

Statistiques pour la psychologie

8 — (Re)présentation d'un couple de variables

Nicolas Gauvrit
 Université de Metz
<http://adems.free.fr/stats.htm>

Année 2003-2004

1 Variables nominales

Exercice 1 Définir les variances marginales et les variances conditionnelles en suivant la logique ci-dessus. ■ La variance marginale est la variance d'une variable calculée sur l'échantillon (ou la population) entière. Une variance conditionnelle de X est la variance de X calculée sur l'ensemble des individus satisfaisant une condition sur une autre variable Y . ■

Exercice 2 On considère le tableau suivant, qui donne les choix des étudiants entrant en première année :

genre\section	sciences	lettres	autres	Σ
Masc	79	64	53	
Fem	13	21	48	
Σ				

1. Identifiez le tableau (de quel type de tableau s'agit-il ?)
2. Complétez le tableau.
3. Calculez les profils ligne et les profils colonnes et interprétez.
4. Représentez les données.
5. Divisez tous les effectifs par l'effectif total. Que donne alors le tableau ? ■ (1) Il s'agit d'un tableau de contingence. (2)

genre\section	sciences	lettres	autres	Σ
Masc	79	64	53	196
Fem	13	21	48	82
Σ	92	85	101	278

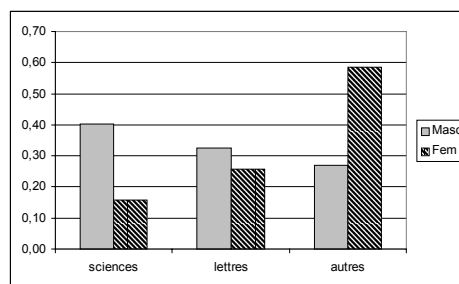
(3) les profils lignes sont les suivants :

genre\section	sciences	lettres	autres
Masc	.40	.33	.27
Fem	.16	.26	.59

où l'on voit que les garçons choisissent de préférence les sciences, alors que les filles préfèrent d'autres matières. Les profils colonnes sont les suivants :

genre\section	sciences	lettres	autres
Masc	.86	.75	.52
Fem	.14	.25	.48

où l'on voit que dans toutes les sections les garçons sont plus nombreux, surtout dans les sections scientifiques et littéraires. Cela est très probablement dû à un problème d'échantillonnage : il y a deux fois plus de garçons que de filles. On préférera donc les profils lignes pour conclure quoi que ce soit. (4) On peut représenter les profils lignes :



(5) on obtient alors la distribution de la variable couple $M \times S$ où S est la variable sexe et M la variable "matière" :

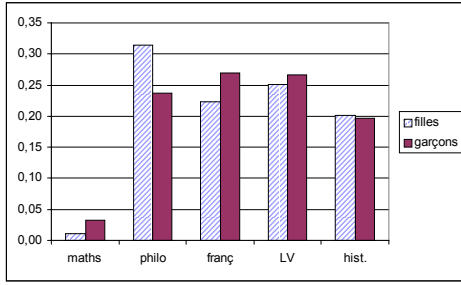
genre\section	sciences	lettres	autres
Masc	.28	.23	.19
Fem	.05	.08	.17

Exercice 3 On demande à des lycéens de donner leur matière préférée, parmi les 5 proposées.

	maths	philo	franç	LV	hist.
filles	4	123	87	98	79
garçons	12	88	100	99	73

Traitez les données. ■ Les effectifs fille et garçon étant différents, il est préférable d'étudier les profils lignes (ici en pourcentage) :

	maths	philo	franç	LV	hist.
filles	1	31	22	25	20
garçons	3	24	27	27	20



Et l'on voit que les maths et le français semblent attirer plus les garçons, à l'inverse de la philosophie. Quant à l'histoire, elle est autant aimée des garçons que des filles. Globalement, seules les mathématiques se distinguent (très peu appréciées des deux sexes). ■

Exercice 4 On demande à de jeunes volontaires français de répondre à quelques questions concernant le chômage. À la question "quelle est selon vous la cause principale du chômage", on obtient les réponses suivantes :

	filles	garçons
fatalité économique	77	75
peu d'emplois	48	59
licenciements	10	34
manque de volonté	23	57

Les deux variables (quelles sont-elles ?) semblent-elles indépendantes ? Décrire les résultats de ce tableau.

■ On s'aperçoit en calculant les profils colonnes (ici en pourcentage) que les filles donnent plus souvent les réponses "fatalité économique", que "peu d'emplois" est au même niveau pour les deux sexes, et que "manque de volonté" (cause interne) et "licenciement" (cause externe) sont plus souvent données par les garçons. Pour résumer, disons que globalement c'est la fatalité économique qui est souvent invoquée par les jeunes, indépendamment du sexe, mais que les filles semblent plus fatalistes que les garçons en ce qui concerne le chômage.

	filles	garçons
fatalité économique	49	33
peu d'emplois	30	26
licenciements	6	15
manque de volonté	15	25

Exercice 5 On reprend l'étude de l'exercice précédent, mais on suppose que la question était "quelles sont selon vous les causes profondes du chômage ?". En supposant que les résultats sont les mêmes que dans l'exercice précédent, quel est le type du tableau ? Comment peut-on décrire la situation statistique ?

