

Statistique univariée

Serge Lhomme

Maître de conférences en géographie

<http://sergelhomme.fr/>
serge.lhomme@u-pec.fr

- 1 Introduction
- 2 Les résumés statistiques
- 3 La distribution statistique
- 4 Standardisation
- 5 Profil en ligne et profil en colonne

- 1 Introduction
- 2 Les résumés statistiques
- 3 La distribution statistique
- 4 Standardisation
- 5 Profil en ligne et profil en colonne

Introduction

Le tableau d'information géographique et sa sémantique

Variables Caractères

	AGRI	ARTI	CADRE	PRO INT	EMPLOYE	OUVRIER	RETRAITE	AUTRES
0	4067	16745	36426	70663	77349	78998	30417	29910
1	5201	11414	18629	48889	71578	78906	32063	42108
2	6159	9396	11842	31102	46196	42101	24218	21344
3	2027	6387	6951	16140	20885	15560	10687	10278
4	2040	5286	5883	15687	19707	12415	8865	7222
5	1918	39514	73884	117652	160574	87227	54801	68990
6	4132	10758	12461	31859	39088	37199	21660	19063
7	3362	6520	9771	25487	36326	42618	16464	23226
8	2348	4888	5537	14442	20788	16635	10379	9741
9	4786	7198	12499	28889	40121	43224	18616	18373
10	5474	12045	13486	31504	47297	34501	25584	25482
11	11100	10116	9937	25717	33970	29390	18550	13039
12	4508	52145	135584	225526	268192	168920	100790	155952
13	6081	18540	36155	72740	94384	82643	41776	34559
14	7431	4976	4547	12517	19508	17183	10386	7786
15	5756	10236	14308	32982	46169	46763	24653	20416
16	9061	21025	24906	57199	84800	65577	46140	35151
17	4025	8183	12967	29897	42604	39209	22057	18361
18	4930	7553	9273	23538	33050	28048	16824	12970
19	5352	13503	32185	62097	71854	61993	30098	23303
20	12517	17359	24766	54660	70213	68805	43023	29034
21	5259	3696	3669	9387	16039	12764	9811	6780
22	7694	15460	14125	34308	54103	47943	31651	24558
23	4123	12196	29138	57994	65018	77588	28449	27692
24	6106	14998	23027	52929	59618	57491	28522	30398
25	4064	14930	28312	62716	74944	85833	34937	33452
26	3848	10110	22675	47678	58632	57416	25285	22370
27	11054	23043	46057	94797	116573	101817	59116	46699

Entités
Individus
Objets
Unités

Valeur
Modalité

Introduction

Variables et valeurs

Les variables d'un tableau d'information géographique peuvent être quantitatives ou qualitatives.

Une variable qualitative est composée de valeurs qui ne sont pas des quantités (ou des rapports de quantités). Les valeurs prises par cette variable ne peuvent pas faire l'objet d'opérations mathématiques simple comme une addition (en tout cas ça n'a pas de sens). Ces valeurs peuvent être simplement nominales (lettres, mots, chiffres) ou ordinales c'est-à-dire qu'on peut les ordonner par ordre croissant (les classements : 1er, 2ème, 3ème, ... ; les jugements : bons, moyens, mauvais...).

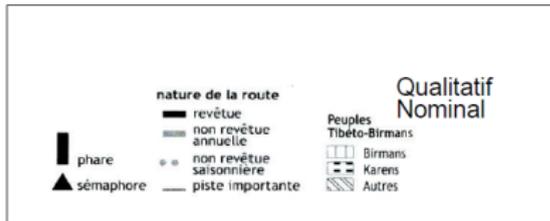
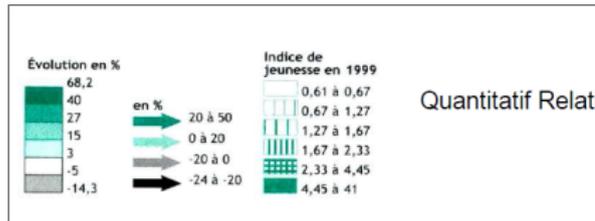
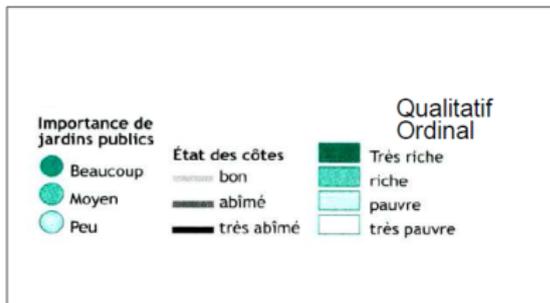
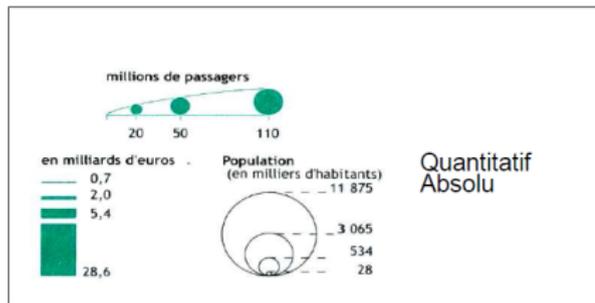
On va dans ce cours plutôt se focaliser sur des variables quantitatives. Les valeurs sont des quantités (des nombres d'individus). Toutes les opérations peuvent être opérés sur les valeurs d'une variable.

Introduction

Des règles cartographiques pour chaque type de valeurs

En cartographie, on différencie notamment les variables (quantitatives) de stock et celle de taux.

On distinguera aussi les variables de flux (tableau d'échange).



Introduction

La cartographie c'est de la statistique : la discrétisation

La discrétisation consiste à découper des données en classes homogènes.

C'est une opération de simplification.

Pour réaliser une discrétisation, il faut déterminer le nombre de classes et les bornes des classes.

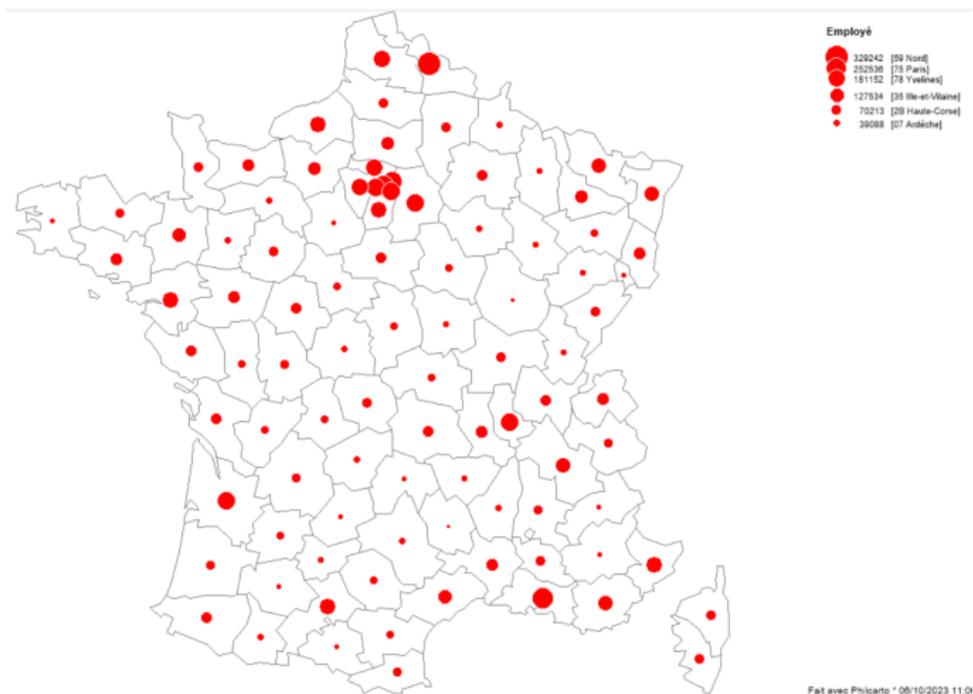
Pour réaliser une bonne discrétisation, il faut justifier à la fois le nombre de classes et les bornes de ces classes.

