

STATISTIQUES



Auteur
Dessins
Scénario
Studio

SHIN TAKAHASHI
IROHA INOUE
RE_AKINO
TREND-PRO

Traduction

GUILLAUME BATOG
KEVIN DESTAGNOL
WAHB ETTOUMI
JEAN-YVES FÉVRIER
TRISTAN POUALLOUEC
VINCENT PUYHAUBERT
WAHB ETTOUMI
JEAN-YVES FÉVRIER

professeur agrégé
enseignant-chercheur
analyste quantitatif
agrégé
professeur agrégé
professeur agrégé
analyste quantitatif
agrégé

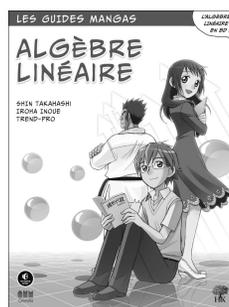
Harmonisation
Révision



Manga de wakaru : toukeigaku est le titre japonais du *Guide manga des statistiques*. L'ouvrage a été créé par Ohmsha, l'un des éditeurs scientifiques les plus anciens et les plus respectés du Japon, et par Trend-Pro, un studio spécialisé dans les mangas. Ils ont réuni pour ce projet un scientifique très pédagogue, un dessinateur très expressif et un scénariste. Une version en anglais a ensuite été co-publiée par Ohmsha et l'éditeur américain No Starch Press. L'adaptation française a été effectuée par des scientifiques accomplis, tous normaliens, et révisée par un relecteur spécialisé.

Shin Takahashi est diplômé de l'université de Kyushu. Il est spécialisé dans l'écriture d'ouvrages scientifiques et techniques.

D'autres titres sont disponibles ! Analyse, algèbre, mécanique...



Misa, étudiante douce et émouvante, trouve l'algèbre linéaire vraiment trop abstraite. Son camarade Reiji, un virtuose de la pédagogie concrète, accepte de l'aider si le grand frère de Misa, le mythique « Marteau de Hanamichi », l'intègre à son cours de karaté.

1. Qu'est-ce que l'algèbre linéaire ?
 2. Les fondamentaux
 3. Introduction aux matrices
 4. Plus de matrices
 5. Introduction aux vecteurs
 6. Plus de vecteurs
 7. Applications linéaires
 8. Valeurs propres, vecteurs propres
- Index

Vous avez trouvé une erreur ? Gasp ! Vous pouvez nous en faire part à l'adresse errare.humanum.est@H-K.fr

© H&K, 68 boulevard de Port-Royal, 75005 Paris

© 2009 (*The Manga Guide to Statistics*) Shin Takahashi, Trend-Pro, éd. No Starch Press et Ohmsha

© 2004 (*Manga de wakaru : toukeigaku*) Shin Takahashi, Trend-Pro, éd. Ohmsha

ISBN 978-2-35141-377-7

ISSN 2558-4375

Dépôt légal juin 2022

Avant-propos

Ce livre est une introduction aux statistiques. Il conviendra aux personnes qui ont envie, ou besoin, de les découvrir, de les apprendre ou de les réviser, pour un examen ou pour les mettre en œuvre dans leur travail.

Les statistiques sont probablement le domaine des mathématiques qui est le plus utile dans la vie de tous les jours. On les utiliserait par exemple pour...

- estimer le nombre de portions de nouilles sautées que l'on pourrait vendre sur un stand lors d'une kermesse ;
- évaluer la probabilité de réussir un examen ;
- comparer l'efficacité d'un médicament à celle d'un placebo.

Chacun des sept chapitres de cet ouvrage comporte quatre temps :

- pages de manga ;
- compléments d'explication ;
- exercices et solutions ;
- résumé.

Les pages de manga vous apprendront déjà beaucoup. Les autres sections vous apporteront des connaissances supplémentaires mais aussi une compréhension plus fine.

Une chose qui me ferait vraiment plaisir, c'est que les statistiques vous apparaissent à la fois utiles et amusantes une fois ce livre refermé.

Je souhaite remercier l'équipe d'Ohmsha de m'avoir donné l'occasion d'écrire ce manuel. Ma gratitude va également à Trend-Pro, qui a transformé mon manuscrit en bande dessinée ; au scénariste, re_akino ; et au dessinateur, Iroha Inoue. Enfin, je remercie D^r Sakaori Fumitake, du département de sociologie de l'université de Rikkyo, qui m'a donné d'excellents conseils pendant l'écriture de ce livre.

Shin Takahashi

TABLE DES MATIÈRES

Prologue

♥ Les statistiques le cœur battant ♥

7

1

Déterminer les types de données

19

- 1. Données qualitatives et quantitatives 20
- 2. Un exemple délicat de donnée catégorielle 26
- 3. Comment traiter un QCM 34
 - Exercice corrigé 35
 - Résumé 35

2

Saisir la vue d'ensemble : interpréter des données numériques

37

- 1. Distribution de fréquence et histogramme 38
- 2. Moyenne 46
- 3. Médiane 50
- 4. Écart-type 54
- 5. Largeur de classe d'une table des fréquences 60
- 6. Théorie de l'estimation et statistiques descriptives 63
 - Exercice corrigé 63
 - Résumé 64

3

Saisir la vue d'ensemble : interpréter des données qualitatives

65

- 1. Tableau des effectifs 66
 - Exercice corrigé 70
 - Résumé 70

4**Score standard et score de déviation**

71

1. Normalisation et score standard	72
2. Caractéristiques du score standard	79
3. Score de déviation	80
4. Interprétation du score de déviation	82
Exercice corrigé	84
Résumé	86

5**Passons aux probabilités**

87

1. Densité de probabilité	88
2. Loi normale	92
3. Loi normale standard	95
4. Loi du khi-deux	105
5. Loi du t de Student	112
6. Loi du F de Fischer	112
7. Lois de probabilité et tableur	113
Exercice corrigé	114
Résumé	115

6**Examinons les liens entre deux variables**

117

1. Coefficient de corrélation	122
2. Rapport de corrélation	127
3. Coefficient de Cramer	133
Exercice corrigé	144
Résumé	148

7**Explorons les tests statistiques**

149

1. Tests d'hypothèse	150
2. Test d'indépendance du khi-deux	157
3. Hypothèses nulle et alternative	176
4. p -valeur et marche à suivre d'un test d'hypothèse	181
5. Autre test d'hypothèse : le test d'homogénéité	190
6. Conclusion d'un test d'hypothèse	193
Exercice corrigé	194
Résumé	195



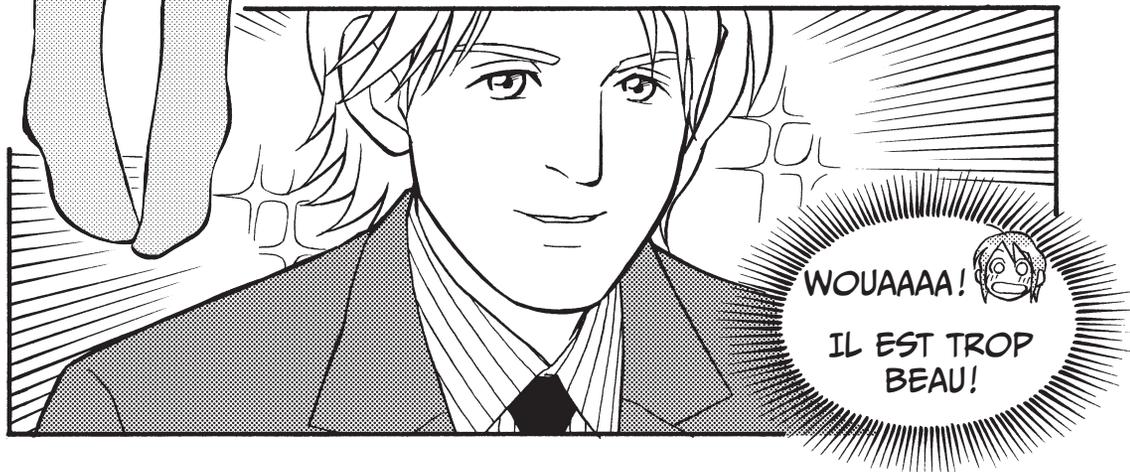
- 1. Moyenne arithmétique, médiane et écart-type.....199
- 2. Faire un tableau des fréquences200
- 3. Faire un tableau croisé201
- 4. Score standard et score de déviation.....202
- 5. Probabilité et distribution normale standard.....203
- 6. Abscisse d'une distribution du khi-deux.....204
- 7. Calcul du coefficient de corrélation204
- 8. Réaliser un test d'indépendance204

