qui assimile diagrammes colonnes et histogrammes.

Par défaut, comme pour tout diagramme à colonnes, Excel® positionne les valeurs *entre* les intervalles.

Pour obtenir le positionnement sur les valeurs entières des abscisses avec Excel, il faut réaliser les manœuvres suivantes : sur une des valeurs de l'axe des abscisses, clic droit / *Mise en forme de l'axe* / *Position de l'axe* / sur les graduations.

Pour obtenir une épaisseur suffisante, il suffit, par la même commande, de donner une dimension d'intervalle satisfaisante, par exemple 50 %.

Pour éviter que les première et dernière colonnes soient tronquées, il peut être nécessaire d'étendre la sélection des données (pour bâtir le graphique), après avoir ajouté une ligne vide au-dessus et en dessous du tableau de données.

[Retour au chapitre "Variables quantitatives"](#)

[Retour au chapitre "Diagramme à bâtons"](#)

III - PLACER LES COLONNES D'UN DIAGRAMME ENTRE LES GRADUATIONS

Par défaut, Excel donne ce type de positionnement. S'il est nécessaire de le rétablir : sur une des valeurs de l'axe des abscisses, clic droit / *Mise en forme de l'axe* / *Position de l'axe* / entre les graduations.

[Retour au chapitre "Variables quantitatives discrètes"](#)

IV - RÉALISER UN DIAGRAMME À PALIERS AVEC EXCEL®

Choisir « histogramme » : le graphique représentant les données s'affiche.

Clic droit sur une des colonnes : « *Mettre en forme une série de données* ».

Options des séries : « *Superposition* » : 0 ; « *Largeur d'intervalles* » : 0.

Remplissage : aucun.

Couleur de bordure : trait plein. Transparence 100 %.

Style de bordure : choisir l'épaisseur.

Ombre : décalage vers le haut ; transparence : 0 ; taille : 100 ; flou : 1 pt ; angle : 100 ; distance : 0 pt.

Eclats et contours adoucis : Variation de lumière : choisir une première ligne ; couleur : noir ; taille 2 pt ; transparence : 0%

☞ *Ne pas oublier d'enregistrer ce type de diagramme comme modèle pour éviter d'avoir à répéter ces opérations !*

[Retour au chapitre "Diagramme à paliers"](#)

V - UTILISER LES FONCTIONS DE TRAITEMENT DE CHAÎNES DE CARACTÈRE POUR FACILEMENT RÉALISER LA LÉGENDE DE L'AXE DES ABCISSES POUR UN HISTOGRAMME

On veut réaliser la figure A.12 dans laquelle les données sont regroupées de millimètre en millimètre : [5-6[; [6-7[, etc, sans avoir à saisir cette succession fastidieuse de signes.

Dans une feuille Excel, on place dans une première colonne la succession des valeurs inférieures de l'intervalle de classe (dans cet exemple, de 5 à 15). Nous supposons dans cet exemple que le tableau est commencé en haut et à gauche de la feuille, ce qui fait que la première ligne comporte les titres de colonnes, et la deuxième ligne la première ligne des données ; la première colonne est numérotée A.

La colonne suivante (B) comportera l'indication de l'intervalle de classe, grâce à la formule suivante que l'on saisit dans la cellule B de la ligne 2 :

CONCATENER("["; A2 ; "-" ; A3 ; "[")

Les guillemets sont indispensables pour introduire un caractère dans la formule.

La cellule A2 est la première de la ligne : elle donne la limite inférieure de l'intervalle.

La cellule A3 est la première de la ligne suivante : elle donne la limite supérieure de l'intervalle. On recopie cette formule vers le bas.

[Retour au chapitre "Comment réaliser un histogramme avec Excel"](#)

[Retour au corrigé de l'exercice "Poids de carcasses de sangliers"](#)

VI - DANS UNE FIGURE, SÉPARER LE SIGNE « % » DES VALEURS CHIFFRÉES CORRESPONDANTES

Par convention dans un texte rédigé en français, le signe « % » doit être séparé des chiffres par une espace : on doit écrire « 10 % » et non « 10% ». En anglais, la convention ne prévoit pas d'espace. Ce qui fait que, par défaut, les logiciels sont paramétrés selon cette dernière conception. Pour respecter la convention en vigueur en français, il faut donc procéder aux réglages nécessaires correspondant à cette mise en forme.

Pour une figure, c'est simple... à condition de savoir comment faire !

Pour mettre en forme un axe en vue de respecter cette convention :

- pointer sur l'axe à mettre en forme / clic droit : « *mise en forme l'axe* » ;
- choisir « *Nombre / personnalisé* » ;
- choisir le format suivant (ou le saisir puis « *ajouter* ») : 0" "% (soit zéro, guillemet, espace, guillemet, %).

Pour mettre en forme les étiquettes de valeur, au cas où elles ont été insérées dans le graphique :

- pointer sur une des valeurs / clic droit : « *mettre en forme les étiquettes de données* »
- même procédure que précédemment.

[Retour au chapitre mise en forme d'une figure ou d'une carte](#)

Version de travail. Ne pas diffuser

LIVRE TROISIEME : EXERCICES ET CORRIGES

Version de travail. Ne pas diffuser

Version de travail. Ne pas diffuser

CHAPITRE I - ENONCES

Ces exercices permettent de mettre en application les recommandations pour la présentation des tableaux et des figures.

I - TABLEAUX

A - OBJECTIF

Être capable de reconnaître les critères de qualité et les défauts d'un tableau.

Instructions

Tous ces exemples sont empruntés à des thèses de doctorat vétérinaire. Le titre d'origine est reproduit en italique. La référence aux auteurs a été volontairement omise.

Pour chaque sujet, les questions sont les mêmes :

- *Dites pourquoi le tableau est (ou non) adapté ?*
- *Si besoin, apportez les modifications qui vous semblent nécessaires.*

Les tableaux sont numérotés en fonction du numéro de l'exercice.

B - ENONCES

1. Enquête Facco

Tableau 1 : enquête FACCO 2012, nombre d'individus, proportion d'animaux de race et taux de stérilisation

	Chiens	Chats
Nombre d'individus (en millions)	7,42	11,41
Proportion d'animaux de race	70 %	25 %
Taux de stérilisation	37 %	77,2 %

2. Etude multivariée des résultats d'une enquête sur l'attachement des propriétaires de chat à leur animal

245 propriétaires ont accepté de répondre à un questionnaire de 27 questions (dérivé du modèle « *Pet attachment Scale* » ou PAS ; réponses selon une échelle de Likert), permettant de leur attribuer un score d'attachement, et de mettre ce score global en relation avec différentes variables (démographiques, comportementales). Une analyse en composantes principales (ACP) a été réalisée, qui a conduit à identifier les 10 premiers axes (« valeurs propres ») permettant de synthétiser au mieux l'ensemble des données.

Tableau 2 : Valeurs propres, valeurs propres cumulées, pourcentages de variance et pourcentages de variance cumulés pour les 10 premiers facteurs pour les chats

	Valeur propre	% total de variance	Valeur propre cumulée	% de variance cumulé
Facteur 1	16.20924	27.01540	16.20924	27.0154
Facteur 2	5.04417	8.40695	21.25341	35.4224
Facteur 3	3.51841	5.86401	24.77182	41.2864
Facteur 4	2.18075	3.63458	26.95257	44.9209
Facteur 5	2.06445	3.44074	29.01701	48.3617
Facteur 6	1.82318	3.03864	30.84020	51.4003
Facteur 7	1.68485	2.80809	32.52505	54.2084
Facteur 8	1.60604	2.67674	34.13110	56.8852
Facteur 9	1.48191	2.46985	35.61301	59.3550
Facteur 10	1.33719	2.22865	36.95019	61.5837

3. Description d'un échantillon d'enquête (une variable)

Une enquête a été réalisée sur 556 personnes. La description porte sur la répartition par sexe.

Tableau 3 : Pourcentage de femmes et d'hommes parmi les répondants

Sexe	Effectif	%
Femmes	242	44%
Hommes	314	56%
Total	556	100%

4. Description d'un échantillon d'une enquête (deux variables)

550 vétérinaires ont accepté de répondre à une enquête professionnelle. Le tableau 4 décrit leur répartition entre secteurs publique et privé en fonction du sexe.

Tableau 4 : Répartition des genres par secteur

	Privé hommes	Privé femmes	Public hommes	Public femmes
Nombre	152	97	158	143
	249		301	
Pourcent	61%	39%	52%	48%

II - FIGURES

A - OBJECTIF 1 :

Être capable de choisir un type de diagramme pour une illustration donnée.

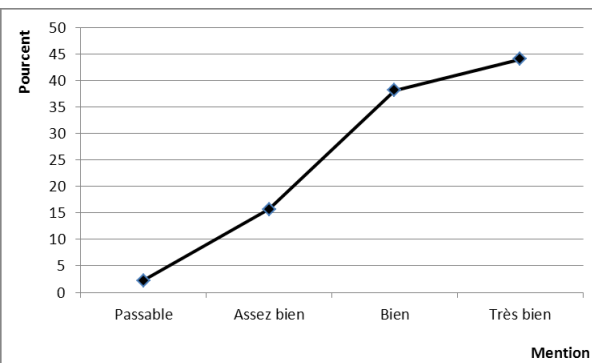
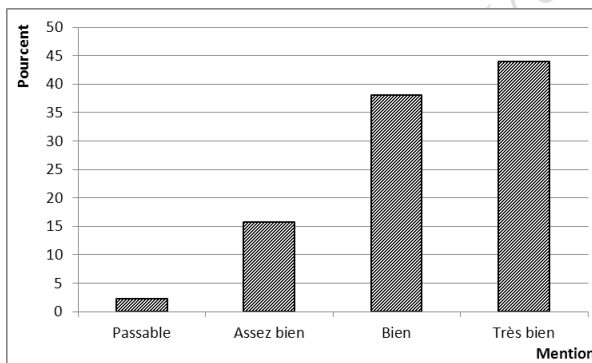
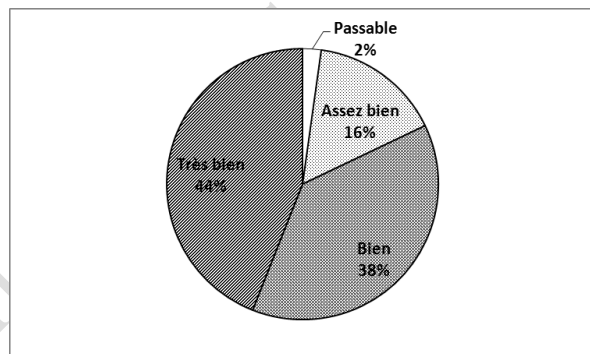
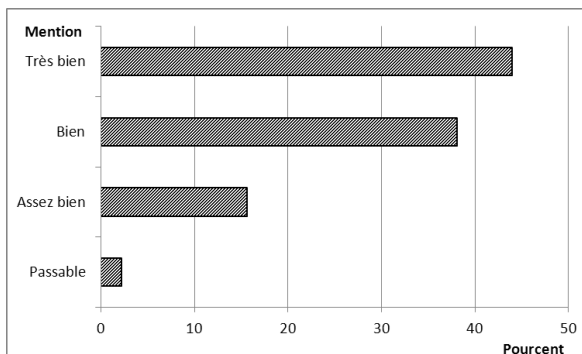
Instructions

Pour tous les exercices, quatre (parfois six) figures sont proposées. Elles sont implicitement référencées de la façon suivante : ligne 1 : a et b ; ligne 2 : c et d, *etc* (lecture de gauche à droite).

Il conviendra de choisir la figure qui semble convenir pour illustrer le propos et de justifier votre réponse.

1. Mentions au bac

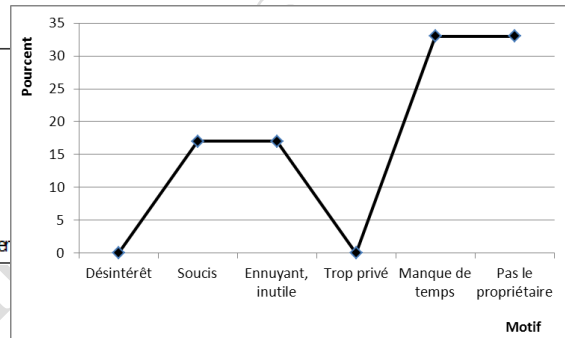
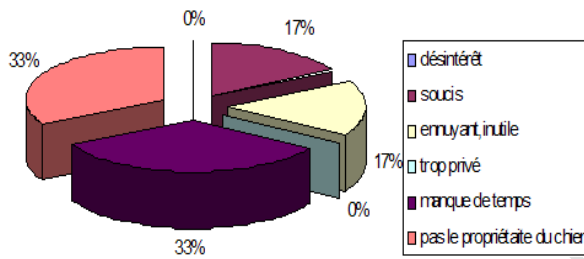
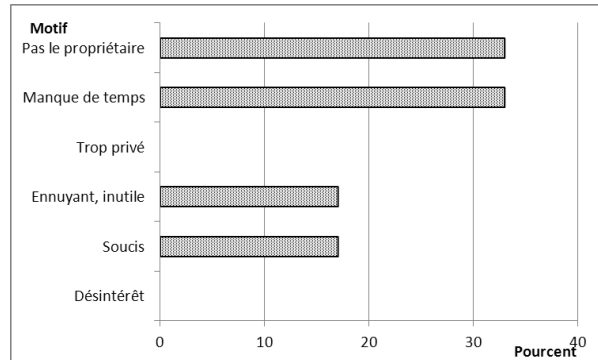
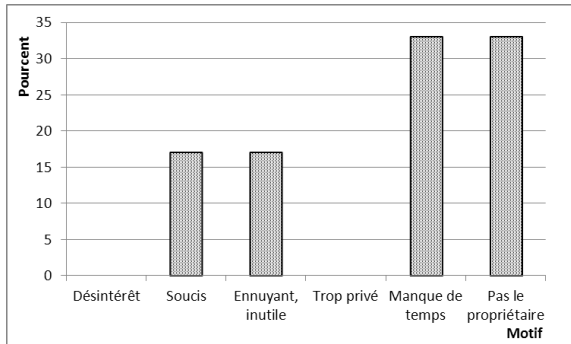
Laquelle des figures suivantes illustre le mieux la distribution des étudiants vétérinaires selon leur mention au bac ?



2. Motifs de non-participation à une enquête

Dans une enquête, l'auteur a illustré les motifs de non-participation à une enquête à l'aide d'un diagramme circulaire.

A vous de dire si vous êtes d'accord avec son choix, ou si vous choisiriez un autre type de diagramme, et bien sûr pourquoi !

