

# Statistique pour psychologues


Cours, QCM  
et exercices corrigés

Nicolas Guéguen

4<sup>e</sup> édition

**DUNOD**

Mise en pages : Nord Compo

<p>Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.</p> <p>Le Code de la propriété intellectuelle du 1<sup>er</sup> juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements</p>		<p>d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée. Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).</p>
--	---	---

© Dunod, 2022

11 rue Paul Bert – 92240 Malakoff

ISBN 978-2-10-081761-0

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.



# *Table des matières*

AVANT-PROPOS .....	XI
--------------------	----

## ***COURS***

CHAPITRE 1 DE L'OBSERVATION DES FAITS PSYCHOLOGIQUES À LEUR RÉSUMÉ STATISTIQUE .....	1
--	---

<b>1. Typologie des variables et représentation graphique des données</b> .....	4
1.1. Les échelles de mesure nominale .....	4
1.1.1. Définition et exemples .....	4
1.1.2. Présentation et représentation graphique de distributions de mesures nominales .....	6
1.2. Les échelles de mesure ordinale .....	8
1.2.1. Définition et exemples .....	8
1.2.2. Des rangs ordonnés .....	9
1.2.3. Présentation et représentation graphique de distributions de mesures ordinales .....	9
1.3. Les échelles de mesure d'intervalles .....	11
1.3.1. Définition et exemples .....	11
1.3.2. Présentation et représentation graphique de distributions de mesures d'intervalles .....	12
1.4. Les échelles de mesure de rapports .....	17
1.5. La combinaison de variables .....	17
1.5.1. La combinaison d'échelles d'intervalles .....	17
1.5.2. La combinaison d'échelles ordinales .....	18

1.5.3.	La combinaison d'échelles nominales. . . . .	20
1.5.4.	La représentation graphique : un puissant instrument d'analyse . . . . .	21
1.6.	La difficile distinction des variables : comment s'y retrouver? . . . . .	21
1.7.	Les changements d'échelles . . . . .	23
1.8.	Les emboîtements d'échelles . . . . .	24
<b>2.</b>	<b>Résumé et description numérique des données. . . . .</b>	<b>25</b>
2.1.	Valeur centrale et dispersion dans le cas d'échelles d'intervalles . . . . .	26
2.1.1.	L'indicateur de tendance centrale . . . . .	26
2.1.2.	Mesure de la dispersion. Exemple d'une évaluation contrastée. . . . .	26
2.1.3.	L'estimation sans biais de la variation : approche de la notion de degrés de liberté . . . . .	29
2.1.4.	Le coefficient de variation : un indicateur de comparaison de dispersion . . . . .	31
2.1.5.	Mesure de l'aplatissement d'une courbe : le coefficient « Kurtosis » . . . . .	32
2.2.	Valeur centrale et dispersion dans le cas d'échelles ordinaires . . . . .	35
2.2.1.	L'indicateur de tendance centrale . . . . .	35
2.2.2.	La dispersion des échelles ordinaires . . . . .	37
2.3.	Valeur centrale et dispersion dans le cas d'échelles nominales . . . . .	41
2.3.1.	L'indicateur de tendance centrale . . . . .	41
2.3.2.	Mode et échelles de mesure d'intervalle . . . . .	42
2.3.3.	La comparaison de fréquences modales . . . . .	43
2.3.4.	L'indicateur de dispersion . . . . .	43
2.4.	Moyenne, médiane, mode : quand les utiliser? . . . . .	44
2.5.	Analyse complémentaire de la dispersion : analyse de l'asymétrie d'une distribution de données . . . . .	46
2.5.1.	Le coefficient « Skewness » pour une échelle d'intervalles. . . . .	46
2.5.2.	Le coefficient d'asymétrie pour échelle de mesure ordinale . . . . .	49
<b>3.</b>	<b>Applications avec traitement informatique . . . . .</b>	<b>49</b>
<b>CHAPITRE 2 DE L'ÉCHANTILLON À LA POPULATION : L'INFÉRENCE STATISTIQUE . . . . .</b>		<b>57</b>
<b>1.</b>	<b>Approche conceptuelle de la notion d'échantillon et de population en statistique . . . . .</b>	<b>59</b>
1.1.	Une population en statistique . . . . .	59
1.2.	Échantillon et population . . . . .	61
1.2.1.	Les échantillons indépendants . . . . .	61
1.2.2.	Les échantillons appariés. . . . .	62
1.3.	Le prélèvement d'échantillons. . . . .	63

1.3.1. Les méthodes d'échantillonnage probabiliste . . . . .	63
1.3.2. Les méthodes d'échantillonnage empirique . . . . .	65
<b>2. Probabilités et distribution de probabilités :     approche empirique.</b> . . . . .	<b>67</b>
2.1. Pile ou face : l'implacable distribution des probabilités . . . . .	67
2.2. Et si tous les échantillons du monde faisaient une loi normale : le principe de l'urne . . . . .	70
2.3. Distribution d'un caractère continu : principe de distribution de la loi normale . . . . .	74
2.4. Principe de l'urne sur des mesures d'intervalles continus ou discontinus . . . . .	74
2.5. Lecture de la loi normale centrée-réduite : détermination des diverses probabilités. . . . .	76
2.6. Une même loi pour plusieurs lois : principe de l'écart centré-réduit	79
2.6.1. Évaluation de la probabilité d'apparition d'une valeur dans une distribution . . . . .	81
<b>3. Les fluctuations de moyennes et de proportions.</b> . . . . .	<b>84</b>
3.1. Les intervalles de fluctuation de moyennes et de proportions provenant d'une population . . . . .	85
3.1.1. Intervalle de fluctuation d'une moyenne de population . . . . .	87
3.1.2. Intervalle de confiance d'une proportion. . . . .	93
3.2. Les intervalles de fluctuation de moyenne et de proportions provenant d'un échantillon : approche de l'inférence statistique. . .	96
3.2.1. Inférence sur les moyennes. . . . .	97
3.2.2. Inférence sur les proportions. . . . .	100
3.3. Population finie et infinie . . . . .	102
3.3.1. Population finie . . . . .	102
3.3.2. Population infinie . . . . .	103
3.3.3. Correction de l'erreur-type dans le cas d'une population finie	103
3.3.4. Détermination et contrôle des paramètres de fluctuation . . . . .	106
3.3.5. Limite du facteur de correction pour une population finie . . . . .	107
<b>4. Théorie des tests en statistique.</b> . . . . .	<b>107</b>
4.1. Hypothèses et tests d'hypothèses. . . . .	108
4.2. Tests paramétriques et non paramétriques . . . . .	109
<b>5. Applications avec traitement informatique.</b> . . . . .	<b>111</b>