Intuitivement, un bon découpage correspond à des classes homogènes et séparées, ce qui correspond respectivement aux notions statistiques de faible variance intraclasse et de forte variance interclasse.

Et oui, pour réaliser une simple carte représentant une seule et pauvre variable, il faut des connaissances statistiques de base solides !

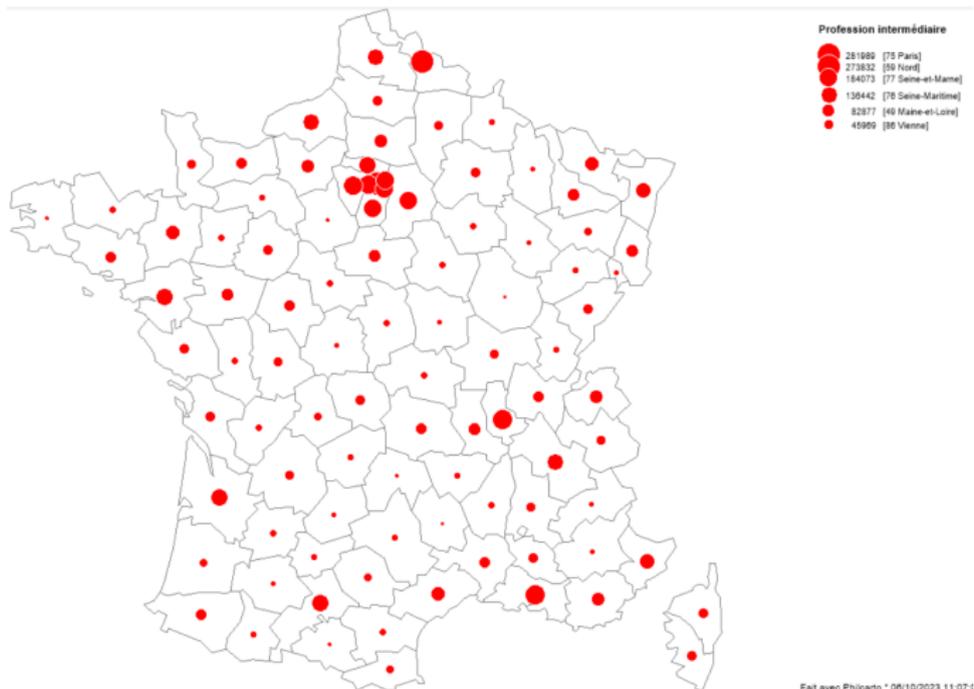
Introduction

Valeur absolue Vs relative



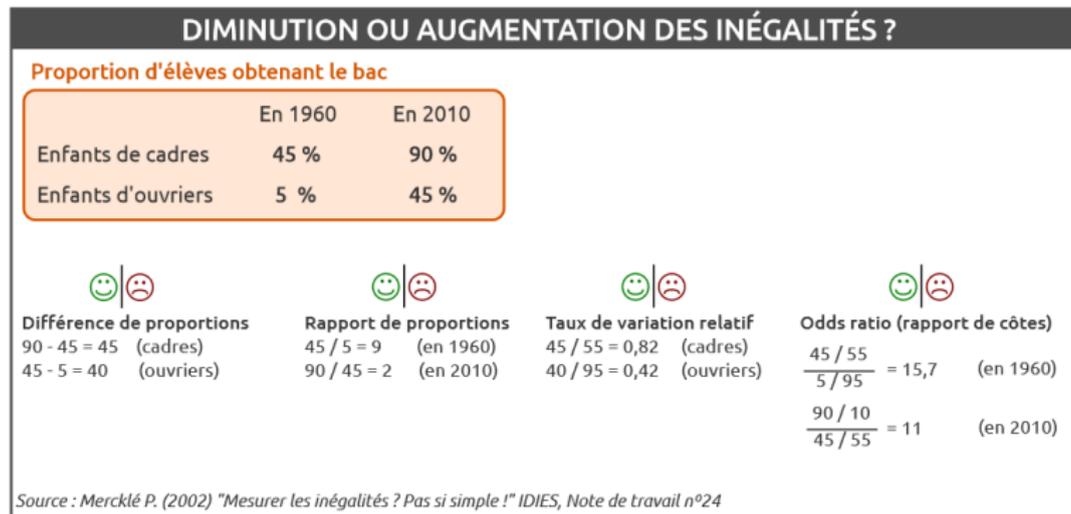
Introduction

Valeur absolue Vs relative



Introduction

Le relatif c'est complexe



Sans oublier que la moyenne des taux, le taux moyen, n'est forcément la meilleure tendance centrale...

- 1 Introduction
- 2 Les résumés statistiques**
- 3 La distribution statistique
- 4 Standardisation
- 5 Profil en ligne et profil en colonne

Les résumés statistiques

On les connaît tous...

La moyenne :

$$\bar{X} = \frac{1}{N} \times \sum_{i=1}^N X_i$$

L'écart-type :

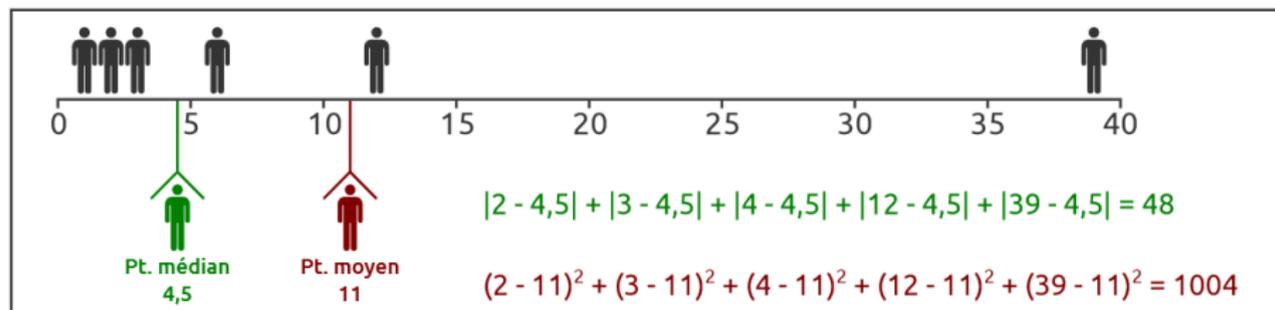
$$\sigma_X = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}$$

Les caractéristiques de position fondées sur l'ordonnement (les modes) :

Minimum ; Maximum ; Médiane ; Quantiles ; Déciles...

