Statistiques pour la psychologie
6 — Paramètres

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

Année 2003-2004

1 Interprétation

Exercice 1 *Afin de connaître[1] l'attitude des étudiants de première année face à leur discipline, on pose aux volontaires les questions "la psychologie est-elle une science" (variable S) et "la psychanalyse fait-elle partie de la psychologie" (variable P). Les réponses sont codées de 1 (pas du tout) à 5 (tout à fait) en passant par 3 (sans opinion, ou neutre). On trouve*

	moyenne	écart type
S	1.1	0.8
P	1.2	2.7

Traitez les données. ■ *Les moyennes ne semblent pas significativement différentes (l'écart entre moyennes étant très inférieur au plus petit écart type. Les étudiants semblent rejeter globalement (en moyenne) les deux hypothèses correspondant à S et P. En revanche, les écarts types sont très différents, et il semble que la question sur la psychanalyse, contrairement à celle sur la science, ne fasse pas consensus.* ■

Exercice 2 *Dans le cadre d'une étude sur les femmes alcooliques et leur familles, on mesure par un score numérique R la plus ou moins bonne réussite scolaire d'enfants de mère alcoolique ou non-alcoolique. Les résultats sont les suivants :*

	effectif	moyenne	écart type
mère alcoolique	45	5	9
non-alcoolique	100	6	4

1. *Qu'est-ce qui différencie les deux échantillons ?*
2. *Calculez la moyenne et l'écart type de X sur l'échantillon total des 145 familles étudiées.* ■ *(1) Les moyennes sont peu différentes par rapport aux écarts types. En revanche, les écarts types diffèrents, les valeurs sont plus dispersées dans le groupe "mère alcoolique". On peut imaginer que certains enfants sont handicapés dans ce groupe, alors que d'autres "compensent" en réussissant particulièrement bien à l'école. (2) On a*

$$E(x) = \frac{5 \times 45 + 6 \times 100}{145} = 5.69$$

On détermine ensuite $E(x^2) = \operatorname{var}(x) + E(x)^2$ dans chaque groupe, ce qui donne

$$9^2 + 5^2 = 106$$

et

$$4^2 + 6^2 = 52$$

D'où

$$\begin{aligned} \operatorname{var}(x) &= 106 + 52 - 5.69^2 \\ &= 125.62. \\ \sigma_x &= \sqrt{125.62} \\ &= 11.21. \end{aligned}$$ ■

Exercice 3 *Un test de rapidité de lecture est effectué chez des personnes ayant consommé du café (groupe test) ou un placebo (groupe témoin). L'écart type de la rapidité de lecture est de 2 dans le groupe témoin, et de 1.3 dans le groupe test. Proposez une hypothèse qui pourrait expliquer ce phénomène.* ■ *La dispersion est supérieure dans le groupe témoin. On peut penser que l'effet placebo diffère plus d'une personne à l'autre que l'effet de la caféine.* ■

Exercice 4 *Qu'appelle-t-on une "classe faible" et une "classe hétérogène" en termes statistiques ? (donnez des équivalents statistiques de ces expressions portant sur la variable note).* ■ *Si X est la variable note, une classe faible est une classe pour laquelle $E(X)$ est faible, une classe est hétérogène si σ_X est grand.* ■

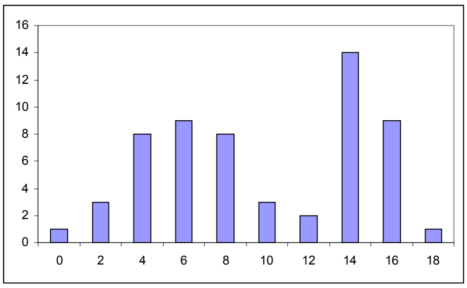
Exercice 5 *Dans une classe, les notes (moyenne par élève) sont (arrondies au nombre pair inférieur) :*

note	0	2	4	6	8	10	12	14	16	18
freq.	1	3	8	9	8	3	2	14	9	1

1. *Représentez la distribution de la note.*
2. *Calculez la moyenne et l'écart type.*

[1] Enquête par questionnaire effectuée à Metz en février 2001.