CHAPITRE 3	LA COMPARAISON DE MOYENNES :	
	LE HASARD DES DIFFÉRENCES (ACTE I).....	115
<b>1.</b>	<b>Principe des tests de comparaison de moyennes</b> .....	117
<b>2.</b>	<b>La comparaison de deux moyennes</b> .....	119
2.1.	La comparaison d'une moyenne à une norme .....	120
2.1.1.	Méthode de calcul .....	120
2.1.2.	Hypothèses .....	122
2.1.3.	Comparaison dans le cas où la variance de population est connue .....	122
2.1.4.	Comparaison dans le cas où la variance de population est inconnue .....	123
2.1.5.	Facteur de correction dans le cas d'une population finie.....	127
2.2.	La comparaison de deux moyennes observées.....	128
2.2.1.	Comparaison de deux moyennes provenant d'échantillons indépendants .....	130
2.2.2.	Comparaison de deux moyennes provenant d'échantillons appariés .....	134
2.3.	Comparaison de moyennes : autre méthode .....	139
2.3.1.	Méthode de l'intervalle de fluctuation pour la comparaison d'une moyenne à une norme.....	139
2.3.2.	Méthode de l'intervalle de fluctuation d'une moyenne : cas de deux moyennes .....	140
2.4.	Test à une issue ou à deux issues : fondements pratiques .....	142
<b>3.</b>	<b>La comparaison de trois moyennes et plus :</b> <b>l'analyse de variance</b> .....	144
3.1.	Principe de la méthode.....	145
3.2.	L'analyse de variance sur des moyennes d'échantillons indépendants .....	148
3.3.	L'analyse de variance sur des moyennes d'échantillons appariés... ..	154
3.4.	Extension du principe de l'analyse de variance : l'analyse de plans factoriels .....	159
3.4.1.	L'analyse de variance d'un plan factoriel utilisant des échantillons indépendants .....	160
3.4.2.	Plans en tous genres .....	165
<b>4.</b>	<b>Applications avec traitement informatique</b> .....	166

CHAPITRE 4	LA COMPARAISON DE FRÉQUENCES :	
	LE HASARD DES DIFFÉRENCES (ACTE II) . . . . .	173
<b>1.</b>	<b>Principe des tests de comparaison de fréquences</b>	
	<b>ou de proportions</b> . . . . .	176
<b>2.</b>	<b>La comparaison de deux proportions :</b>	
	<b>approche paramétrique</b> . . . . .	178
2.1.	La comparaison d'une proportion observée à une norme . . . . .	178
2.2.	La comparaison de deux proportions observées . . . . .	181
2.3.	Cas des échantillons appariés . . . . .	183
2.4.	Hypothèse bilatérale ou unilatérale . . . . .	185
2.5.	Comparaison de deux proportions :	
	méthode par l'intervalle de fluctuation . . . . .	185
2.5.1.	Comparaison d'une proportion observée à une norme . . . . .	186
2.5.2.	Comparaison de deux proportions par la méthode	
	de l'intervalle de fluctuation . . . . .	187
2.6.	Limites du test de comparaison de proportions . . . . .	189
<b>3.</b>	<b>La comparaison de plusieurs fréquences :</b>	
	<b>le test du <math>\chi^2</math></b> . . . . .	178
3.1.	Méthode générale de calcul . . . . .	190
3.2.	Comparaison d'une répartition d'effectifs observée	
	à une répartition théorique : le $\chi^2$ d'ajustement . . . . .	191
3.2.1.	Comparaison d'une répartition d'effectifs observée	
	à une répartition d'effectifs uniforme . . . . .	191
3.2.2.	Approfondissement de l'analyse . . . . .	195
3.2.3.	Comparaison d'une répartition d'effectifs observée	
	à une répartition d'effectifs théoriques indépendants . . . . .	197
3.3.	Le $\chi^2$ d'indépendance . . . . .	200
3.3.1.	Cas des échantillons indépendants . . . . .	200
3.3.2.	Cas des échantillons appariés : le $\chi^2$ de Mac Nemar . . . . .	210
3.4.	Après le texte, l'analyse décomposée du $\chi^2$ . . . . .	212
3.5.	Facteurs empiriques agissant sur la valeur du $\chi^2$ . . . . .	215
3.5.1.	Amplitude de l'écart entre effectifs observés	
	et effectifs théoriques . . . . .	215
3.5.2.	Importance du volume d'observations . . . . .	217
3.6.	Limites d'utilisation du test du $\chi^2$ . . . . .	218
3.6.1.	Un nombre minimal d'observations . . . . .	219
3.6.2.	L'évaluation du nombre de degrés de liberté . . . . .	219
3.7.	Limites à l'interprétation du test du $\chi^2$ . . . . .	223
<b>4.</b>	<b>Applications avec traitement informatique</b> . . . . .	223

CHAPITRE 5	LA CORRÉLATION :	
	LA RECHERCHE DES PARALLÈLES . . . . .	233
<b>1.</b>	<b>Principe des tests de corrélation</b> . . . . .	235
<b>2.</b>	<b>La corrélation linéaire</b> . . . . .	240
2.1.	Relation entre deux variables : approche graphique . . . . .	240
2.2.	Méthode de calcul . . . . .	241
2.3.	Corrélations en tous genres . . . . .	245
2.4.	Corrélations contagieuses. . . . .	246
2.5.	Bien voir ! Bien vu ! Analyse graphique de la corrélation . . . . .	247
2.6.	La corrélation partielle : la recherche des facteurs communs . . . . .	250
2.7.	Après la corrélation . . . . .	252
2.7.1.	L'appréciation de l'influence de la relation. . . . .	252
2.7.2.	$r$ ou $r^2$ ? . . . . .	254
2.8.	La régression linéaire : un outil de pronostic . . . . .	255
2.8.1.	Régression et corrélation . . . . .	257
2.8.2.	Régression et prédiction . . . . .	258
2.8.3.	Prudence et prédiction . . . . .	261
<b>3.</b>	<b>La corrélation des rangs</b> . . . . .	262
3.1.	Méthode de calcul . . . . .	263
3.2.	L'ajustement à un modèle théorique . . . . .	265
3.3.	Les rangs ex aequo dans la corrélation de Spearman . . . . .	266
3.3.1.	Détermination des rangs ex aequo. . . . .	267
3.3.2.	La présence de rangs ex aequo en grand nombre . . . . .	267
3.4.	Test de corrélation bilatéral ou unilatéral . . . . .	269
3.5.	Combinaison d'échelles d'intervalles et d'échelles ordinaires. . . . .	271
<b>4.</b>	<b>Les coefficients d'association de variables nominales</b> 273	
4.1.	L'association de variables dichotomiques : le coefficient Phi . . . . .	274
4.2.	L'association de variables polychotomiques : le coefficient de Cramer . . . . .	279
<b>5.</b>	<b>Conclusion : corrélation et causalité</b> . . . . .	282
<b>6.</b>	<b>Applications avec traitement informatique</b> . . . . .	283