Les résumés statistiques

La tendance centrale : Médiane vs Moyenne



H. Commenges

La moyenne est une valeur agrégée, elle est donc sensible aux valeurs "extrêmes", aux valeurs "atypiques".

Une valeur centrale, c'est une valeur située au plus près de toutes les valeurs prises par la variable, une valeur pour laquelle la somme des écarts entre cette valeur et toutes les valeurs prises par la variable est minimale.

Les résumés statistiques

La dispersion : la variance

$$V_X = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2$$

Avoir une moyenne de 10/20 en ayant eu deux fois 10/20 ou en ayant eu 15/20 et 5/20, ce n'est pas la même chose.

La variance est simplement une mesure de dispersion. Pour cela, pour chaque individu, on mesure l'écart (la différence) entre sa valeur et la moyenne. On moyenne alors tous les carrés de ces écarts.

A noter, qu'il existe peu d'alternatives à ce calcul de variation. On pourrait calculer tous les écarts entre tous les individus, mais cela reviendrait en fait au même... Plutôt que de prendre les carrés, on pourrait choisir les valeurs absolues (c'est l'écart moyen).

L'écart-type est simplement la racine carrée de la variance. Comme son nom l'indique, c'est un écart, il peut être comparé à la moyenne.

Les résumés statistiques

L'intervalle de confiance

Un intervalle de confiance encadre une valeur réelle que l'on cherche à estimer à l'aide de mesures prises par un procédé aléatoire. En particulier, cette notion permet de définir une marge d'erreur entre les résultats d'un sondage et un relevé exhaustif de la population totale.

De manière plus simple, on pourra utiliser cette notion statistique pour identifier des valeurs "normales" (proches de la moyenne) et des valeurs "atypiques" (éloignées de la moyenne).

Si on considère que la distribution statistique étudiée est "normale" alors :

- la moyenne plus ou moins ECT définit un intervalle de confiance de X à environ 68 %
- la moyenne plus ou moins $2 \times$ ECT définit un intervalle de confiance de X à environ 95 %
- la moyenne plus ou moins $3 \times$ ECT définit un intervalle de confiance de X à environ 99,7 %

TP1

Résumé statistique

Vous disposez d'un fichier Excel décrivant la répartition de différentes catégories socioprofessionnelles au sein des départements de France métropolitaine.

- 1 Ouvrez le fichier, puis calculez la somme, la moyenne et la médiane de chaque CSP.
- 2 Déterminez la variance, puis l'écart-type pour chaque CSP.
- 3 Rapportez les moyennes par rapport aux écart-types, puis les moyennes par rapport aux médianes.
- 4 Déterminez les bornes d'un intervalle de confiance à 95% pour la variable des agriculteurs. Mettez en évidence les départements dont les valeurs sont en dehors de cet intervalle de confiance.
- 5 Calculez le rapport interquartile et interdécile de chaque CSP.

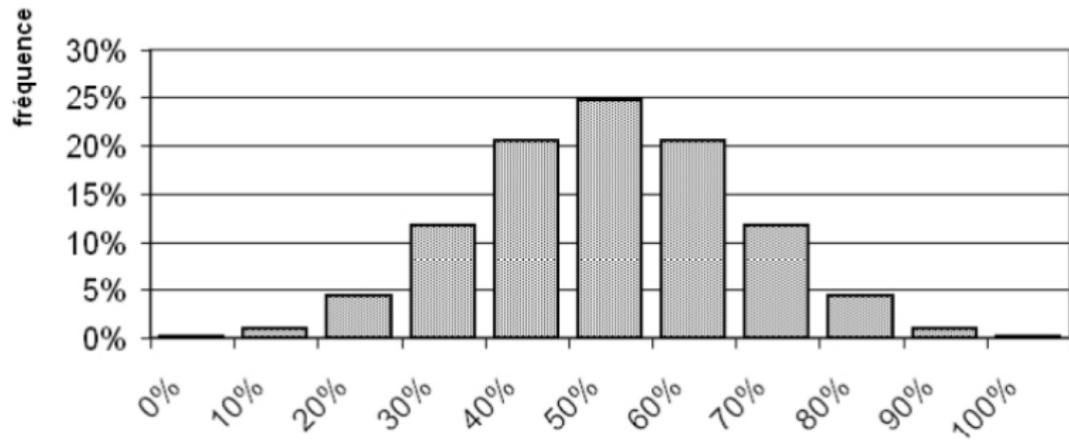
- 1 Introduction
- 2 Les résumés statistiques
- 3 La distribution statistique**
- 4 Standardisation
- 5 Profil en ligne et profil en colonne

La distribution statistique

Définition

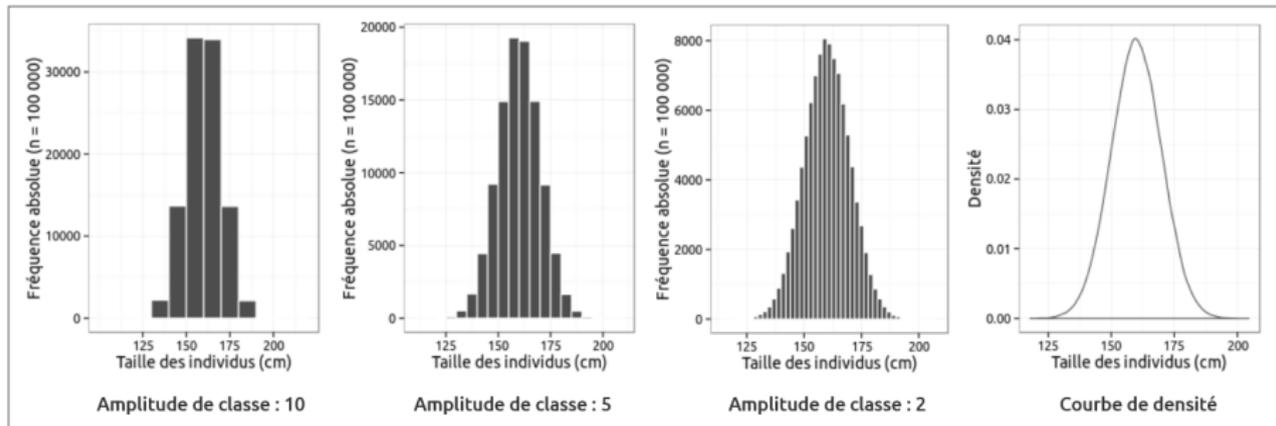
La distribution statistique (distribution des fréquences) est un tableau qui associe des classes de valeurs obtenues lors d'une expérience à leurs fréquences d'apparition (tableau de dénombrement). Il peut notamment se représenter sous la forme d'un histogramme. C'est une classification (discrétisation) par amplitude égale.