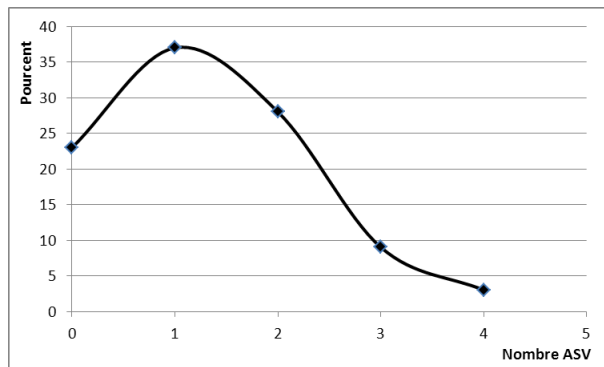
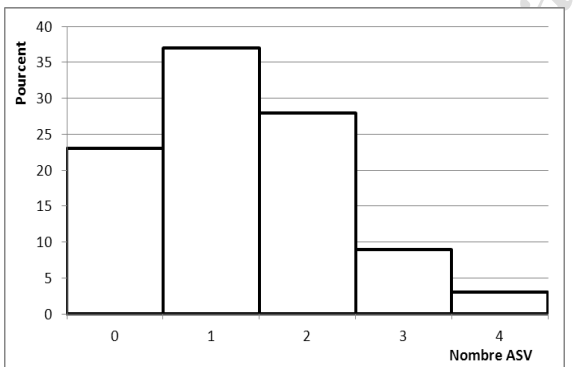
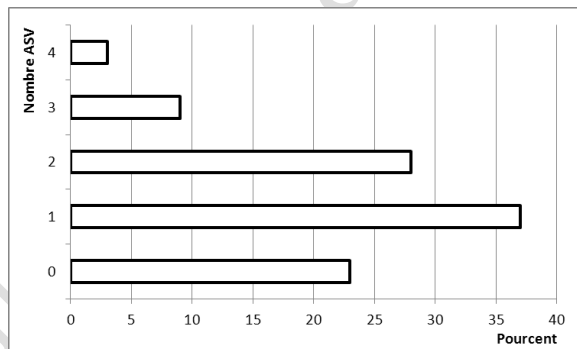
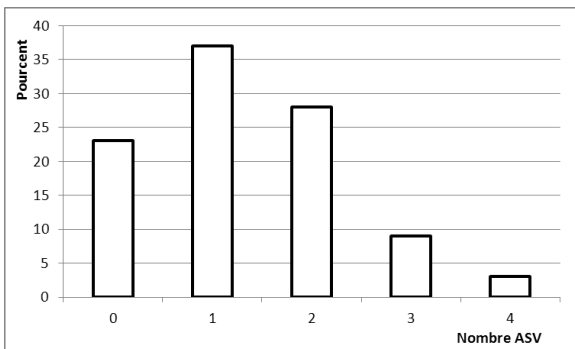
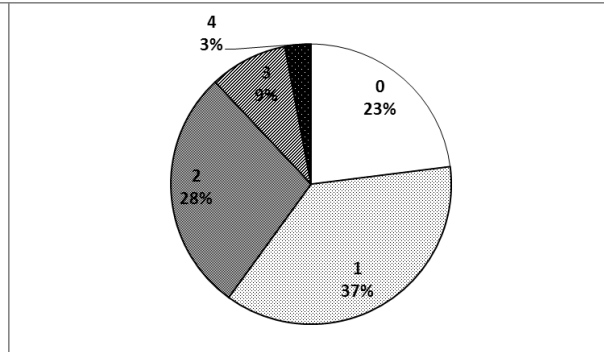
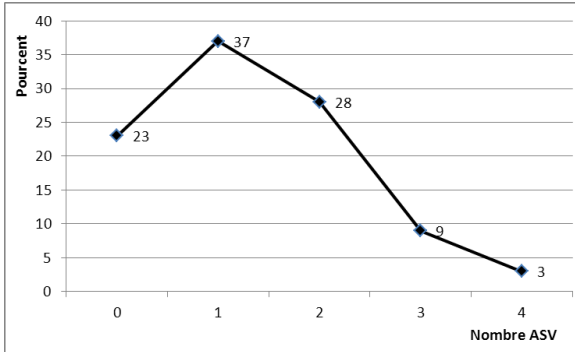


Version de travail

3. Nombre d'ASV par cabinet vétérinaire

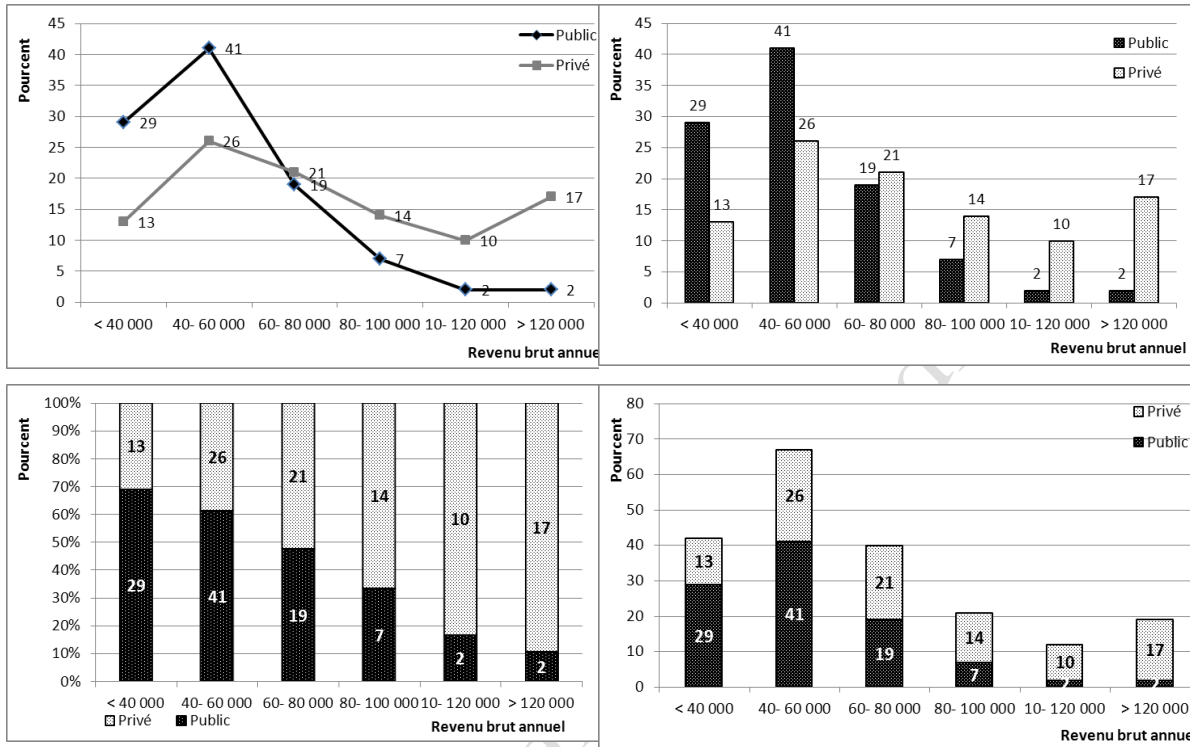
Une enquête a porté, entre autres, sur le nombre de collaborateurs par cabinet vétérinaire, dans notre exemple, celui des auxiliaires de soins vétérinaires (ASV).

Quelle figure vous semble convenir ?



4. Revenus des vétérinaires du public et du privé

C'est connu, les revenus du public et du privé ne sont pas de même niveau, ce qui mérite comparaison. Mais, quel graphique conviendrait le mieux pour illustrer cette comparaison ?



B - OBJECTIF 2 :

Être capable de reconnaître les caractéristiques de différentes figures.

Instructions

Tous ces exemples sont empruntés à des thèses de doctorat vétérinaire, dont la référence aux auteurs a été volontairement omise. Les figures qui étaient en couleur ont été converties en noir et blanc. Le titre d'origine est reproduit en italique.

Les figures portent le même numéro que celui de l'exercice. Les données nécessaires pour la réalisation des figures sont disponibles dans la feuille Excel « Figures », à l'onglet portant le numéro du sujet.

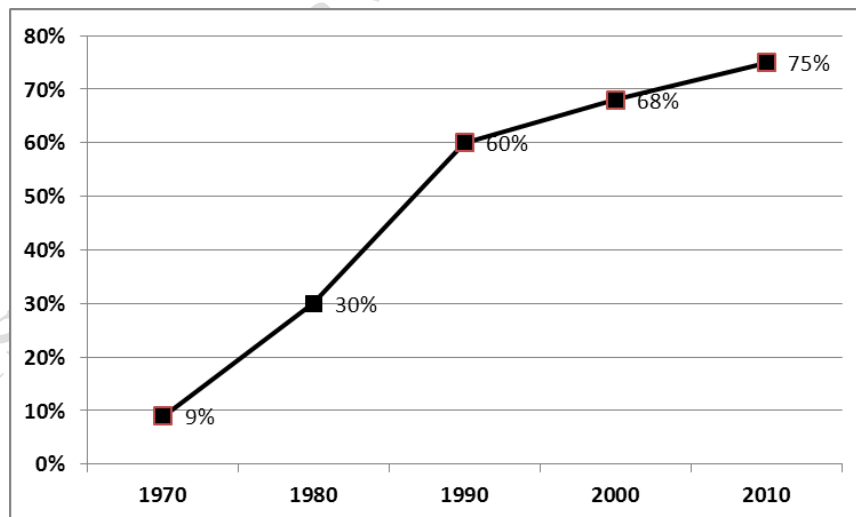
Pour tous ces exercices, les questions sont les mêmes :

- *De quel type de diagramme s'agit-il ?*
- *Dites pourquoi il est (ou non) adapté ?*
- *Dites si le titre est correctement formulé ou non. Faites des propositions d'amélioration si besoin.*
- *Réalisez la figure qui serait la mieux adaptée à partir des données de l'onglet correspondant au numéro de la figure.*

1. La proportion de femmes parmi les étudiants vétérinaires

La figure 1 a été utilisée pour représenter l'évolution du pourcentage de femmes parmi les étudiants vétérinaires au cours des quarante dernières années.

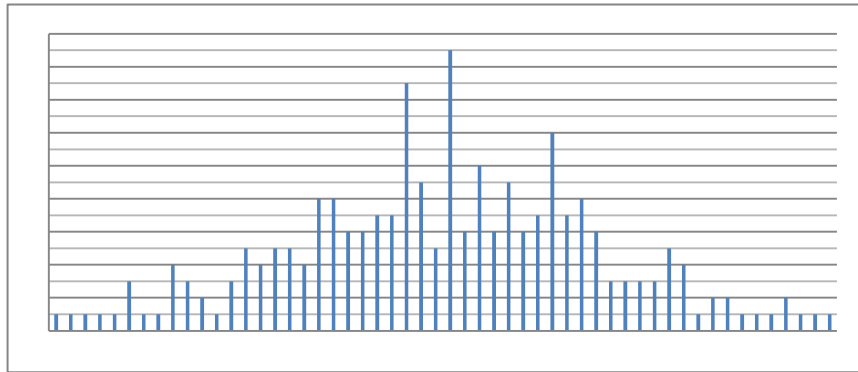
Figure 1 : Evolution du pourcentage de femmes à l'école vétérinaire depuis les années 70



2. Etude de l'attachement des propriétaires de chats

245 propriétaires de chats ont répondu à un questionnaire, qui a permis d'évaluer le degré d'attachement à leur animal avec un test reposant sur 27 questions (*Pet attachment scale, PAS*), auxquelles ils ont répondu à l'aide d'une échelle de Likert en 4 niveaux (1 : jamais, 2 : parfois, 3 : souvent, 4 : toujours). Un score global a été calculé par sommation des réponses à chaque question. Les valeurs possibles allaient de 27 à 108. La figure 2 représente les résultats obtenus.

Figure 2 : Histogramme de distribution de la variable score PAS

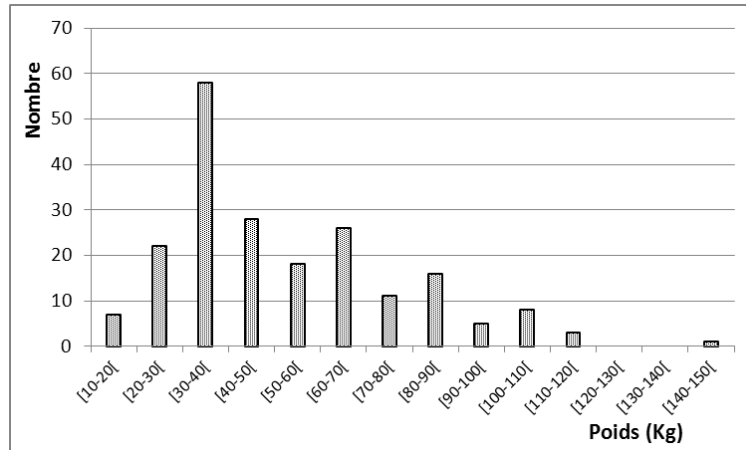


Version de travail. Ne pas utiliser

3. Poids de carcasses de sangliers

La tuberculose a fait l'objet d'une surveillance en forêt de Brotonne (Seine-Maritime). Pendant la saison 2012-2013, les carcasses de 205 sangliers abattus ont été pesées (Figure 3).

Figure 3 : Répartition des sangliers inspectés en fonction de la classe de poids

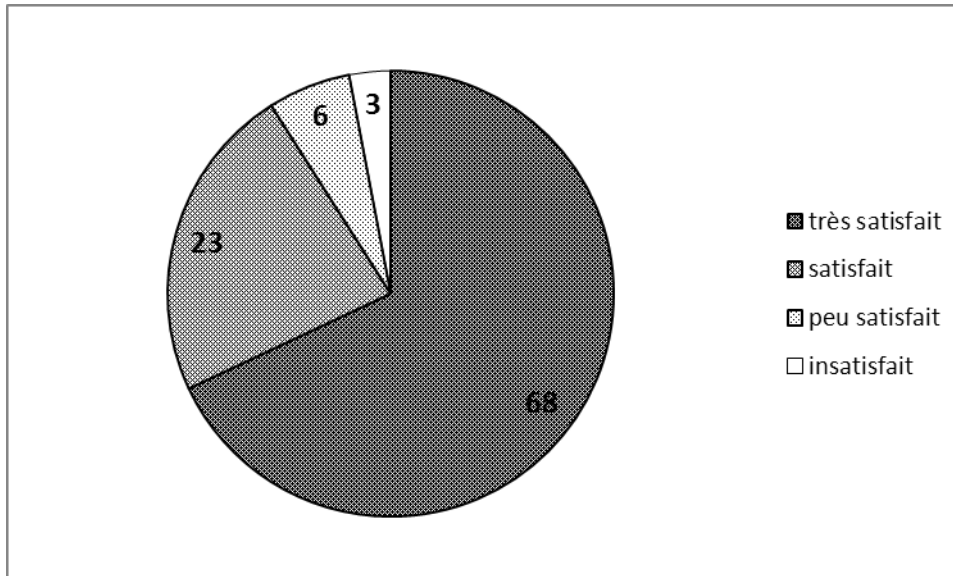


Version de travail. Ne pas diffuser

4. Satisfaction des propriétaires concernant la consultation

Sur l'ensemble des 97 propriétaires ayant fréquenté une consultation spécialisée, 78 d'entre eux ont pu être contactés par téléphone et ont accepté de répondre à un questionnaire, qui portait, entre autres, sur leur degré de satisfaction (figure 4)

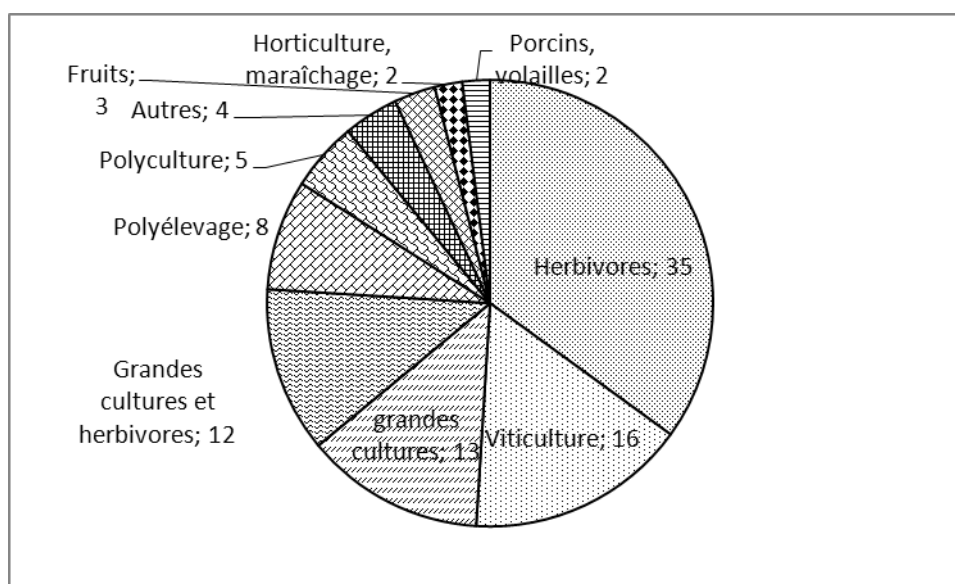
Figure 4 : Satisfaction des propriétaires concernant la consultation



5. Description des cultures dans un département

Dans la présentation d'un département, le graphique de la figure 5 a été utilisé. Dans sa version originale, il est en couleur.

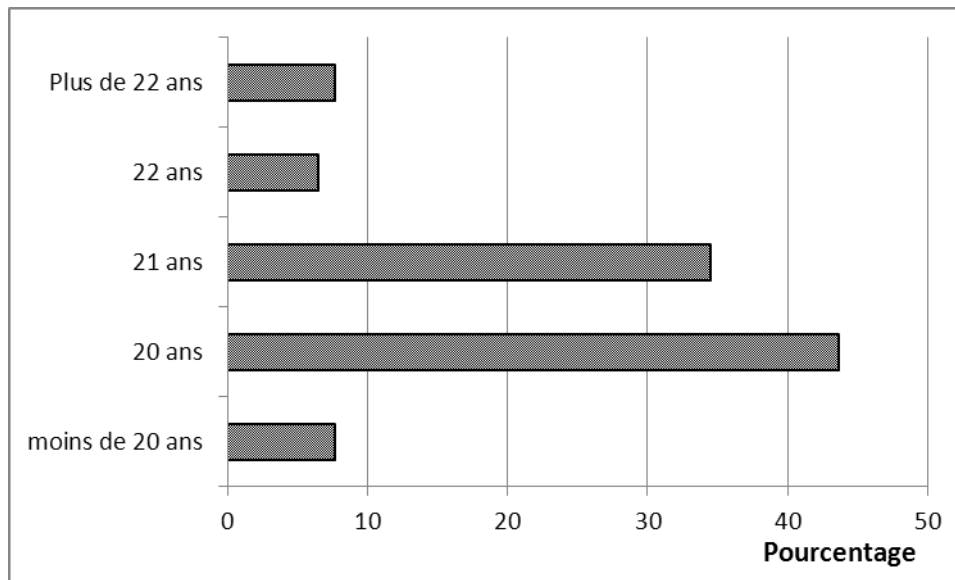
Figure 5 : Répartition des exploitations par filière dans le département (Agreste, 2008)



6. Age des étudiants

Une enquête a été réalisée sur la population des étudiants d'un établissement d'enseignement supérieur (figure 6).