## **QCM, EXERCICES ET CORRIGÉS**

EXERCICE 1 : LA BONNE ODEUR DE L'ARGENT.....	291
EXERCICE 2 : POURBOIRES POUR MANGER.....	295
EXERCICE 3 : THÉRAPIE ASSISTÉE PAR ANIMAUX .....	298
EXERCICE 4 : PIED-DANS-LA-PORTE AMÉRICAIN CONTRE PIED-DANS-LA-PORTE FRANÇAIS : MATCH NUL .....	304
EXERCICE 5 : SUICIDES ET CHÔMAGE.....	313
EXERCICE 6 : LE TÉLÉPHON QUI SON .....	317
EXERCICE 7 : IL FAUT BOIRE POUR LE CROIRE.....	323
EXERCICE 8 : LA MUSIQUE ADOUCIT LE STRESS .....	328
EXERCICE 9 : UN « BON » COUPABLE.....	340
EXERCICE 10 : FAITES L'HUMOUR, PAS LA GUERRE.....	345
INTERNET ET LA STATISTIQUE.....	351
<b>1. Pas de statistique sans statistiques :         rapide aperçu des sites de données.</b> .....	353
<b>2. Les cours en ligne : formation et entraînement         à distance.</b> .....	354
<b>3. Les logiciels de traitement statistique</b> .....	355
3.1. Les logiciels « Free ».....	355
3.2. Les logiciels commerciaux .....	356
3.2.1. Les logiciels en langue française .....	356
3.2.2. Les logiciels en langue anglaise .....	357

<b>4. Les logiciels de traitement « On-Line »</b> .....	357
<b>5. Package d'applications Excel</b> .....	358
5.1. Comparaison de moyennes grands échantillons .....	358
5.2. Comparaison de moyennes : petits échantillons .....	360
5.3. Le test du khi-deux d'homogénéité .....	362
5.4. Le test du khi-deux d'indépendance .....	364
<b>TABLES DE PROBABILITÉS</b> .....	367
<b>BIBLIOGRAPHIE</b> .....	379
<b>INDEX</b> .....	389



# *Avant propos*

« *La statistique sert à démontrer qu'il y a plus de morts en temps de guerre qu'en temps de paix<sup>1</sup>.* »

La statistique : voilà un mot qui, dès qu'il est évoqué, doit bien faire prendre, de manière significative, quelques points supplémentaires de tension artérielle à de nombreux étudiants en psychologie. Il est vrai que l'on a parfois l'impression que cette discipline, obligatoire, enseignée dans le cadre de leur cursus universitaire, réactualise chez eux d'effroyables souvenirs des difficultés que certains ont eu avec les mathématiques. Sans vouloir nier la filiation directe entre la statistique et les mathématiques, nous avons opté pour un mode de démonstration différent. Plutôt que d'emprunter à des modèles de démonstration mathématique, nous avons essayé, dans la mesure du possible, d'illustrer certains concepts et méthodes de manière empirique. La simulation de lancers de pièces par ordinateur constitue une de ces illustrations de même que la sélection d'échantillons de sujets dans une population. Nous avons même poussé le vice à ne pas employer la syntaxe habituelle de la statistique pour désigner certains indicateurs ou paramètres. Aussi, que l'on ne s'étonne pas de voir écrit « Nombre de degrés de liberté = Nombre de modalités de la variable - 1 » plutôt que «  $v = k - 1$  ». Il apparaissait en effet important d'appeler les choses par leur nom et d'employer le plus simple des langages : un langage, en tout cas, que les étudiants de psychologie, comme les psychologues, connaissent parfaitement.

L'objectif principal de ce livre est de faire connaître les premiers principes et les premières méthodes d'une discipline dont l'objet essentiel est d'extraire et de

---

1. Citation de Alphonse Allais, célèbre humoriste français du XIX<sup>e</sup> siècle.

séparer des faits numériques, issus d'observations, les informations essentielles et la part de hasard qui pourrait les expliquer.

Deux aspects importants de la démarche statistique vont être étudiés ici : la statistique descriptive et la statistique inférentielle. La première présente les méthodes destinées à résumer le mieux possible une masse d'informations numériques qui, laissées en l'état, ne pourraient être interprétées, en raison, principalement, de leur nombre et de leur hétérogénéité. La seconde présente les outils qui permettent d'extrapoler ces résultats. On verra en effet que l'on part, souvent, d'un nombre restreint d'informations obtenues auprès d'un nombre restreint d'individus pour extrapoler ces résultats. Par extrapoler, on entend considérer que l'information obtenue auprès d'un nombre restreint de personnes est l'expression de tous et, de fait, permet de tirer des conclusions générales.

L'utilisation d'un outil statistique est étroitement liée à ce que l'on veut montrer. Pour cette raison, on a essayé, autant que faire se peut, d'appliquer chacune des méthodes statistiques étudiées ici à des données issues de recherches propres à la psychologie. Bien sûr, certains auteurs vous diront que les principes et les méthodes de la statistique moderne s'appliquent au traitement de nombreux phénomènes numériques et qu'il importe peu que les données proviennent de mesures correspondant au public auquel l'ouvrage s'adresse. Pourtant il est fréquent, pour ne pas dire systématique, que les statisticiens en sciences économiques proposent, dans leurs ouvrages, des analyses de données... économiques, que les statisticiens des sciences agronomiques présentent des données issues de... l'agronomie, que les statisticiens des sciences médicales, quant à eux, s'appuient sur des exemples de médecine ou de biologie. Il en va de même pour de nombreuses disciplines qui utilisent la statistique comme moyen d'étude et d'approfondissement théorique et méthodologique (sociologie, sciences de l'éducation...). On peut donc penser qu'un ouvrage de statistique destiné aux étudiants de psychologie et aux psychologues contient des exemples et des exercices d'application issus de recherches et d'études effectuées en psychologie. Aussi, dans ce manuel, les « cas » analysés statistiquement proviennent, la plupart du temps, de recherches fondamentales et appliquées issues des revues et ouvrages de psychologie (psychologie clinique, cognitive, développementale, sociale, éthologie...) mais aussi, pour certains, de mémoires, de thèses ainsi que de rapports d'activité rédigés par des praticiens (psychologues du marketing, de la publicité, de la Sécurité sociale, de l'évaluation, de l'enfance inadaptée, de la petite enfance, du nourrisson, des hôpitaux spécialisés, de la psychopathologie libérale et institutionnelle...). Si les distributions de données auxquelles les méthodes statistiques présentées ici sont appliquées, ont été reconstituées et inventées<sup>1</sup>, on a fait en sorte – par

---

1. En effet, par souci de clarté et de concision, les articles ou les rapports scientifiques présentent uniquement les résumés des distributions.

souci de transparence scientifique et parce que cet ouvrage s'adresse à des psychologues – que les résumés de ces distributions et les résultats des différents tests pratiqués soient conformes à ceux des recherches précitées.