3. *Décrivez cette distribution et interprétez.* ■ *(1)*



(2) $m = 9.79, \sigma = 4.87$ *(3) La distribution est bimodale. Il semble y avoir deux groupes dans la classe : l'un de niveau modeste, l'autre de bon niveau.* ■

Exercice 6 ♠ *On forme 4 groupes de volontaires qui participent à un exercice de vitesse. Les conditions de la passation sont plus ou moins déteriorées selon les groupes. Ainsi, dans le groupe 1 (G1), les participants sont au calme dans une pièce particulière. Dans le groupe 2, un bruit de fond continu peut les déranger. Dans le groupe 3, les lumières clignotent. Dans le groupe 4, un bruit irrégulier perturbe les sujets, et la lumière clignote de manière imprévisible. On relève les temps T mis pour réaliser la tâche :*

G1	G2	G3	G4
2	1	3	15
3	3	4	16
3	4	6	16
3	4	7	3
4	5	7	4
2	6	8	5
5	5	9	6
8	12	9	9
6	9	11	11

1. *Décrivez la situation statistique.*
2. *Calculez les moyennes et les variances de T pour chaque groupe.*
3. *Quelle est la variance moyenne (moyenne des quatre variances de la question précédente) ? On l'appellera la variance intra-groupe.*
4. *Calculez la variance et la moyenne générale.*
5. *On suppose que si le facteur groupe n'a pas d'effet sur T, alors chaque individu donnerait pour T la valeur moyenne de son groupe. Quelle serait alors la moyenne générale et la variance générale ? (on l'appellera désormais variance inter-groupes).*
6. *Comparez les variances inter-groupe, intra-groupe, et totale (c'est-à-dire générale).*
7. *Concluez.* ■ *(1) Les individus sont des personnes, on a deux variables : le groupe (VI nominale) et le temps T (VD numérique). (2)*

	G1	G2	G3	G4
moy	4.00	5.44	7.11	9.44
var	3.56	9.58	5.65	24.69

(3)

$$var_{intra} = 10.87$$

(4)

$$E(T) = 6.50 \; ; \; var(T) = 14.97$$

(5) La moyenne serait 6.50, *et la variance*

$$var_{inter}(T) = 4.10$$

(6) On a

$$var(T) = var_{inter}(T) + var_{intra}(T)$$

(7) On peut dire que le groupe "explique" une proportion

$$\frac{4.10}{14.97} = 27.4\%$$

de la variance totale. ■

2 Compléments

Exercice 7 *Un rat étudié dans un labyrinthe se déplace à une vitesse a (en m/s), sur vingt centimètres, puis à une vitesse b sur 20 centimètres, et enfin à une vitesse c sur la même distance de 20cm. Quelle est sa vitesse moyenne ?* ■ *La distance parcourue est de 60cm, soit 0.6m. Le temps mis au total est*

$$t = \frac{.2}{a} + \frac{.2}{b} + \frac{.2}{c}$$

d'où la vitesse moyenne :

$$v = \frac{0.6}{0.2 \sum \frac{1}{v_i}}$$

où v_i prend les valeurs a, b, et c. En simplifiant par 0.6, on trouve

$$v = \frac{1}{E\left(\frac{1}{v}\right)}$$

qui est la moyenne harmonique des vitesses. ■

Exercice 8 *Un enseignement A permet d'augmenter le résultat à une épreuve cognitive de 5%. Un enseignement B permet de l'augmenter de 8% et un enseignement C de l'augmenter de 19%. Un sujet suit les trois enseignement. De combien est augmenté le résultat à l'épreuve considérée ? Quelle augmentation a (en proportion), appliquée trois fois, aurait donné la même augmentation finale ?* ■ *Au total, on passe d'un score X à un score de*

$$1.05 \times 1.08 \times 1.19 \times X = 1.35 \times X,$$

soit une augmentation de 35%. On doit avoir

$$1.35 = (1+a)^3$$

soit

$$\begin{aligned} 1 + a &= (1.35)^{\frac{1}{3}} \\ &= 1.11 \end{aligned}$$

et il faut donc poser

$$a = 11\%.$$ ■

Exercice 9 ♠ *On étudie le résultat X à un test de reconnaissance spatiale sur une population de 100 sujets.*

	effectif	*moyenne*	*écart type*
hommes	*40*	*6.8*	*1.2*
femmes	*60*	*4.5*	*0.9*

.

Déterminez la moyenne de X sur la population toute entière, ainsi que de l'écart type. ■ *On trouve d'abord la somme en faisant* $6.8 \times 40 + ...$ *et on en déduit*

$$\begin{aligned} E(X) &= \frac{40 \times 6.8 + 60 \times 4.5}{100} \\ &= 5.42. \end{aligned}$$

Grâce aux écarts type, on peut ensuite déterminer $E\left(X^2\right)$, *et le tour est joué.* ■