Distribution des réussites (10 questions)



La distribution statistique

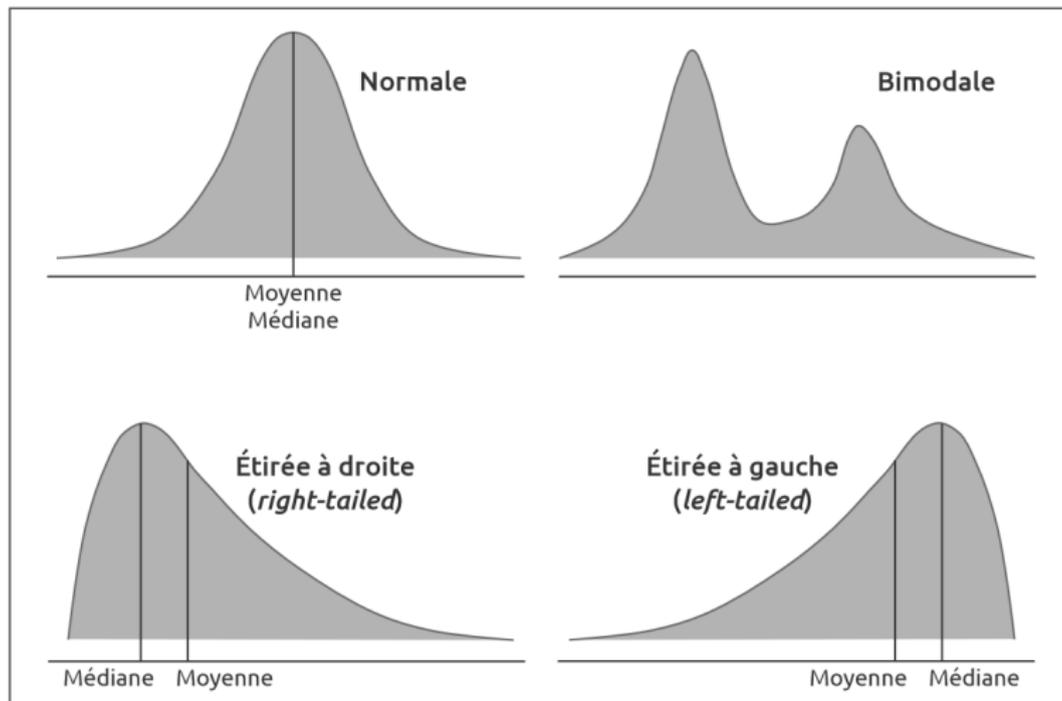
La forme dépend du nombre de classes, de l'amplitude des classes



Le nombre de classes cherche un compromis entre simplification de la forme et exactitude de cette forme.

La distribution statistique

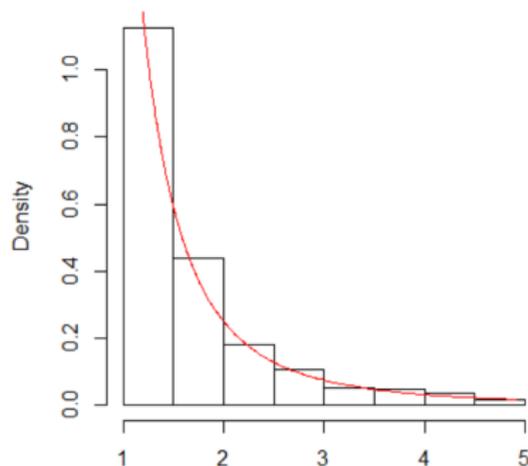
Les différentes forme de distribution



On parlera aussi de distribution dissymétrique/asymétrique à droite/gauche.

La distribution statistique

Les formes souvent oubliées



Empiriquement, et souvent négligées par le passé, de nombreuses distributions apparaissent suivre une loi de puissance : loi de parato, loi de Zipf, loi rang-taille, réseaux scale-free...

Plus rare, une distribution peut aussi être uniforme/équivalente.

TP2

Distribution et discrétisation

Vous disposez d'un fichier Excel décrivant les taux des différentes catégories socioprofessionnelles au sein des départements de France métropolitaine.

- 1 Réalisez des histogrammes de distribution statistique pour les T_x _Cadre et T_x _Profession intermédiaire. Testez pour cela plusieurs amplitudes de classe.
- 2 Quelles sont les formes de ces deux distributions statistiques ?

- 1 Introduction
- 2 Les résumés statistiques
- 3 La distribution statistique
- 4 Standardisation**
- 5 Profil en ligne et profil en colonne

Standardisation

Un outil puissant

$$Z(i) = \frac{X(i) - \bar{X}(i)}{\sigma(X)}$$

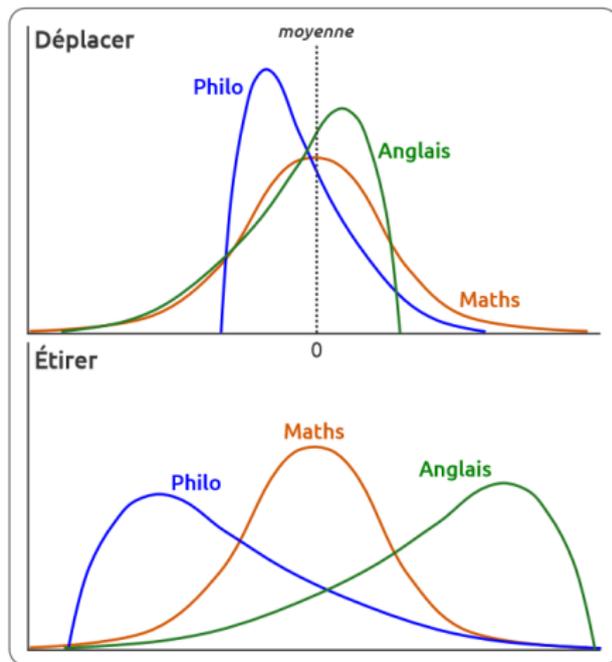
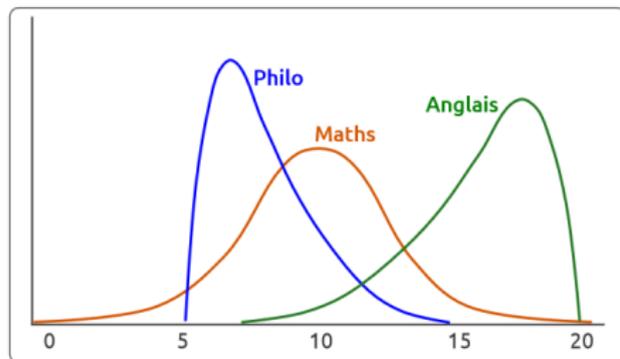
La standardisation consiste à centrer-réduire une distribution. Centrer consiste à faire en sorte que la moyenne soit égale à zéro. Réduire consiste à faire en sorte que l'écart-type soit égale à 1.

Cette transformation permet ainsi de facilement identifier les individus situés en dessous ou au dessus de la moyenne ($Z < 0$ ou $Z > 0$) et d'identifier les éléments qui s'éloignent de la moyenne (par exemple $Z < -1$ ou $Z > 1$).

Néanmoins, la standardisation est souvent utilisée pour comparer différentes distributions.