On a essayé, dans la mesure du possible, de détailler ces diverses recherches afin que le lecteur puisse évaluer leurs enjeux théoriques et pratiques, les intentions expérimentales de ceux et celles qui les ont menées, et d'identifier, clairement, les raisons qui ont conduit à utiliser tel test, tel mode de résumé, à produire telle inférence... Nous pensons réellement que bien connaître des recherches ayant abouti au prélèvement de données numériques et notamment le détail des procédures statistiques utilisées permet d'acquérir un réel savoir-faire méthodologique et de transférer ces connaissances lorsque, un jour ou l'autre, on est mis en situation d'explorer, par soi-même, des données que l'on a obtenues.

Enfin, nous n'avons pas oublié que nous sommes résolument à l'ère du traitement informatique des données. Aussi avons-nous présenté les solutions aux différents exercices d'application donnés à la fin de chacun des chapitres, sous la forme des éditions de résultats tels que les fournissent, aujourd'hui, les logiciels de traitement statistique disponibles sur micro-ordinateur. Le développement de la micro-informatique ne fait plus du traitement statistique des données un espace réservé à certaines grandes structures organisationnelles ou à des laboratoires de recherche disposant de confortables moyens. Il est aujourd'hui possible d'acquérir, parfois pour quelques dizaines d'euros, un logiciel de traitement statistique qui permet de traiter les données de la plupart des études de cas présentées dans cet ouvrage. Vous verrez par vous-même que la fastidieuse tâche de calcul est laissée à la machine mais que les conclusions et l'interprétation relèvent exclusivement de la responsabilité de celui qui fait la demande de traitement. La difficulté de la statistique réside, aujourd'hui, plus dans la compréhension des concepts qu'elle manipule que dans la mise en œuvre des méthodes dont elle dispose.

L'ambition de ce livre est donc de fournir au psychologue et à l'étudiant en psychologie des méthodes susceptibles d'être utilisées pour élaborer des théories expliquant les phénomènes psychologiques qu'ils ont observés. Bien entendu, ce livre n'a qu'une fonction d'initiation. Il ne couvre donc qu'un champ très limité des méthodes statistiques. Toutefois, si les quelques méthodes présentées ici permettent d'apporter un petit plus à la connaissance du fonctionnement psychologique, ce sera déjà une réussite.





# *Cours*




## **Sommaire**

- ▶ **Chapitre 1. De l'observation des faits  
psychologiques à leur résumé statistique.... Page 1**
- ▶ **Chapitre 2. De l'échantillon à la population :  
l'inférence statistique ..... Page 57**
- ▶ **Chapitre 3. La comparaison des moyennes :  
le hasard des différences (acte I) ..... Page 115**
- ▶ **Chapitre 4. La comparaison des fréquences :  
le hasard des différences (acte II)..... Page 173**
- ▶ **Chapitre 5. La corrélation : la recherche  
des parallèles..... Page 233**






CHAPITRE  
1



*De l'observation  
des faits psychologiques  
à leur résumé statistique*



## ***Sommaire***

- ▶ ***1. Typologie des variables et représentation graphique des données ..... Page 4***
- ▶ ***2. Résumé et description numérique des données..... Page 25***
- ▶ ***3. Applications avec traitement informatique ..... Page 49***

« Il existe trois sortes de mensonges :  
les mensonges, les affreux mensonges et les statistiques<sup>1</sup>. »

En psychologie, comme dans de nombreuses autres disciplines scientifiques, il n'y a pas d'instrument à partir duquel on peut tout mesurer et une condition et une seule où tout se mesure. La complexité des faits psychologiques ne saurait s'appréhender par le biais d'un ou de quelques instruments de mesure et d'une ou de quelques conditions de prélèvement. Dans ce domaine, force est de constater que l'imagination des psychologues semble sans limite, et les subtilités et stratagèmes qu'ils emploient pour faire émerger les phénomènes psychologiques montrent bien que leur ingéniosité n'a d'égale que la pluralité des phénomènes psychologiques qu'ils désirent observer<sup>2</sup>. Toutefois, la multitude des instruments de mesure de ces phénomènes – on pourrait même dire ces faits psychologiques – ne rend pas la tâche de description de ceux-ci surréaliste. En effet, si l'instrument de mesure est varié, les caractéristiques idiosyncrasiques des mesures le sont moins et, dans les faits, il n'existe guère plus de trois, voire quatre catégories différentes de mesures. Si ces catégories – que l'on appelle, en statistique, des échelles de mesures – sont peu nombreuses, il est important de pouvoir toujours situer la catégorie à laquelle appartient la mesure que l'on a utilisée pour « capter » le ou les phénomènes psychologiques recherchés. Le degré de précision statistique d'une mesure – le statisticien entend par précision le type d'opérations mathématiques que l'on pourra faire sur ces mesures – est dépendant du type d'échelle utilisé et les tests statistiques (cf. chapitres 3, 4 et 5) que l'on pourra employer ne portent, la plupart du temps, que sur des échelles bien déterminées. Toutes les mesures n'ont pas le même poids informatif et il est essentiel de pouvoir situer avec précision le niveau atteint par celles-ci. Cette étape d'évaluation effectuée, on pourra accéder à la phase de

---

1. Citation de Benjamin Disraeli, comte de Baeconsfield et homme politique britannique de talent.

2. Pour s'en convaincre, on pourra lire un ouvrage présentant différentes « technologies » psychosociales mises en œuvre sur le terrain pour répondre à des préoccupations d'entreprises, de collectivités : G. Guingouain, F. Le Poutier, *À quoi sert aujourd'hui la psychologie sociale ? Demandes actuelles et nouvelles réponses*, Rennes, Presses universitaires de Rennes, 1994.

description proprement dite des informations (les données) que l'on a recueillies. La statistique descriptive offre un éventail de méthodes graphiques et numériques dont l'emploi est aujourd'hui facilité par la micro-informatique, et qui permettent de présenter ces données sous une forme lisible, synthétique et surtout, signifiante.

## **1. Typologie des variables et représentation graphique des données**

---

Comme on vient de le dire, situer le niveau de mesure du phénomène que l'on étudie constitue une étape fondamentale, peut-être la plus importante, dans le processus d'analyse et d'inférence statistique des données dont on dispose. Pour cette raison, il convient de déterminer le type d'échelle de mesure utilisée afin d'évaluer les possibilités de traitement et de description qui seront susceptibles d'être opérées à partir des données obtenues.

### **1.1. Les échelles de mesure nominale**

#### **► 1.1.1. Définition et exemples**

Les échelles de mesure nominale – appelées plus généralement échelles nominales – caractérisent les variables dont les modalités n'ont qu'une visée descriptive des phénomènes sur lesquels elles portent. On cherche ici, tout simplement, à donner un nom et non une valeur numérique aux phénomènes que l'on observe. Les modalités de la variable sont de simples catégories non hiérarchisées. Ces catégories peuvent être des valeurs (codes) mais, dans ce cas, il n'y a pas de modalité qui puisse être considérée comme inférieure ou supérieure à une autre modalité.

La plus petite variable permettant de rendre compte d'un phénomène, ou, comme on le dit parfois, d'un observable, doit impérativement comprendre deux modalités, c'est-à-dire permettre de qualifier deux états possibles. En dessous de deux modalités il n'y a plus de variable puisque rien ne peut varier. Ces variables sont dites dichotomiques – on dit aussi parfois qu'elles sont binaires – et sont fréquemment employées en psychologie.

