

Fiche TD avec le logiciel  : tdr206

Croisement de deux variables quantitatives

A.B. Dufour & M. Royer

L'objectif de cette séance est d'étudier la relation entre deux variables quantitatives. Les notions abordées sont détaillées sur le site dans le cours BS02.

Table des matières

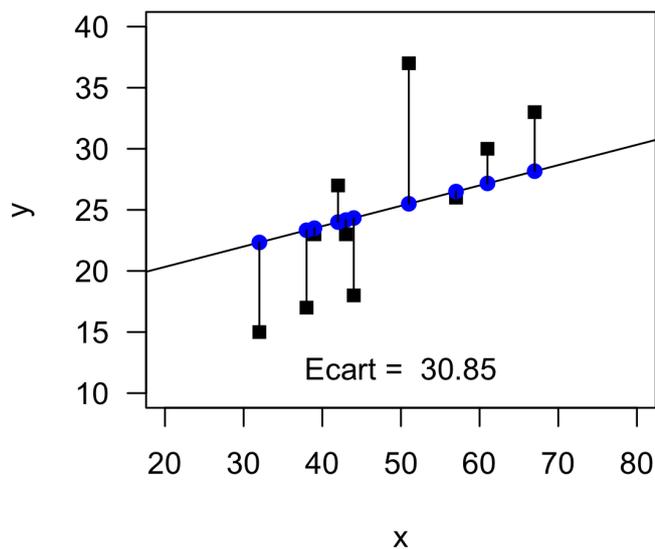
Principe Général	2
Exercice 1	3
Exercice 2	4
Exercice 3	5
Annexe : éléments d'interprétation de la fréquence cardiaque	6

Principe Général

Recherche des paramètres qui minimisent les écarts entre les valeurs observées et les valeurs modélisées. Pour cela, nous donnons un exemple de relation entre deux variables X et Y . Après avoir construit le nuage de points, nous proposons comme paramètres initiaux, pour la droite de régression, une pente $b = \frac{1}{6}$ et une ordonnée à l'origine $a = 17$. Puis nous calculons la moyenne des carrés des écarts entre les valeurs observées et les valeurs modélisées.

```
x <- c(61, 67, 32, 43, 57, 44, 39, 38, 42, 51)
y <- c(30, 33, 15, 23, 26, 18, 23, 17, 27, 37)
plot(x, y, xlim=c(20, 80), ylim=c(10,40), pch=15, las = 1, main = "Le critère d'ajustement")
b <- 1/6
a <- 17
ytheo <- b*x + a
abline(a,b)
points(x, ytheo, pch=20, col="blue", cex=1.5)
segments(x,ytheo,x,y)
mess <- paste("Ecart = ",round(mean((y-ytheo)^2),dig=2))
text(50,12,mess)
```

Le critère d'ajustement



- 1) Faire varier les paramètres de la pente b et de l'ordonnée à l'origine a .
- 2) Regarder la valeur des écarts entre les valeurs observées et les valeurs modélisées.

Les énoncés des exercices suivants sont tirés du livre *"Statistique Vraiment appliquée au sport. Cours et Exercices."* écrit par *S. Champely* aux éditions **De boeck et Larcier** en 2003.

Exercice 1

La Fréquence Cardiaque Maximum, notée FCM , est un paramètre essentiel pour permettre au coureur de fond d'élaborer des plans d'entraînement efficaces. Cette fréquence peut se mesurer, soit en laboratoire sur tapis roulant, soit sur le terrain à l'aide d'un cardio-fréquencemètre.

- 1) Une première étude a été faite auprès de 13 hommes s'entraînant régulièrement (2 à 4 fois par semaine), et participant à de petites compétitions. On a mesuré leur fréquence cardiaque maximum. On souhaite étudier une relation éventuelle entre l'âge d'un individu et sa fréquence cardiaque maximum. Voici pour chaque individu son âge et sa fréquence cardiaque maximum.

Age	40	36	51	49	47	51	32	55	55
FCM	187	195	180	190	185	183	195	185	189

23	49	52	35
201	189	185	195

- a) Tracer le nuage de points, avec l'âge en abscisse et la FCM en ordonnée.
- b) Calculer le coefficient de corrélation linéaire entre la variable \hat{age} et la variable FCM .
Que constatez-vous? Peut-on dire que la fréquence cardiaque est fortement corrélée à l'âge des sportifs?
- c) Calculer l'équation de la droite de régression linéaire. Tracer cette droite sur le même graphe que le nuage de points.
- d) Utiliser cette relation pour donner une estimation de **votre** FCM .
- e) Peter Snell (Nouvelle-Zélande) a été 6 fois recordman du monde en demi-fond, du 800m au mile (1609m), à l'âge de 26 ans (26,6 exactement). A l'époque, sa FCM était de 192. Placer ce point sur le graphique.
D'après la droite de régression, quelle FCM devrait-il avoir pour son âge? Commenter.
- 2) Lors d'un entraînement en endurance, la fréquence cardiaque du sportif se situe entre 70 % et 80 % de la FCM .
- a) Pour vous entraîner en endurance, dans quel intervalle doit se situer, lors de l'effort, votre fréquence cardiaque?
- b) Pour s'entraîner en endurance, dans quel intervalle doit se situer, lors de l'effort, la fréquence cardiaque d'un sportif qui a une FCM de 185?
- 3) Dans la littérature, on trouve une formule appelée **formule d'Astrand** qui relie la FCM à l'âge :

$$FCM = 220 - \hat{age}$$

- a) Ajouter cette autre droite sur le graphe.
- b) Utiliser la formule d'Astrand pour prédire **votre** FCM .