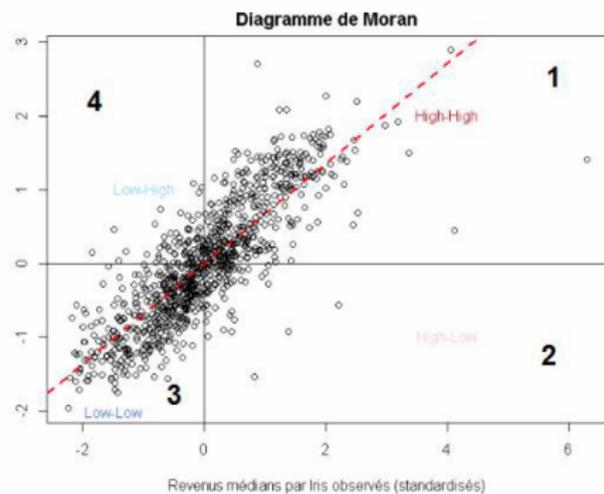
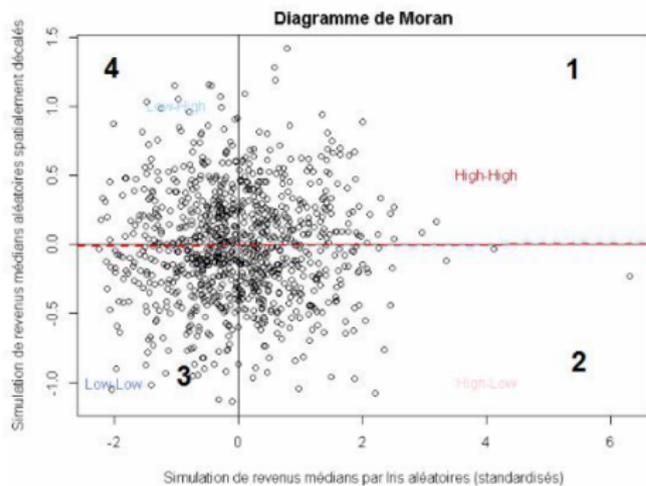
Standardisation

Un pas vers la statistique multivariée



Standardisation

Un pas vers la statistique multivariée



TP3

Standardisation

Vous disposez d'un fichier Excel décrivant la répartition de différentes catégories socioprofessionnelles au sein des départements de France métropolitaine.

- 1 Standardisez les valeurs des variables suivantes : les Professions intermédiaires et les employés.
- 2 Représentez les valeurs standardisées de ces deux variables à l'aide d'un nuage de points.
- 3 Identifiez les quatre grande catégories d'objets géographiques issus de ces standardisations.
- 4 Faites de même pour les variables suivantes : les Professions intermédiaires et les agriculteurs.

- 1 Introduction
- 2 Les résumés statistiques
- 3 La distribution statistique
- 4 Standardisation
- 5 Profil en ligne et profil en colonne**

Tableau de contingence

Poids des entités géographiques et des variables

<i>Effectifs observés (N_{ij})</i>									
1963	ALIM	TEXT	BOIS	EDIT	CHIM	CONS	META	EQUIP	Total
BULGARIE	130	128	39	14	29	47	21	151	559
HONGRIE	144	241	53	28	77	61	91	423	1118
POLOGNE	380	612	164	84	222	199	147	881	2689
R.D.A.	206	451	119	118	308	142	109	1056	2509
ROUMANIE	136	305	244	41	76	114	106	366	1388
TCHECO.	185	412	130	63	139	151	177	883	2140
YUGOSL.	126	223	132	58	76	78	69	307	1069
Total	1307	2372	881	406	927	792	720	4067	11472

Tableau de contingence

Poids des entités géographiques et des variables

Candidats	Agriculteurs	Artisans, Commerçants et chefs d'entreprise	Professions libérales, Cadres Supérieurs	Professions intermédiaires	Employés	Ouvriers	Étudiants	Chômeurs	Total
Schivardi	0	0	0	0	12	10	0	0	21*
Laguiller	0	0	0	9	23	29	0	4	65
Besancenot	0	0	4	53	58	78	24	32	249
Buffet	5	0	4	9	23	20	3	24	87
Bové	5	0	7	18	12	10	8	4	64
Royal	14	40	110	276	289	205	85	128	1148
Voynet	4	0	7	36	12	10	5	4	77
Nihous	0	0	0	9	12	20	0	4	44
Bayrou	32	64	103	178	185	157	59	88	865
Sarkozy	64	117	103	231	335	205	56	76	1189
Villiers	20	5	7	18	35	10	5	0	100
Le Pen	34	40	11	53	162	225	21	36	582
Total	178	267	356	889	1156	978	267	400	4492

* Avertissement : Le tableau donne les effectifs de vote aux dix millièmes (4492 au lieu de 44.920.000 individus). Les effectifs sont exprimés sans aucune décimale ce qui conduit à des approximations quant aux calculs des effectifs marginaux. Par exemple, le nombre de votes pour le candidat Schivardi a été estimé à 21 (soit 210.000) électeurs et non à 22 (soit 22.000) électeurs (12+10). Ce constat est généralisable à l'ensemble des tableaux de résultats. Cette approximation n'interfère en aucun cas sur le résultat de l'AFC.

Tableau de contingence

Poids des entités géographiques et des variables

	Chimie	Economie	Littérature	Medecine	Paix	Physique	Somme
Allemagne	24	1	8	18	5	24	80
Canada	4	3	2	4	1	4	18
France	8	3	11	12	10	9	53
GB	23	6	7	26	11	20	93
Italie	1	1	6	5	1	5	19
Japon	6	0	2	3	1	11	23
Russie	4	3	5	2	3	10	27
USA	51	43	8	70	19	66	257
Somme	121	60	49	140	51	149	570

Tableau de contingence

Deux profils possibles : le profil en ligne

Profils en ligne (Nij/Ni.)									
1963	<i>ALIM</i>	<i>TEXT</i>	<i>BOIS</i>	<i>EDIT</i>	<i>CHIM</i>	<i>CONS</i>	<i>META</i>	<i>EQUIP</i>	Total
<i>BULGARIE</i>	23%	23%	7%	3%	5%	8%	4%	27%	100%
<i>HONGRIE</i>	13%	22%	5%	3%	7%	5%	8%	38%	100%
<i>POLOGNE</i>	14%	23%	6%	3%	8%	7%	5%	33%	100%
<i>R.D.A.</i>	8%	18%	5%	5%	12%	6%	4%	42%	100%
<i>ROUMANIE</i>	10%	22%	18%	3%	5%	8%	8%	26%	100%
<i>TCHECO.</i>	9%	19%	6%	3%	6%	7%	8%	41%	100%
<i>YOUGOSL.</i>	12%	21%	12%	5%	7%	7%	6%	29%	100%
Total	11%	21%	8%	4%	8%	7%	6%	35%	100%

Tableau de contingence

Deux profils possibles : le profil en colonne

Profils en colonne (Nij/N.j)									
1963	ALIM	TEXT	BOIS	EDIT	CHIM	CONS	META	EQUIP	Total
BULGARIE	10%	5%	4%	3%	3%	6%	3%	4%	5%
HONGRIE	11%	10%	6%	7%	8%	8%	13%	10%	10%
POLOGNE	29%	26%	19%	21%	24%	25%	20%	22%	23%
R.D.A.	16%	19%	14%	29%	33%	18%	15%	26%	22%
ROUMANIE	10%	13%	28%	10%	8%	14%	15%	9%	12%
TCHECO.	14%	17%	15%	16%	15%	19%	25%	22%	19%
YOUGOSL.	10%	9%	15%	14%	8%	10%	10%	8%	9%
Total	100%								