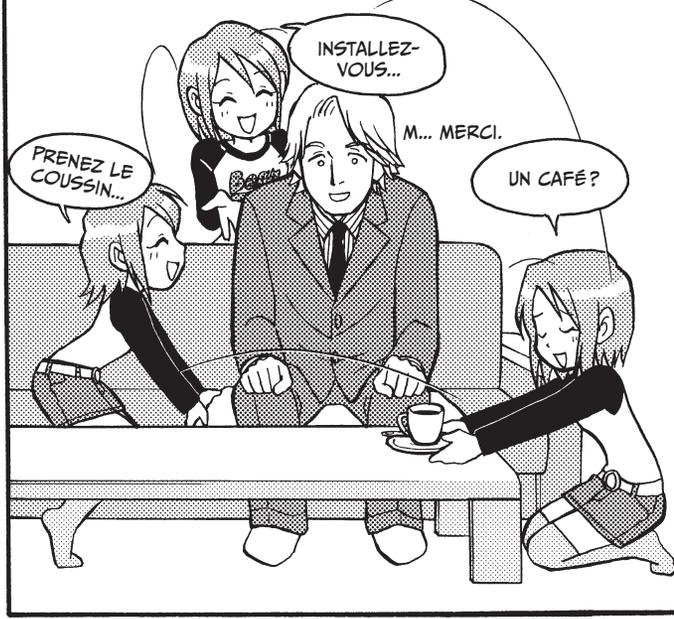
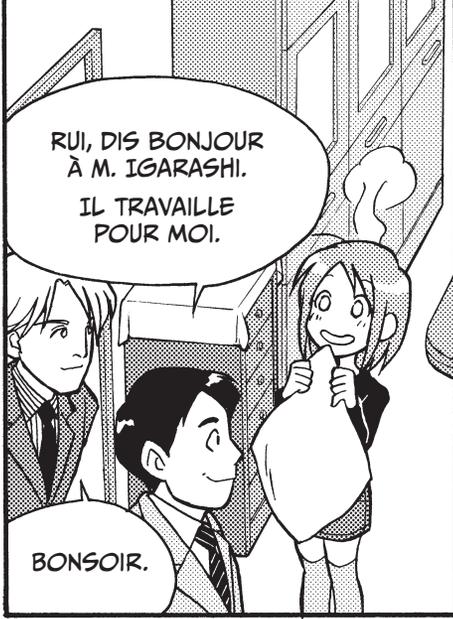
Index.....206

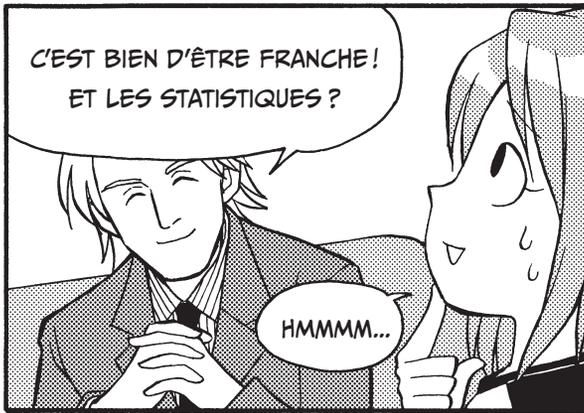
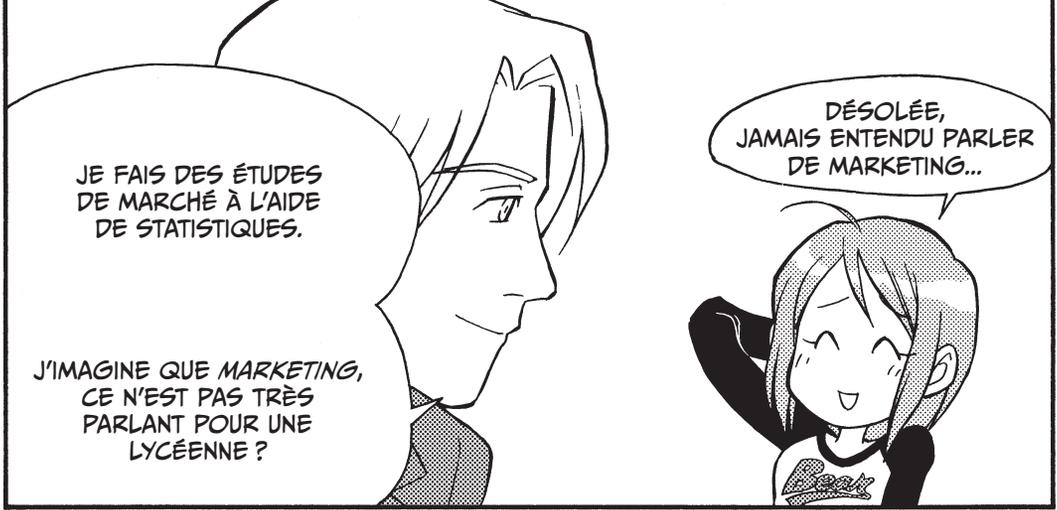


Prologue

Les statistiques
♥ le cœur battant ♥









« D'APRÈS UN SONDAGE DE CHOMAI MATIN, LA COTE DE POPULARITÉ DU GOUVERNEMENT EST DE 39 %. »

ET ALORS ?



CHOMAI MATIN NE M'A JAMAIS DEMANDÉ MON AVIS SUR LE GOUVERNEMENT.

ET VOUS, M. TAKATSU ?



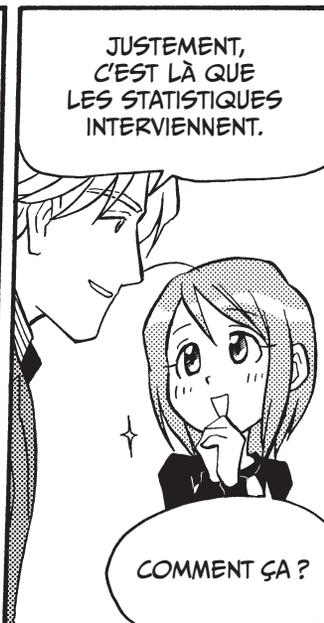
NON.

PAS UNE SEULE FOIS.



HMM... AUCUN DE VOUS DEUX N'A ÉTÉ INTERROGÉ, MAIS LA COTE DE POPULARITÉ EST BEL ET BIEN DANS LE JOURNAL.

C'EST BIZARRE. VOUS AURIEZ DÙ ÊTRE SONDÉS PUISQUE VOUS ÊTES ADULTES, NON ?



JUSTEMENT, C'EST LÀ QUE LES STATISTIQUES INTERVIENNENT.

COMMENT ÇA ?



RUI, SAIS-TU COMBIEN IL Y A D'ADULTES AU JAPON ?

HEU... BEAUCOUP !

JE SAIS !



TU AS RAISON...

EN INTERROGEANT TOUS LES ADULTES, ON CONNAÎTRAIT LA COTE DE POPULARITÉ EXACTE DU GOUVERNEMENT.

MAIS...

OUI ?



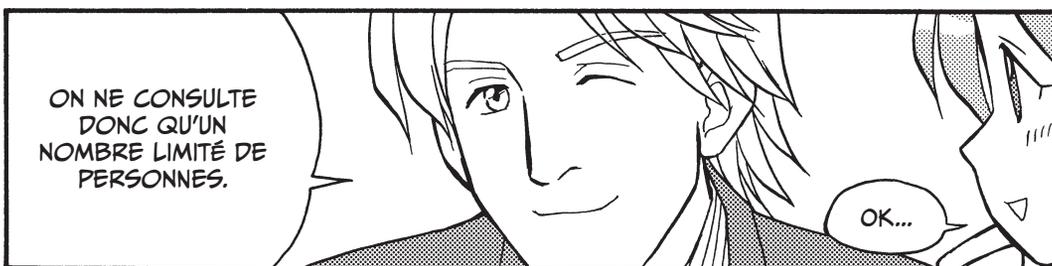
...C'EST INFAISABLE ! ON NE PEUT PAS SONDER DES MILLIONS DE PERSONNES.

BIEN SÛR.

J'ABANDONNE !

C'EST ABSURDE !

JE SUIS D'ACCORD.



ON NE CONSULTE DONC QU'UN NOMBRE LIMITÉ DE PERSONNES.

OK...



TU VOIS, RUI ?

L'ENSEMBLE DES PERSONNES QU'IL FAUDRAIT INTERROGER EST LA POPULATION.

LES PERSONNES DE LA POPULATION QUE L'ON INTERROGE RÉELLEMENT FORMENT L'ÉCHANTILLON.