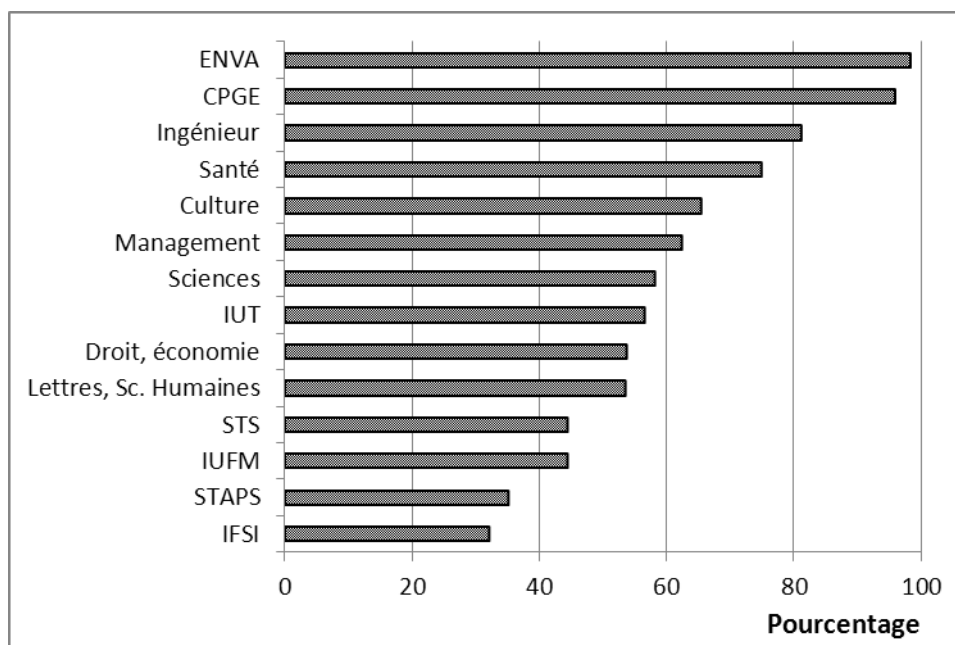
Figure 6 : Age à l'entrée dans l'établissement



7. Mentions au bac

Une enquête réalisée en 2010 par l'Observatoire national de la vie étudiante a permis de comparer les pourcentages d'étudiants ayant obtenu une mention au baccalauréat à l'ENV d'Alfort (ENVA) et dans différents établissements d'enseignement supérieur (figure 7).

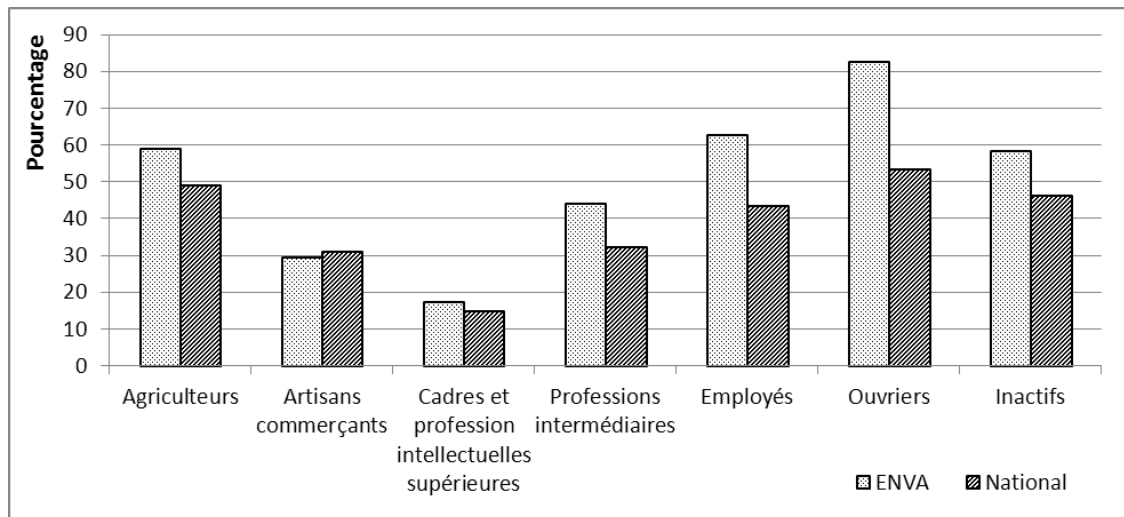
Figure 7 : Proportion d'étudiants de l'ENVA ayant obtenu une mention au baccalauréat en comparaison avec d'autres types d'études (en %)



8. Origine sociale des parents

Une enquête réalisée en 2010 par l'Observatoire national de la vie étudiante a permis de décrire les origines sociales du père des étudiants boursiers dans les établissements d'enseignement supérieur. Ces données ont été comparées à celle d'une enquête réalisée à l'Ecole nationale vétérinaire d'Alfort (ENVA) (figure 8).

Figure 8 : Répartition des étudiants boursiers selon la catégorie socioprofessionnelle du père (en %)

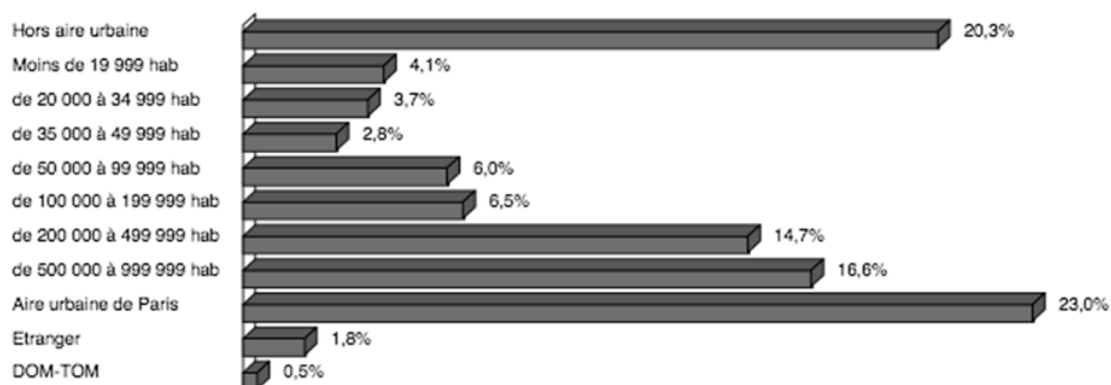


Version de travail. N

9. Commune d'origine des étudiants vétérinaires

Une thèse de doctorat vétérinaire s'est intéressée à la taille de la commune où les étudiants vétérinaires ont passé leur enfance (Figure 9).

Figure 9 : Taille de la commune dans laquelle les étudiants entrés dans les ENV françaises en 2005 ont passé la majorité de leur enfance

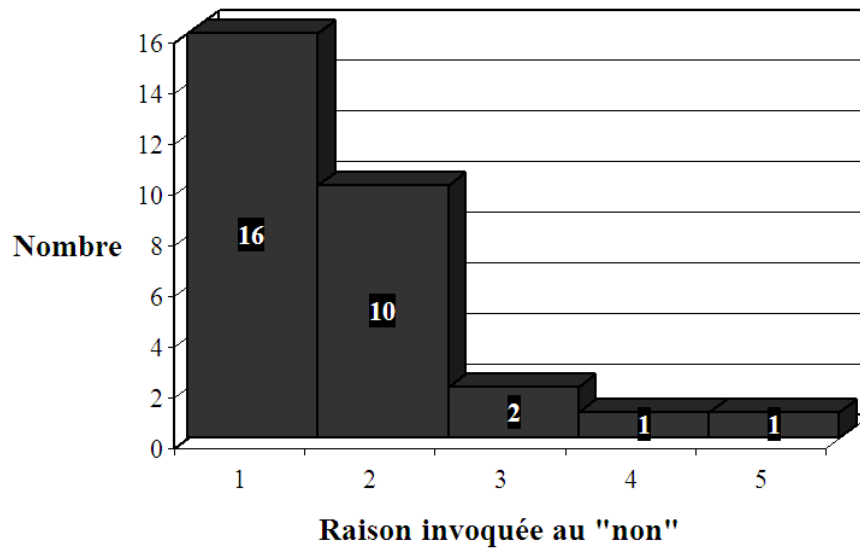


Version de travail. Ne pas

10. Raisons invoquées pour ne pas participer à une étude

Une étude sur les cancers du chien a été réalisée. Sur 108 cliniques vétérinaires contactées, 82 ont accepté de participer, 26 ont refusé. Les raisons du refus ont été prospectées par une question ouverte (figure 10) : de la plus fréquente à la moins fréquente, le manque de temps (raison 1), le manque d'intérêt (raison 2), une clientèle comprenant peu ou pas de chiens (raison 3), le refus systématique de répondre à des questionnaires téléphoniques (raison 4), et enfin une clientèle formée essentiellement de cas référés d'ophtalmologie (raison 5).

Figure 10 : Raisons invoquées pour justifier le refus de répondre au questionnaire



11. Valeurs propres

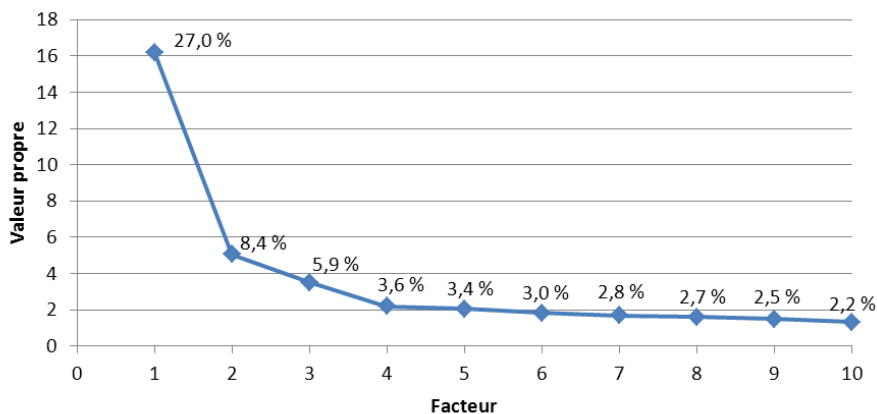
La description statistique d'un fichier de données comportant un grand nombre de variables doit aller plus loin que l'approche uni- ou bi-variée, par une description multivariée, comme l'ACP (analyse en composantes principales).

Celle-ci calcule des variables synthétiques qui permettent au mieux de rendre compte du nuage à n dimensions, appelées valeurs propres ; la qualité de cet axe est appréciée par la proportion de la variance qu'il porte ; les projections sur un plan associent ces valeurs propres deux à deux : la qualité de la projection est appréciée par la somme des valeurs propres.

On cherche à associer des axes deux par deux, pour obtenir une représentation plane combinant deux axes donnant la meilleure représentation possible, ce que l'on juge, d'une part, selon la part de variance prise en compte, d'autre part, selon le pouvoir discriminant de la représentation.

Une étude sur l'attachement des propriétaires à leur chat a mis en relation le score de réponse des propriétaires à un questionnaire standardisé (*Pet Attachment Scale, PAS* ; réponses selon une échelle de Likert), et leurs réponses à diverses questions (démographiques, comportementales). La figure 11 représente les valeurs des 10 premiers axes.

Figure 11 : Valeurs propres et pourcentages de variance expliquée par les 10 premiers facteurs pour les chats

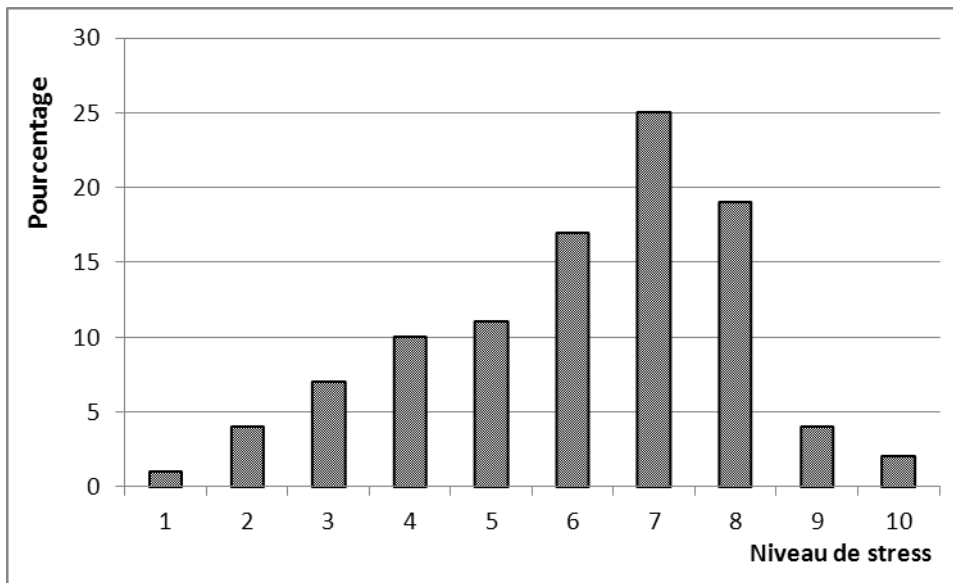


12. Étude du stress au travail dans les cabinets vétérinaires

Le stress au travail dépend de l'activité et des circonstances. Une enquête a été conduite auprès de 494 vétérinaires qui ont accepté de répondre à un questionnaire afin de faire l'étude du stress au travail dans un cabinet vétérinaire.

La question illustrée par la figure 12 est relative au niveau de stress en relation avec les exigences du travail (horaires par exemple), évalué sur une échelle de 1 à 10 par les personnes enquêtées.

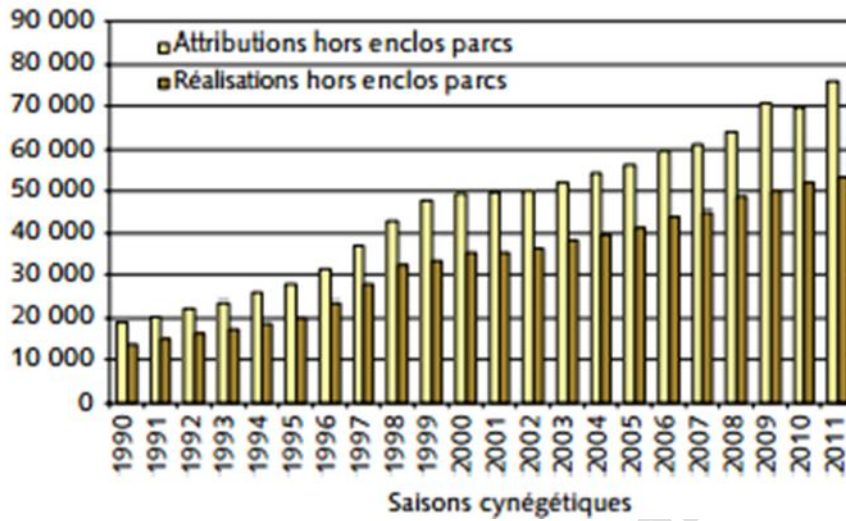
Figure 12 : Niveau de stress en lien avec les exigences du travail (n = 494)



13. Tableau de chasse des cerfs en France

La population d'animaux sauvages a considérablement augmenté au cours des vingt dernières années, comme en témoignent les tableaux de chasse réalisés (figure 13).

Figure 13 : *Évolution annuelle du tableau de chasse national cerf*



CHAPITRE II - CORRIGES

Les tableaux et figures sont numérotés en fonction du numéro de l'exercice. Le libellé de l'exercice a été repris pour faciliter la lecture du corrigé. Le titre est en italique lorsqu'il est celui formulé par l'auteur.

Le corrigé comporte des commentaires et éventuellement une version corrigée du tableau ou de la figure, portant le même numéro assorti d'une apostrophe (').

I - TABLEAUX

A - OBJECTIF

Être capable de reconnaître les critères de qualité et les défauts d'un tableau.

B - CORRIGÉS

1. Enquête Facco

Tableau 2 : enquête FACCO 2012, nombre d'individus, proportion d'animaux de race et taux de stérilisation

	Chiens	Chats
Nombre d'individus (en millions)	7,42	11,41
Proportion d'animaux de race	70 %	25 %
Taux de stérilisation	37 %	77,2 %

Commentaires

- Le titre n'est pas suffisamment explicite et n'est pas conforme : il devrait commencer par une majuscule ; il manque de clarté parce qu'il ne répond pas aux questions *quoi, où, comment, quand, etc.*
- Trop de lignes délimitant les contours.
- Les deux dernières lignes auraient dû être organisées comme la première ligne, qui précise les unités, sans les rappeler dans le corps du tableau.
- Les chiffres devraient être alignés sur le dernier chiffre avant la virgule décimale.

Proposition de modifications.