Profil en ligne

Indice de spécialisation

	Chimie	Economie	Littérature	Medecine	Paix	Physique	Somme
Allemagne	24	1	8	18	5	24	80
Canada	4	3	2	4	1	4	18
France	8	3	11	12	10	9	53
GB	23	6	7	26	11	20	93
Italie	1	1	6	5	1	5	19
Japon	6	0	2	3	1	11	23
Russie	4	3	5	2	3	10	27
USA	51	43	8	70	19	66	257
Somme	121	60	49	140	51	149	570

Profil en ligne

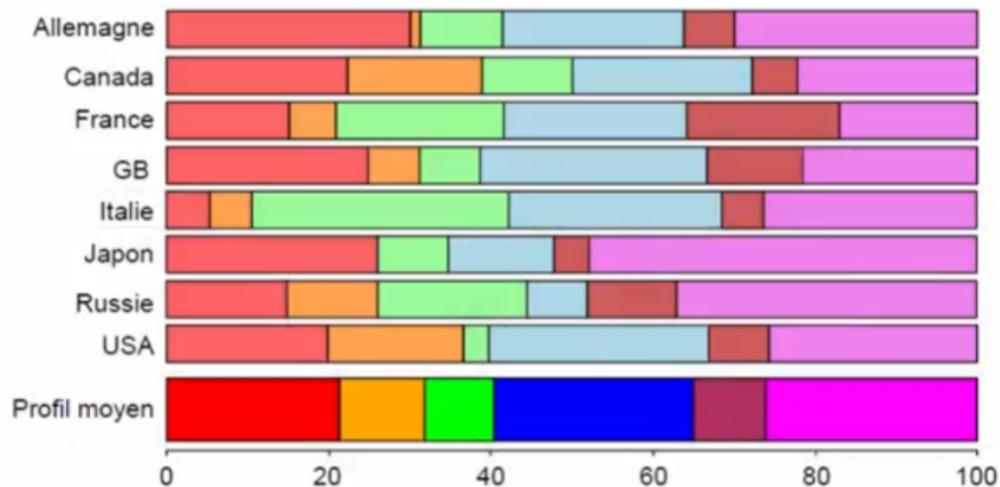
Indice de spécialisation

	Chimie	Economie	Littérature	Medecine	Paix	Physique	Somme
Allemagne	0,300	0,013	0,100	0,225	0,063	0,300	1
Canada	0,222	0,167	0,111	0,222	0,056	0,222	1
France	0,151	0,057	0,208	0,226	0,189	0,170	1
GB	0,247	0,065	0,075	0,280	0,118	0,215	1
Italie	0,053	0,053	0,316	0,263	0,053	0,263	1
Japon	0,261	0,000	0,087	0,130	0,043	0,478	1
Russie	0,148	0,111	0,185	0,074	0,111	0,370	1
USA	0,198	0,167	0,031	0,272	0,074	0,257	1
Somme	0,212	0,105	0,086	0,246	0,089	0,261	1

Profil en ligne

Indice de spécialisation

	Chimie	Eco	Lit.	Médecine	Paix	Physique	Somme
Allemagne	30.0	1.2	10.0	22.5	6.2	30.0	100
Canada	22.2	16.7	11.1	22.2	5.6	22.2	100
France	15.1	5.7	20.8	22.6	18.9	17.0	100
GB	24.7	6.5	7.5	28.0	11.8	21.5	100
Italie	5.3	5.3	31.6	26.3	5.3	26.3	100
Japon	26.1	0.0	8.7	13.0	4.3	47.8	100
Russie	14.8	11.1	18.5	7.4	11.1	37.0	100
USA	19.8	16.7	3.1	27.2	7.4	25.7	100
Profil moyen	21.2	10.5	8.6	24.6	8.9	26.1	100



Profil en ligne

Indice de spécialisation

	Chimie	Economie	Littérature	Medecine	Paix	Physique	Somme
Allemagne	0,300	0,013	0,100	0,225	0,063	0,300	1
Canada	0,222	0,167	0,111	0,222	0,056	0,222	1
France	0,151	0,057	0,208	0,226	0,189	0,170	1
GB	0,247	0,065	0,075	0,280	0,118	0,215	1
Italie	0,053	0,053	0,316	0,263	0,053	0,263	1
Japon	0,261	0,000	0,087	0,130	0,043	0,478	1
Russie	0,148	0,111	0,185	0,074	0,111	0,370	1
USA	0,198	0,167	0,031	0,272	0,074	0,257	1
Somme	0,212	0,105	0,086	0,246	0,089	0,261	1

On appelle INDICE DE SPECIALISATION (S_i) l'écart entre le profil d'une unité spatiale et le profil moyen de l'ensemble de référence.

$$S_i = \sum_{j=1}^n \left| \frac{X_{ij}}{X_{i.}} - \frac{X_{.j}}{X_{..}} \right| = \sum_{j=1}^n |PO_{ij} - PM_{ij}|$$

Profil en ligne

Indice de spécialisation

	Chimie	Economie	Littérature	Medecine	Paix	Physique	Somme
Allemagne	0,300	0,013	0,100	0,225	0,063	0,300	1
Canada	0,222	0,167	0,111	0,222	0,056	0,222	1
France	0,151	0,057	0,208	0,226	0,189	0,170	1
GB	0,247	0,065	0,075	0,280	0,118	0,215	1
Italie	0,053	0,053	0,316	0,263	0,053	0,263	1
Japon	0,261	0,000	0,087	0,130	0,043	0,478	1
Russie	0,148	0,111	0,185	0,074	0,111	0,370	1
USA	0,198	0,167	0,031	0,272	0,074	0,257	1
Somme	0,212	0,105	0,086	0,246	0,089	0,261	1

$$S_{(Allemagne)} = |0.300 - 0.212| + |0.013 - 0.105| + |0.100 - 0.086| + |0.225 - 0.246| + |0.063 - 0.089| + |0.300 - 0.261| = 0.280$$

Profil en ligne

Indice de spécialisation

Allemagne	0,281
Canada	0,193
France	0,442
GB	0,196
Italie	0,498
Japon	0,533
Russie	0,471
USA	0,178

Profil en colonne

Indice de localisation

	Chimie	Economie	Littérature	Medecine	Paix	Physique	Somme
Allemagne	24	1	8	18	5	24	80
Canada	4	3	2	4	1	4	18
France	8	3	11	12	10	9	53
GB	23	6	7	26	11	20	93
Italie	1	1	6	5	1	5	19
Japon	6	0	2	3	1	11	23
Russie	4	3	5	2	3	10	27
USA	51	43	8	70	19	66	257
Somme	121	60	49	140	51	149	570