CE SONT DES TERMES STATISTIQUES.

POPULATION ?

ÉCHANTILLON ?



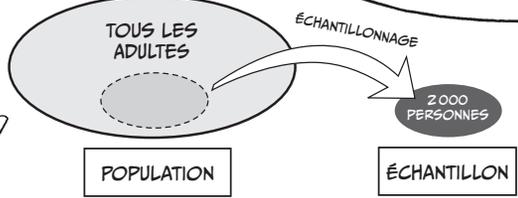
PAPA ME TORTURE AVEC DES MOTS COMPLIQUÉS !!

QUIN !

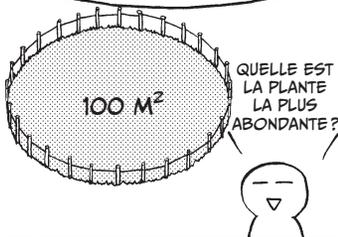
MAIS... JE NE TE TORTURE PAS !

PAR EXEMPLE, POUR LE SONDAGE DU CHOMAI MATIN, LA POPULATION EST L'ENSEMBLE DES ADULTES.

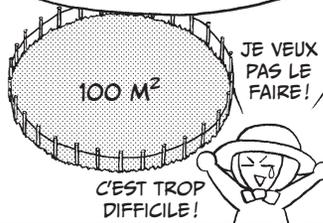
ET LES ADULTES RÉELLEMENT INTERROGÉS PAR LE JOURNALISTE, DISONS 2 000 PERSONNES, FORMENT L'ÉCHANTILLON.



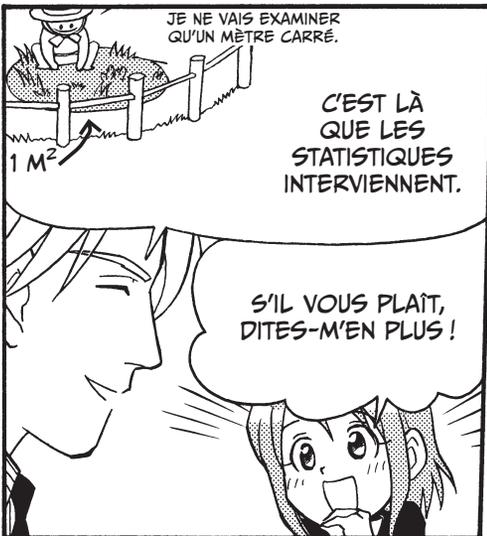
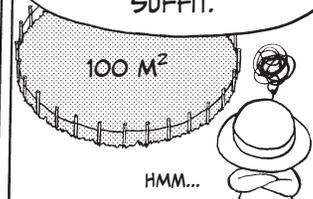
IDÉALEMENT, ON EXAMINERAIT TOUTE LA POPULATION...



...MAIS CE SERAIT TROP LONG. QUE FAIRE ?

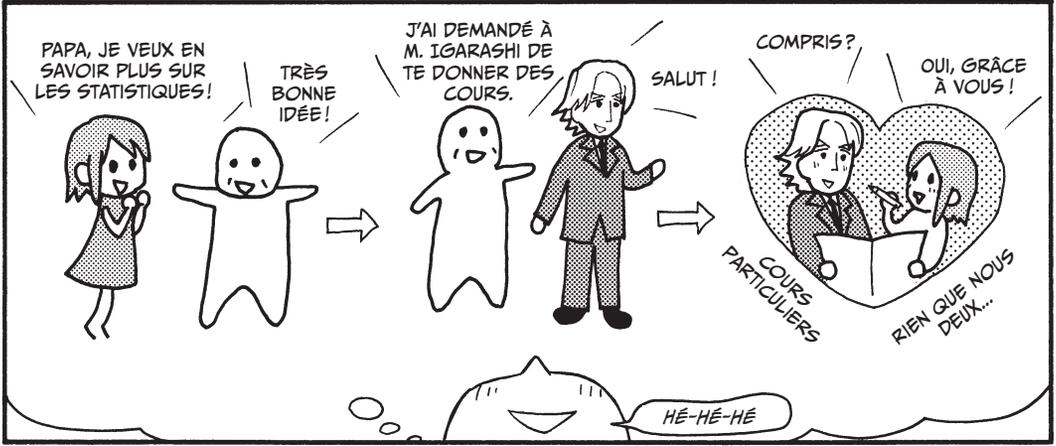
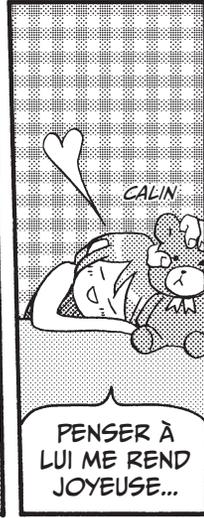


EN FAIT, ON N'A PAS BESOIN D'UNE RÉPONSE EXACTE. UNE BONNE APPROXIMATION SUFFIT.



LE
LENDEMAIN

LA-LA-LA





IL EST LÀ!

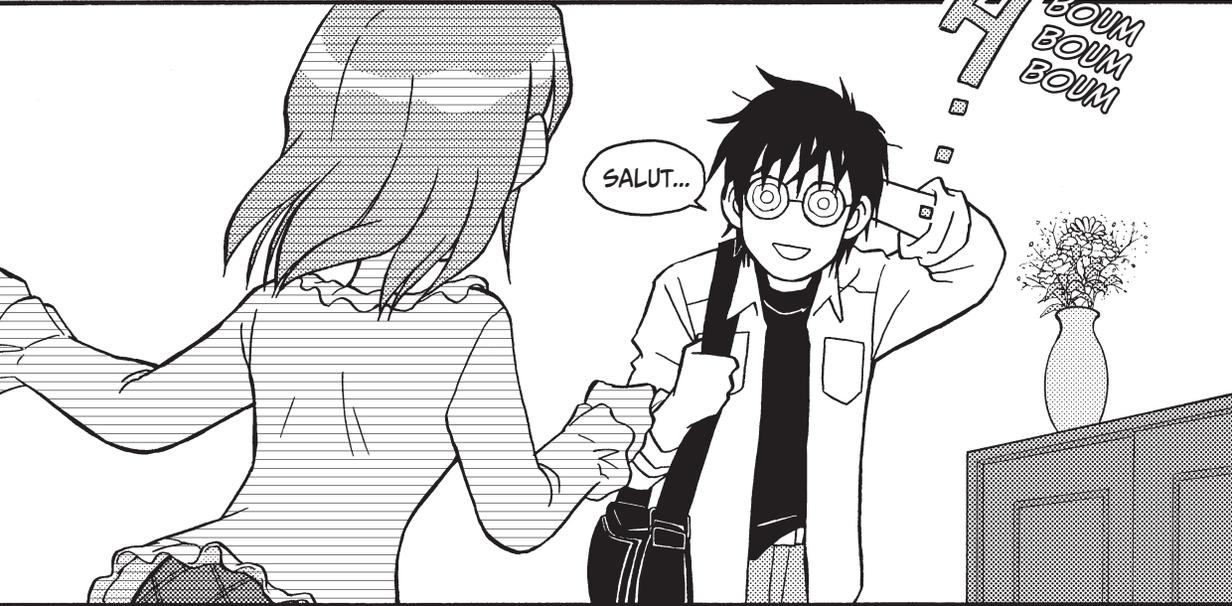


RUI!
TON
PROFESSEUR
EST LÀ!



J'ARRIVE!

A!
A!



SALUT...

BOUM
BOUM
BOUM

C'EST QUI
CE TYPE?!

RUI, VOICI
MON EMPLOYÉ,
MAMORU
YAMAMOTO.

COMMENT
VAS-TU?

