Exercices récapitulatifs :Mesures d'association : *corrigé*

Le garage L. Harnack dispose du tableau suivant qui résume l'état des ventes de voitures l'an dernier en fonction de leur prix de vente et de leur cylindrée.

Tableau : répartition des ventes de voitures
(nombre de voitures vendues)

		Prix (10^3€)			
		]6, 10]	**]10, 20]**	**]20, 30]**	**Total**
Cyl. (10^2 cm^3)	**]9, 15]**	**35**	**10**	**5**	**50**
	]15, 19]	**10**	**60**	**20**	**90**
	]19, 21]	**0**	**5**	**25**	**30**
	Total	**45**	**75**	**50**	**170**

On vous demande de :

1. *Calculer toutes les mesures d'association spécifiques aux variables quantitatives.*

Pour ce faire, il faut d'abord disposer des *centres de classe* des deux variables :

- pour le prix : 8, 15, 25 (10^3€)
- pour la cylindrée : 12, 17, 20 (10^2 cm^3)

On calcule alors les *moyennes marginales* des deux variables :

Le prix moyen sera calculé comme :

$$\overline{\text{px}} = [(45*8)+(75*15)+(50*25)]/170 = 16{,}0882 \approx 16\ 10^3\text{Euros}$$

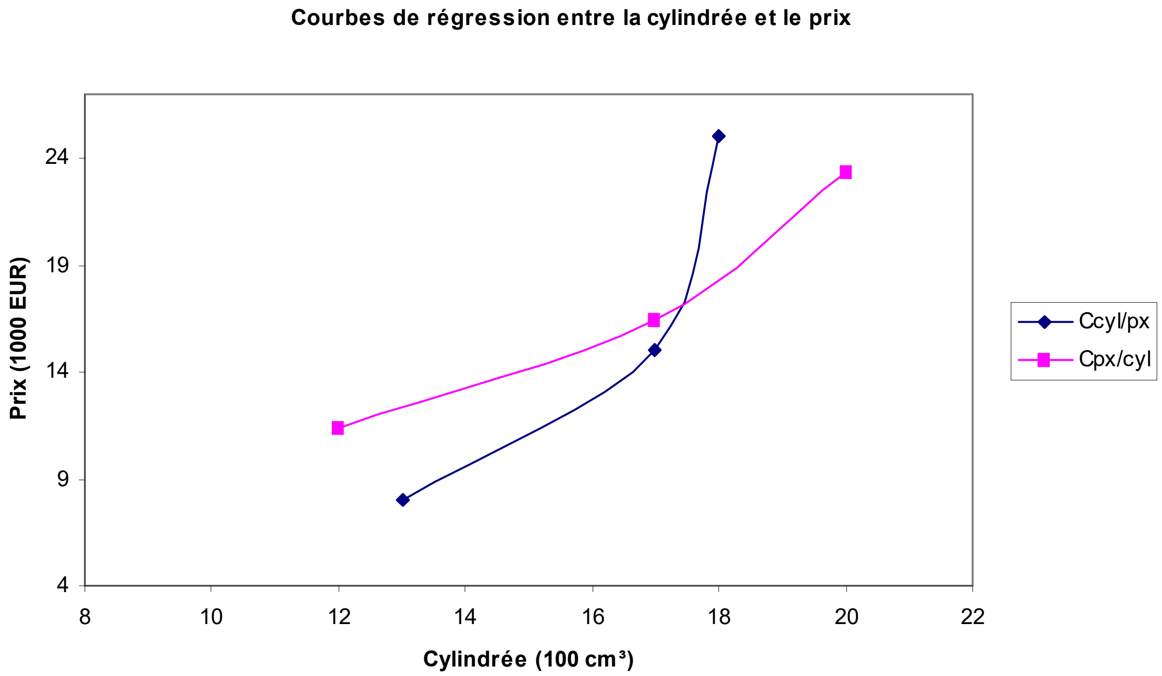
$$\overline{\text{cyl}} = [(50*12)+(90*17)+(30*20)]/170 = 16{,}0588 \approx 16\ 10^2\text{cm}^3$$