Profil en colonne

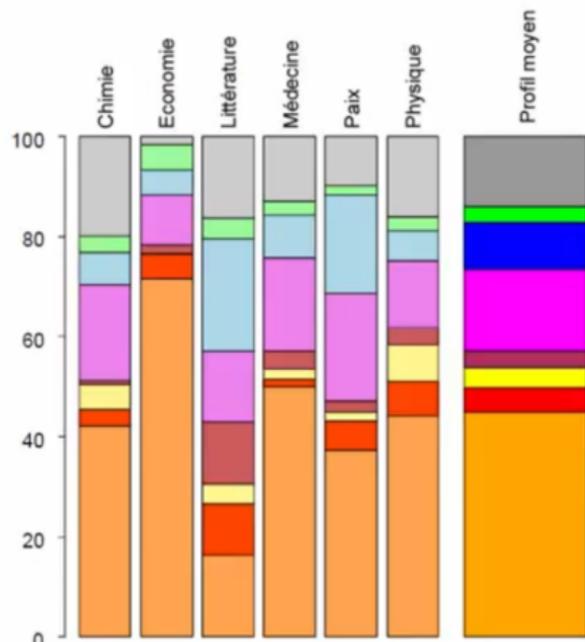
Indice de localisation

	Chimie	Economie	Littérature	Medecine	Paix	Physique	Somme
Allemagne	0,198	0,017	0,163	0,129	0,098	0,161	0,140
Canada	0,033	0,050	0,041	0,029	0,020	0,027	0,032
France	0,066	0,050	0,224	0,086	0,196	0,060	0,093
GB	0,190	0,100	0,143	0,186	0,216	0,134	0,163
Italie	0,008	0,017	0,122	0,036	0,020	0,034	0,033
Japon	0,050	0,000	0,041	0,021	0,020	0,074	0,040
Russie	0,033	0,050	0,102	0,014	0,059	0,067	0,047
USA	0,421	0,717	0,163	0,500	0,373	0,443	0,451
Somme	1	1	1	1	1	1	1

Profil en colonne

Indice de localisation

	Chimie	Eco	Lit	Méd	Paix	Phys	Profil moyen
Allemagne	19.8	1.7	16.3	12.9	9.8	16.1	14.0
Canada	3.3	5.0	4.1	2.9	2.0	2.7	3.2
France	6.6	5.0	22.4	8.6	19.6	6.0	9.3
GB	19.0	10.0	14.3	18.6	21.6	13.4	16.3
Italie	0.8	1.7	12.2	3.6	2.0	3.4	3.3
Japon	5.0	0.0	4.1	2.1	2.0	7.4	4.0
Russie	3.3	5.0	10.2	1.4	5.9	6.7	4.7
USA	42.1	71.7	16.3	50.0	37.3	44.3	45.1
Somme	100	100	100	100	100	100	100



Profil en colonne

Indice de localisation

	Chimie	Economie	Littérature	Medecine	Paix	Physique	Somme
Allemagne	0,198	0,017	0,163	0,129	0,098	0,161	0,140
Canada	0,033	0,050	0,041	0,029	0,020	0,027	0,032
France	0,066	0,050	0,224	0,086	0,196	0,060	0,093
GB	0,190	0,100	0,143	0,186	0,216	0,134	0,163
Italie	0,008	0,017	0,122	0,036	0,020	0,034	0,033
Japon	0,050	0,000	0,041	0,021	0,020	0,074	0,040
Russie	0,033	0,050	0,102	0,014	0,059	0,067	0,047
USA	0,421	0,717	0,163	0,500	0,373	0,443	0,451
Somme	1	1	1	1	1	1	1

On appelle INDICE DE LOCALISATION (L_j) l'écart entre le profil d'une catégorie et le profil moyen de l'ensemble de référence.

$$L_j = \sum_{i=1}^n \left| \frac{N_{ij}}{N_{.j}} - \frac{N_{i.}}{N_{..}} \right| = \sum_{i=1}^n |PO_{ij} - PM_{ij}|$$

Profil en colonne

Indice de localisation

	Chimie	Economie	Littérature	Medecine	Paix	Physique	Somme
Allemagne	0,198	0,017	0,163	0,129	0,098	0,161	0,140
Canada	0,033	0,050	0,041	0,029	0,020	0,027	0,032
France	0,066	0,050	0,224	0,086	0,196	0,060	0,093
GB	0,190	0,100	0,143	0,186	0,216	0,134	0,163
Italie	0,008	0,017	0,122	0,036	0,020	0,034	0,033
Japon	0,050	0,000	0,041	0,021	0,020	0,074	0,040
Russie	0,033	0,050	0,102	0,014	0,059	0,067	0,047
USA	0,421	0,717	0,163	0,500	0,373	0,443	0,451
Somme	1	1	1	1	1	1	1

$$L_{(Chimie)} = |0.198 - 0.140| + |0.033 - 0.032| + |0.066 - 0.093| + |0.190 - 0.163| + |0.008 - 0.033| + |0.050 - 0.040| + |0.033 - 0.047| + |0.421 - 0.451| = 0.192$$

Profil en colonne

Indice de localisation

Chimie	Economie	Littérature	Medecine	Paix	Physique
0,191	0,574	0,616	0,148	0,334	0,148

Quotient de localisation

Explication

Le quotient de localisation est un indicateur de « concentration », de spécialisation.

Il donne une mesure de l'importance relative d'un effectif pour une unité spatiale, comparée à son poids dans les autres unités spatiales.

Le quotient de localisation est un outil d'analyse spatiale, car il permet de caractériser le degré de concentration d'une sous-population dans une unité spatiale en le comparant à toutes les autres unités spatiales d'un même ensemble territorial.

Il permet de mener cette comparaison en faisant abstraction des inégalités de poids entre les unités spatiales et les différentes catégories.

Quotient de localisation

Présentation

	C1	C2	C3	Total
B1	x_{11}	x_{12}	x_{13}	$x_{1\bullet}$
B2	x_{21}	x_{22}	x_{23}	$x_{2\bullet}$
B3	x_{31}	x_{32}	x_{33}	$x_{3\bullet}$
Total	$x_{\bullet 1}$	$x_{\bullet 2}$	$x_{\bullet 3}$	$x_{\bullet \bullet}$

$$Q(x_{ij}) = (x_{ij}/x_{\bullet j}) / (x_{i\bullet}/x_{\bullet \bullet}) = (x_{ij} \times x_{\bullet \bullet}) / (x_{i\bullet} \times x_{\bullet j})$$

Quotient de localisation

Exemple

Branche				
Zone	B1	B2	B3	Total
Z1	48	325	287	660
Z2	27	185	148	360
Z3	45	90	45	180
Total	120	600	480	1200

Quotient de localisation

Exemple

Branche				
Zone	B1	B2	B3	Total
Z1	48	325	287	660
Z2	27	185	148	360
Z3	45	90	45	180
Total	120	600	480	1200

$$Q(X[z2b1]) = \frac{(27/120)}{(360/1200)} = \frac{0,225}{0,300} = 0,75$$

Quotient de localisation

Exemple

Branche	B1	B2	B3
Zone			
Z1	0,727	0,985	1,087
Z2	0,750	1,028	1,028
Z3	2,500	1,000	0,625

TP4

Profil en ligne et profil en colonne

Vous disposez d'un fichier Excel qui contient le nombre d'habitants appartenant à une catégorie socioprofessionnelle donnée pour chaque département français. On est alors en droit de se poser des questions du type : Quelles sont les « spécialités » de chacun des départements ? Le département du Nord est-il un département « ouvrier » ? Paris est-elle une ville de cadres ?