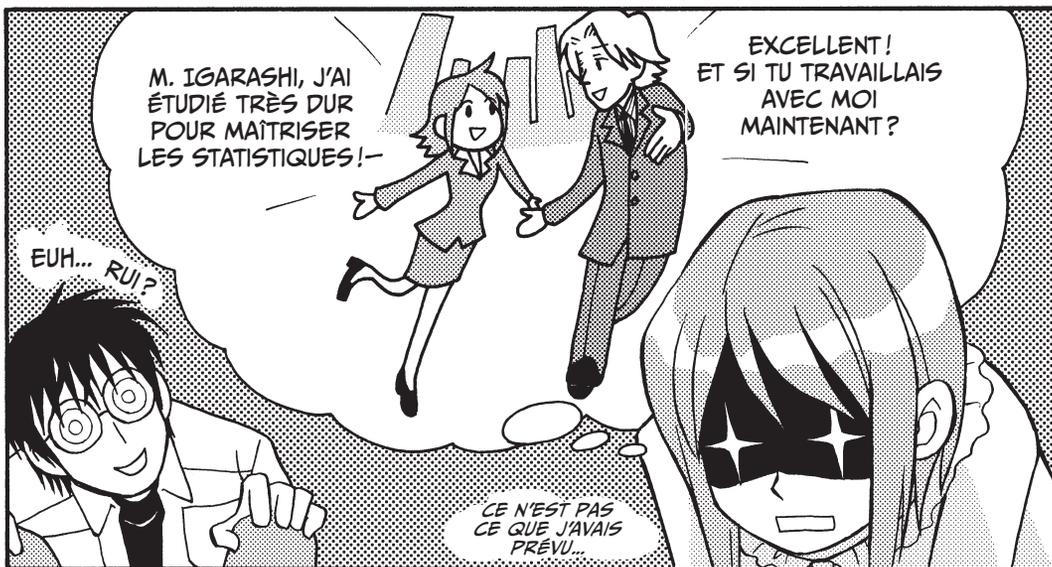
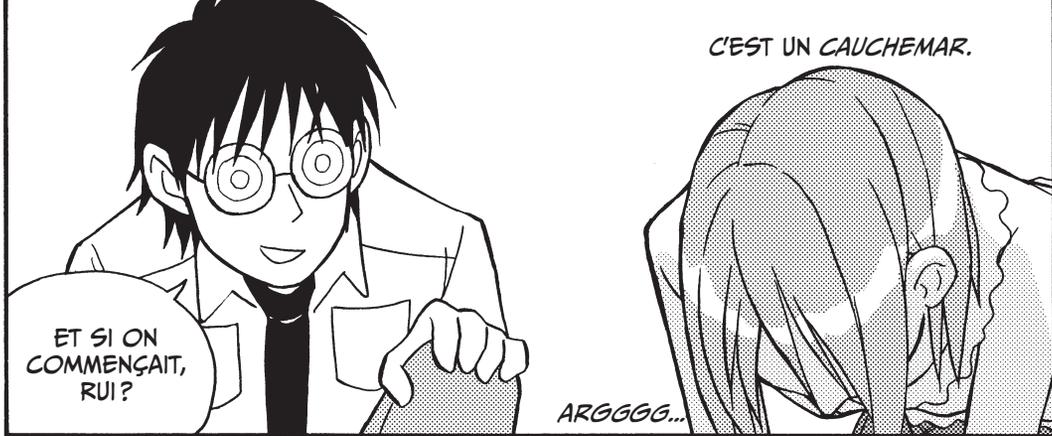
PAPA...
M. IGARASHI
NE VIENT PAS?

IGARASHI?
MAMORU
HABITE PLUS
PRÈS DE CHEZ
NOUS.

ET C'EST UN
TRÈS BON
ENSEIGNANT.

ÉTUDIE
BIEN!

HA ~ HA ~ HA!



chapitre 1

Déterminer
les types de données

chapitre 2

Saisir la vue d'ensemble :
interpréter des données
numériques

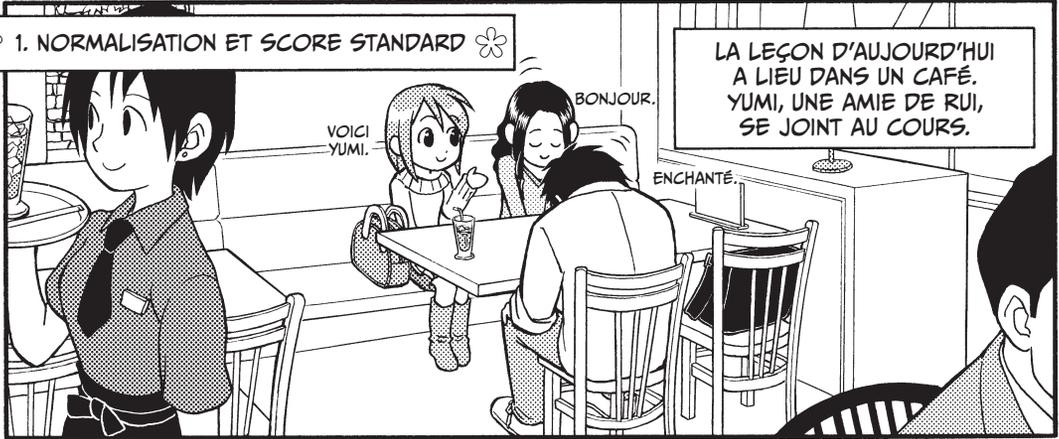
chapitre 3

Saisir la vue d'ensemble :
interpréter des données
qualitatives

chapitre 4

Score standard et
score de déviation

1. NORMALISATION ET SCORE STANDARD



VOICI YUMI.
BONJOUR.
ENCHANTÉ.

LA LEÇON D'AUJOURD'HUI A LIEU DANS UN CAFÉ. YUMI, UNE AMIE DE RUI, SE JOINT AU COURS.



EXCUSE-MOI...
JE VAIS VOUS LAISSER, LES TOURTERAUX.

NON, C'EST PAS ÇA DU TOUT !
MONSIEUR YAMAMOTO M'APPREND LES STATISTIQUES.

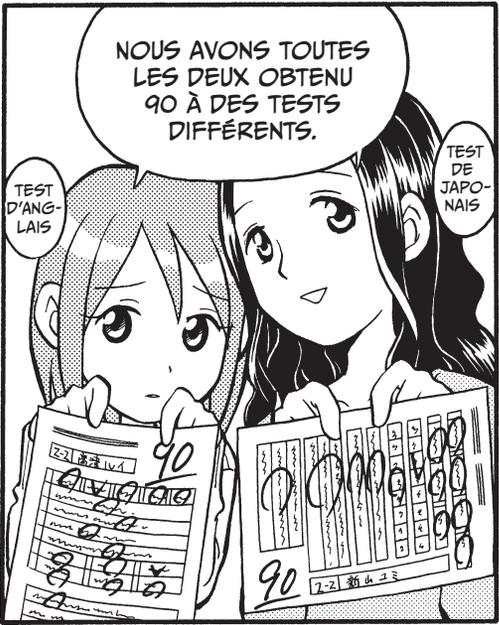


RESTE AVEC NOUS YUMI ! QUE VOULEZ-VOUS ÉTUDIER AUJOURD'HUI ?

NOS PROFS NOTENT AVEC LE SCORE STANDARD, MAIS ON N'Y COMPREND RIEN !



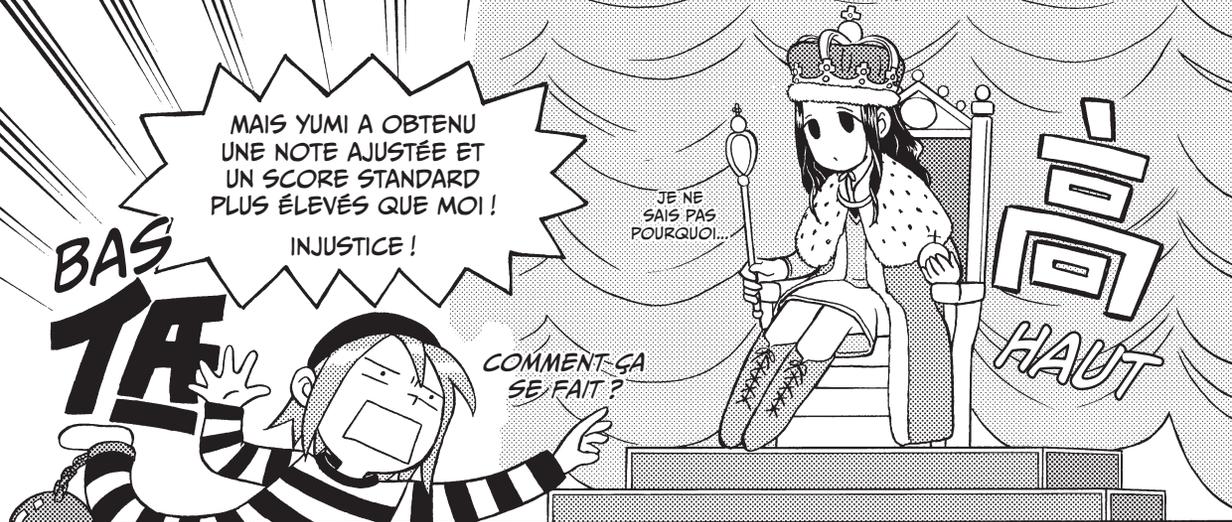
LE SCORE STANDARD ?!
BONNE IDÉE !



