

# Statistiques pour la psychologie — Licence

## Corrélation

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

22 décembre 2004

**Exercice 1** *On pose la question suivante à des personnes : "Une fille tombe amoureuse d'un homme à l'enterrement de sa mère, mais elle n'a pas le temps de lui parler. Deux jours plus tard, elle tue sa sœur. Essayez de deviner pourquoi." Sur 40 psychopathes interrogés, 31 répondent "Pour revoir l'homme qu'elle aime". Sur 60 sujets normaux, 12 répondent de cette manière. Cette question permet-elle de confirmer un diagnostic ? Si une personne répond "Pour revoir l'homme qu'elle aime", doit-on en conclure qu'elle est (probablement) psychopathe ?*

**Éléments de réponse.** Non. Cela serait négliger le "taux de base" (la plupart des gens ne sont pas psychopathes). ■

**Exercice 2** *On demande à des étudiants en lettres de rappeler une série de faux mots appris dans la première phase de l'expérience (exemple : synotropol). Dans un groupe ( $G = 0$ ), on leur dit qu'il s'agit de mots techniques décrivant l'usinage de pièces métalliques ; dans un autre groupe ( $G = 1$ ), on leur dit qu'il s'agit de mots décrivant des procédés littéraires. On mesure le nombre de mots correctement rappelés  $X$ . On trouve un coefficient  $\tau = .24$ . Interprétez.*

□ Ce coefficient de corrélation, calculé à partir d'une variable numérique  $X$  et d'une VI dichotomique (codée par 0 et 1)  $G$ , s'appelle *coefficient bisériel de points*. On pourrait donc écrire

$$r_{bp} = 0.24,$$

ce qui donne immédiatement

$$r_{bp}^2 = 5.8\%.$$

Cela signifie entre autres choses ( $r > 0$ ) que le groupe 1 rappelle en moyenne plus de mots que le groupe 0, même si cette différence n'est peut-être pas significative<sup>1</sup>.

On pourrait expliquer ce résultat (à supposer qu'il se généralise) en faisant l'hypothèse que la motivation est plus élevée dans le groupe 1, parce que les sujets

<sup>1</sup>Nous ne pouvons pas savoir si cette valeur est significative, parce que cela dépend de la taille  $n$  de l'échantillon, que nous n'avons pas.

pensent travailler sur un matériel ayant trait à un domaine qu'ils connaissent et apprécient.

La valeur de  $r_{bp}^2$ , si l'on confond le groupe et la motivation, indique alors que la motivation explique, dans notre expérience, 5.8% des variations du taux de rétention. □

**Exercice 3** *Une agence de telemarketing téléphone à des abonnés du téléphone pour une entreprise de services financiers<sup>2</sup>. Le but de cet appel est de faire venir les clients potentiels dans les bureaux de l'entreprise, où on pourra plus facilement les convaincre d'acheter les produits vendus par l'entreprise. Deux techniques sont utilisées : soit on demande à la personne si elle souhaite prendre rendez-vous (situation  $S = 0$ ), soit on lui demande "Quand êtes-vous disponible ?" sans lui demander si elle souhaite prendre rendez-vous (situation  $S = 1$ ). Si elle donne une date, on lui annonce alors "Notre conseiller financier vous attendra". On note ensuite la réussite éventuelle  $R$  du coup de téléphone ( $R = 0$  si la personne ne se déplace pas,  $R = 1$  sinon). On trouve sur un échantillon assez grand les résultats suivants :*

	$R = 0$	$R = 1$
$S = 0$	144	16
$S = 1$	117	38

Traitez les données.

**Éléments de réponse.** Le mieux est de calculer un coefficient  $\phi$ , qui donne la force du lien entre les variables situation et réponse. On remarque alors, si  $\phi$  est significatif, que le taux de réponse positive est supérieur dans la situation où on force la main des gens. Forcer la main des clients est donc efficace. ■

**Exercice 4** *Deux médecins<sup>3</sup> qui ont étudié les questions relatives aux drogues jugent la dangerosité de quelques drogues courantes au moyen des critères*

<sup>2</sup>La psychologie a grandement aidé au développement de techniques de vente efficace. On trouvera des informations sur ce sujet dans le divertissant *Petit traité de manipulation à l'usage des honnêtes gens* de \*\*\* paru aux Presses Universitaires de Grenoble (\*\*\*).