- 1 Pour chaque département, calculez les quotients de localisation pour les ouvriers et pour les cadres.
- 2 A l'aide d'un profil en ligne, déterminez le département français le plus "spécialisé".
- 3 A l'aide d'un profil en colonne, déterminez la CSP la plus "localisée".

- 1 Introduction
- 2 Les résumés statistiques
- 3 La distribution statistique
- 4 Standardisation
- 5 Profil en ligne et profil en colonne

Affectation et calcul

R fonctionne un peu comme une calculatrice. Si vous tapez $2 + 3$, le logiciel vous retournera la valeur 5. Néanmoins, on utilisera R davantage comme un langage de programmation en suivant les principes de l'affectation informatique.

Exemple d'affectation avec R

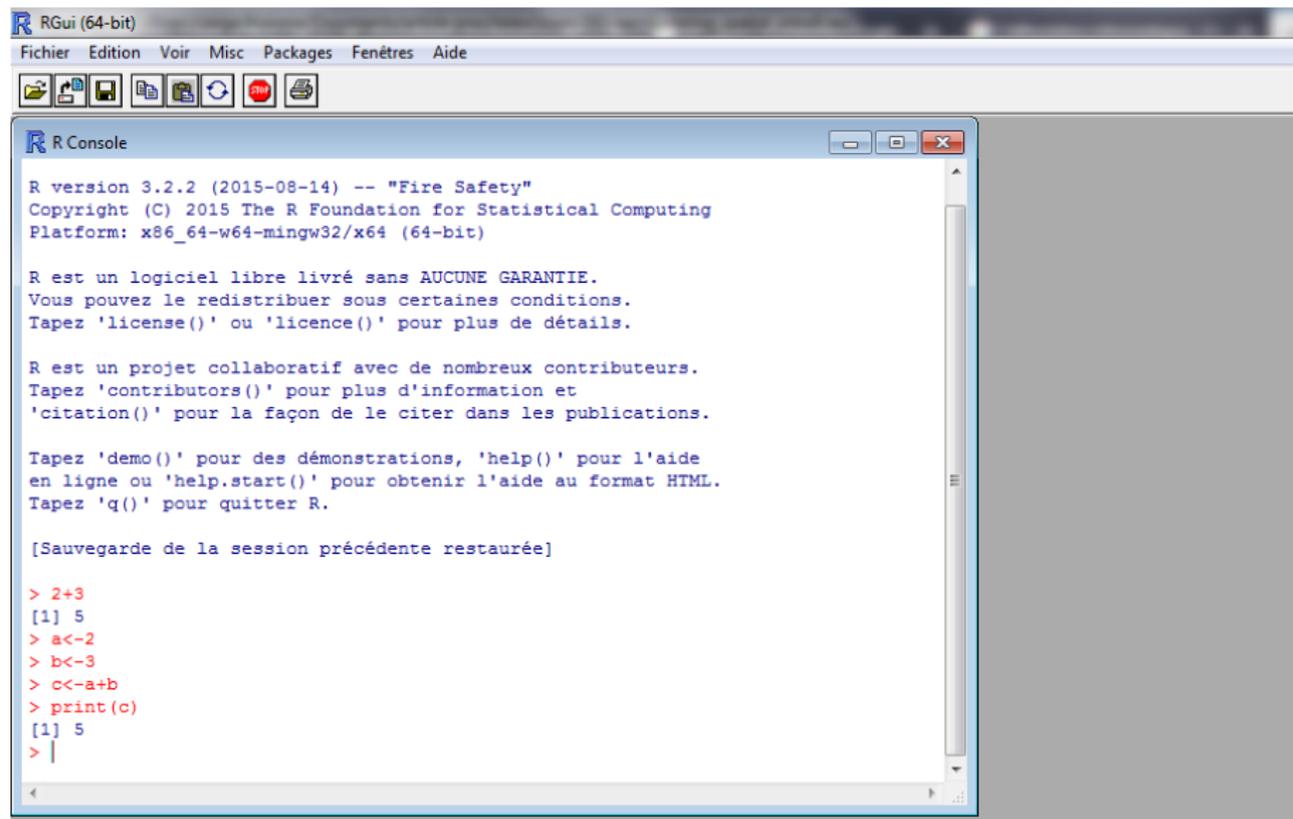
```
a <- 2  
b <- 3  
c <- a + b
```

L'affichage des résultats se fera alors en utilisant une fonction : « `print()` ».

Affichage d'une variable avec R

```
print(c)
```

Affectation et calcul



```
RGui (64-bit)
Fichier  Edition  Voir  Misc  Packages  Fenêtres  Aide

R Console

R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

[Sauvegarde de la session précédente restaurée]

> 2+3
[1] 5
> a<-2
> b<-3
> c<-a+b
> print(c)
[1] 5
> |
>
```

Les types de données

Il existe de nombreux types de variables dans R.

Les variables de type texte

```
a <- "Texte"
```

Ces variables peuvent être ordonnées dans une liste (un vecteur) ou dans plusieurs listes pour former une matrice (un tableau de valeurs).

Les vecteurs et les matrices

```
b <- c(18, 182, 1.5, 15, 200, 5)
```

```
c <- matrix(c(18, 182, 1.5, 15, 200, 5), nrow = 2)
```

```
d <- matrix(c(18, 182, 1.5, 15, 200, 5), ncol = 2)
```

Les types de données

Pour accéder à une valeur ou à un ensemble de valeurs, il faut utiliser les index des vecteurs ou des matrices.

Accès aux valeurs des vecteurs et des matrices

```
e <- b[2] + b[3]
f <- c[1,2] + c[2,3]
col <- c[,1]
ligne <- c[1,]
```

Accès avancé aux valeurs des vecteurs et des matrices

```
e <- b[c(2,4)]
f <- c[(c<15)]
g <- b[2 :5]
```

Les types de données

Les data frames permettent de manipuler des tableaux bien structurés. Ce type de données est particulièrement bien adapté aux importations de fichiers textes.

Les Data Frames

```
articles <- c( "un", "le", "la", "les")
sujets <- c( "mot", "terme", "chose", "images")
dfmots <- data.frame(articles, sujets)
dfmots2 <- data.frame(col1 = articles, col2 = sujets)
```

Appel des valeurs des Data Frames

```
print(dfmots$sujets)
print(dfmots[,1])
```

L'import de données et premières fonctions

Importation de fichiers textes

```
MyTexte <- read.table(file="c :/TheData.csv", header=TRUE, sep=",")
MyData <- read.csv(file="c :/TheData.csv", header=TRUE, sep=",")
adresse <- file.choose()
MyData <- read.csv(file=adresse, header=TRUE, sep=",")
```

Fonctions de base

```
res <- summary(b)
plot(d[,1],d[,2])
hist(b)
reg <- lm(d[,1] ~d[,2])
res3 <- summary(reg)
t.test(d[,1], d[,2])
```