### **Quelques exemples de variables dichotomiques**

– Un questionnaire d'enquête d'opinion

*Seriez-vous favorable, aujourd'hui, à la restauration de la peine de mort dans notre pays ?*

Oui

Non

– Une recherche sur la perception sensorielle chez le nourrisson

*On place deux morceaux de coton de manière aléatoire à gauche et à droite d'un bébé de quelques jours ; l'un d'eux a été frotté sur le corps de la mère du nourrisson tandis que l'autre provient d'une autre maman ayant un bébé du même âge. À différents intervalles temporels définis, on compte le nombre de mouvements de tête du bébé vers l'un ou l'autre des cotons.*

*Deux états peuvent donc être observés : il s'agit bien d'une variable à deux modalités que l'on peut dénommer de la manière suivante :*

- Retournement vers l'odeur maternelle
- Retournement vers l'odeur étrangère

Toutes les recherches où seuls deux états sont possibles (oui/non, d'accord/pas d'accord, réussite/échec...) peuvent être qualifiées de variables dichotomiques. Elles se distinguent en cela des variables nominales qui permettent de qualifier plus de deux états. Par distinction, on dit de ces variables qu'elles sont polychotomiques.

### Quelques exemples de variables polychotomiques

– Les enquêtes qui présentent des questions à choix fermés (réponses définies à l'avance) comprenant plus de deux modalités de réponse. C'est le cas, par exemple, des catégories socioprofessionnelles qui présentent parfois jusqu'à 500 modalités soit 500 appartenances professionnelles distinctes.

– La nosographie des maladies mentales constitue une variable où il y a, selon les cas d'ailleurs, différentes catégories de maladies dans lesquelles on affecte, après examen, un individu présentant des troubles d'ordre psychologique (exemple : paranoïa, schizophrénie, mégalomanie, psychose maniaco-dépressive...).

– La production d'un certain nombre de comportements supposés caractériser telle activité et définis à l'avance dans le cadre d'un protocole d'observation. On peut évoquer, par exemple, une grille de comportements caractérisant l'autonomie d'un individu en termes d'hygiène quotidienne. Cette grille contient un ensemble de comportements définis à l'avance : se lave tous les matins, change ses sous-vêtements chaque jour, se brosse les dents le soir avant de se coucher...

Les exemples de variables nominales sont légion et nous ne pouvons en énumérer davantage. L'important est de retenir le principe essentiel de l'attribution du statut de ces variables : l'existence de catégories (deux au minimum) et l'absence de hiérarchie entre ces catégories. Il est donc important de considérer, malgré l'interprétation évaluative que l'on peut parfois être amené à faire, les modalités comme équivalentes. En effet, pour le statisticien, ce n'est pas mieux ni plus mal d'être **pour** la restauration de la peine de mort, c'est simplement différent que d'être **contre**.

### Phases de construction des variables nominales

Qu'il s'agisse de variables nominales à deux ou plus de deux modalités, il n'est pas nécessaire de définir *a priori* les modalités de ces variables. Celles-ci peuvent, en effet, être déduites. C'est le cas, particulièrement, des recherches sur l'analyse du discours ou l'analyse de contenu dans lesquelles les catégories sont construites après

examen du matériel récolté (contenu d'entretiens, d'articles, de notes de service...). Par la suite, les informations recueillies sont placées dans les catégories ainsi constituées.

### ► 1.1.2. Présentation et représentation graphique de distributions de mesures nominales

#### Application • Phobies en tout genre

■ Un psychologue décide d'étudier la répartition du nombre de patients phobiques qu'il a reçus en psychothérapie depuis deux ans. Les résultats qu'il obtient, pour les cas les plus courants de phobies, sont les suivants :

Type de phobie	Nombre de consultations
Agoraphobie	18
Claustrophobie	37
Éreutophobie	4
Hypocondrie	16
Nosophobie	12

Bien évidemment, la façon la plus habile de présenter ces différentes données – on les appelle généralement des fréquences absolues ou encore, des effectifs – est de les traduire en pourcentages. On les appelle alors des fréquences relatives. Le pourcentage est une représentation que l'on manipule bien mentalement et qui permet, en outre, de se fixer sur une échelle. On peut ainsi comparer des variables de même nature mais avec des effectifs différents provenant d'autres populations ou d'autres échantillons. De cette manière, notre psychologue pourrait effectuer la comparaison de la distribution des pathologies avec celle d'un collègue ou, éventuellement, les fréquences régionales ou nationales si elles existent.

La traduction des effectifs en fréquences relatives – le pourcentage est une fréquence relative ramenée à une échelle de 100 – est aisée puisqu'il suffit de faire le rapport entre l'effectif de chacune des modalités et l'effectif total, cet effectif total étant constitué par le nombre total de patients « phobiques » du psychologue. Le tableau obtenu est alors le suivant :

Type de phobie	Nombre de consultations en %
Agoraphobie	20.69 %
Claustrophobie	42.53 %
Éreutophobie	4.60 %
Hypocondrie	18.39 %
Nosophobie	13.79 %

Ces répartitions peuvent maintenant être représentées graphiquement. L'idéal est de prendre les fréquences relatives comme des surfaces représentant la part de chacune des modalités de la variable. À titre d'illustration, nous donnons trois représentations graphiques possibles des données ci-dessus :

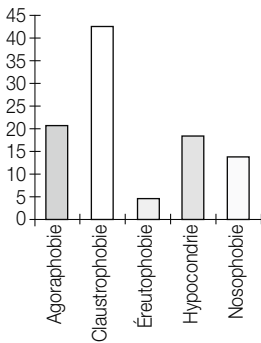


Diagramme en barres

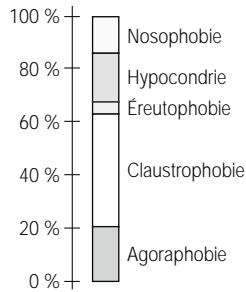
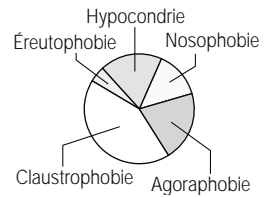


Diagramme en tronçons

Diagramme circulaire  
ou « camembert »

D'autres représentations sont possibles, notamment avec les logiciels de traitement de données modernes (diagramme en bâtons, visualisation en trois dimensions...), mais toutes ne sont pas applicables pour les données nominales. Nos exemples graphiques ne présentent pas de courbe. Dans la mesure où les modalités sont indépendantes les unes des autres, et c'est justement cela qui caractérise les modalités des variables nominales, cette courbe n'aurait pas de signification.