NOUS AVONS TOUTES LES DEUX OBTENU 90 À DES TESTS DIFFÉRENTS.

TEST D'ANG-LAIS

TEST DE JAPONAIS



NOTES (SUR 100)

ÉLÈVE	ANGLAIS	JAPONAIS	ÉLÈVE	ANGLAIS	JAPONAIS
RUI	90	71	H	67	85
YUMI	81	90	I	87	93
A	73	79	J	78	89
B	97	70	K	85	78
C	85	67	L	96	74
D	60	66	M	77	65
E	74	60	N	100	78
F	64	83	O	92	53
G	72	57	P	86	80

JE VOIS...

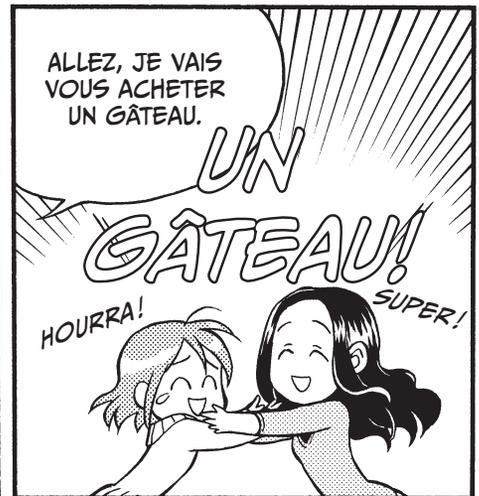
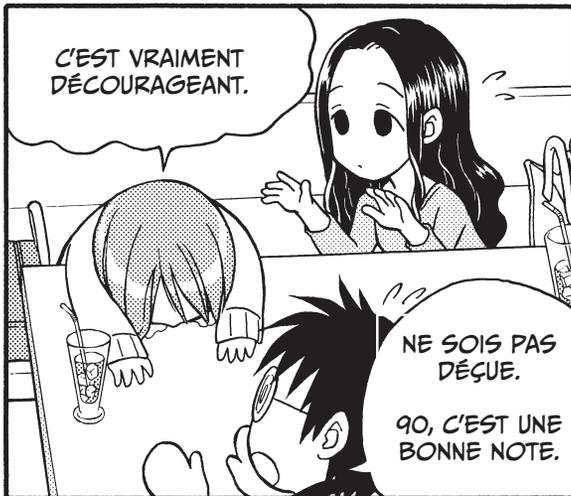
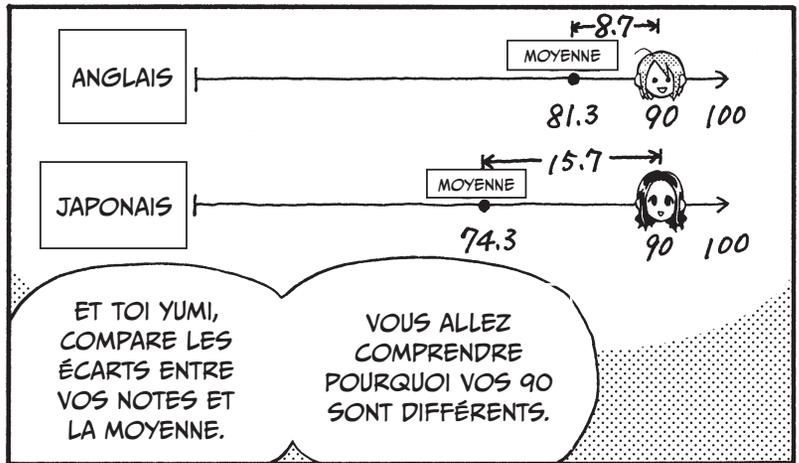


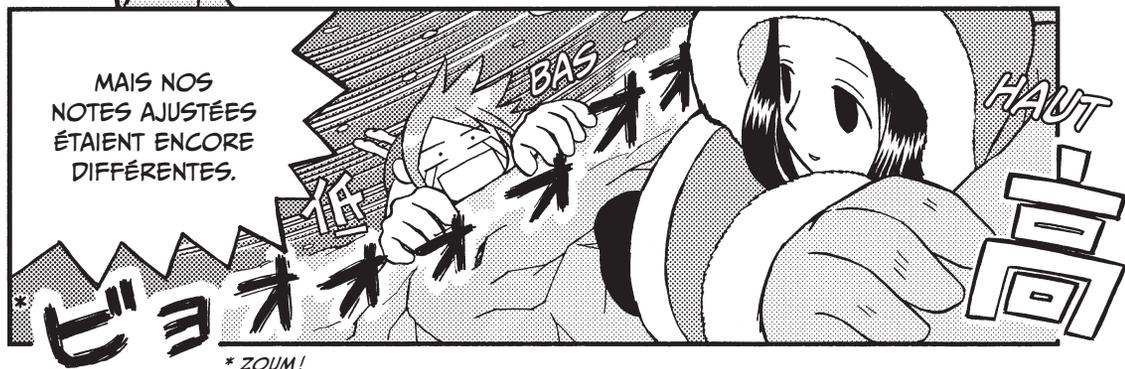
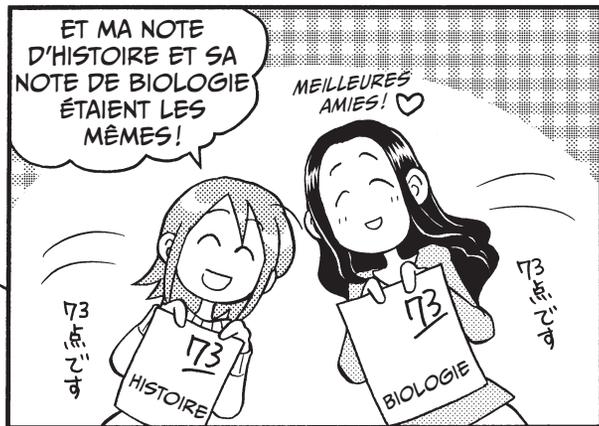
VOILÀ!

MOYENNE

ANGLAIS = 81.3

JAPONAIS = 74.3

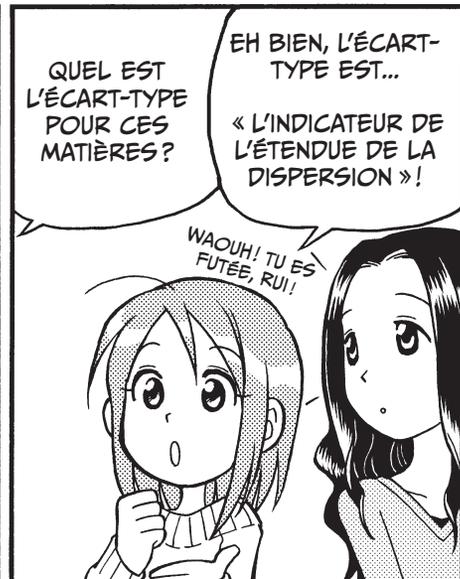




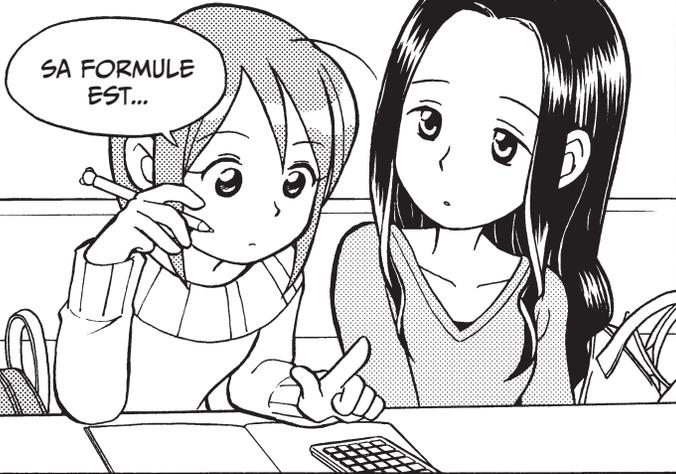
POURTANT LES ÉCARTS ENTRE NOS NOTES ET LES MOYENNES SONT IDENTIQUES!

HMM...