Exercice 10 ♠ *On appelle* moyenne quadratique *le paramètre définie par*

$$mq(x) = \sqrt{\frac{1}{n}\sum x_i^2}.$$

1. *Exprimez $mq(x)$ en fonction de la moyenne arithmétique (standard) E.*
2. *Où avez-vous déjà rencontré cette moyenne quadratique ?* ■ *(1)*

   $$mq\,(x) = \sqrt{E\,(x^2)}$$

   (2) dans l'expression de l'écart type. ■

Exercice 11 ♠ *On appelle* moyenne harmonique *le paramètre définie par*

$$mh(x) = \frac{1}{\frac{1}{n}\sum \frac{1}{x_i}}.$$

1. *Exprimez $mh(x)$ en fonction de la moyenne arithmétique E.*
2. *La moyenne harmonique est-elle définie pour toutes les variables ?*
3. *Où avez-vous déjà rencontré cette moyenne ?* ■ *(1)*

   $$mh\,(x) = \frac{1}{E\left(\frac{1}{x}\right)}$$

   (2) il faut que x ne s'annule pas (3) plus haut, dans l'exrcice sur les vitesses. ■

Exercice 12 ♠ *On appelle* moyenne géométrique *le paramètre défini par*

$$mg(x) = \sqrt[n]{x_1 ... x_n}.$$

1. *Exprimer $mg(x)$ en fonction de la moyenne standard E.*
2. *Pour quelles variables mg est-elle définie ?*
3. *Quelle est la moyenne géométrique d'une variable qui s'annule ?* ■ *(1) si x ne s'annule pas*

   $$mg\,(x) = \exp\left(E\,(\ln x)\right)$$

   (2) il faut (pas tout à fait : précisez) que x reste positif (3) 0. ■

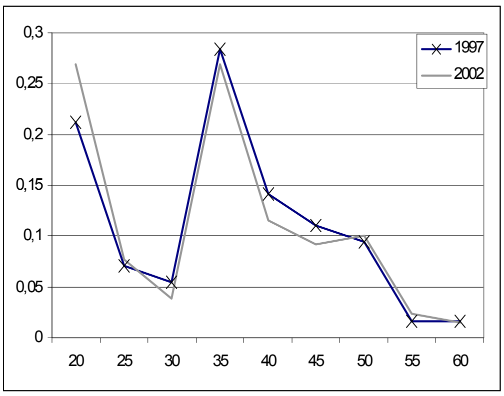
3 Petits problèmes

Exercice 13 *On relève le nombre d'heures*[2] *effectivement travaillées par salarié en 1997 et en 2002. Les données sont arrondies. On trouve sur un échantillon les valeurs qui suivent.*

	20	25	30	35	40	45	50	55	60
1997	27	9	7	36	18	14	12	2	2
2002	35	10	5	35	15	12	13	3	2

.

1. *Décrivez la situation statistique. De quel type de tableau s'agit-il ?*
2. *Calculez une approximation du temps de travail hebdomadaire en 1997 et en 2002 sur les échantillons considérés. Pourquoi est-ce une approximation ?*
3. *Construisez les graphiques correspondants aux deux distributions du temps de travail.*
4. *Déterminez pour chaque année la médiane, le mode, et les premiers et troisième quartiles.*
5. *Calculez les écarts types.*
6. *La distribution du temps de travail a-t-elle changée entre 1997 et 2002 ? Dans quel sens ?* ■
   (1) On pourra considérer que les individus sont les salariés et qu'on étudie deux populations, qui sont celles de 1997 et de 2002. L'unique variable est alors le temps de travail, que nous noterons T. Il manque une légende au tableau, mais il ne peut être qu'un tableau statistique, car la première ligne contient des valeurs de T.
   (2) Les moyennes que nous calculons seront des approximations car les données sont arrondies. Si elles ne l'étaient pas, nous aurions alors une valeur exacte et non une approximation de la moyenne d'échantillon. Cette moyenne d'échantillon pourrait pourtant être utilisée comme estimation *pour la moyenne de la population. La moyenne en 1997 est de 34.76. Elle est de 33.85 en 2002.*
   (3)



[2] "Contraintes de temps dans le travail et risques pour la santé en Europe", *Quatre pages du CEE*, 11-02.

(4)

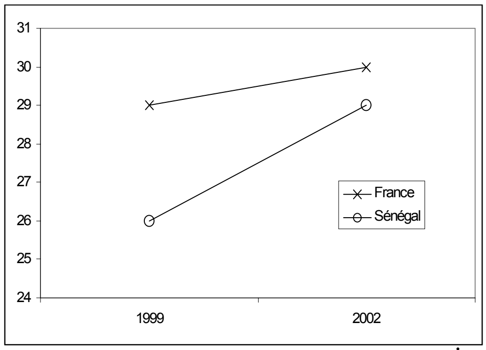
	méd	mode	Q1	Q3
1997	*35*	*35*	*25*	*40*
2002	*35*	*20 ; 35*	*20*	*40*

(5) L'écart-type de T est de 10.40 en 1997 sur notre échantillon, et de 11.02 en 2002.
(6) Le graphique semble montrer que la distribution du temps de travail n'a presque pas changé entre 1997 et 2002. On observe cependant un léger décalage : en 2002, les 40-45 heures sont en légère baisse, alors que les 20 heures sont en hausse. La baisse de moyenne constatée pourrait ainis être due à la multiplication des contrats à temps partiel, plus qu'à la mise en place des 35heures habdomadaires. ■