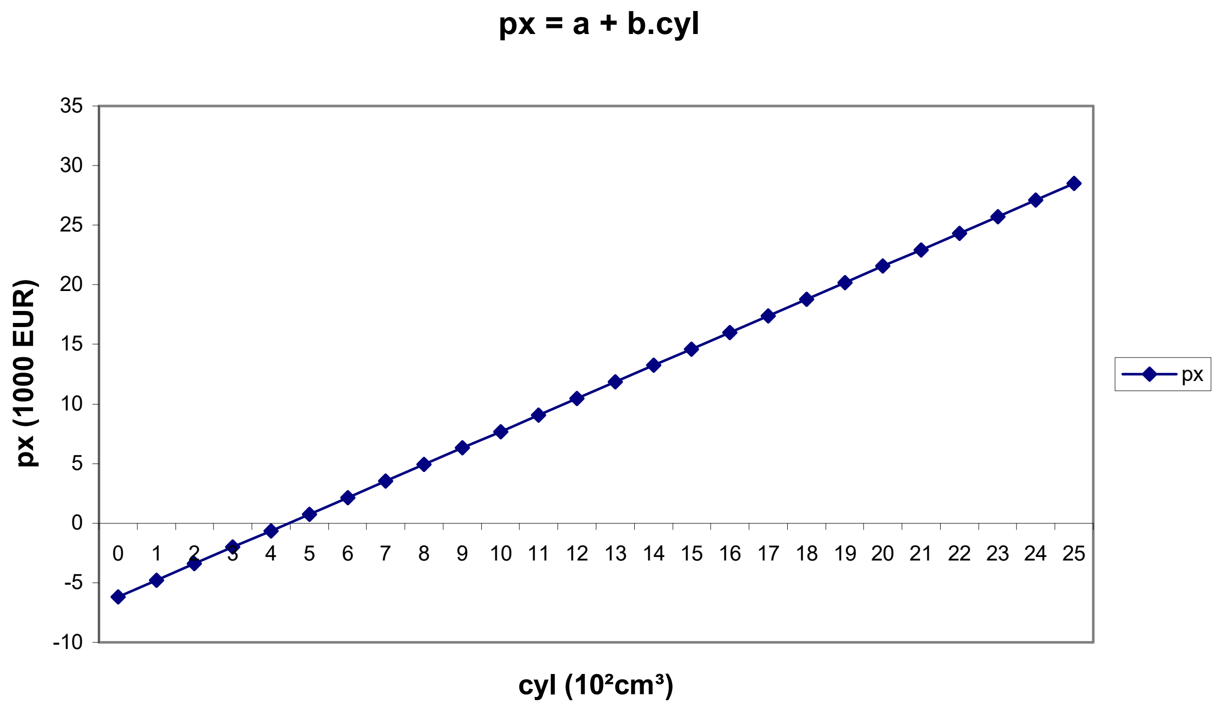
donc le *centre de gravité* G = (16,16)

Ce qui permet de calculer les écarts centrés (ec) et quadratiques (eq) des centres de classe reportés dans le tableau suivant :

(N.B. On prendra pour raisons pédagogiques les moyennes arrondies plutôt que celles calculées avec leurs décimales, ceci facilite la présentation des écarts. Il va de soi qu'un calcul professionnel DOIT conserver les décimales tout au long des calculs.)

Exercices récapitulatifs : mesures d’association :corrigés


Courbes de régression entre la cylindrée et le prix
Prix (1000 EUR)
24
19
14
9
4
8
10
12
14
16
18
20
22
Cylindrée (100 cm³)
Ccyl/px
Cpx/cyl


px = a + b.cyl
px (1000 EUR)
35
30
25
20
15
10
5
0
-5
-10
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
cyl (10²cm³)
px

Exercices récapitulatifs : mesures d’association :corrigés

Exercices récapitulatifs : mesures d’association :corrigés

				Prix			
			eq	64	1	81	
			ec	-8	-1	9	
	eq	ec		**8**	**15**	**25**	Tot.
Cyl.	16	-4	**12**	35	10	5	50
	1	1	**17**	10	60	20	90
	16	4	**20**	0	5	25	30
			Tot.	45	75	50	170

Il est donc possible de calculer la *variance marginale* de chaque variable ainsi que la *covariance* entre elles :

$$\sigma^2_{px} = [(45*64)+(75*1)+(50*81)]/170 = 41{,}2059 \text{ Mio Euros}^2$$

$$\sigma^2_{cyl} = [(50*16)+(90*1)+(30*16)]/170 = 8{,}0588 \ (10^2\text{cm}^3)^2$$

Donc les *écart-types* des deux variables sont calculés :

$$\sigma_{px} = \sqrt{41{,}2059} = 6{,}4192 \quad 10^3 \text{ Euros}$$

$$\sigma_{cyl} = \sqrt{8{,}0588} = 2{,}8388 \ 10^2 \text{ cm}^3$$

et la covariance vaut :

$$\sigma_{cyl,px} = \{[(-8.-4).35]+[(-1.-4).10]+[(9.-4).5] + [(-8.1).10]+ \dots\} = 11{,}1765$$

On peut donc conclure que la liaison est positive entre les deux variables, mais nous ne savons encore rien de *l'intensité* de la relation entre elles.