ÉLÈVE	HISTOIRE	BIOLOGIE	ÉLÈVE	HISTOIRE	BIOLOGIE
RUI	73	59	H	7	50
YUMI	61	73	I	53	41
A	14	47	J	100	62
B	41	38	K	57	44
C	49	63	L	45	26
D	87	56	M	56	91
E	69	15	N	34	35
F	65	53	O	37	53
G	36	80	P	70	68
			MOYENNE	53	53



$$\sqrt{\frac{\text{SOMME DE (CHAQUE VALEUR - MOYENNE)}^2}{\text{NOMBRE DE VALEURS}}}$$

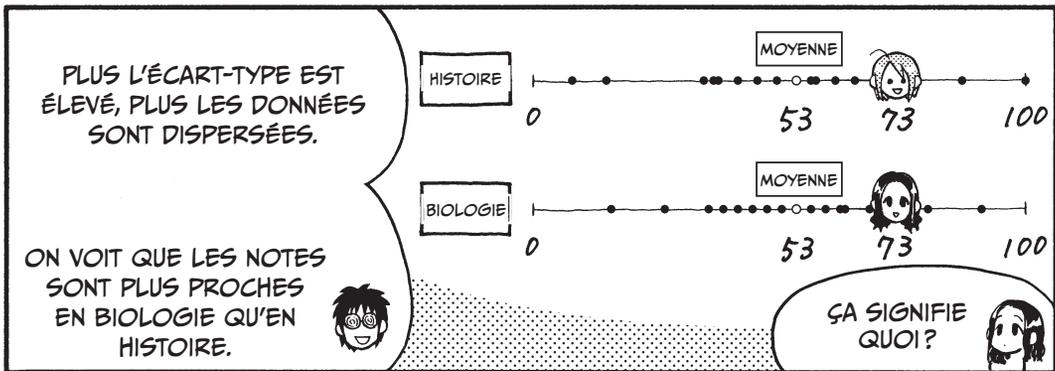


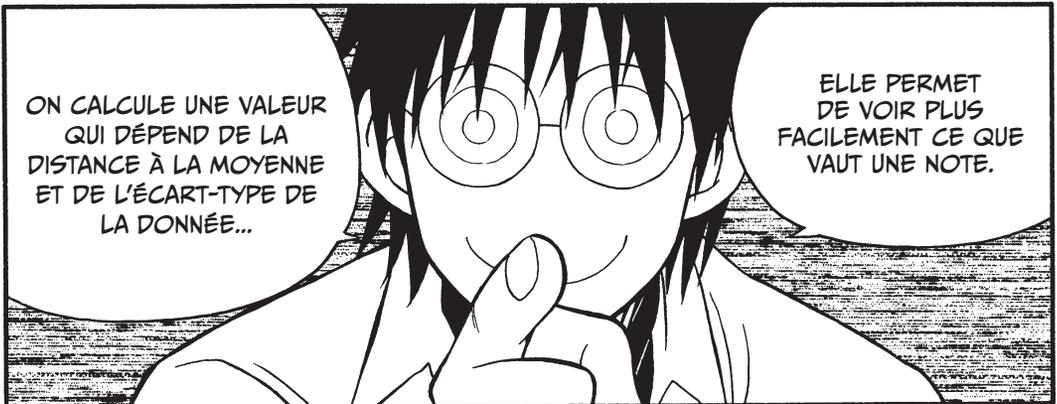
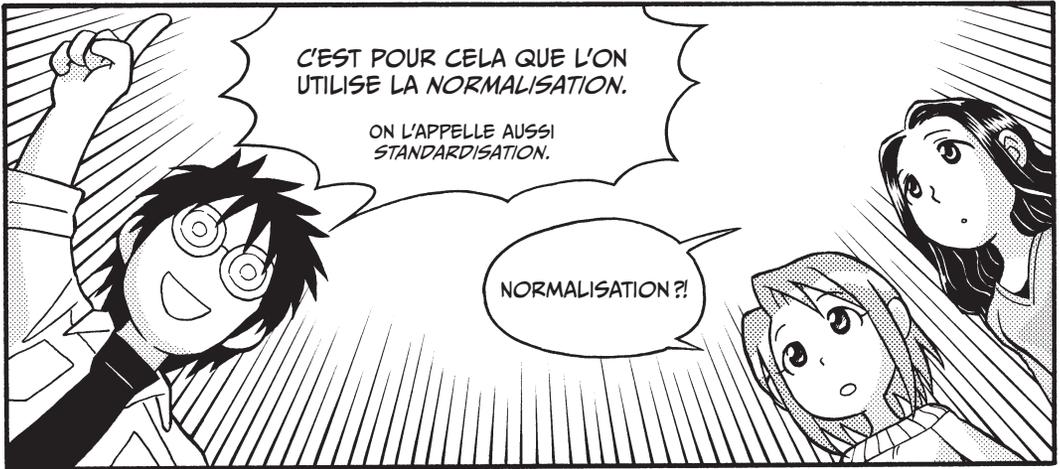
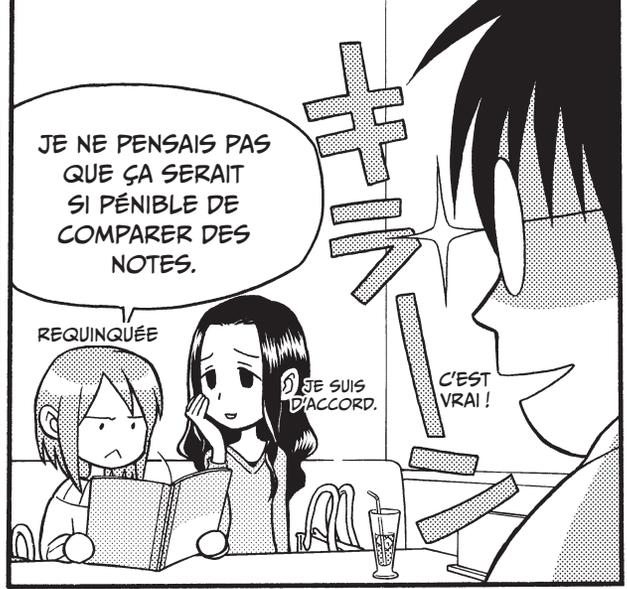
ÉCART-TYPE

HISTOIRE = $\underline{22.7}$

BIOLOGIE = $\underline{18.3}$

VOILÀ!





LA DONNÉE STANDARDISÉE S'APPELLE LE SCORE STANDARD*.

$$\frac{\text{DONNÉE} - \text{MOYENNE}}{\text{ÉCART-TYPE}} = \text{SCORE STANDARD}$$



C'EST L'ÉCART ENTRE LA DONNÉE ET LA MOYENNE MESURÉ EN ÉCARTS-TYPE : AVEC UN SCORE STANDARD DE 2, LA DONNÉE VAUT « MOYENNE + 2 ÉCARTS-TYPE ».



* OU Z-SCORE.



UN SCORE STANDARD DE -1 SIGNIFIE 1 ÉCART-TYPE DE MOINS QUE LA MOYENNE. VOYONS VOS SCORES STANDARDS.

COMPRIS!



RÉSULTATS ET SCORES STANDARDS DES TESTS D'HISTOIRE ET DE BIOLOGIE

ÉLÈVE	HISTOIRE	BIOLOGIE	SCORE STANDARD EN HISTOIRE	SCORE STANDARD EN BIOLOGIE
RUI	73	59	0.88	0.33
YUMI	61	73	0.35	1.09
A	14	47	-1.71	-0.33
B	41	38	-0.53	-0.82
C	49	63	-0.18	0.55
D	87	56	1.49	0.16
E	69	15	0.70	-2.08
F	65	53	0.53	0
G	36	80	-0.75	1.48
H	7	50	-2.02	-0.16
I	53	41	0	-0.66
J	100	62	2.07	0.49
K	57	44	0.18	-0.49
L	45	26	-0.35	-1.48
M	56	91	0.13	2.08
N	34	35	-0.84	-0.98
O	37	53	-0.70	0
P	70	68	0.75	0.82
MOYENNE	53	53	0	0
ÉCART-TYPE	22.7	18.3	1	1

VOICI LES VALEURS.



SCORE STANDARD DU TEST D'HISTOIRE DE RUI

$$\frac{73 - 53}{22.7} = \frac{20}{22.7} = 0.88$$

SCORE STANDARD DU TEST DE BIOLOGIE DE YUMI

$$\frac{73 - 53}{18.3} = \frac{20}{18.3} = 1.09$$

2. CARACTÉRISTIQUES DU SCORE STANDARD

CONCRÈTEMENT, À QUOI SERVENT CES NOMBRES ?

0,88 ?
1,09 ?

LE SCORE STANDARD A DES PROPRIÉTÉS UTILES POUR COMPARER LES DONNÉES.

Quel que soit ce que l'on mesure,

- la moyenne des scores standards vaut 0
- l'écart-type des scores standards vaut 1



TU PEUX COMPARER LES SCORES DE DEUX TESTS, MÊME SI L'UN EST NOTÉ SUR 100 ET L'AUTRE SUR 200...