Exercice 14 *On relève l'âge auquel des individus accèdent à leur premier emploi, en France et au Sénégal.*

	1999			*2002*		
	$\overline{x}$	n	σ_x	$\overline{x}$	n	σ_x
France	29	100	3.6	30	100	2.4
Sénégal	26	100	2.8	29	100	3

1. *Décrivez la situation statistique*
2. *Représentez les données*
3. *Semble-t-il que les Sénégalais accèdent plus tôt au premier emploi ?*
4. *Semble-t-il y avoir une interaction entre le pays et la date, visible sur l'âge du premier emploi ? Interprétez.* ■ *(1) On pourrait considérer qu'il y a deux populations, l'une en 1999 et l'autre en 2002. Cependant, on s'interroge à la question 4 sur l'interaction entre l'année et le pays, si bien qu'il est préférable de considérer l'année comme une variable. A cause de cela, on doit préciser que les individus sont des personnes* à une date donnée, *ou, ce qui revient au même, des "premiers emplois". Ainsi, le même sujet étudié en 1999 et en 2002 correspond à deux individus distincts. La population et formée par les Français et les Sénégalais en 1999 et 2002. Les variables étudiées sont le pays P, l'année A et l'âge d'accès au premier emploi X.*
(2) Le plus simple est de représenter une courbe par pays, représentant les moyennes de X. Cela permet de lire simultanément X, A et P sur le graphique :



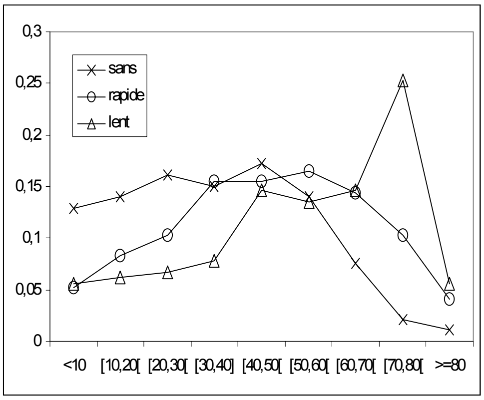
âge moy. d'accès au 1er emploi
en fonction de P et A

(3) La valeur moyenne de X est inférieure au Sénégal, en 1999 et en 2002. Il semble donc bien que les Sénégalais accède plus tôt au premier emploi en moyenne, bien que l'écart se réduise.
(4) Il s'agit d'une question qui oblige à considérer trois variables simultanément et non les variables une par une. Les deux courbes du diagramme ne sont par parallèles. Cela signifie qu'il y a bien, entre P et A (les deux VI) une interaction, et que cette interaction a un effet sur X. En termes concrets, cela se traduit par le fait que l'évolution *de X n'est pas la même en France et au Sénégal. On peut même préciser cela : en 3 ans, le X moyen a augmenté de 1 an en France, mais il a augmenté de 3 ans au Sénégal.* ■

Exercice 15 *Dans une étude sur l'effet de la musique dans les supermarchés, on mesure dans trois situations (pas de musique / musique rapide / musique lente) les dépenses des clients (en €) :*

dépense	*sans*	*rapide*	*lent*
< 10	12	5	10
$10 - 20$	13	8	11
$20 - 30$	15	10	12
$30 - 40$	14	15	14
$40 - 50$	16	15	26
$50 - 60$	13	16	24
$60 - 70$	7	14	26
$70 - 80$	2	10	45
≥ 80	1	4	10

■ *Les individus sont les clients. On a deux variable. Le contexte C, VI qualitative, avec pour modalité "sans", "rapide", et "lent", et la dépense D, VD numérique continue. On représente les distributions* *conditionnelles de D connaissant C.*



Distributions cond. de $D|C$.

Il apparaît queles dépenses semblent plus élevées avec une musique lente, et moins élevées sans musique. On calcule pour formaliser cette idée les moyennes conditionnelles, qui font ici sens puisqu'elles sont liées à la recette du supermarché.

	sans	*rapide*	*lent*
moy de D	*34.89*	*46.13*	*52.86*

Il s'agit d'approximation, car nous n'avons pas les valeurs de D (Nous avons supposé que la rubrique ">=80" était en réalité $[80, 90[$*, ce qui est peut-être faux). Cependant, ces estimations vont dans le sens attendu après lecture du graphique. La musique pousserait donc à la consommation, et plus spécialement les musiques lentes. Une musique lente contribue peut-être à la sensation de bien-être qui poussent les clients à rester plus longtemps chez le commerçant.* ■