Pour calculer l'intensité de la liaison entre les variables, nous disposons du *coefficient de corrélation* :

$$\rho = \frac{\sigma_{cyl,px}}{\sigma_{px}.\sigma_{cyl}} = \frac{11{,}1765}{6{,}4192.\,2{,}8388} = \frac{11{,}1765}{18{,}2228} = 0{,}6133$$

Le coefficient de corrélation étant *normé* entre –1 et +1, cette valeur de 0,6133 indique une liaison positive significative mais pas complète entre les deux variables

2. *A partir des valeurs calculées pour ces mesures d'association de répondre aux deux questions suivantes :*
   a. *La cylindrée d'une voiture vendue est-elle un bon prédicteur du prix de cette même voiture ?*
   b. *Le prix d'une voiture vendue est-il un bon prédicteur de la cylindrée de cette même voiture ?*

Nous cherchons des prédicteurs, c'est-à-dire des variables qui, si nous connaissons leur valeur, pourraient nous aider à prévoir la valeur d'une autre variable. La *régression linéaire* peut nous aider.

Pour la réponse à la question 2.a., on va supposer le *modèle linéaire* suivant :

$$\text{px} = \text{a} + \text{b cyl}$$

et estimer les valeurs de a , *l'intercept* de la droite de régression et b, la *pente* de cette même droite.

$$\text{b} = \frac{\sigma_{\text{cyl,px}}}{\sigma^2_{\text{cyl}}} = \frac{11{,}1765}{8{,}0588} = 1{,}3869 \text{ et}$$

$$\text{a} = \overline{\text{px}} - \text{b}\,.\,\overline{\text{cyl}} = 16{,}088 - 1{,}3869\,.\,16{,}0588 = -\,6{,}1838$$

ainsi que son *coefficient de détermination* :

$$\rho^2 = (0{,}6133)^2 = 0{,}3761$$

Pour la réponse à la question 2.b., on va supposer le *modèle linéaire* suivant :

$$\text{cyl} = \text{c} + \text{d px}$$

et estimer les valeurs de c , *l'intercept* de la droite de régression et d, la *pente* de cette même droite.

$$\text{d} = \frac{\sigma_{\text{cyl,px}}}{\sigma^2_{\text{px}}} = \frac{11{,}1765}{41{,}2059} = 0{,}2712 \text{ et}$$

$$\text{c} = \overline{\text{cyl}} - \text{b}\,.\,\overline{\text{px}} = 16{,}0588 - 0{,}2712\,.\,16{,}088 = 11{,}6957$$

ainsi que son *coefficient de détermination* :

$$\rho^2 = (0{,}6133)^2 = 0{,}3761$$

Le coefficient de détermination nous indique qu'un peu plus du tiers de la variance de la variable dépendante de chacune des régression est « expliquée » par la relation linéaire la liant à l'autre variable, la variable « explicative ».

Il existe donc d'autres explications à trouver dans la théorie économique, dans l'observation, etc.

La cylindrée et le prix sont bien des prédicteurs de l'autre variable mais loins d'être parfaits.

Interprétation de a, b, c, d :

- a est l'INTERCEPT de la régression 2.a., il signifie que le prix moyen d'une voiture, indépendamment de sa cylindrée est de -6,1838 10^3€, …

- b, la PENTE de la régression 2.a., représente l'effet marginal de la cylindrée sur le prix d'une voiture, sa valeur de 1,3869 par unité de cylindrée ($10^2 cm^3$) nous dit qu'en moyenne, quand la cylindrée varie de 100 cm^3, le prix de la voiture varie dans le même sens de 1,3869 10^3€.

- c, INTERCEPT de la régression 2.b., vaut 11,6957. Cette valeur signifie qu'indépendamment du prix, la cylindrée moyenne des voitures vendues vaut 11,6957 $10^2 cm^3 \cong 1200\ cm^3$.

- d, la PENTE de la régression 2.b., représente l'effet marginal du prix sur la cylindrée d'une voiture, sa valeur de 0,2712 par unité de prix (10^3€) nous dit qu'en moyenne, quand le prix varie de 1000 €, la cylindrée de la voiture varie dans le même sens de 0,2712 $10^2 cm^3$.