Tableau 1' : Résultats principaux de l'enquête nationale sur les propriétaires d'animaux de compagnie en France (Facco, 2012)

	Chiens	Chats
Nombre d'individus (en millions)	7,4	11,4
Proportion d'animaux de race (%)	70	25
Taux de stérilisation (%)	37	77,2

2. Etude multivariée des résultats d'une enquête sur l'attachement des propriétaires de chat à leur animal.

245 propriétaires ont accepté de répondre à un questionnaire de 27 questions (dérivé du modèle « *Pet attachment Scale* » ou PAS ; réponses selon une échelle de Likert), permettant de leur attribuer un score d'attachement, et de mettre ce score global en relation avec différentes variables (démographiques, comportementales). Une analyse en composantes principales (ACP) a été réalisée, qui a conduit à identifier les 10 premiers axes (« valeurs propres ») permettant de synthétiser au mieux l'ensemble des données.

Tableau 2 : Valeurs propres, valeurs propres cumulées, pourcentages de variance et pourcentages de variance cumulés pour les 10 premiers facteurs pour les chats

	Valeur propre	% total de variance	Valeur propre cumulée	% de variance cumulé
Facteur 1	16.20924	27.01540	16.20924	27.0154
Facteur 2	5.04417	8.40695	21.25341	35.4224
Facteur 3	3.51841	5.86401	24.77182	41.2864
Facteur 4	2.18075	3.63458	26.95257	44.9209
Facteur 5	2.06445	3.44074	29.01701	48.3617
Facteur 6	1.82318	3.03864	30.84020	51.4003
Facteur 7	1.68485	2.80809	32.52505	54.2084
Facteur 8	1.60604	2.67674	34.13110	56.8852
Facteur 9	1.48191	2.46985	35.61301	59.3550
Facteur 10	1.33719	2.22865	36.95019	61.5837

Commentaires

Les lignes de contour sont limitées au strict minimum, conformément aux recommandations. Le fond grisé, qui vise à distinguer une ligne sur deux est en trop. En langue française, le séparateur décimal devrait être la virgule, non le point. Le nombre de valeurs après la virgule est totalement excessif : un seul chiffre serait largement suffisant pour permettre la distinction entre les différentes valeurs. Le titre devrait apporter plus d'information.

Proposition de modifications.

Tableau 2' : Valeurs propres, pourcentages de variance, valeurs propres cumulées, et pourcentages de variance cumulés pour les 10 premiers facteurs d'une ACP réalisée sur les réponses de propriétaires de chats à un questionnaire PAS (N = 245).

	Valeur propre	% total de variance	Valeur propre cumulée	% de variance cumulé
Facteur 1	16,2	27,0	16,2	27,0
Facteur 2	5,0	8,4	21,3	35,4
Facteur 3	3,5	5,9	24,8	41,3
Facteur 4	2,2	3,6	27,0	44,9
Facteur 5	2,1	3,4	29,0	48,4
Facteur 6	1,8	3,0	30,8	51,4
Facteur 7	1,7	2,8	32,5	54,2
Facteur 8	1,6	2,7	34,1	56,9
Facteur 9	1,5	2,5	35,6	59,4
Facteur 10	1,3	2,2	37,0	61,6

3. Description d'un échantillon d'enquête (une variable)

Une enquête a été réalisée sur 556 personnes. La description porte sur la répartition par sexe.

Tableau 3 : Pourcentage de femmes et d'hommes parmi les répondants

Sexe	Effectif	%
Femmes	242	44%
Hommes	314	56%
Total	556	100%

Commentaires

Ce tableau à simple entrée ne devrait pas comporter de couleur.

Les pourcentages devraient être regroupés dans la même case que les effectifs (au passage, remarquez que le signe du pourcentage ne devrait pas être accolé aux chiffres correspondants, mais séparés par une espace).

La dernière ligne n'est pas utile : il suffit de rappeler l'effectif total dans le titre, comme c'est l'usage.

Proposition de modifications.

Tableau 3' : Distribution des effectifs des répondants à l'enquête selon leur sexe (N = 556).

Sexe	Nombre (%)
Femmes	242 (44)
Hommes	314 (56)

4. Description d'un échantillon d'une enquête (deux variables)

550 vétérinaires ont accepté de répondre à une enquête professionnelle. Le tableau 4 décrit leur répartition entre secteurs publique et privé en fonction du sexe.

Tableau 4 : Répartition des genres par secteur

	Privé / hommes	Privé / femmes	Public / hommes	Public / femmes
Nombre	152	97	158	143
	249		301	
Pourcent	61%	39%	52%	48%

Commentaires

Les couleurs sont en trop, ainsi que les lignes de contour, qui sont trop nombreuses.

Les pourcentages ne devraient pas faire l'objet d'une ligne distincte : les données sont présentées comme s'il s'agissait d'un tableau à simple entrée, alors qu'il conviendrait de l'organiser selon une double entrée (Secteur / sexe).

Le mot « genre », qui vise à différencier un homme d'une femme selon leurs caractéristiques psycho-sociales et non à désigner un rang taxonomique regroupant diverses espèces, n'est pas pertinent : il s'agit bien ici du sexe, tel que défini d'après les caractéristiques biologiques.

Proposition de modifications.

Tableau 4' : Répartition des vétérinaires ayant répondu à l'enquête selon leur secteur d'activité (public, n = 249 ; privé, n = 301) et leur sexe

Sexe	Secteur privé effectif (%)	Secteur public effectif (%)
Homme	152 (61)	158 (52)
Femme	97 (39)	143 (48)

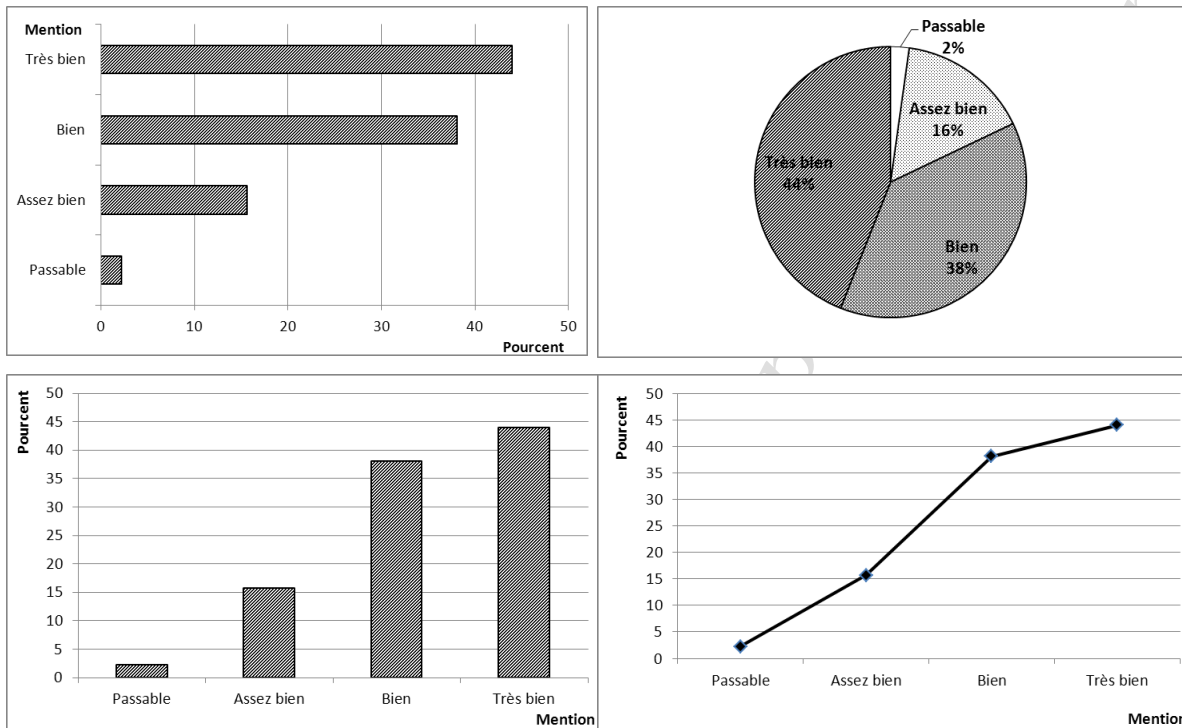
II - FIGURES

A - OBJECTIF 1

Être capable de choisir un type de diagramme pour une illustration donnée.

1. Mentions au bac

Les figures suivantes illustrent la distribution des étudiants vétérinaires selon leur mention au bac.



Commentaires

Réglons tout de suite le sort du diagramme circulaire : il n'y a aucune raison de l'utiliser, car la perception des secteurs n'est pas aussi facile que des longueurs et il n'y a aucune raison de se préoccuper d'éviter de froisser des susceptibilités qui résulteraient de l'ordre des items.

Le diagramme à lignes n'est pas non plus justifié ici, car il ne s'agit pas d'une variable continue, ce que pourrait suggérer la continuité de la ligne.

Les mentions (passable, assez bien, bien, très bien) peuvent être considérées comme des dénominations et par conséquent être traitées comme une **variable nominale**. C'est pourquoi certains choisissent un **diagramme à barres** (graphique a).

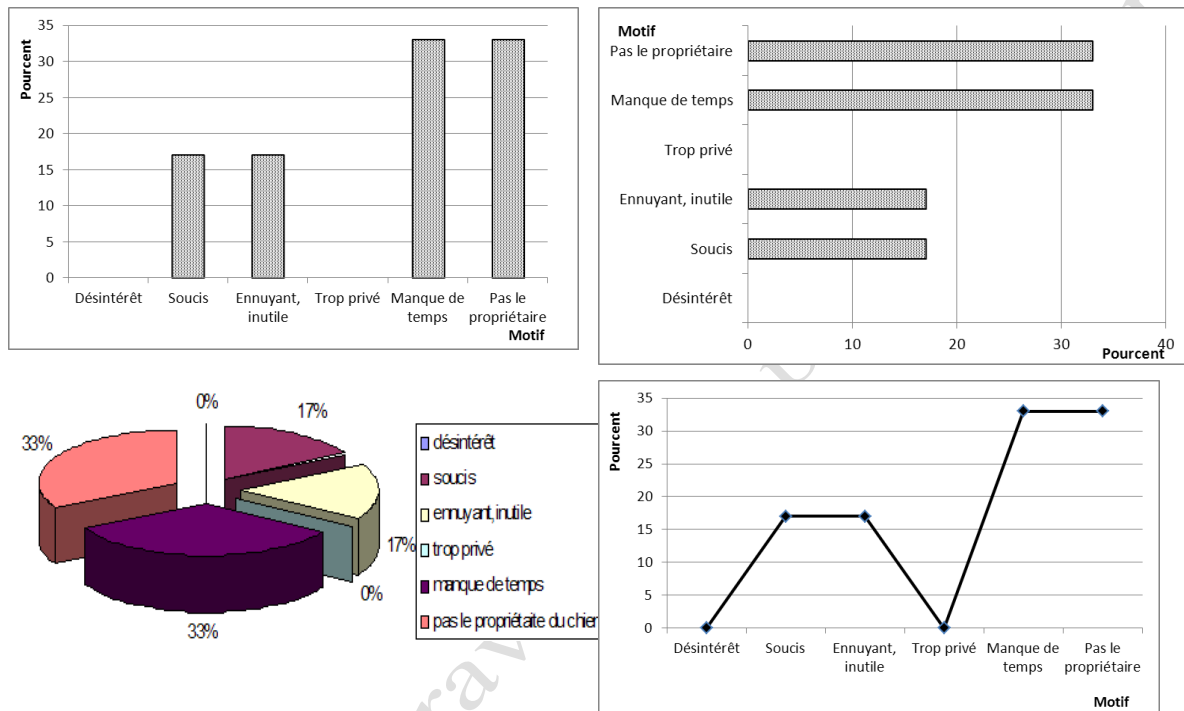
Toutefois, ce choix n'est pas conforme aux usages associant le choix d'un diagramme à la nature de la variable. En effet, les mentions peuvent être *ordonnées* : elles pourraient être codées respectivement par les valeurs 1, 2, 3 ou 4, pour en faciliter l'ordonnement: par conséquent, il s'agit bien d'une **variable ordinale**, qui justifie un **diagramme à colonnes** (graphique a).

Cela étant, certains pourront toujours exprimer un choix sur des considérations esthétiques personnelles, et préférer telle figure, tout en évitant les grisés (principe d'économie d'encre).

2. Motifs de non-participation à une enquête

Une enquête a porté, entre autres, sur le nombre de collaborateurs par cabinet vétérinaire, dans notre exemple, celui des auxiliaires de soins vétérinaires (ASV).

Quelle figure vous semble convenir ?



Commentaires

Le diagramme circulaire a divers inconvénients qui doivent en restreindre l'usage aux seules circonstances où l'on veut éviter de suggérer un classement de valeurs. Ce n'est pas le cas ici. En plus, le choix de dispositifs graphiques en usage dans le milieu du marketing (3D, éclatement de la figure) ne convient pas pour une communication scientifique.

Pour choisir le bon graphique, il suffit d'identifier correctement la catégorie de la variable.

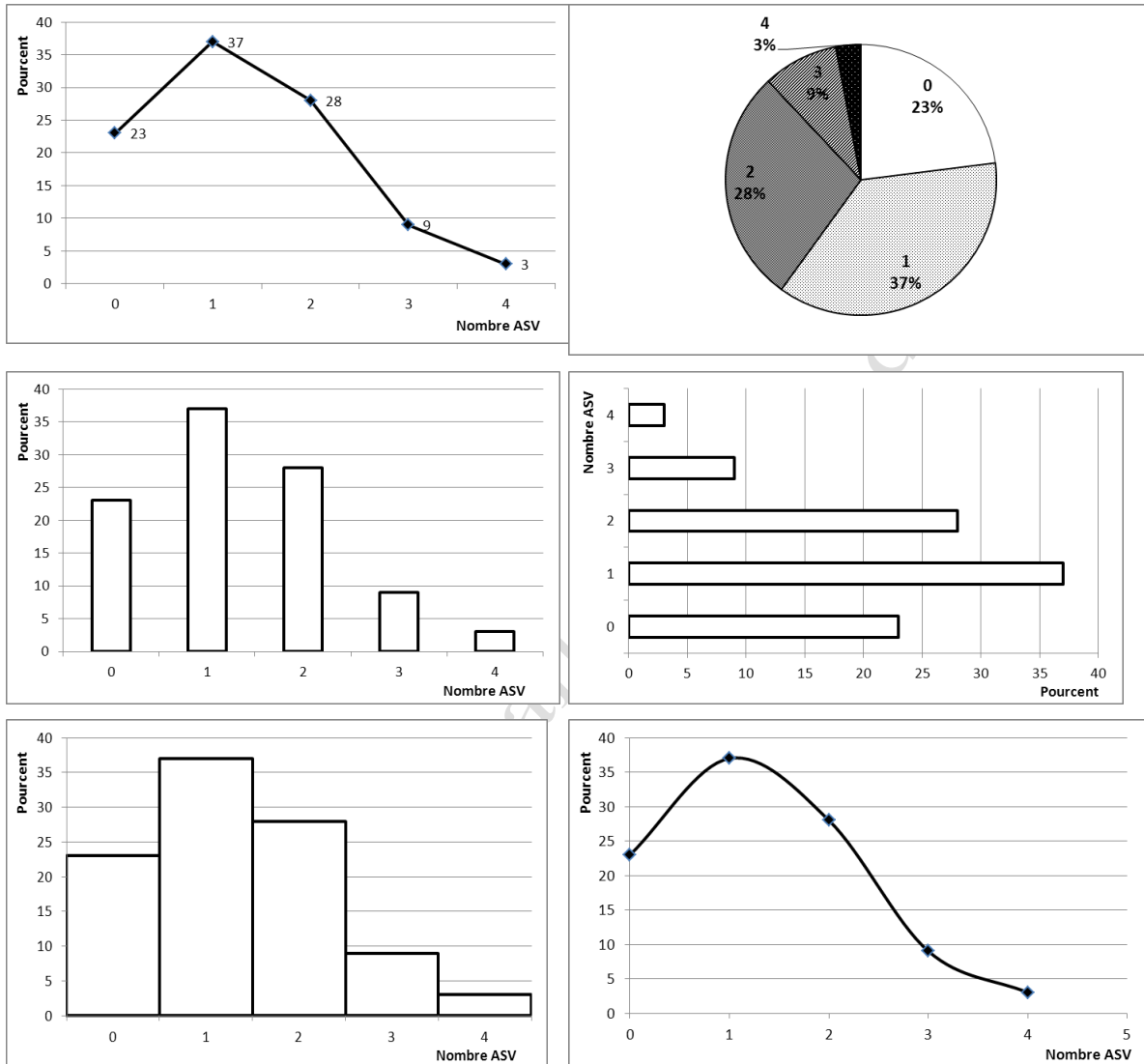
Le répondant disposait de diverses propositions de motif de refus : c'est donc une **variable qualitative**. On ne peut pas les classer de façon évidente : c'est donc une **variable nominale**.

Donc, le **diagramme à barres** est tout indiqué (b), mais il faudrait également respecter le principe d'économie d'encre et s'abstenir des grisés.