### La réduction de l'étendue de variables nominales polychotomiques

Il peut arriver, parfois, que l'on ait effectué une mesure à l'aide d'une échelle comprenant un certain nombre de catégories, et que l'on souhaite réduire ce nombre. Cette opération est tout à fait possible, puisqu'il suffira, dans ce cas, de construire une catégorie qui intègre les deux ou plus des deux catégories que l'on souhaite assimiler. L'effectif de cette nouvelle catégorie sera alors constitué par la somme des effectifs respectifs de chacune des catégories assimilées.

Il va sans dire qu'une telle assimilation de catégories doit se faire avec logique. On suppose que les catégories fusionnées ont un « petit quelque chose » en commun. C'est le cas, notamment, lorsqu'on assimile les différents partis qui composent respectivement la « gauche » et la « droite » en politique. Plusieurs catégories (les partis) sont intégrées dans deux catégories (la « gauche » et la « droite »). Ainsi, si nous reprenions notre application ci-dessus, nous pourrions facilement assimiler la catégorie « Hypocondrie » avec la catégorie « Nosophobie ». L'assimilation serait fondée sur une correspondance étiologique. Assimiler la catégorie « Agoraphobie » et « Hypocondrie » constituerait certainement une erreur grave. Il n'y a, toutefois, pas de règle théorique en la matière. Seule la responsabilité de celui qui opère les transformations peut être engagée.

## 1.2. Les échelles de mesure ordinale

### ► 1.2.1. Définition et exemples

Comme pour les variables nominales, il s'agit de catégoriser les phénomènes relevant de la variable étudiée mais en leur donnant, cette fois, un ordre de grandeur.

L'échelle ordinale est une échelle que tout un chacun a déjà manipulée dans son existence lorsque les crises romantiques nous conduisaient à effeuiller une fleur tout en répétant inlassablement : « Je t'aime : un peu, beaucoup, passionnément... ».

En psychologie, les variables ordinales les plus connues sont les fameuses échelles d'opinion ou d'attitude. Bien que les modalités qui les définissent peuvent être différentes et en nombre inégal, celles-ci se caractérisent par la même structure. Les catégories qui les composent sont ordonnées par des gradients qui permettent au lecteur d'apprécier l'ordre.

Selon le cas, ces catégories peuvent être définies sémantiquement...

Êtes-vous pour bla bla bla bla...?

Absolument opposé	Opposé	Indifférent	D'accord	Absolument d'accord
[ ]	[ ]	[ ]	[ ]	[ ]

... ou définies numériquement, comme le sont les fameuses échelles de type « Likert » :

Êtes-vous pour bla bla bla bla...?

Absolument opposé				Absolument d'accord
[1]	[2]	[3]	[4]	[5]

Ce dernier type d'échelle semble être le plus employé aujourd'hui si l'on en croit les publications expérimentales et le contenu des rapports d'évaluation et de recherches appliquées. Quelques exemples illustrent l'emploi de ces échelles numériques avec une amplitude et une polarisation variables.

- Évaluation du degré de certitude d'un sujet à propos d'une décision ou d'un choix antérieurs  
Pouvez-vous estimer le degré de confiance que vous accordez au jugement que vous venez de produire ?

Pas du tout confiance 

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

 Confiance absolue

La polarité des valeurs est quelquefois clairement indiquée sur ce type d'échelle.

- Étude où l'on demande à des instituteurs de juger du caractère désirable ou indésirable de certains comportements supposés avoir été émis par des élèves de cours moyen première année (enfants de 9 ans)

Regarde sa montre presque toutes les cinq minutes :

Indésirable				Désirable
-2	-1	0	+1	+2



### ► 1.2.2. Des rangs ordonnés

Bien que ces échelles de type « Likert » soient « monnaie courante » dans la recherche et la pratique en psychologie, elles ne sont pas les seules qui caractérisent les échelles de mesures ordinales dans cette discipline. Il est possible, en effet, de concevoir une variable dépendante où les modalités ne sont pas évaluées de manière ordinale les unes indépendamment des autres mais les unes par rapport aux autres. C'est le cas, notamment, du classement en ordre croissant ou décroissant selon un critère défini.

### Exemples

- On présente à des sujets 10 annonces d'offres d'emploi. On leur demande de classer celles-ci dans l'ordre croissant allant de l'annonce pour laquelle le sujet pense qu'elle présente l'emploi pour lequel il a le plus de chances d'être retenu jusqu'à celle qui présente l'emploi pour lequel il pense qu'il a le moins de chances.
- On fait passer une épreuve psychométrique à des sujets et on range ceux-ci dans l'ordre croissant de la réussite. Le niveau de réussite est apprécié par le score obtenu à l'épreuve. S'il y a des ex æquo, on attribue le rang moyen correspondant à l'intervalle des rangs.

### ► 1.2.3. Présentation et représentation graphique de distributions de mesures ordinales

#### Application • Satisfait, pas satisfait ?

■ Dans le cadre d'un dispositif d'évaluation d'un plan de formation, on souhaite mesurer, parmi d'autres indicateurs, le degré de satisfaction de stagiaires ayant suivi cette formation. Celui-ci est mesuré à l'aide de l'échelle ci-après.

Quelle appréciation porteriez-vous à l'égard de la formation que vous venez d'achever ?

Pas du tout satisfait    

1	2	3	4	5	6	7
---	---	---	---	---	---	---

    Très satisfait

Les effectifs de chaque catégorie de réponses obtenues auprès de 85 stagiaires interrogés sont les suivants :

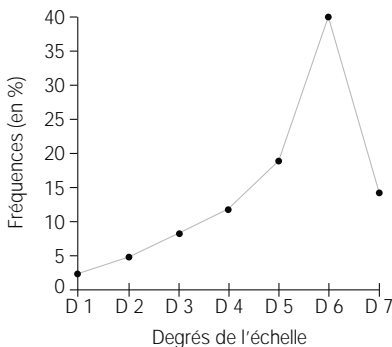
Degré de l'échelle	1	2	3	4	5	6	7
Effectif	2	4	7	10	16	34	12

Comme il s'agit ici d'effectifs, nous pourrions représenter nos données comme nous l'avons fait précédemment (calcul des pourcentages, présentation par diagramme circulaire...). Toutefois, en procédant de cette manière, nous enlevons une information importante qui distingue les variables ordinales des variables nominales : l'ordre des modalités. Aussi, plutôt que de ne représenter que les fréquences des modalités, convient-il aussi de présenter les fréquences cumulées. Celles-ci présentent l'avantage d'apprécier l'allure générale de la progression. Elles sont obtenues en sommant, successivement, les fréquences de chacune des modalités. Cette sommation doit s'effectuer dans l'ordre défini des catégories.

