

MAXI FICHES

S tatistique en 80 fiches pour les scientifiques

F. Bertrand

M. Maumy-Bertrand

DUNOD

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.

Le Code de la propriété intellectuelle du 1^{er} 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

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).



© Dunod, Paris, 2011
ISBN 978-2-10-055882-7

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	IX
STATISTIQUE DESCRIPTIVE	
1 Concepts fondamentaux de la statistique	2
2 Modalités, classes et tableaux statistiques	6
3 Statistique descriptive univariée	8
4 Représentations graphiques pour les séries non groupées	10
5 Représentations graphiques pour les séries statistiques groupées	12
6 Représentations graphiques pour les séries statistiques qualitatives	14
7 Caractéristiques de position	16
8 Caractéristiques de dispersion	18
9 Boîte à moustaches	20
10 Caractéristiques de forme	22
11 Statistique descriptive bivariée	24
12 Représentations graphiques bivariées	26
13 Mesures de liaison entre deux caractères	28
PROBABILITÉS	
14 Langage des ensembles	34
15 Combinatoire et dénombrement	36
16 Fonction caractéristique et fonction génératrice	40
17 Introduction aux probabilités	44
18 Espaces probabilisés	46
19 Probabilité conditionnelle et indépendance en probabilité	48

20	Variables aléatoires réelles	52
21	Variables aléatoires discrètes	54
22	Moments et fonctions génératrices	57
23	Couples de variables aléatoires discrètes – Indépendance	60
24	Lois discrètes usuelles 1	64
25	Lois discrètes usuelles 2	67
26	Variables aléatoires continues	70
27	Loi normale ou de Laplace-Gauss	75
28	Lois dérivées de la loi normale	79
29	Lois continues	85
30	Fonctions caractéristiques	91
31	Convergences et théorèmes limites	93
32	Approximations	96
33	Vecteurs aléatoires	98
34	Espérance, variance et covariance – Conditionnement	100
35	Vecteurs aléatoires gaussiens	102
36	Simulation d'une expérience aléatoire	104
STATISTIQUE INFÉRENTIELLE		
37	Échantillonnage – Modèles statistiques	108
38	Estimation ponctuelle – Estimateurs	110
39	Méthode des moments	112
40	Méthode du maximum de vraisemblance	114
41	Comment choisir un estimateur ?	116
42	Comment améliorer un estimateur ?	118
43	Exemples d'estimateurs de caractéristiques de position	120

44	Exemples d'estimateurs de caractéristiques de dispersion	122
45	Estimation par intervalle de confiance	124
46	Intervalles de confiance pour une espérance	126
47	Intervalles de confiance pour une variance	128
48	Intervalles de confiance pour une proportion	130
49	Intervalles de confiance pour un coefficient de corrélation linéaire	132
50	Introduction à la théorie des tests d'hypothèses	134
51	Tests de conformité d'une espérance à une norme	138
52	Test des signes	140
53	Test des rangs signés de Wilcoxon	142
54	Test de conformité d'une variance à une norme	146
55	Test de conformité d'une proportion à une norme	148
56	Tests de comparaison de deux espérances	150
57	Test de Mann-Whitney	154
58	Test de la médiane de Mood	158
59	Test de Wilcoxon	160
60	Test de comparaison de deux variances	162
61	Test de comparaison de deux variances : échantillons appariés	164
62	Test d'Ansari-Bradley de comparaison de deux variances	166
63	Test de comparaison de deux proportions	168
64	Test du coefficient de corrélation linéaire	170
65	Test du coefficient de corrélation de Spearman	172
66	Test du coefficient de corrélation de Kendall	176
67	Test du Khi-deux : adéquation	180
68	Test du Khi-deux : indépendance	182

69	Tests de Mac-Nemar et d'homogénéité marginale	184
70	Test exact de Fisher-Freeman-Halton	186
71	Test de normalité	189
72	Analyse de la variance à un facteur	191
73	Pratique de l'analyse de la variance à un facteur	196
74	Test de Bartlett d'égalité des variances	198
75	Test de Kruskal-Wallis	200
76	Tests de Levene et de Fligner d'égalité des variances	204
77	Test de Friedman	206
78	Régression linéaire simple	208
79	Intervalles de confiance et régression linéaire simple	212
80	Tests sur les paramètres de la régression linéaire simple	214
	Index alphabétique	216

Avant-propos

L'organisation en crédits d'enseignement entraîne des variations dans les contenus des enseignements entre les différentes universités françaises. Les trois années de Licence ont cependant suffisamment de points communs pour proposer des livres utiles à tous les étudiants de Licence.

Avec la collection Maxi Fiches, le cours est exposé de manière claire et synthétique. Chaque fiche présente les points essentiels du cours pour vous permettre de ne retenir que les notions du moment, semestre après semestre. Un index détaillé permet d'effacer rapidement un malencontreux trou de mémoire.

Le contenu de ce Maxi Fiches est le résultat d'une sélection de ce qui est le plus fréquemment enseigné lors des trois années des licences scientifiques. La quasi-totalité des fiches est conçue pour être abordable par des étudiants des licences de biologie, psychologie, chimie et une très grande partie d'entre elles pourra également être utile aux étudiants de la première année commune des études de santé, concours de médecine, d'odontologie, de pharmacie et de sage-femme. Quelques fiches présentent plus en détail la théorie sur laquelle repose la statistique. Elles sont principalement destinées aux étudiants de mathématique, informatique, sciences physiques, mathématiques et économie et des cycles préparatoires intégrés.