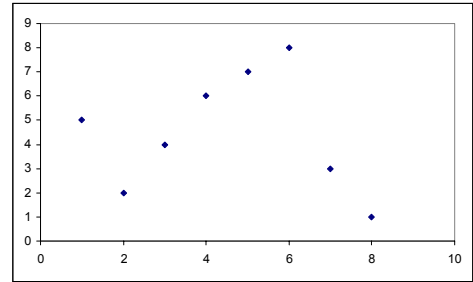
■ Les jeunes peuvent désormais donner plusieurs réponses. Les individus sont les sujets, et les variables sont au nombre de 5 : le sexe (VI), et les réponses à chaque proposition (4 VD dichotomiques). Par exemple "fatalité économique" est une variable dichotomique, avec pour modalité oui ou non selon le choix de l'individu. Le tableau est la juxtaposition de

4 tables de contingence croisant le sexe et chacune des 4 autres variables. ■

Exercice 6 On demande à deux membres du CNU de classer huit universités, selon la valeur du département de psychologie. On donne les numéros d'ordre obtenus :

univ	A	B	C	D	E	F	G	H
rang 1	3	1	6	5	4	2	7	8
rang 2	4	5	8	7	6	2	3	1

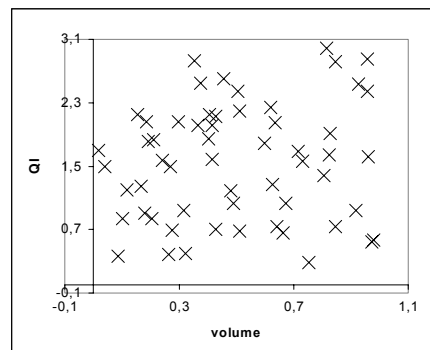
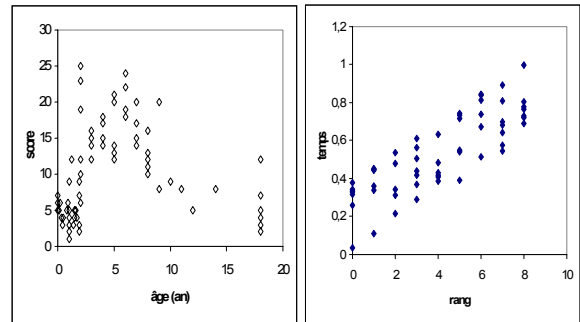
Représentez et interprétez les données. ■ On a deux variables ordinales rang1 et rang2. Un diagramme de dispersion (en abscisse rang1, en ordonné rang2) permet d'étudier le lien entre les rangs :



Et il semble donc que les deux membres du CNU ne soient pas d'accord du tout sur les classement des universités A, G et H, alors que le reste des universités est classé dans le même ordre. ■

2 Deux variables numériques

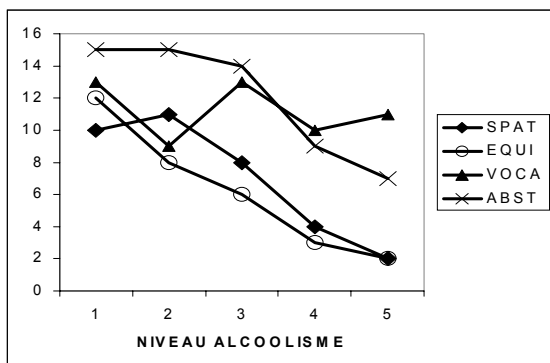
Exercice 7 Discutez le lien entre X et Y dans les cas suivants :



■ (1) Un lien mais non linéaire (2) lien linéaire fort (3) pas de lien du tout. ■

3 Cas mixte

Exercice 8 On étudie l'effet de l'alcoolisme sur une population féminine. Les femmes sur lesquelles porte l'étude sont classées en plusieurs groupes selon la gravité de l'alcoolisme, par un jury de plusieurs thérapeutes. On relève un score de mémoire spatiale (SPAT), un score d'équilibre (EQUI), de vocabulaire (VOCA) et de mémoire à court terme sur du matériel abstrait (ABST).



Interprétez. ■ L'alcool ne semble pas avoir d'effet notable sur le vocabulaire. En revanche, il aurait un effet négatif sur l'équilibre et la mémoire spatiale, et, à haute dose, sur la MCT. Il faut cependant faire attention qu'il ne s'agit peut-être pas de causalité! ■

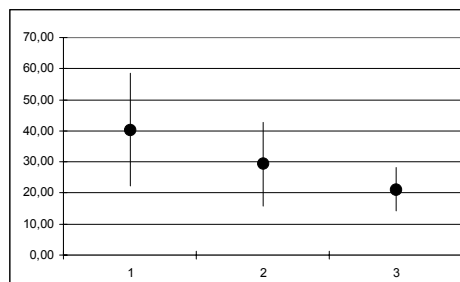
Exercice 9 On étudie grâce à une mesure de stress X l'effet de la cigarette sur l'anxiété. La mesure X est d'autant plus forte que le stress est élevé. On a formé trois groupes de fumeurs, soumis aux mêmes conditions. Dans le groupe témoin 0, les fumeurs ne peuvent pas fumer pendant l'expérience. Dans le groupe 1, ils peuvent fumer des cigarettes sans nicotine (ils ne sont pas informés du fait que les cigarettes proposées n'en contiennent pas : on étudie l'effet placebo) et dans le groupe 2, ils peuvent fumer des cigarettes ordinaires, également proposées par l'expérimentateur. On trouve :

groupe 0	groupe 1	groupe 2
34	33	30
29	28	26
76	51	23
36	38	17
48	14	10
19	12	

Représentez les données, interprétez le graphique. ■
On trouve

	0	1	2
\bar{x}	40.33	29.33	21.20
σ	18.14	13.51	7.03

et un diagramme de variation donne



Il montre l'effet placebo, moins fort que l'effet de la nicotine. Il se peut que la réduction de l'écart type dans le groupe 3 ne soit qu'un effet de bord (le stress arrivant à des niveaux limites ne peut plus tellement varier). ■