- c) Le calcul de la droite de régression a-t-il donné le même genre de relation ?
- d) Pour un coureur de 105 ans, quelle est la *FCM* prédite ? Que pensez-vous de cette prédiction ?
- 4) La régression linéaire de l'âge par rapport à la fréquence cardiaque a-t-elle un sens biologique ?

Exercice 2

On rappelle que l'indice de Quételet, appelé encore "indice de masse corporelle" (IMC) est donné par : $IMC = \frac{poids}{(taille)^2}$ où la taille est exprimée en mètre et le poids en kilogrammes.

Il est défini comme le standard pour évaluer les risques liés au surpoids chez l'adulte. (18 à 65 ans). Attention ! Il ne prend pas en compte la proportion de masse musculaire ni de masse osseuse. Il est donc inadapté sur certaines populations et en particulier les sportifs qui se retrouvent alors très souvent en surpoids alors que leur forme physique est souvent meilleure que la moyenne des individus (http://fr.wikipedia.org/wiki/Indice_de_masse_corporelle).

- 1) On va utiliser les données qui se trouvent dans le fichier `L3APA06.txt`, c'est-à-dire les réponses au questionnaire des étudiants de la licence "Activités Physiques Adaptées", promotion 2006.
- a) Calculer l'IMC des étudiants et étudier cette variable.
- b) Tracer l'IMC en fonction de la taille de chaque étudiant. Est-il pertinent de faire une régression de l'IMC par rapport à la taille ?
- c) Tracer l'IMC en fonction du poids de chaque étudiant. Que pouvez-vous remarquer ?
- d) D'après la droite de régression de l'IMC par rapport au poids, quel IMC devriez-vous avoir pour votre poids ? Qu'est-ce que cela signifie ?
- 2) Considérons maintenant un échantillon très particulier, réunissant les médaillés (or) masculins français lors des Jeux Olympiques de Sydney en 2000. Pour les épreuves par équipe, on a retenu un seul médaillé, le premier dans l'ordre alphabétique.

Nom	Taille	Poids	Spécialité
Andrieux	192	97	Aviron
Asloum	165	63	Boxe
Bette	186	70	Aviron
Douillet	196	125	Judo
Dumoulin	171	64	Tir
Estanguet	182	75	Canoë-Kayak
Ferrari	187	83	Escrime
Gané	176	79	Cyclisme sur Piste
Martinez	164	50	VTT
Rousseau	182	85	Cyclisme sur Piste

- Calculer l'IMC de ces champions olympiques.
- Calculer la droite de régression de l'IMC par rapport au poids de ces sportifs. Cette droite de régression est-elle pertinente ?
- Quel est le point le moins bien représenté par la droite de régression ?

Exercice 3

Attention !!!
Corrélation n'est pas Causalité.

Il existe de nombreuses situations où, malgré un coefficient de corrélation élevé entre 2 variables quantitatives, on ne peut pas établir de relation de cause à effet entre elles. L'exemple suivant illustre ces propos.

On a relevé la production de bière alsacienne et le nombre de licences sportives des fédérations françaises entre 1960 et 1995.

<i>Années</i>	<i>Nombre de licences (en milliers)</i>	<i>Production de bière en Belgique (en millions d'hectolitres)</i>
1960	1640	3.3
1960	1640	3.3
1965	2220	4.1
1970	3240	5.6
1975	4620	8.0
1980	6300	9.6
1985	8340	10.2
1990	8980	11.3
1995	9210	11.2

Calculer la corrélation entre les variables *Licences* et *Bières*. Peut-on en déduire que la pratique du sport conduit à boire de la bière ?

Annexe : éléments d'interprétation de la fréquence cardiaque

Afin d'interpréter judicieusement l'exercice 1, nous donnons quelques extraits d'un document rédigé par Michel LESBATS (<http://hse.iut.u-bordeaux1.fr/lesbats/physiosite/physiowhtm/PhyW3.htm>).

...A partir de 40 ans la fréquence cardiaque de repos s'accroît et le débit cardiaque diminue (baisse d'efficacité du travail cardiaque). Pour une même contrainte musculaire, le coeur bat d'autant plus vite que le sujet est plus âgé. Or la fréquence cardiaque maximale décroît avec l'âge (195 Rc/mn à 20 ans, 185 Rc/mn à 30 ans, 165 Rc/mn à 60 ans).

La conjonction de ces divers phénomènes entraîne une sérieuse limitation de la puissance maximale fournie...

...Les individus âgés ont une fréquence cardiaque de repos (f_{crA}) plus élevée que celle d'un individu jeune (f_{crJ}). Les individus âgés ont une fréquence cardiaque maximale (f_{cmA}) plus faible que celle d'un individu jeune (f_{cmJ}).