Dans la collection Maxi Fiches, vous trouverez donc l'essentiel, sauf votre propre travail, bien sûr. Alors courage !

Toutes vos remarques, vos commentaires, vos critiques, et mêmes vos encouragements seront accueillis avec plaisir aux adresses électroniques suivantes :

fbertran@math.unistra.fr
mmaumy@math.unistra.fr

Enfin, les tables mentionnées dans les différents tests exposés dans cet ouvrage sont consultables en ligne : <http://www.dunod.com>

1

Concepts fondamentaux de la statistique

Mots clés

Population, individus, échantillon, observations, caractères, données brutes, statistique descriptive, statistique inférentielle.

1. POPULATION – INDIVIDUS – ÉCHANTILLON

a) Population

L'ensemble sur lequel porte l'activité statistique s'appelle la **population**. Elle est généralement notée Ω pour rappeler la notation des probabilités mais par exemple dans la théorie des sondages elle est notée U , U comme Univers.

Remarques

1. Le terme de « population » est employé aussi bien lorsqu'il s'agit d'un ensemble d'êtres humains que d'un ensemble d'objets inanimés.
2. Il est recommandé de toujours bien définir la population étudiée.
3. Le nombre d'éléments contenus dans Ω est généralement noté N , ce qui s'écrit également $\text{Card}(\Omega) = N$ ou encore $|\Omega| = N$.

Exemple

Dans les Fiches 7, 8, 9 et 10, le temps de travail moyen a été recensé sur l'ensemble des pays européens. Il faudra donc bien faire attention : dans cette application, tout ce qui est calculé l'est sur la **population** et non sur un **échantillon**.

b) Individus ou unités statistiques

Les éléments qui constituent la population sont appelés les **individus** ou encore les **unités statistiques**. Un individu est noté ω lorsque la population est notée Ω ou u lorsque la population est notée U .

Remarque

Ces « individus » peuvent être de natures très diverses.

Exemples

Ensemble de personnes, mois d'une année, pièces produites par une usine, résultats d'expériences répétées un certain nombre de fois...

c) Échantillon

Un **échantillon**, noté généralement S , S comme « *sample* », qui signifie échantillon en anglais, est une partie de la population prélevée soit de façon aléatoire, soit de façon déterministe. Les trois notions de population, d'individus et d'échantillon sont représentées sur la Figure 1.1.

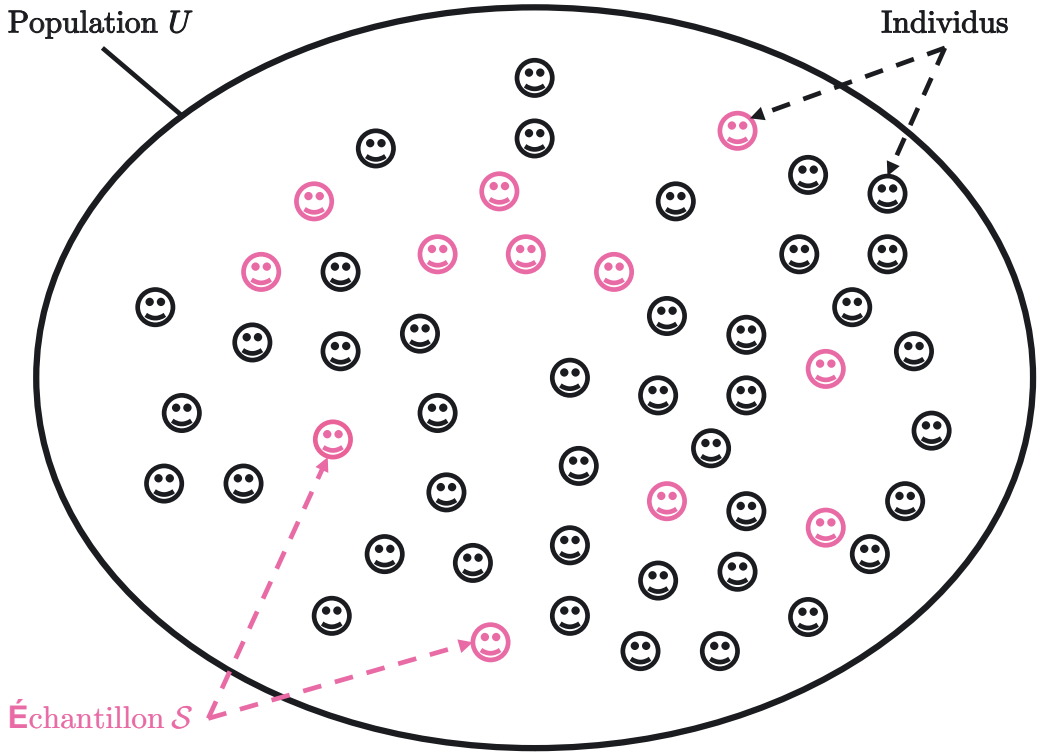


Figure 1.1 – Population, individus et échantillon.

2. OBSERVATIONS – CARACTÈRES – DONNÉES BRUTES

a) Observations

Le statisticien fait des relevés sur les unités statistiques : ce sont les **observations**.

b) Caractères

Les caractéristiques étudiées sur les individus d'une population sont appelées les caractères. Un **caractère** est une application χ d'un ensemble Ω fini de cardinal N (la population) dans un ensemble C (l'ensemble des valeurs possibles du caractère), qui associe à chaque individu ω de Ω la valeur $\chi(\omega)$ que prend ce caractère sur l'individu ω . Nous considérons deux types de caractères :