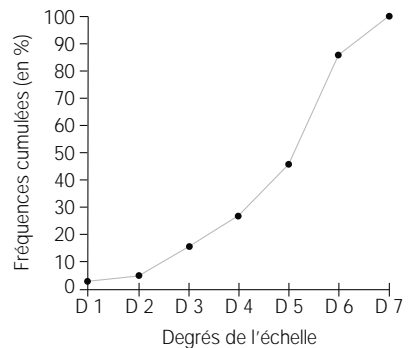
Avec les données du tableau ci-dessus, nous obtenons les résultats suivants :

Degré de l'échelle	Effectif	Effectif cumulé	Fréquence (en %)	Fréquence cumulée (%)
1	2	2	2.35	2.35
2	4	6	4.71	7.06
3	7	13	8.24	15.30
4	10	23	11.76	27.06
5	16	39	18.82	45.88
6	34	73	40.00	85.88
7	12	85	14.12	100.00

Les fréquences par catégorie, c'est-à-dire par degré d'échelle, et les fréquences cumulées peuvent maintenant être représentées graphiquement. Un des graphiques les plus utilisés est celui faisant apparaître, dans un plan à deux dimensions, la courbe de ces fréquences. La représentation obtenue pour nos données est la suivante :



*Courbe des fréquences relatives*



*Courbe des fréquences relatives cumulées*

D'autres représentations sont possibles (histogramme cumulé, aires cumulées...) mais on verra plus loin que la représentation de la courbe des fréquences cumulées facilite la recherche graphique de certains indicateurs destinés à résumer une distribution de mesures ordinales. Toutefois, à ce niveau de l'analyse, la courbe de ces fréquences cumulées peut être un bon outil de repérage des catégories privilégiées par les sujets. On observe, ici, une élévation « brutale » pour les catégories situées dans la seconde moitié de l'échelle. Si la représentation permet cette analyse, elle est aussi particulièrement utile dans le cadre de la comparaison de deux et plus de deux distributions de mesures provenant d'échelles de même nature.

### **La réduction de l'étendue de variables ordinales**

Comme on l'a vu pour les variables nominales, il est possible d'assimiler entre elles des catégories. Par exemple, on pourrait logiquement assimiler la catégorie « pas du tout satisfait » avec la catégorie « pas satisfait ». Dans ce cas, comme il s'agit de

catégories ordonnées, il conviendra de fusionner des catégories qui se retrouvent les unes à la suite des autres. On ne peut, en effet, assimiler deux catégories entre elles si elles sont séparées par une catégorie intermédiaire. Il faudra, dans ce cas, assimiler les trois. Comme précédemment, l'effectif des catégories assimilées résultera de la somme des effectifs de chacune de ces catégories. Ici aussi, on n'oubliera pas qu'en réduisant le nombre de modalités, on crée une variable nouvelle qui ne correspond pas à celle que l'on a utilisée initialement.

### 1.3. Les échelles de mesure d'intervalles

#### ► 1.3.1. Définition et exemples

Ces échelles caractérisent les variables dont les modalités peuvent être représentées par des nombres, mais ce ne sont toutefois pas des catégories numériques comme on a pu le voir précédemment. Pour cette raison, ces variables sont souvent qualifiées de variables quantitatives ou numériques. D'un point de vue théorique, leur statut repose sur le fait que deux intervalles prélevés renvoient toujours à la même quantité : la différence entre 1 et 2 est la même qu'entre 17 et 18 et la différence entre 2 et 6 est deux fois plus grande qu'entre 25 et 27. Les opérations arithmétiques que nous connaissons tous (addition, soustraction, division et multiplication) peuvent, dans ce cas, être appliquées auprès des valeurs appartenant à ce type d'échelle.

Ces nombres peuvent être des entiers et, dans ce cas, la variable est qualifiée de variable discontinue ou discrète.

#### Exemples d'échelles d'intervalles discontinues

- *Les scores d'épreuves psychométriques (discrimination auditive, de visualisation spatiale, de vocabulaire, de fluidité verbale...), les questionnaires à choix multiples (QCM), les scores de mémorisation de liste d'items... s'expriment le plus souvent dans cette échelle car ils résultent, la plupart du temps, des scores obtenus à partir de chacun des items. Le plus souvent, ceux-ci sont constitués selon le mode tout ou rien (réussite/échec à l'item, mémorisation/non mémorisation de chaque mot...). On le voit, on ne peut obtenir que des mesures sous forme d'entiers. Il ne peut pas y avoir de demi-comportement, pas de quart d'item mémorisé...*
- *Les recherches faisant appel à des listes d'items (comportements, traits de personnalité...) pour produire un jugement permettent, au-delà de l'analyse item par item, d'aboutir à des scores qui ont le statut de mesure d'intervalles (nombre total de traits sélectionnés, nombre de traits négatifs, nombre de traits positifs...).*
- *Des paramètres d'ordre physiologique comme les battements du cœur dans les études sur l'émotion. On voit, là encore, que l'on ne peut obtenir de mesures que sous forme d'entiers. On ne peut pas, en effet, produire une moitié de battement de cœur.*

Les nombres caractérisant les échelles d'intervalles peuvent être des réels et, dans ce cas, la variable est qualifiée de variable continue. Pour ce type de variable, le nombre possible de valeurs doit ne pas être défini à l'avance : avec un score à un

QCM par exemple, on ne sait pas quel peut être le score d'un sujet mais on sait quel est le nombre de scores possibles qu'il peut obtenir. Pour que la variable soit continue, il faut que, entre deux mesures effectuées et donc deux valeurs obtenues, il y ait de la place pour une valeur quelconque. Par exemple entre 1 et 2, il est possible d'obtenir 1.2, et entre 1 et 1.2, il est encore possible d'obtenir 1.15, 1.155, 1.1555... Entre deux valeurs quelconques, il y a donc toujours de la place pour une valeur intermédiaire quelconque.

### Exemples d'échelles d'intervalles continues

- *Le temps mis pour réaliser une épreuve, juger d'un stimulus, comme, par exemple, le temps de réaction verbale, le temps d'identification...*<sup>1</sup>
- *L'appréciation de la vitesse lors d'une étude en éthologie sur les comportements de fuite.*
- *La distance dans le cas d'études sur les distances d'intimité, de communication...*
- *La mesure de la quantité de certains paramètres biologiques dans le cas d'études psycho-pharmacologiques (concentration de neuromédiateurs...) ou psychophysiologiques (analyse de certaines ondes cérébrales...).*

Ici encore, la liste des paramètres mesurables, que l'on pourrait qualifier de mesure d'intervalles, ne saurait être exhaustive. L'important est que le principe de la correspondance des intervalles des valeurs soit le même à quelque endroit que l'on soit sur l'échelle de mesure. La différence, pour une même échelle, entre 12.3 et 10.8 millisecondes est bien la même qu'entre 25.6 et 24.1 millisecondes et la différence entre 15.35 et 12.63 est bien deux fois plus grande que la différence observée entre 31.17 et 29.81.