3. Nombre d'ASV par cabinet vétérinaire

Une enquête a porté, entre autres, sur le nombre de collaborateurs par cabinet vétérinaire, dans notre exemple, celui des auxiliaires de soins vétérinaires (ASV).

Quelle figure vous semble convenir ?



Commentaires

Dans le questionnaire, le choix du nombre d'ASV possible était limité : 0, 1, 2, 3, 4.

On peut nommer les différentes issues. On peut les classer par ordre croissant. S'agissant de chiffres, on pourrait penser qu'il est possible de faire des calculs. Serait-ce donc une variable quantitative ? Si je réunis deux cabinets, l'un avec, par exemple, 1 ASV et l'autre avec 2, nous aurons bien ensemble comportant 3 ASV. Mais, on ne peut pas dénombrer de morceaux d'ASV : le dénombrement ne peut se faire que par unités entières.

Le dénombrement des ASV ne peut comporter de fraction de personne, car il s'agit de personnes *physiques*, et non d'un dénombrement théorique en *équivalent temps plein*, comme on peut le faire en gestion de ressources humaines. Mais cette précision est indispensable dans la récolte des informations !

Donc, il s'agit d'une **variable quantitative discrète**, ce qui n'empêchera pas de faire des calculs de moyenne dont le résultat pourra aboutir à des quantités décimales indiquant des fractions d'ASV : ce n'est pas choquant, car il ne s'agit que d'un *indicateur* statistique et non d'un dénombrement.

Nous éliminons le diagramme circulaire : ce diagramme est indiqué lorsqu'on souhaite éviter de suggérer une idée de classement. Ce n'est pas le cas ici.

Le diagramme à lignes (droite ou courbe) ne convient pas non plus, car la jonction entre deux points crée une information qui ne correspond à aucune réalité. Notez que la ligne courbe constitue manifestement une interprétation totalement débridée !

Le diagramme à barres n'est pas approprié non plus : il convient pour les variables nominales, ce qui n'est pas le cas ici.

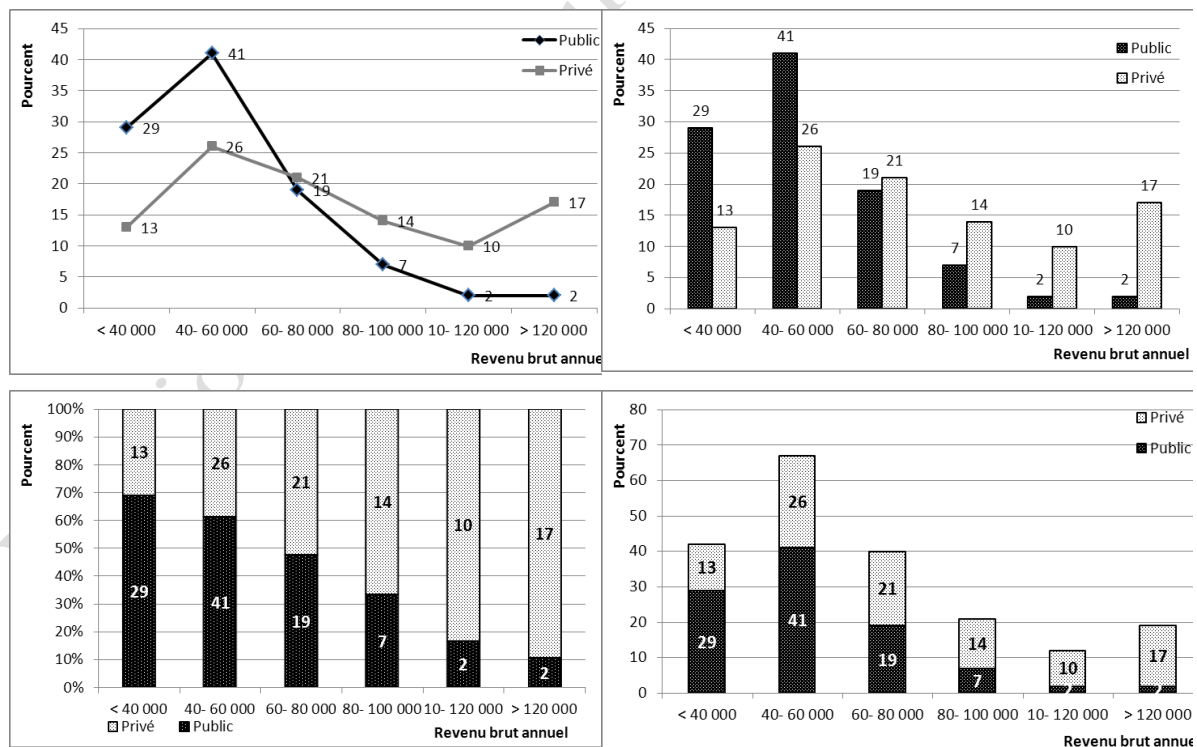
Restent l'histogramme et le diagramme à colonnes. Comme la variable n'est pas continue, ce ne peut être l'histogramme (même s'il est régulier, comme ici).

Donc, c'est bien le **diagramme à colonnes** (c).

Notez que les diagrammes à colonnes, à barres et l'histogramme représentés ici respectent bien le principe d'économie d'encre

4. Revenus des vétérinaires du public et du privé

C'est connu, les revenus du public et du privé ne sont pas de même niveau, ce qui mérite comparaison. Mais, quel graphique conviendrait le mieux pour illustrer cette comparaison ?



Commentaires

Les revenus sont une **variable quantitative continue**. Ici, elle est mise en classe (20 000 €). Un **diagramme à colonnes** est donc indiqué.

Le but est de comparer deux séries : le **diagramme à lignes** pourrait être utilisé. Soulignant une tendance, il permet de souligner le tassement des effectifs dans le secteur public pour les hauts revenus. Mais, si on considère qu'il s'agit de statistiques descriptives, décidément, le **diagramme à colonnes** serait plus indiqué. Le plus simple, est celui de la **juxtaposition** des deux séries. Il faut reconnaître que l'on perçoit quand même bien l'écart dans les hauts revenus, les colonnes de la série du privé semblant écraser les toutes petites colonnes du public.

Plus élaborés, les **diagrammes à colonnes superposées** : notez que les valeurs reproduites sont les mêmes. Pour la figure d, la longueur des colonnes correspond à la valeur des pourcentages, tout comme dans le graphique b, où les colonnes sont juxtaposées. Mais si les rapports sont conservés, les longueurs sont différentes, du fait du changement d'échelle pour pouvoir faire tenir le total des *deux* séries sur une seule colonne. Donc, finalement, les écarts sont tassés, et par conséquent moins visibles. C'est l'inverse avec la figure c : les valeurs portées ne sont pas représentées telles quelles, mais systématiquement en proportion *relative* des deux valeurs d'une même colonne. De ce fait, l'écart relatif est amplifié. On voit ainsi que c'est pour le revenu compris entre 60 et 80 000 € que les proportions sont équilibrées entre les deux séries. Puis, la proportion des revenus du public continue de baisser pour les revenus plus élevés : $1/3$ pour 80 à 100 000 €, un peu moins de $1/5$ pour 100 à 120 000 € et enfin environ $1/10$ pour les revenus supérieurs à 120 000 €. Mais, pour être plus informatif, il aurait fallu reporter ces valeurs de proportions relatives plutôt que celles du pourcentage total de la série.

Au total, les deux premiers graphiques (a et b) illustrent le propos de l'auteur de façon satisfaisante, la figure b donnant une image plus adaptée à une vision de statistique descriptive, tandis que la a est plus orienté vers une interprétation : le choix dépend de l'intention de l'auteur.

Le graphique d est peu lisible.

Le graphique c illustre un propos plus élaboré, différent de la simple description. Il demanderait un commentaire spécifique et surtout, si des données devaient y être reportées, celles-ci devraient être calculées en proportion relative de la colonne.

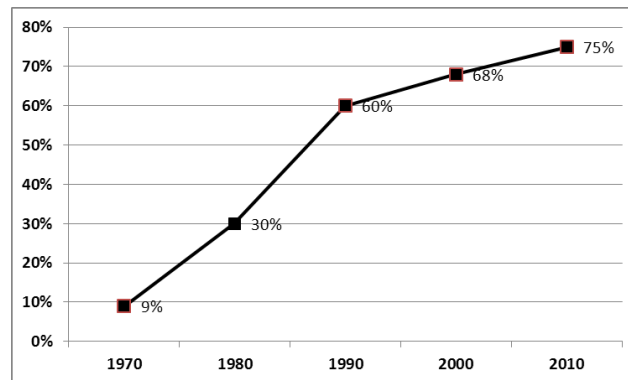
B - OBJECTIF 2 :

Reconnaître les caractéristiques de différentes figures

1. La proportion de femmes parmi les étudiants vétérinaires

La figure 1 a été utilisée pour représenter l'évolution du pourcentage de femmes parmi les étudiants vétérinaires au cours des quarante dernières années.

Figure 1 : Évolution du pourcentage de femmes à l'école vétérinaire depuis les années 70



Commentaires

Il s'agit d'une évolution dans le temps, l'axe des abscisses est donc correct.

Mais on ne dispose que de 5 points : les lignes qui joignent les points constituent une interprétation particulièrement libre de l'évolution réelle. S'il y avait eu un point par année, ce choix aurait été justifié, du fait de la continuité temporelle, et si, justement, la tendance générale de l'évolution se prêtait à une représentation linéaire. En fait, le faible nombre de points correspond plutôt à des images prises tous les 10 ans, sans qu'on sache ce qui se passe dans l'intervalle.

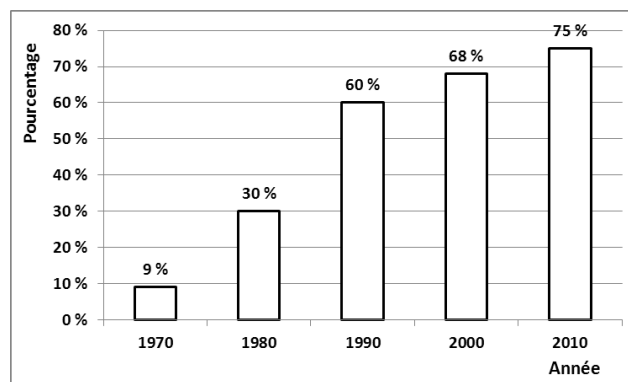
Un diagramme à colonnes aurait été plus indiqué.

Le signe de pourcentage (%) devrait être séparé du nombre par une espace.

Le titre devrait être adapté en conséquence.

Proposition de modifications

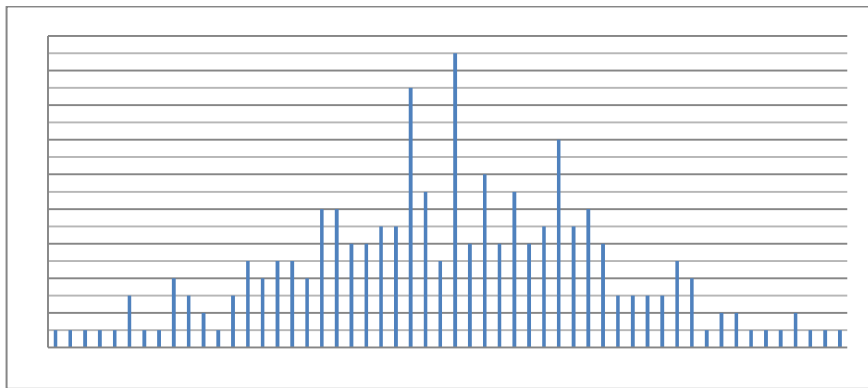
Figure 1' : Évolution décennale du pourcentage de femmes parmi les étudiants vétérinaires en France à partir de 1970



2. Etude de l'attachement des propriétaires de chats

245 propriétaires de chats ont répondu à un questionnaire, qui a permis d'évaluer le degré d'attachement à leur animal avec un test reposant sur 27 questions (*Pet attachment scale, PAS*), auxquelles ils ont répondu à l'aide d'une échelle de Likert en 4 niveaux (1 : jamais, 2 : parfois, 3 : souvent, 4 : toujours). Un score global a été calculé par sommation des réponses à chaque question. Les valeurs possibles allaient de 27 à 108. La figure 2 représente les résultats obtenus.

Figure 2 : Histogramme de distribution de la variable score PAS



Commentaires

Il s'agit d'un diagramme à bâtons.

Il s'agit bien d'une variable discrète, puisqu'elle ne peut prendre que des valeurs entières. Le choix d'un diagramme à bâtons était donc justifié.

La représentation fait figurer les bâtons sur les graduations, ce qui est conforme.

Mais le nombre de valeurs possibles est élevé (de 46 à plus de 100, soit plus de 50 valeurs possibles). D'un point de vue statistique, la représentation d'une distribution de fréquence doit être adaptée au nombre total des valeurs collectées, non à celui des valeurs possibles. Une des formules permettant de donner un ordre de grandeur du nombre de valeurs est $5 \times \log_{10}(N)$, ce qui, pour 245 observations, donne 12, soit, pour une étendue d'environ 60 (de 45 à 100 environ), des intervalles de 5.

Le terme « histogramme » est totalement inapproprié.

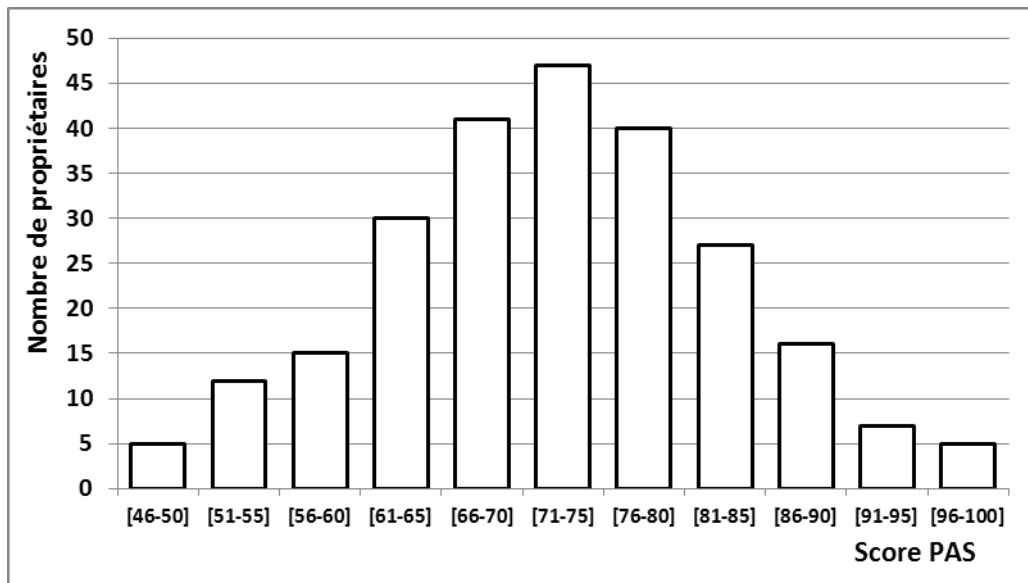
La distribution concerne des propriétaires, et non des chats, comme indiqué dans la légende de l'axe des ordonnées. Il s'agit donc d'une « distribution d'effectifs de propriétaires selon leur score obtenu au test « PAS » ».

Proposition de modifications

La réalisation de classes d'une amplitude 5 permet d'aboutir à la figure 2'.

Cette fois, les colonnes, du fait de la représentation de classes, devront être placées entre les graduations (ce qu'Excel fait par défaut).

Figure 2' : Distribution des effectifs de propriétaires selon leur score obtenu au test PAS, répartis en classes de 5.

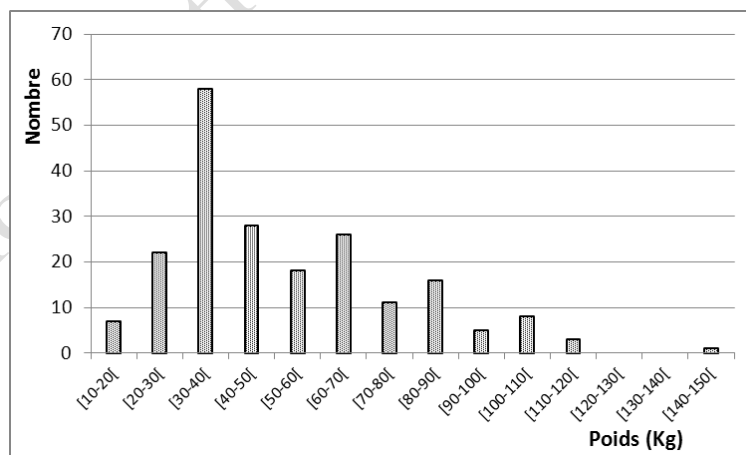


La figure donne une bien meilleure représentation : la distribution est centrée sur la colonne [71-75], ce qui correspond bien à la moyenne calculée (72,6), et sa répartition de chaque côté est bien plus parlante que telle qu'elle apparaissait dans la figure d'origine.