Cela reste vrai quelle que soit l'unité des données :

- mètre
- kilogramme
- seconde
- ...



...OU DES DONNÉES DE NATURE DIFFÉRENTE, COMME LE NOMBRE DE STRIKES ET LE NOMBRE DE GUILLES RENVERSÉES.

RUI, TU AS UN SCORE STANDARD DE 0,88 EN HISTOIRE.

TOI YUMI, TON SCORE STANDARD EN BIOLOGIE EST 1,09.

ON SAIT DONC QUELLE NOTE EST LA MEILLEURE DANS SA DISCIPLINE.

YUMI EST LA MEILLEURE...



3. SCORE DE DÉVIATION

LE SCORE DE DÉVIATION EST UN AJUSTEMENT DU SCORE STANDARD. IL EST CENTRÉ SUR 50 ET SON ÉCART-TYPE VAUT 10.

OH!

VOICI LA FORMULE...

$$\text{SCORE DE DÉVIATION} = (\text{SCORE STANDARD} \times 10) + 50$$

ON MULTIPLIE PAR 10 PUIS ON AJOUTE 50 ?

VOILÀ DES MATHS QUE J'AIME !

...ET VOS SCORES DE DÉVIATION.

RUI
(HISTOIRE)

$$0.88 \times 10 + 50 = 8.8 + 50 = 58.8$$

YUMI
(BIOLOGIE)

$$1.09 \times 10 + 50 = 10.9 + 50 = 60.9$$

OH, CE SONT NOS NOTES ! LES PROFESSEURS ONT DONC CALCULÉ NOS SCORES DE DÉVIATION.

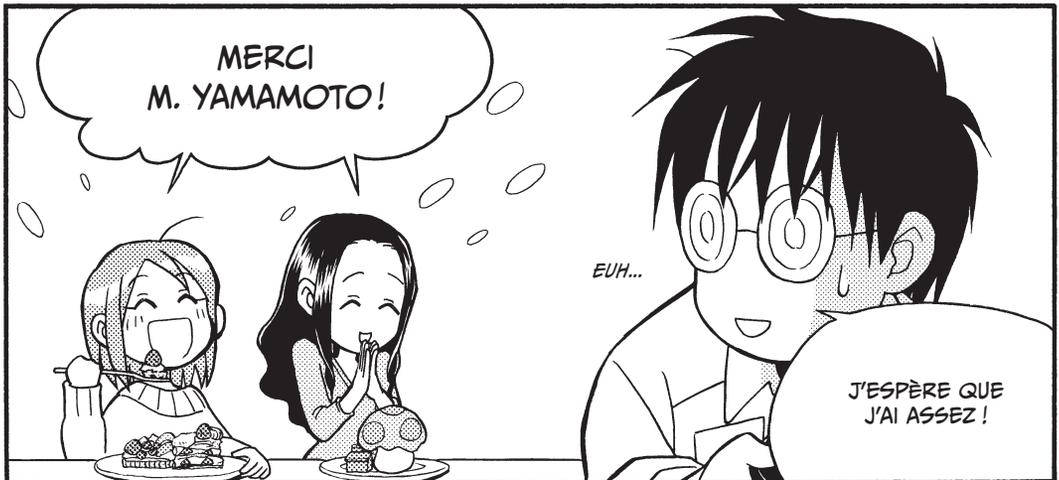
LES DEUX SCORES ONT DES PROPRIÉTÉS SIMILAIRES.

SCORE STANDARD

- (1) Quel que soit ce que l'on mesure, les scores standards ont toujours une **moyenne** de **0** et un **écart-type** de **1**.
- (2) Cela reste vrai quelle que soit l'unité de la donnée.

SCORE DE DÉVIATION

- (1) Quel que soit ce que l'on mesure, les scores standards ont toujours une **moyenne** de **50** et un **écart-type** de **10**.
- (2) Cela reste vrai quelle que soit l'unité de la donnée.





Il faut faire attention en interprétant un score de déviation.
Nous avons vu page 80 qu'il se définit ainsi :

$$\begin{aligned} \text{score de déviation} &= \text{score standard} \times 10 + 50 \\ &= \frac{\text{donnée} - \text{moyenne}}{\text{écart-type}} \times 10 + 50 \end{aligned}$$

Nous savons que la classe de Rui comprend 40 élèves (page 67).
Pour aller plus vite, j'ai triché en calculant les scores standards (page 78)
car je n'ai pris en compte que les notes de 18 élèves.

Si les calculs avaient été effectués avec les 40 notes, la moyenne et
l'écart-type auraient été différents. Les scores de déviation de Rui et Yumi
n'auraient donc pas été les mêmes. La page suivante donne les 40 notes
dans les deux disciplines. Si vous calculez les scores standards réels, vous
constatarez que Rui a finalement été meilleure que Yumi : elle a obtenu
59,1 en histoire alors que son amie n'a eu que 56,7 en biologie.

Voyons un exemple.

En début de seconde, tous les élèves passent le même test. On calcule
les scores de déviation de chaque classe. Deux élèves de deux classes
différentes obtiennent un score identique. Ont-ils le même niveau ?
Eh bien non : les deux scores étant calculés à partir de moyennes et
écarts-type différents, ils ne sont pas comparables.

Imaginons deux classes, C1 (niveau très faible) et C2 (élèves excellents).
L'élève 1 (de la classe C1) et l'élève 2 (de C2) ont eu exactement la note
moyenne de leur classe, à savoir respectivement 40/100 et 70/100.
L'élève 2 a 30 points de plus que l'élève 1. Il a donc nettement mieux
réussi. Pourtant, les deux ont un score de déviation de $0 \times 10 + 50 = 50$.

Voici un autre exemple.

En avril, un élève obtient un score de déviation de 54 à l'examen d'entrée
d'une école. Durant l'été, il révise beaucoup et passe l'examen d'une autre
école en septembre. Son score de déviation vaut 69.

Il est très fier d'être passé de 54 à 69. Il a beaucoup progressé... Non ?

En fait, il est impossible de le savoir. Les examens et les candidats sont
différents donc les moyennes et les écarts-type des deux épreuves n'ont
pas la même valeur. La comparaison de ces deux scores de déviation n'a
donc aucun sens.

TABLEAU 4-1 : RÉSULTATS DES TESTS D'HISTOIRE ET DE BIOLOGIE
(TOUS LES ÉLÈVES DE LA CLASSE DE RUI)

Élève	Histoire	Biologie	Élève	Histoire	Biologie
Rui	73	59	a	54	2
Yumi	61	73	b	93	7
A	14	47	c	91	98
B	41	38	d	37	85
C	49	63	e	44	100
D	87	56	f	16	29
E	69	15	g	12	57
F	65	53	h	44	37
G	36	80	i	4	95
H	7	50	j	17	39
I	53	41	k	66	70
J	100	62	l	53	14
K	57	44	m	14	97
L	45	26	n	73	39
M	56	91	o	6	75
N	34	35	p	22	80
O	37	53	q	69	77
P	70	68	r	95	14
			s	16	24
			t	37	91
			u	14	36
			v	88	76
Moyenne de la classe				48,0	54,9
Écart-type de la classe				27,5	26,9



EXERCICE

Voici les résultats d'un 100 mètres de quelques élèves. La moyenne et l'écart-type des temps ont été calculés par l'entraîneur.

Élève	temps au 100 m (en secondes)
A	16,3
B	22,4
C	18,5
D	18,7
E	20,1
Moyenne	19,2
Écart-type	2,01

1. Montrer que la moyenne des scores standards du 100 mètres vaut 0.
2. Montrer que l'écart-type des scores standards du 100 mètres vaut 1.

CORRIGÉ

1. Moyenne des scores standards pour le 100 mètres.

$$\begin{aligned} & \left(\frac{16,3 - 19,2}{2,01} \right) + \left(\frac{22,4 - 19,2}{2,01} \right) + \left(\frac{18,5 - 19,2}{2,01} \right) + \left(\frac{18,7 - 19,2}{2,01} \right) + \left(\frac{20,1 - 19,2}{2,01} \right) \\ &= \frac{(16,3 - 19,2) + (22,4 - 19,2) + (18,5 - 19,2) + (18,7 - 19,2) + (20,1 - 19,2)}{2,01 / 5} \\ &= \frac{16,3 + 22,4 + 18,5 + 18,7 + 20,1 - 19,2 - 19,2 - 19,2 - 19,2 - 19,2}{0,402} \\ &= \frac{96 - 19,2 \times 5}{0,402} \\ &= \frac{0}{5} \\ &= 0 \end{aligned}$$

Le numérateur est réorganisé pour regrouper les « -19,2 ». Cette astuce évite les calculs superflus !

