

# Statistiques pour la psychologie — Licence

## Régression linéaire

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

10 octobre 2004

**Exercice 1** Joseph Klatzmann<sup>1</sup> a relevé, pour différents pays, des indices de consommation d'alcool par habitant ( $A$ ) et l'espérance de vie ( $E$ ) (base 100 pour le Japon). Les résultats sont les suivants :

pays	Nigéria	Brésil	Chine
$A$	2	5	6
$E$	56	68	71

Pays	Italy	Soudan	Australie
$A$	15	1	18
$E$	94	50	99

pays	Japon	USA	France
$A$	9	15	16
$E$	100	98	95

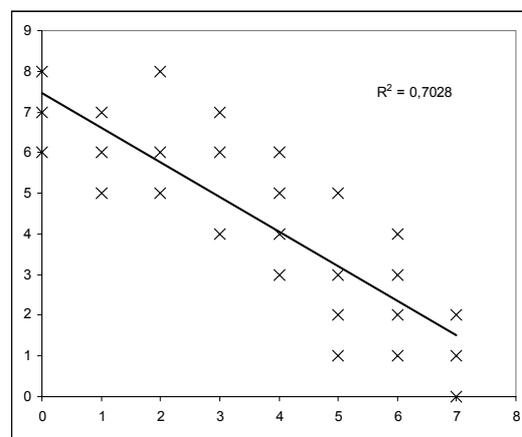
Pays	Uruguay	Luxembourg	Suède
$A$	6	14	13
$E$	60	95	96

1. Décrivez la situation statistique
2. Représentez les données
3. Les deux variables  $A$  et  $E$  sont-elles liées ?
4. La consommation d'alcool a-t-elle un effet positif sur l'espérance de vie ? Expliquez le lien entre les variables.
5. En vous appuyant sur cet exemple, discutez la distinction que l'on pourrait faire entre facteur explicatif et facteur prédictif.

**Exercice 2** On relève sur tous les élèves d'une école le nombre  $X$  de fautes d'orthographe à une dictée (la même dictée pour tous) et la peinture  $Y$  (taille des chaussures, unité arbitraire). On trouve les valeurs

<sup>1</sup>dans *Attention! statistiques*, éd. La découverte, 1997.

représentés ci-dessous :



Concluez (on admettra que le lien entre  $x$  et  $y$  est négatif). Construisez un diagramme bayésien (causal) rendant compte de la situation.

**Exercice 3** Dans une étude portant sur des jugements de beauté, on demande à des volontaires de noter, par une note  $X$ , six portraits ( $A, B, \dots, F$ ). Deux groupes sont formés. Dans le groupe 1, la consigne est de donner une note comprise en 0 à 10, avec 0 pour "très moche" et 10 pour "très beau". Dans le groupe 2, la consigne est de noter entre -5 (très moche) et 5 (très beau). On notera pour chaque visage représenté,  $X_1$  la note obtenue dans le groupe 1 et  $X_2$  celle obtenue dans le groupe 2.

1. Sous certaines hypothèses à préciser, on s'attend à avoir  $r_{X_1 X_2} = 1$ . Expliquez.
2. On trouve en réalité  $r = 0.91$  et l'équation de régression

$$\hat{X}_2 = 1.3X_1 - 5.$$

Concluez.

**Exercice 4** On relève<sup>2</sup> chez des élèves l'image de soit par une note  $X$  et les résultats scolaires par une note  $Y$ . On trouve  $r_{XY} = .59$ . Concluez. Pourquoi  $Y$  est-elle la variable dépendante ?

<sup>2</sup>Monteil, J.-M. & Hughet, P. (2002). *Réussir ou échouer à l'école : une question de contexte ?* Grenoble : Pug.

**Exercice 5** Des chercheurs<sup>3</sup> ont étudié les résultats en sciences (S) et en orthographe (O) d'élèves de primaire, pour les comparer selon le genre<sup>4</sup> G<sup>5</sup>. Ils trouvent les résultats suivants

G	S	O	G	S	O
0	3	4	1	2	3
0	4	5	1	5	4
0	5	5	1	6	5
0	6	5	1	7	4
0	7	6	1	8	6
0	7	7	1	8	9
0	6	7	1	9	4
0	8	8	1	8	2

1. Décrivez la situation statistique de deux manières différentes :

- (a) en supposant qu'il y a une variable "matière" avec pour modalité S et O
- (b) en supposant que les élèves sont les individus statistiques

2. Concluez.

**Exercice 6** On étudie le lien entre le stress et la santé au moyen d'une mesure numérique X de stress et d'une mesure de l'importance des symptômes Y, également numérique. Sur un échantillon de  $n = 107$  personnes, on trouve  $r_{xy} = .506$ . Calculez le coefficient de corrélation ajusté

$$r_{aj} = \sqrt{1 - \frac{(1 - r^2)(n - 1)}{n - 2}}$$

et concluez. Cela prouve-t-il que le stress est mauvais pour la santé ?

**Exercice 7** On demande à 35 volontaires de dessiner un chapeau de la manière suivante :

tel que la hauteur  $h$  soit égale à la largeur  $l$ . Un régres-

<sup>3</sup>Bacharach, V. R. *et al.* (2003). Racial and gender science achievement gaps in secondary education. *The Journal of Genetic Psychology*, 144(1)

<sup>4</sup>Les auteurs étudient aussi les différences blancs/noirs, que nous n'évoquons pas ici.

<sup>5</sup>On notera 0 pour "fille" et 1 pour "garçon".

sion linéaire de  $l$  en  $h$  donne les résultats suivants :

où la courbe en pointillés est la droite d'équation  $l = h$ . Le coefficient de détermination est .7785 et l'équation de régression est

$$\hat{l} = .92h + 8.34.$$

Concluez.

**Exercice 8** <sup>6</sup>Des auteurs anglais mesurent par un grandeur X numérique les capacités de déductions sur les rapports humains (theory of mind) chez des sujets d'âge A. On trouve les moyennes conditionnelles suivantes :

A	60	65	70	75	80
X	4	4.1	3.7	3.3	2.9

Une régression linéaire vous paraît-elle judicieuse ? Traitez les données et fournissez une conclusion en supposant que l'on a 10 sujets par âge, et un écart type conditionnel  $\sigma_{x|A}$  constamment égal à 1.

**Exercice 9** On reprend le thème de l'exercice 7, mais en demandant aux sujets de dessiner le chapeau couché, c'est-à-dire de la manière suivante :

On appelle alors  $l$  la dimension verticale et  $h$  la dimension horizontale du chapeau. On trouve les résultats suivants :

$$r = .90$$

$$\hat{y} = .92x + 6.65.$$

Concluez.

<sup>6</sup>Taylor, E. A. *et al.* (2003). Does performance on theory of mind decline in old age? *British Journal of Psychology*, 43