#### ► 1.3.2. Présentation et représentation graphique de distributions de mesures d'intervalles

On verra, dans la deuxième partie de ce chapitre, qu'il existe des indicateurs chiffrés très puissants qui permettent de décrire une distribution de données provenant d'une échelle de mesures d'intervalles. En ce qui concerne la représentation graphique, il est difficile de procéder, comme on l'a fait précédemment, en représentant, par exemple, l'ensemble des données obtenues. On imagine sans mal la difficulté qu'il y aurait à construire et à interpréter la courbe des résultats aux tests psychométriques des armées que l'on fait passer à tous les jeunes hommes dans notre pays. Le test porte sur plusieurs centaines d'items. On ne peut donc

1. On comprendra aisément qu'entre deux mesures, il n'y a pas toujours de place pour une nouvelle valeur. C'est le cas, par exemple, du temps où, même si l'on dispose d'une précision temporelle de l'ordre de celle donnée par une horloge atomique, il y a bien un moment où l'on ne peut plus effectuer de mesure. En fait, l'important est que l'éventail des valeurs possibles soit suffisamment grand pour que la répartition de ces valeurs présente l'aspect d'une répartition théorique continue. De la même manière, on peut trouver certaines échelles d'intervalles discontinues (fondées sur des mesures s'exprimant sous forme d'entiers) qui sont, bien souvent, suffisamment étendues pour que leur courbe de répartition ait l'aspect d'une courbe continue.

construire une courbe où tous les scores obtenus sont représentés. Aussi, pour cette raison, convient-il souvent d'opérer une certaine transformation de l'échelle où l'étendue de la dispersion est divisée en un nombre défini d'intervalles qui permettront de « trouver une place » à chacune des données de la distribution. Le nombre d'intervalles possibles observés est parfois large avec une variable continue, il est donc parfois nécessaire de le réduire.

### Application • Les identificateurs

■ On a mesuré, en millisecondes, les seuils d'identification (vitesse à laquelle on identifie un stimulus) pour des mots simples (poule, lapin, champignon...) auprès de 50 sujets. Les résultats obtenus sont les suivants :

<b>Sujet</b>	1	2	3	4	5	6	7	8	9	10
<b>Seuil d'identification</b>	19	36	27	31	15	29	21	33	27	24

<b>Sujet</b>	11	12	13	14	15	16	17	18	19	20
<b>Seuil d'identification</b>	23	18	22	32	24	21	25	27	30	27

<b>Sujet</b>	21	22	23	24	25	26	27	28	29	30
<b>Seuil d'identification</b>	23	20	17	31	28	23	25	20	15	35

<b>Sujet</b>	31	32	33	34	35	36	37	38	39	40
<b>Seuil d'identification</b>	24	21	30	16	34	24	29	32	27	21

<b>Sujet</b>	41	42	43	44	45	46	47	48	49	50
<b>Seuil d'identification</b>	19	15	18	20	23	17	25	28	24	27

On voit peu d'intérêt à présenter toutes ces valeurs sur un graphique sauf si, éventuellement, nous avons moins de sujets ou si ceux-ci avaient passé plusieurs épreuves différentes ou encore s'ils avaient passé la même épreuve dans des conditions différentes (voir ci-dessous « La combinaison de variables »). Comme ce n'est ni l'un ni l'autre de ces cas, il n'y a pas d'intérêt à les présenter de cette manière. Toutefois, il y a beaucoup de valeurs différentes dans cette distribution. L'idéal serait de pouvoir étudier cette distribution en segmentant l'étendue de ces valeurs.

Pour cela, il suffit de constituer ce que l'on appelle des classes de partition qui séparent tout simplement l'intervalle maximum observé entre les données, c'est-à-dire la différence entre la plus grande valeur et la plus petite, en un certain nombre de classes équivalentes.

Dans notre distribution, on observe ainsi que le seuil le plus élevé observé est de 36 tandis que le seuil d'identification le plus bas est de 15. L'amplitude de la distribution est donc de 21 (36-15). Nous pouvons utiliser, par exemple, six classes de partition. L'intervalle de partition est donc de  $(21/6 = 3.5)$  soit 4 puisqu'il n'y a pas de mesure du temps inférieure à la milliseconde. On peut commencer la partition en utilisant un pas de 4 à partir du seuil d'identification immédiatement inférieur à la plus petite valeur observée.

Les six partitions de la distribution deviennent donc :

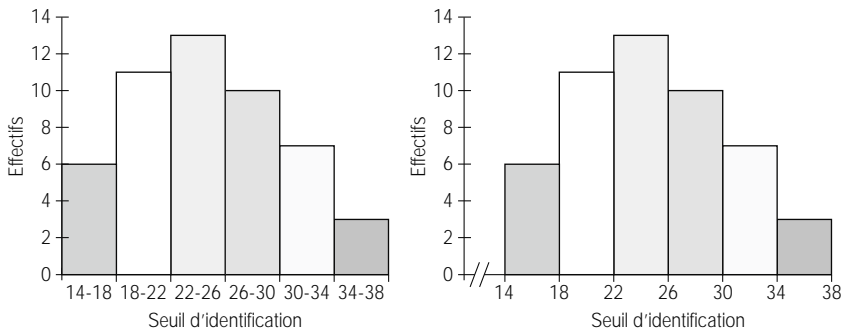
N° de la classe	1	2	3	4	5	6
Intervalle de partition	14-18	18-22	22-26	26-30	30-34	34-38

Chaque partition intègre les valeurs supérieures ou égales à la borne inférieure et les valeurs inférieures à la borne supérieure. Souvent, d'ailleurs, cela est indiqué dans les distributions réparties en classes de partition (exemple 14-18 est dit 14 et moins de 18 ou encore  $14 \leq x < 18$ ). Parfois on trouve la spécification des inclusions à l'aide d'accolades (18-22 devient [18-22]). Rien ne nous empêche aussi, comme on le voit parfois, de faire des partitions entièrement inclusives (14-17 ; 18-21... ; 34-37).

L'affectation de chacun des seuils d'identification de notre distribution à sa classe d'appartenance donne, pour les cinquante sujets, le tableau d'effectifs suivant :

N° de la classe	1	2	3	4	5	6
Intervalle de partition	14-18	18-22	22-26	26-30	30-34	34-38
Effectif correspondant	6	11	13	10	7	3

Il n'y a aucune difficulté, à présent, pour rendre compte graphiquement de cette répartition d'effectifs. Il suffira d'utiliser les représentations que l'on vous a données précédemment : la représentation la plus connue étant l'histogramme des fréquences. Voici deux représentations obtenues sur logiciel à partir de la même distribution de données :



À première vue, il y a peu de différences avec un diagramme en barres (cf. ci-dessus paragraphe 1.1.2.). Toutefois, les barres sont accolées et cela, afin de marquer le caractère continu des données de cette distribution. On remarquera que le graphique de gauche donne l'intervalle des classes de partition correspondant à chacune des barres tandis que celui de droite délimite, sur une ligne continue, les pas de progression des valeurs. En règle générale, la première méthode est utilisée lorsqu'il y a peu de classes de partition et lorsqu'elles sont équivalentes tandis que la seconde est plutôt utilisée dans le cas inverse. Là encore, il n'y a pas de représentation unique. Toutefois, on peut observer aujourd'hui que la plupart des logiciels de traitement statistique empruntent à l'une ou l'autre de ces représentations.