3. Poids de carcasses de sangliers

La tuberculose a fait l'objet d'une surveillance en forêt de Brotonne (Seine-Maritime). Pendant la saison 2012-2013, les carcasses de 205 sangliers abattus ont été pesées (Figure 3).

Figure 3 : Répartition des sangliers inspectés en fonction de la classe de poids



Commentaires

C'est un diagramme à colonnes.

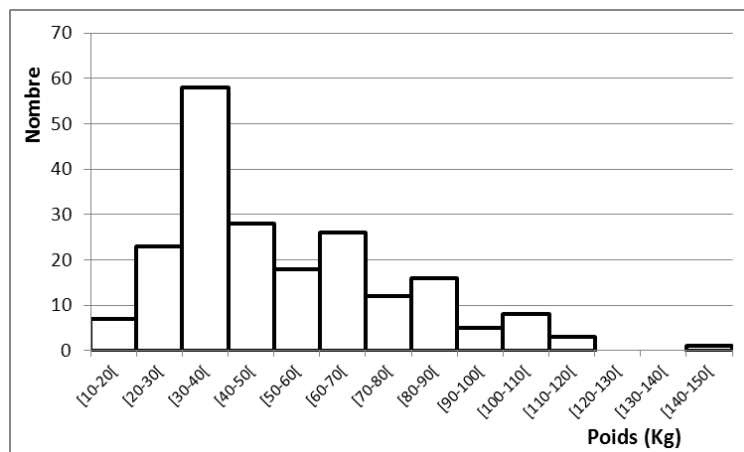
Il s'agit d'une variable quantitative continue répartie en classes égales d'amplitude de 10 kg. La représentation devrait suggérer la continuité de la variable : histogramme (qui serait ici régulier du fait que les classes sont d'amplitude égale), ou diagramme à colonnes jointives.

Pour reproduire la légende, il faut utiliser la fonction « Concatener » ([voir comment](#)).

Selon le principe de l'économie d'encre, il conviendrait de ne pas mettre de couleur (noir ou gris) dans les colonnes, puisqu'il n'y a qu'une seule distribution et donc nul besoin de la distinguer d'une autre.

Le titre devrait être plus complet. D'abord, il s'agit d'une "distribution" et non d'une répartition, et préciser le lieu et le temps. En revanche, il n'est pas nécessaire de préciser qu'il s'agit d'une distribution en classes de poids : la figure l'explique suffisamment sans qu'il soit nécessaire de le préciser (figure 3').

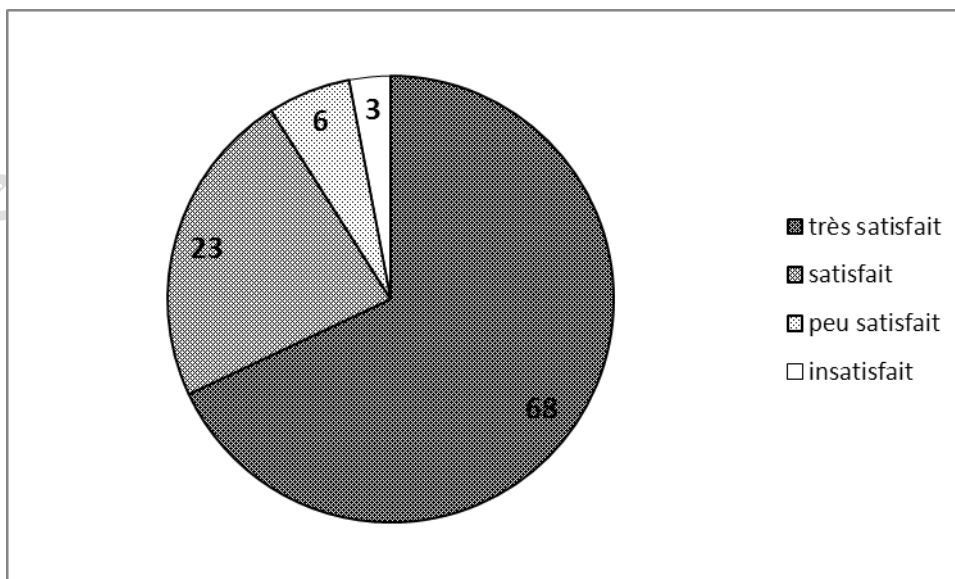
Figure 3' : Distribution du nombre de sangliers selon leur poids (forêt de Brotonne, 2012-2013)



4. Satisfaction des propriétaires concernant la consultation

Sur l'ensemble des 97 propriétaires ayant fréquenté une consultation spécialisée, 78 d'entre eux ont pu être contactés par téléphone et ont accepté de répondre à un questionnaire, qui portait, entre autres, sur leur degré de satisfaction (figure 4)

Figure 4 : Satisfaction des propriétaires concernant la consultation



Commentaires

Il s'agit d'un diagramme circulaire.

Il n'est pas adapté : il s'agit d'une variable qualitative ordinale, qui justifie un diagramme à colonnes.

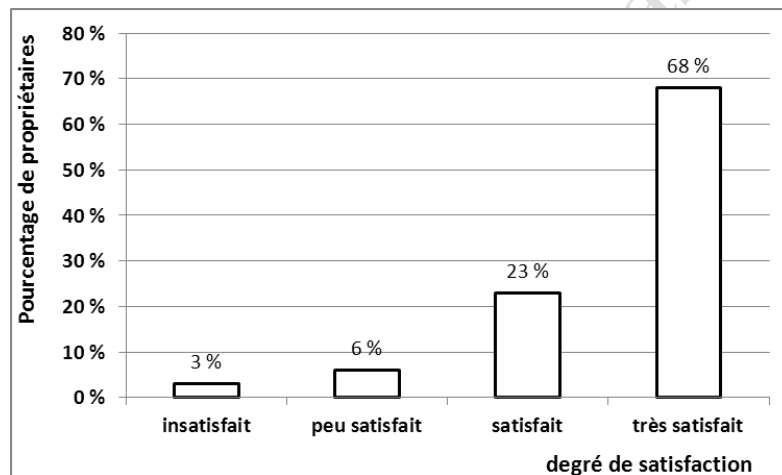
Les diagrammes circulaires sont d'un usage à limiter aux circonstances où l'on veut éviter de suggérer un ordre entre les différents items, ce qui n'est justement pas le cas ici.

Le titre n'est pas assez précis.

Plutôt que des dégradés de gris, il aurait mieux valu choisir des motifs, de densité croissante selon le degré de satisfaction. Mais, cette remarque n'a pas d'intérêt autre que de souligner le principe d'économie d'encre, car de toute façon, il n'y aura qu'un seul motif à représenter (figure 4').

Proposition de modifications

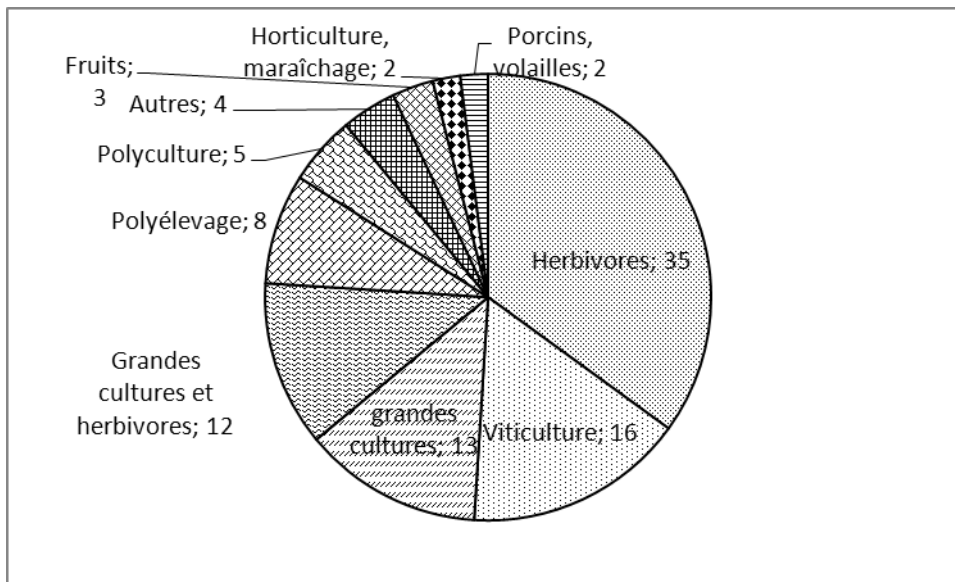
Figure 4' : Distribution du pourcentage de propriétaires ayant participé à l'enquête selon leur degré de satisfaction



5. Description des cultures dans un département

Dans la présentation d'un département, le graphique de la figure 5 a été utilisé. Dans sa version originale, il est en couleur.

Figure 5 : Répartition des exploitations par filière dans le département (Agreste, 2008)



Commentaires

C'est un diagramme circulaire.

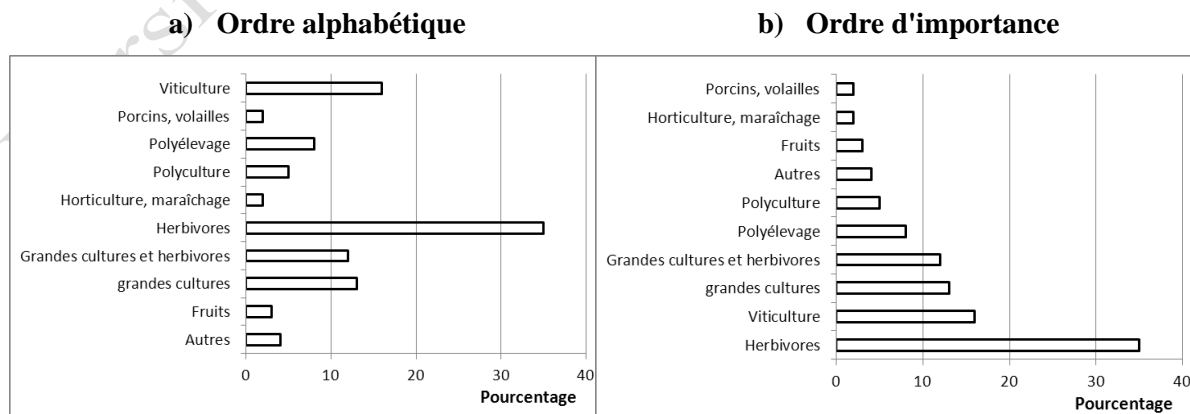
Il n'est pas adapté : il s'agit d'une variable qualitative nominale, qui justifie un diagramme à barres.

Les diagrammes circulaires sont d'un usage à limiter aux circonstances où l'on veut éviter de suggérer un ordre entre les différents items, ce qui n'est pas le cas ici.

Proposition de modifications

Pour le meilleur usage d'un diagramme à barres, les données ont été classées par ordre d'importance, ce qui facilite la perception, en choisissant une disposition évoquant une pyramide (figure 5'b) : la version classée par ordre alphabétique est effectivement nettement moins parlante (figure 5'a) !

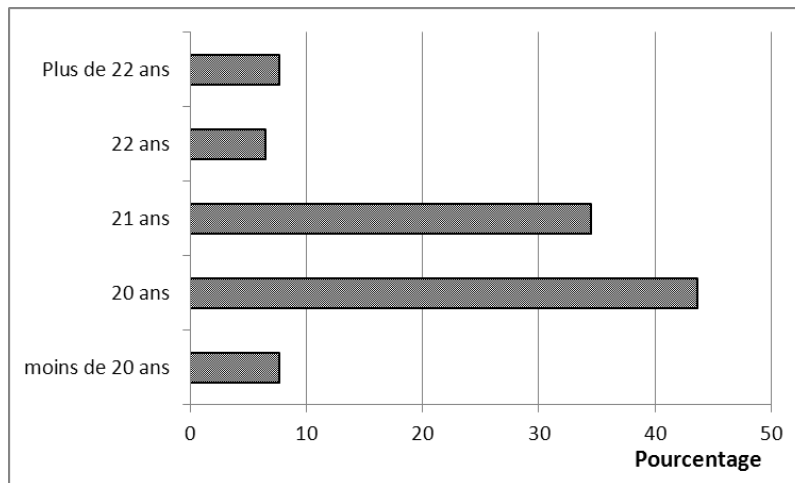
Figure 5' : Répartition des exploitations par filière dans le département (Agreste, 2008)



6. Age des étudiants

Une enquête a été réalisée sur la population des étudiants d'un établissement d'enseignement supérieur. Le résultat comportait la figure 6.

Figure 6 : Age à l'entrée dans l'établissement



Commentaires

Il s'agit d'un diagramme à barres.

Il n'est pas adapté : il s'agit d'une variable qualitative ordinale, qui justifie un diagramme à colonnes.

Selon le principe d'économie d'encre, il vaudrait mieux se passer du grisé.

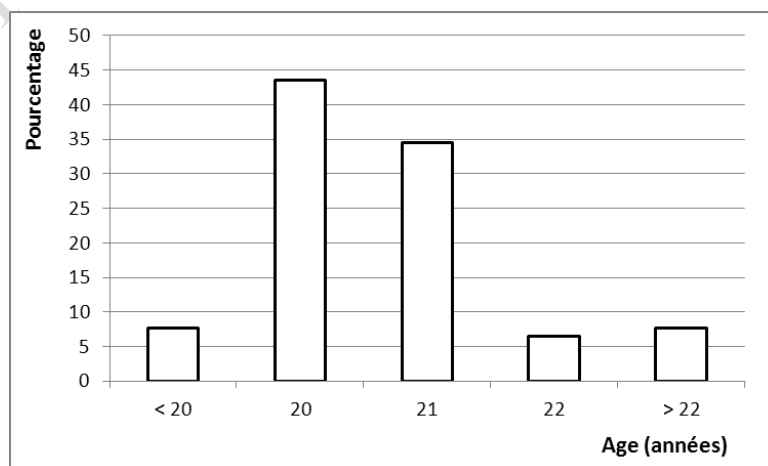
Proposition de modifications (figure 6')

L'indication « ans » a été supprimée, car inutile, la légende de l'axe des abscisses précisant qu'il s'agit d'années.

Le grisé a donc été remplacé par un simple contour avec remplissage en blanc, qui ressort du fait que les lignes de fond, qui permettent de mieux estimer les valeurs des ordonnées, ont été conservées.

Le titre est plus conforme aux exigences de présentation.

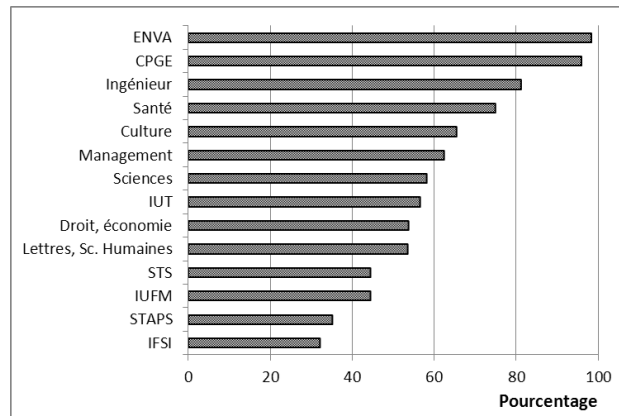
Figure 6' : Distribution de fréquence des étudiants selon leur âge à l'entrée dans l'établissement



7. Mentions au bac

Une enquête réalisée en 2010 par l'Observatoire national de la vie étudiante a permis de comparer les pourcentages d'étudiants ayant obtenu une mention au baccalauréat à l'ENV d'Alfort (ENVA) et dans différents établissements d'enseignement supérieur (figure 7).

Figure 7 : Proportion d'étudiants de l'ENVA ayant obtenu une mention au baccalauréat en comparaison avec d'autre type d'étude (en %)



Commentaires

C'est un diagramme à barres.

Il est adapté : il s'agit d'une variable qualitative nominale, qui justifie un diagramme à barres.

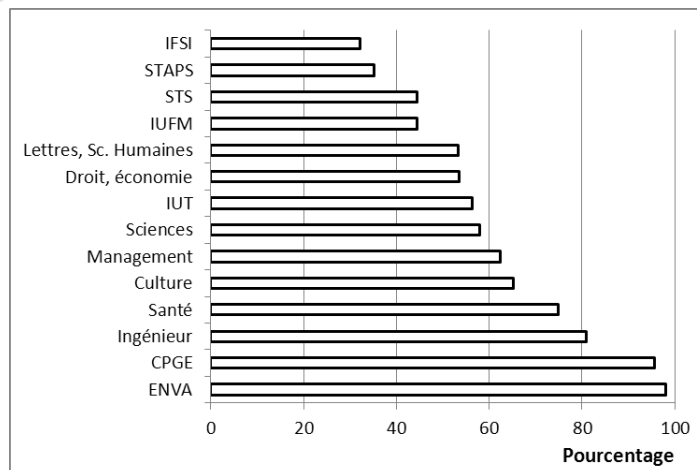
L'ordre choisi, plaçant l'ENVA en haut de cette pyramide inversée, risque toutefois de froisser certaines susceptibilités (d'autant que la couleur d'origine était le rouge pour cette observation !) Il vaudrait mieux disposer les valeurs dans l'autre ordre, suggérant l'image d'une pyramide (partielle).