<sup>3</sup>J.E. Henningfield, *National Institute* & N.L. Benowitz, *University of California San Francisco*

suivants : syndrome de sevrage *SDS* (gravité du contre-coup lorsqu'on arrête la drogue après une utilisation standard), renforcement *RENF* (tendance de la drogue qui fait que les utilisateurs en prennent et en reprennent), tolérance *TOL* (fait que l'utilisateur doit augmenter la dose pour obtenir les mêmes effets), dépendance *DEP* (force qui empêche les utilisateurs d'arrêter), intoxication *INTOX* (degré des effets négatifs de la drogue sur la santé physique). Les deux médecins ont classé quelques drogues selon ces critères, par ordre décroissant (1 correspond à la drogue la plus dangereuse pour le critère selon le médecin), et on trouve

substance	<i>SDS</i>	<i>RENF</i>	<i>TOL</i>	<i>DEP</i>	<i>INTOX</i>
Nicotine	3	4	2	1	5
Héroïne	2	2	1	2	2
Cocaïne	4	1	4	3	3
Alcool	1	3	3	4	1
Caféine	5	6	5	5	6
Marijuana	6	5	6	6	4

(Henn.)

substance	<i>SDS</i>	<i>RENF</i>	<i>TOL</i>	<i>DEP</i>	<i>INTOX</i>
Nicotine	3*	4	4*	1	6
Héroïne	2	2	2	2	2
Cocaïne	3*	1	1	3	3
Alcool	1	3	4*	4	1
Caféine	4	5	3	5	5
Marijuana	5	6	5	6	4

(Beno.)

□ Il est très tentant d'utiliser les avis de ces médecins pour tirer des conclusions sur les drogues. Cependant, les résultats n'ont de sens que s'il y a, au minimum, un consensus entre les deux médecins considérés, et donc si les classements des deux experts coïncident. Dans le cas contraire, il faudra considérer que les classements donnés ne reflètent que deux avis personnels.

Pour mesurer l'accord entre les juges, on peut utiliser par exemple des coefficients de Spearman (qui ne sont que des coefficient  $r$ ). On trouve par ordinateur :

critère	<i>SDS</i>	<i>RENF</i>	<i>TOL</i>	<i>DEP</i>	<i>INTOX</i>
$r_s$	.99	.94	.35	1.00	.94
$p$	< .001	< .005	.499	< .001	< .005

Compte tenu du fait que l'on réalise 5 tests, un risque maximum de 1% par test est déjà grand (dans la mesure où le risque de l'ensemble est alors approximativement de 5%), et c'est celui que nous fixons. On peut donc conclure que les deux médecins sont significativement d'accord pour ce qui concerne *SDS*, *RENF*, *DEP* et *INTOX* — donc que les principes d'évaluation qu'ils utilisent coïncident — mais pas en ce qui concerne *TOL*. Nous pouvons désormais interpréter les deux tableaux, mais en ne tenant pas compte de *TOL*, qui n'est apparemment que le reflet d'un avis personnel.

Une des conclusions attrayantes est bien entendu que l'alcool est bien plus dangereuse (au sens des 4 variables considérées) que la marijuana par exemple, qui obtient des scores plus grands (donc est moins néfaste).

Il ne faut pas pour autant en déduire que la légalisation de la marijuana serait une bonne chose : le fait que l'alcool soit autorisée et bon marché explique certainement pourquoi les doses habituelles d'alcool sont importantes. Si l'alcool était interdite et la marijuana autorisée, peut-être les alcooliques boiraient-ils beaucoup moins et les fumeurs de hachisch fumeraient-ils beaucoup plus, si bien que les dangers des deux substances pourraient être inversées... □