2. Écart-type des scores standards pour le 100 mètres.

$$\begin{aligned} & \sqrt{\frac{\left(\frac{16,3 - 19,2}{2,01} - 0 \right)^2 + \left(\frac{22,4 - 19,2}{2,01} - 0 \right)^2 + \left(\frac{18,5 - 19,2}{2,01} - 0 \right)^2 + \left(\frac{18,7 - 19,2}{2,01} - 0 \right)^2 + \left(\frac{20,1 - 19,2}{2,01} - 0 \right)^2}{5}} \\ &= \sqrt{\frac{\left(\frac{16,3 - 19,2}{2,01} \right)^2 + \left(\frac{22,4 - 19,2}{2,01} \right)^2 + \left(\frac{18,5 - 19,2}{2,01} \right)^2 + \left(\frac{18,7 - 19,2}{2,01} \right)^2 + \left(\frac{20,1 - 19,2}{2,01} \right)^2}{5}} \\ &= \sqrt{\frac{\left\{ (16,3 - 19,2)^2 + (22,4 - 19,2)^2 + (18,5 - 19,2)^2 + (18,7 - 19,2)^2 + (20,1 - 19,2)^2 \right\}}{2,01^2 \times 5}} \\ &= \sqrt{\frac{1}{2,01^2} \times \frac{(16,3 - 19,2)^2 + (22,4 - 19,2)^2 + (18,5 - 19,2)^2 + (18,7 - 19,2)^2 + (20,1 - 19,2)^2}{5}} \\ &= \frac{1}{2,01} \times \sqrt{\frac{(16,3 - 19,2)^2 + (22,4 - 19,2)^2 + (18,5 - 19,2)^2 + (18,7 - 19,2)^2 + (20,1 - 19,2)^2}{5}} \\ &= \frac{1}{\text{écart-type des temps du 100 mètres}} \times \frac{\text{écart-type des temps du 100 mètres}}{1} \\ &= 1 \end{aligned}$$

RÉSUMÉ

- La *standardisation* permet de situer une valeur par rapport à d'autres en utilisant sa distance à la moyenne exprimée en fonction de l'écart-type des données.
- On emploie la standardisation pour :
 - Comparer des variables d'étendues différentes.
 - Comparer des variables exprimées dans des unités différentes.
- Le *score standard* d'une donnée est obtenu en standardisant sa valeur.
- Le *score de déviation* est une transformation du score standard.

chapitre 5

Passons aux
probabilités

chapitre 6

Examinons les liens
entre deux variables

chapitre 7

Explorons les
tests statistiques

Annexe

Calculons avec
un tableur !

INDEX

A

alternative (hypothèse),
Voir hypothèse alternative

C

catégorielle (donnée), 20-35
nuage de points, 120
résultat de sondage, 66-70
rapport de corrélation, 127
tableau (création de), 66-70
classe (étendue), 45, 60-63, 90
règle de Sturges, 61, 62, 64
coefficient de corrélation,
122-126, 212-213
coefficient de Cramer, 133-144
calcul, 136-141, 147
hypothèse alternative, 192
hypothèse nulle, 174, 192
interprétation informelle, 142
variance, 151-156, 163, 192
corrélation, 121, 125
croisé (tableau), *Voir* tableau croisé

D

degré de liberté, 105-111
densité de probabilité, 88-91, 105,
113, 115
descriptive (statistiques),
Voir statistiques(s)
déviation (score),
Voir score de déviation
distribution
F (Fisher), 112-113
khi-carré,
Voir khi-carré (distribution)
normale, 92-97
normale standard,
95-104, 210-2111
t (Student), 112
donnée
catégorielle (qualitative), 20-36
dispersion 55, 64, 75, 76, 86
numérique (quantitative), 20-35
types, 19-35, 123

E

écart-type, 54-59, 76-85
échantillon, 12, 13, 58, 63
estimation (théorie), 63-64
étendue, classe, 45, 60-63, 90
Excel, *Voir* tableur

F

F (Fisher), distribution, 112-113
fréquence
attendue, 136, 137
description, 42
réelle, 136, 137
relative, 42-43, 45
table de distribution, 38-45

G

géométrique (moyenne),
Voir moyenne

H

harmonique (moyenne),
Voir moyenne
histogramme, 44-45
avantage, 89
étendue de classe, 90, 91
exemples, 45, 89, 90, 160
homogénéité (test), *Voir* test
hypothèse alternative, 176-180
exemples, 167, 177-179
P-valeur, 181-185
V de Cramer, 192
hypothèse nulle, 176-180
exemples, 173-180
impossible à rejeter,
156, 173, 184, 185, 193
P-valeur, 181-185
pour les tests de
corrélation, 178
rapport de corrélation, 178
comparaison de proportion
entre deux populations, 179
indépendance, 178
V de Cramer, 174, 192
rejet, 164, 165, 184
hypothèse (test),
Voir test d'hypothèse

I

indépendance, coefficient,
Voir coefficient de Cramer
inter-classe (variance),
Voir variance
intra-classe (variance),
Voir variance

K

khi-carré (distribution), 105-111
calcul, 136-139
degrés de liberté, 105-114
description, 105
test d'indépendance, 157-175

L

LibreOffice, *Voir* tableur
liberté (degré de), 105-114

M

médiane, 50, 51
mesurable, *Voir* donnée
mesurable (non), *Voir* donnée
milieu (classe), 42-45, 60, 62
moyenne
arithmétique, 49, 79, 80
distribution normale, 93-96
géométrique, 49
harmonique, 49

N

normalisation, 77-78
nulle (hypothèse),
Voir hypothèse nulle

P

Pearson (statistique du test du
khi-deux),
138, 158-161, 164
population, 6
écart-type, 58, 151-156, 163,
192
échantillon (lien), 58
test d'hypothèse, 155, 192
V de Cramer, 151-156, 163, 192
pourcentage, 11, 43, 68, 70

- probabilité, 87-115
 associée, 110
 résultat des tests, 89-90
- P-valeur
 hypothèse alternative, 181-185
 hypothèse nulle, 181-185
 test d'hypothèse,
 169, 181-185, 195
 test d'indépendance, 181
- Q**
- qualitative (donnée),
Voir donnée catégorielle
- quantitative (donnée),
Voir donnée numérique
- R**
- rapport de corrélation, 123, 127-133
- région critique, 165, 171-173, 193
- relation entre données
 degré de, 121, 122-126
 linéaire, 126
 non linéaire, 126
 rapport de corrélation, 123,
 127-133
 variables, 118-121
- S**
- score
 déviation, 80-86
 standard, 71-86
- seuil de signification (α), 165, 169
- signification statistique, 193
- standardisation, 77-78, 86
- statistique(s)
 définition, 10
 descriptive, 63-64
 théorie de l'estimation, 10-13
- Sturges (règle de), 61, 62, 64
- T**
- t , distribution, 106
- table
 distribution khi-carré, 109
 distribution normale standard, 98
- tableau
 croisé, 134, 136, 141, 157, 159
 des fréquences, *Voir* fréquence
- tableur,
 liste des fonctions statistiques,
 113
 mise en œuvre, 198-205
- test, 155, 177
 d'homogénéité, 190-192
 d'hypothèse,
Voir test d'hypothèse
 d'indépendance, 214-217
 exemples, 155, 177, 190-192
 khi-carré, 157-175
 P-valeur, 181
 quelle utilisation, 143, 155
- test (résultat)
 distribution normale, 92-95
 distribution normale standard,
 95-104
 fonction de densité de
 probabilité, 89-90
- test d'hypothèse, 150, 156
 comparaison de moyennes entre
 deux populations, 155, 177, 179
 comparaison de proportions entre
 deux populations, 155, 177, 179
 conclusion, 193
 corrélation, 155, 177, 178
 exemples, 155, 174-180
 hypothèse nulle,
Voir hypothèse nulle
 mise en œuvre, 156, 181-185
 population(s), 155, 172
 P-valeur, 169, 181-185, 195
 rapport de corrélation,
 155, 177, 178
 test d'indépendance du khi-deux,
 157-175
- V**
- variables (comparaison), 117-148
 coefficient de corrélation,
 122-126
 degré de dépendance,
 121, 122-126
 rapport de corrélation, 127-132
 V de Cramer, 133-144, 147, 148
- variance (inter-, intra-classe),
 123, 129, 130, 1132
- Z**
- z-score. *Voir* score standard