De même, le titre devrait être plus général, sans mettre l'accent sur un établissement en particulier : c'est le lecteur qui fera la comparaison.

Le grisé devrait être évité.

Proposition de modifications (figure 7')

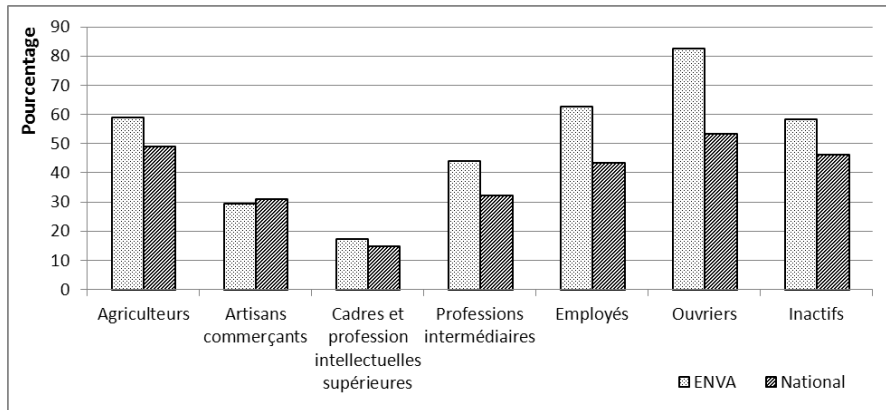
Figure 7' : Distribution des étudiants ayant obtenu une mention au baccalauréat dans différents établissements d'enseignement supérieur (en %)



8. Origine sociale des parents

Une enquête réalisée en 2010 par l'Observatoire national de la vie étudiante a permis de décrire les origines sociales du père des étudiants boursiers dans les établissements d'enseignement supérieur. Ces données ont été comparées à celle d'une enquête réalisée à l'Ecole nationale vétérinaire d'Alfort (ENVA) (figure 8).

Figure 8 : Répartition des étudiants boursiers selon la catégorie socioprofessionnelle du père (en %)



Commentaires

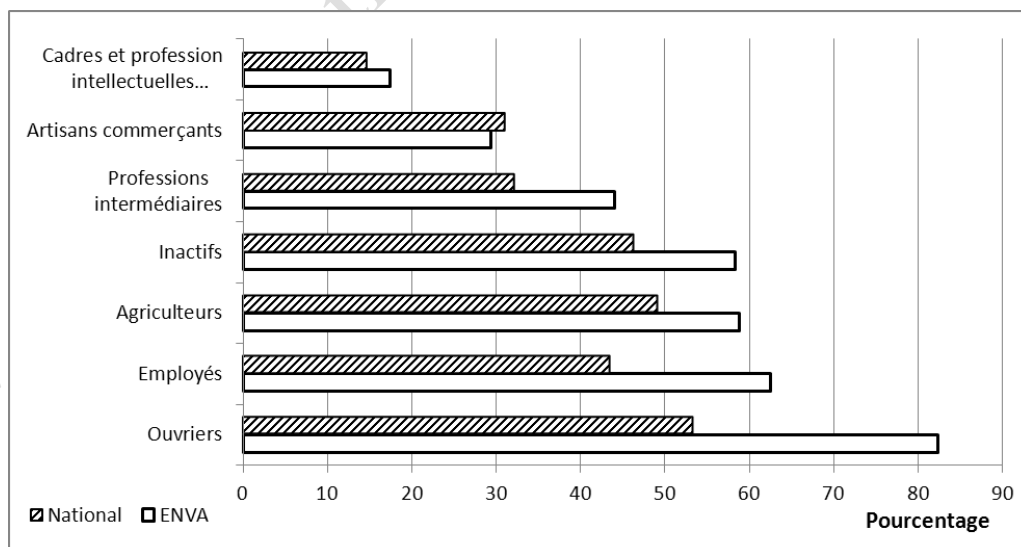
C'est un diagramme à colonnes.

La variable est nominale. Elle aurait dû être représentée par un diagramme à barres, pour éviter de suggérer un ordre hiérarchique.

Les grisés devraient être évités.

Proposition de modifications (figure 8')

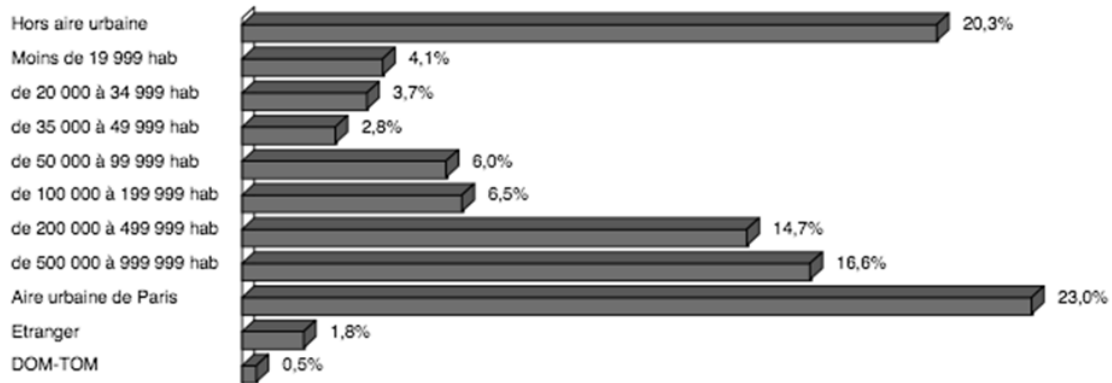
Figure 8' : Répartition des étudiants boursiers selon la catégorie socioprofessionnelle du père (en %)



9. Commune d'origine des étudiants vétérinaires

Une thèse de doctorat vétérinaire s'est intéressée à la taille de la commune où les étudiants vétérinaires ont passé leur enfance (Figure 9).

Figure 9 : Taille de la commune dans laquelle les étudiants entrés dans les ENV françaises en 2005 ont passé la majorité de leur enfance



Commentaires

Il s'agit d'un diagramme à barres.

La pertinence de ce choix se discute.

D'un côté, la répartition en classes selon la taille correspond à une variable quantitative quasi continue, qui aurait justifié un diagramme à colonnes.

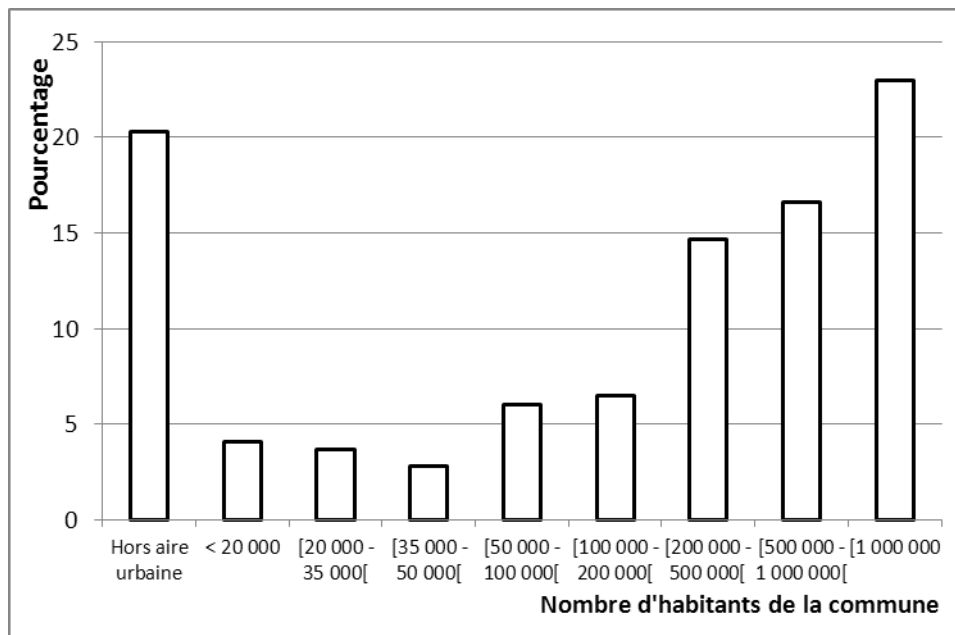
D'un autre, l'hétérogénéité des libellés, combinant des données chiffrées, et des éléments d'ordre nominal (Etranger, DOM-TOM, Paris ayant une population supérieure à un million d'habitants) justifient le traitement par un diagramme à barres.

En fait, qu'y a-t-il d'intéressant à voir ? Indiscutablement, c'est l'information quantitative qui est la plus intéressante : on voit distinctement deux distributions, l'une, décroissante, qui concerne le début de la distribution, l'autre, croissante en fonction de l'augmentation de la taille des communes. Les deux mentions complémentaires (Etranger, DOM TOM) sont là pour le besoin de l'exhaustivité des données... mais pas pour l'intérêt du phénomène à mettre en évidence. L'auteur aurait dû faire le choix d'exclure ces deux catégories, qui ne correspondent pas à l'objet (« taille des communes »), mais à un autre (origine géographique). L'axe ne respecte pas la proportionnalité de taille des communes.

D'autre part, le graphique comporte une orientation en 3D, et l'utilisation du grisé est contestable.

Proposition de modifications (figure 9')

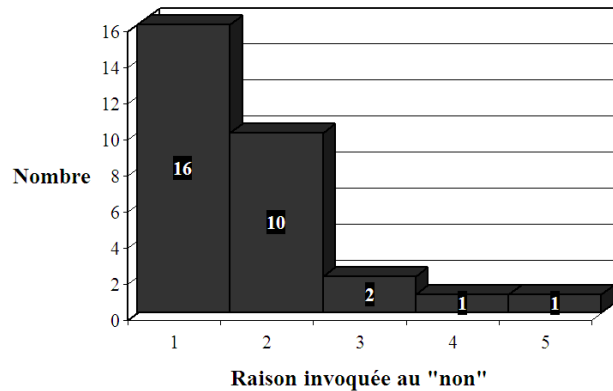
Figure 9' : Distribution des étudiants de l'École vétérinaire d'Alfort selon la taille de leur commune d'origine (France métropolitaine) réparties en classes (amplitude non régulière)



10. Raisons invoquées pour ne pas participer à une étude

Une étude sur les cancers du chien a été réalisée. Sur 108 cliniques vétérinaires contactées, 82 ont accepté de participer, 26 ont refusé. Les raisons du refus ont été prospectées par une question ouverte (figure 10) : de la plus fréquente à la moins fréquente, le manque de temps (raison 1), le manque d'intérêt (raison 2), une clientèle comprenant peu ou pas de chiens (raison 3), le refus systématique de répondre à des questionnaires téléphoniques (raison 4), et enfin une clientèle formée essentiellement de cas référés d'ophtalmologie (raison 5).

Figure 10 : Raisons invoquées pour justifier le refus de répondre au questionnaire



Commentaires

La figure 10 n'est pas compréhensible sans se référer au texte : c'est une infraction majeure aux principes fixés, selon lesquels une figure ou un tableau doivent pouvoir être compris sans l'obligation de devoir consulter le texte.

L'auteur ayant remplacé le libellé des motifs par des codes numériques a été incité à utiliser une représentation les disposant sur un axe horizontal. Il a choisi un histogramme, avec effet 3D (inutile).

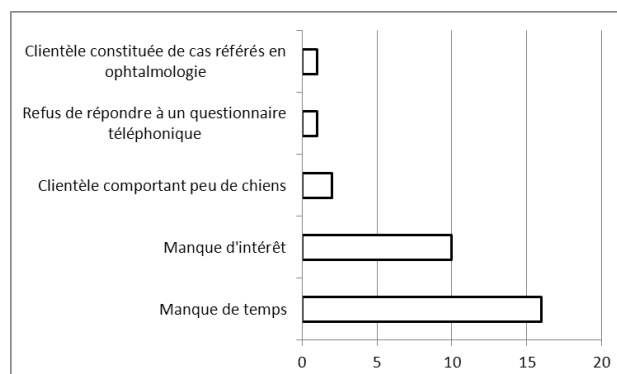
En fait, la variable est **qualitative**, et non quantitative comme le suggèrent l'axe et la représentation par un histogramme, totalement inappropriée. Les motifs en question ne peuvent pas être classés : la variable n'est donc même pas ordinale, mais **nominale**.

Il aurait fallu faire apparaître les motifs de façon explicite, et pour cela les disposer en ligne, et par conséquent utiliser un diagramme à barres.

Le titre devrait faire référence au nombre d'observations.

Proposition de modifications (figure 10').

Figure 10' : Distribution des raisons de refus de participer à l'étude (n = 26)



11. Valeurs propres

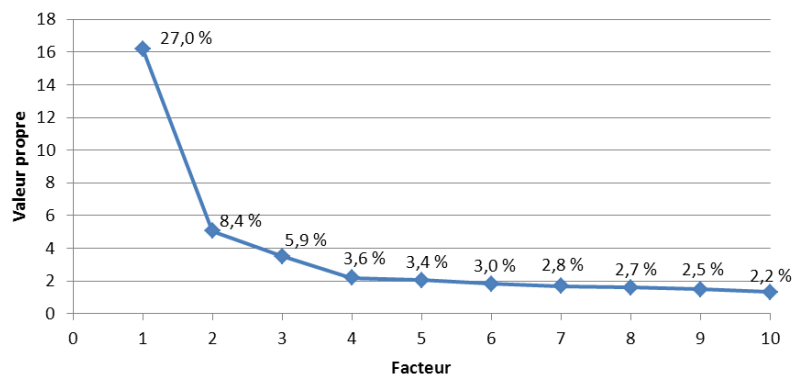
La description statistique d'un fichier de données comportant un grand nombre de variables doit aller plus loin que l'approche uni- ou bi-variée, par une description multivariée, comme l'ACP (analyse en composantes principales).

Celle-ci calcule des variables synthétiques qui permettent au mieux de rendre compte du nuage à n dimensions, appelées valeurs propres ; la qualité de cet axe est appréciée par la proportion de la variance qu'il porte ; les projections sur un plan associent ces valeurs propres deux à deux : la qualité de la projection est appréciée par la somme des valeurs propres.

On cherche à associer des axes deux par deux, pour obtenir une représentation plane combinant deux axes donnant la meilleure représentation possible, ce que l'on juge, d'une part, selon la part de variance prise en compte, d'autre part, selon le pouvoir discriminant de la représentation.

Une étude sur l'attachement des propriétaires à leur chat a mis en relation le score de réponse des propriétaires à un questionnaire standardisé (*Pet Attachment Scale, PAS* ; réponses selon une échelle de Likert), et leurs réponses à diverses questions (démographiques, comportementales). La figure 11 représente les valeurs des 10 premiers axes.

Figure 11 : Valeurs propres et pourcentages de variance expliquée par les 10 premiers facteurs pour les chats

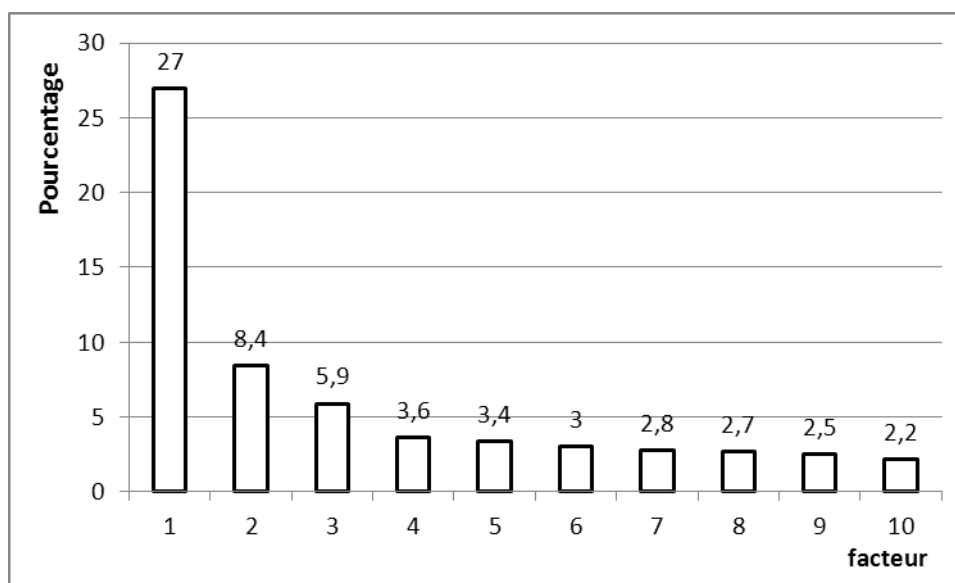


Commentaires

Il s'agit d'un diagramme à lignes. La variable est ordinale, car ce sont les facteurs qui sont représentés selon leur ordre de classement de leur valeur propre.

Une variable ordinale est normalement représentée par un diagramme à colonnes, qui devraient être placées sur la graduation (puisque les valeurs du numéro de facteur sont entières), comme c'est le cas actuellement. Donc, dans l'absolu, la représentation n'est pas correcte et il faudrait choisir un diagramme à colonnes (figure 11').

Figure 11' : Valeurs propres et pourcentages de variance expliquée par les 10 premiers facteurs pour les chats



Mais, l'intention de l'auteur est de trouver une indication graphique du nombre d'axes à prendre en compte. L'usage est de chercher un décrochement, une « épaule » suggérant une discontinuité dans la répartition des valeurs. Il faut reconnaître que la représentation par un diagramme à lignes est beaucoup plus suggestive : le décrochement se situe au niveau du 2^{ème} axe.

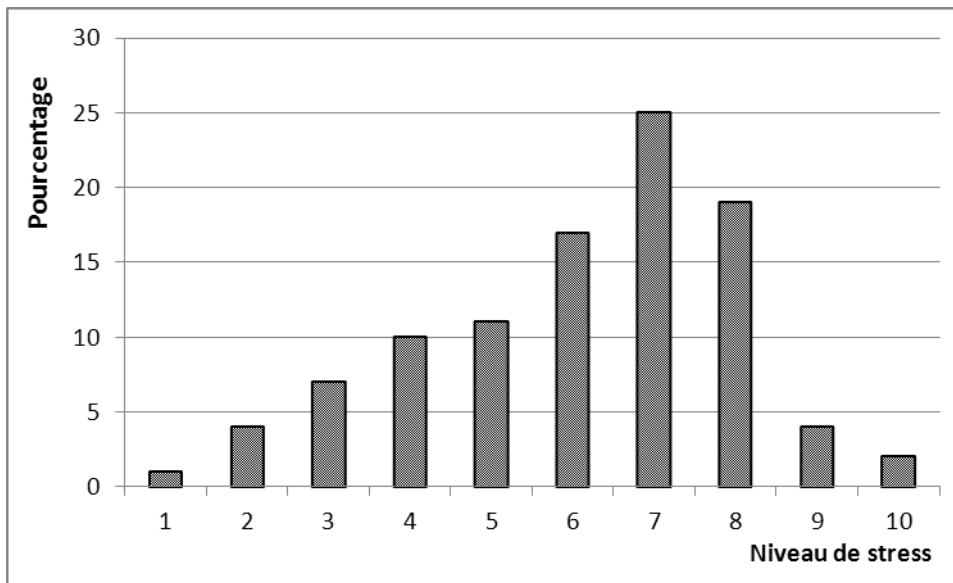
Le diagramme à lignes était donc tout aussi justifié, malgré une entorse à l'orthodoxie, car cette forme sert aussi bien l'intention de l'auteur.

12. Etude du stress au travail dans les cabinets vétérinaires

Le stress au travail dépend de l'activité et des circonstances. Une enquête a été conduite auprès de 494 vétérinaires qui ont accepté de répondre à un questionnaire afin de faire l'étude du stress au travail dans un cabinet vétérinaire.

La question illustrée par la figure 12 est relative au niveau de stress en relation avec les exigences du travail (horaires par exemple), évalué sur une échelle de 1 à 10 par les personnes enquêtées.

Figure 12 : Niveau de stress en lien avec les exigences du travail (n = 494)



Commentaires

Une échelle subjective comportant 10 niveaux... Cela semble être quantitatif ? Oui, mais, on ne peut pas faire la somme des niveaux de stress perçu de deux personnes pour dire que dans un cabinet vétérinaire, le niveau de stress est égal à la somme du niveau du stress ressenti par ces personnes. Donc, si ce n'est pas quantitatif, cela ne peut être que qualitatif, et par conséquent, il s'agit d'une **variable ordinale**.

La représentation qui convient est donc bien un **diagramme à colonnes**.

Maintenant, les colonnes, où doit-on les placer : *sur* les graduations, ou *entre* ?

Sur les graduations, cela correspond à un **diagramme à bâtons**, pour une variable quantitative discrète. Certes, les unités de score de stress sont entières, mais il ne s'agit pas d'une variable quantitative. Par ailleurs, pour la représentation d'une variable ordinale, on place bien les valeurs *entre* les graduations. Ce qui peut nous troubler, c'est que les variables ordinales, comme un niveau de satisfaction, sont généralement exprimées de façon littérale (pas du tout, un peu, *etc.*), et leur représentation est placée *entre* les graduations. Mais, on peut aussi remplacer ces échelles littérales par des chiffres, et par conséquent, c'est bien entre les graduations qu'il faut placer les colonnes.

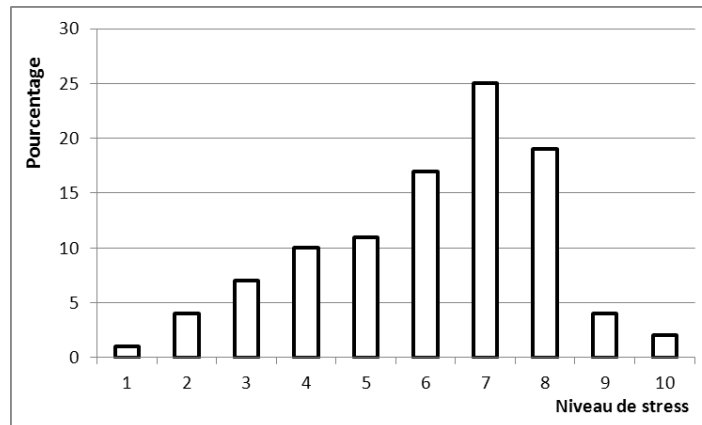
Le choix du graphique est donc correct !

Notons au passage que l'auteur avait également pensé à indiquer le nombre d'individus constituant l'échantillon, comme cela devrait être systématiquement pratiqué.

Proposition de modification

Le titre pourrait être amélioré, et les grisés supprimés :

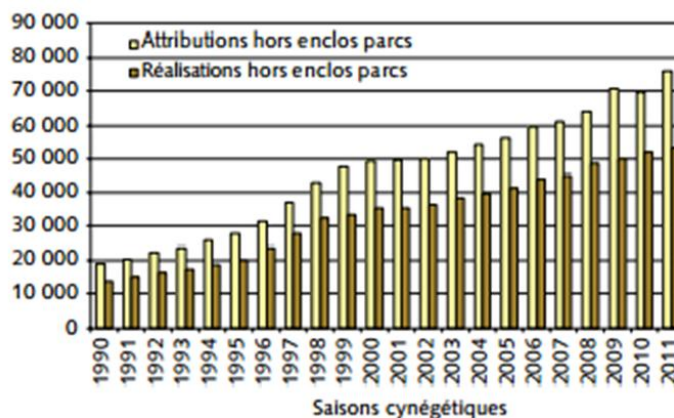
Figure 12' : Distribution des effectifs de vétérinaires ayant participé à l'enquête selon le niveau de stress perçu sur une échelle de 1 à 10 en relation avec les exigences au travail (n = 494).



13. Tableau de chasse des cerfs en France

La population d'animaux sauvages a considérablement augmenté au cours des vingt dernières années, comme en témoignent les tableaux de chasse réalisés (figure 13).

Figure 13 : Évolution annuelle du tableau de chasse national cerf



Commentaires

S'agissant d'un dénombrement de cerfs abattus, il s'agit d'une **variable quantitative**. Les cerfs sont des unités indivisibles, c'est donc une **variable discrète**, mais dont le très grand nombre en fait une **variable quasi-continue**.

Certes, mais pour choisir le type de représentation, c'est l'axe des abscisses qu'il faut considérer.

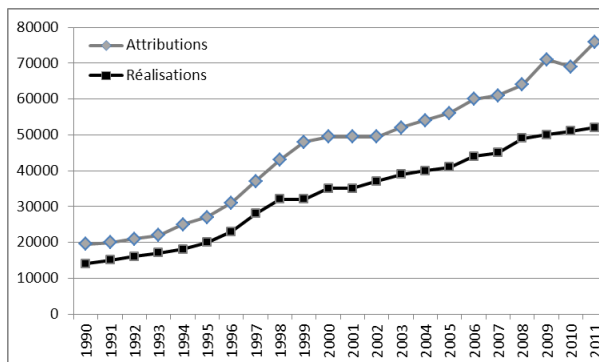
L'axe des abscisses est une représentation des années. Il s'agit d'une **variable temporelle**, continue, mais non quantitative. De ce point de vue, le choix d'une représentation par un diagramme à lignes pourrait être pertinent.

Mais, les données statistiques sont annuelles, ce qui rompt la continuité temporelle. Pour rendre compte de cette rupture, le choix d'un diagramme à colonnes serait plus indiqué.

Les deux modes de représentation étant également possibles, le choix doit résulter de *l'objectif poursuivi* : on veut montrer une *tendance* générale d'augmentation tout en permettant la *comparaison de deux séries*. Si la représentation par un diagramme à colonnes suggère bien cette évolution, on pourrait également choisir le diagramme à lignes, qui renforce la perception d'une tendance et facilite la comparaison de deux séries

Proposition de modifications (figure 13').

Figure 13' : Évolution annuelle du tableau de chasse national cerf en France de 1990 à 2011



La comparaison des deux figures est révélatrice : en supprimant les traits verticaux, qui soulignent la compilation de données annuelles aux dépens de la tendance, on renforce bien la perception de l'évolution. Finalement, c'est le diagramme à lignes qui sert mieux le propos de l'auteur.

Version de travail. Ne pas diffuser

BIBLIOGRAPHIE

- ANCELLE T. – Statistique, épidémiologie. 2011, Paris, Maloine.
- ANDRAUD M., DUMAREST M., CARIOLET R., BARNAUD E., EONO Florent, PAVIO N., ROSE N. – Modélisation de la transmission du virus de l'hépatite E chez le porc à partir de données expérimentales. *Épidémiol. et santé anim.*, 2013, **63**, 39-50.
- AUVIGNE V. et LÉGER L. – La typologie, un outil au service du dialogue entre épidémiologiste et praticien. L'exemple de l'analyse des profils sérologiques complexes. *Épidémiol. et santé anim.*, 2013, **63**, 93-101.
- BEKARA M. et A. – Impact de l'évolution du schéma de prophylaxie et des structures et pratiques d'élevage sur l'évolution de la tuberculose bovine en France entre les années 1965 et 2000 : modélisation de l'incidence cheptel et de la dynamique de transmission intra-élevage de l'infection. Thèse d'Université Paris Sud, 2014, 251 p.
- BÉNET J.-J. et DUFOUR B. – Analyse des données épidémiologiques produites par la lutte contre la tuberculose bovine sur le terrain en France. *Épidémiol. et santé anim.*, 2014, **65**, 53-76.
- BRANGIER B. – Présentation et communication des données : http://www.reseau-naissance.com/joomla/images/tabgraph_donnees_2009.pdf
- CLEVELAND W. S et MCGILL R. Graphical Perception – Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, 1984, **79** (387), 531-554.
- Conseil supérieur de l'Ordre des vétérinaires. Les statistiques de la profession pour 2012. http://www.veterinaire.fr/document/menu_gauche_grand_public/demographie_professionnelle.htm
- CRÉPIN C. – Élaboration d'une méthode d'analyse de la dynamique relationnelle et de la communication entre le vétérinaire et le propriétaire au cours de la consultation. Thèse Doctorat vétérinaire Alfort, 2014, 148 p.
- DELAHOUSSE, A.-L. – L'identité professionnelle du vétérinaire. Thèse Doctorat vétérinaire Alfort, 2014, 247 p.
- du BREIL M. – Enquête nationale tuberculose bovine 2005,2006 et 2007. Rapport de stage de M2 Santé Publique Paris XI et Sciences de la Santé Paris XII, spécialité Surveillance épidémiologique des maladies humaines et animales, 2009, 54 pages.
- FEDIAEVSKY A., BÉNET JJ., BOSCHIROLI M.L. RIVIÈRE J. et HARS J. – La tuberculose bovine en France en 2011, poursuite de la réduction du nombre de foyers. *Bulletin épidémiologique, santé animale et alimentation*, 2012, **54**, 3-12.
- GANE, J. – L'analyse des données de deux études relatives aux tuberculinations. Rapport de stage M1 Méthode en Santé publique Paris XI, 2011, 43 p.
- Interbev, 2008. Liste des types raciaux [de bovins]. <http://www.franceagrimer.fr/content/download/8682/55092/file/races-bovines-v03.pdf>. Consulté le 23/07/2014.
- LAMBERT O. Evaluation d'un nouveau protocole de diagnostic de la tuberculose bovine. Mémoire de stage de Licence professionnelle, Statistiques et informatique décisionnelles pour la santé, IUT Paris Descartes, 2014, 53 p.
- MASSÉ L. et MASSÉ G. – Statistique sanitaire et sociale, 1972, Paris, Foucher éd.
- MOUTOU F., BARRAT J. et BRUYÈRE V. – Virus des chauves-souris. Actualités épidémiologiques en France et dans le monde. *Épidémiol. et santé anim.*, 2000, **38**, 99-107.
- PAYNE A., MILLET L., HARS J., DUFOUR B. et GILOT-FROMONT E. – Interaction entre sangliers, blaireaux et bovins en Côte-d'Or : quelle opportunité pour la transmission de *Mycobacterium bovis* ? *Épidémiol. et santé anim.*, 2013, **63**, 77-92.

- RAMBAUD T. – Etude de la répartition des lésions de tuberculose chez les bovins issus de cheptels infectés en Dordogne. Mémoire de C.E.S. d'épidémiologie animale, Ecole nationale vétérinaire d'Alfort, 2012, 26 p.
- RAUTUREAU S. – Simulations d'épizooties de Fièvre aphteuse et aide à la décision. Approches épidémiologique et économique. Thèse d'Université Paris XI, 2012, 260 p.
- RÉGNIER J.-C. – Histogramme. *Réflexion sur une représentation graphique particulière abusivement utilisée tant dans l'enseignement que dans l'application de la statistique.* In Enseigner la statistique du CM à la seconde. Pourquoi ? Comment ? 1998, 21-42. <http://www.youscribe.com/catalogue/rapports-et-theses/savoirs/sciences-humaines-et-sociales/histogramme-1541848>
- RIVIÈRE J. - Identification et description d'une méthode d'évaluation du dispositif de surveillance de l'infection à *Mycobacterium bovis* dans la faune sauvage libre en France. *Épidémiol. et santé anim.*, 2013, **63**, 25-38.
- SALMI L.R. – Lecture critique et communication médicale scientifique. 3^{ème} édition, Elsevier Masson, 2012, 288 p.
- TOMA B., DUFOUR B., BÉNET J.J., SANAA M., SHAW A. et MOUTOU F. – Epidémiologie appliquée à la lutte contre les maladies animales transmissibles majeures. 3^{ème} éd., Alfort, AEEMA, 2010, 600 p.
- TOMA B. – Le réservoir de la maladie d'Aujeszky. *Épidémiol. et santé anim.*, 2013, **63**, 141-162.
- VALÉRY P. – Œuvres II, Paris, Gallimard, collection Bibliothèque de la Pléiade, 1 726 p..
- VIRIOT D., COCHET A., WATRIN M., BENOIT P., MOYANO M.-B. et GOLLIOT F. – Investigation d'une toxi-infection alimentaire collective à germes multiples. Stade Yves du Manoir, Montpellier (France), février 2010. *BEH* 2011, 25/28 juin, p 289-292.
- VITOUR D, GUILLOTIN J., SAILLEAU C., VIAROUGE C., DESPRAT A., WOLFF F. *et al.* – Interférence des anticorps maternels avec la réponse des veaux à la vaccination par un vaccin inactivé contre le sérotype 8 du virus de la fièvre catarrhale ovine. *Épidémiol. et santé anim.*, 2010, **58**, 123-132

INDEX

Cartes.....	26, 53
Densité.....	55
Limite d'extension	56
Points (à)	54
Diagramme	
Barres (à).....	36
Bâtons (à)	32
Boîtes (à).....	40
Choix	17
Circulaire.....	38
Colonnes (à)	34, 77, 83
Choix selon type de pourcentage.....	79
Colonnes (à), bâtons (à)	
Distinction	78
Colonnes (à), histogramme	
Choix	77
Histogramme	44, 87
Lignes (à).....	30, 75, 76
Paliers (à).....	48
Points (à)	28, 73, 74, 76
Pyramide.....	42
Radar (en).....	46
Diagrammes	26
Excel	
Diagramme à paliers (comment faire).....	96
Eviter format texte pour variable numérique	94
Légende des abscisses	96
Placer les valeurs sur graduations	95
Séparer "%" par une espace.....	96
Figures	26
Graphiques	26
Memento.....	15
Recommandations	
Générales	57
Compréhension.....	58
Esthétique	58
Numérotation.....	57
Spécifiques pour un tableau	59
Spécifiques pour une figure.....	60
Schémas.....	26, 50
Tableau synthétique des diagrammes.....	15

Tableaux	24, 67
Variable	20
Qualitative	20
Binaire	22
Nominale	22
Ordinale	22
Quantitative	20
Continue	21
Discrète.....	21
Quantitative et qualitative	
Distinguer	63
Temporelle	22
Texte.....	23, 65

Version de travail. Ne pas diffuser