

# CORRÉLATION-RÉGRESSION, COEFFICIENT DE CORRÉLATION

## Coefficient de corrélation

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{(n-1) s_x s_y}$$

$x_i$	$y_i$	$x_i y_i$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
42	106	4 452	68,0625	462,25
46	88	4 048	18,0625	12,25
47	97	4 559	10,5625	156,25
49	84	4 116	1,5625	0,25
52	85	4 420	3,0625	0,25
53	74	3 922	7,5625	110,25
55	78	4 290	22,5625	42,25
58	64	3 712	60,0625	420,25
<b><math>\Sigma</math> 402</b>	<b>676</b>	<b>33 519</b>	<b>191,50</b>	<b>1 204,00</b>

$\bar{x} = \frac{402}{8} = 50,25$      
 $\bar{y} = \frac{676}{8} = 84,5$      
 $s_x^2 = \frac{191,50}{7} \approx 27,36$      
 $s_y^2 = \frac{1\,204,00}{7} = 172,00$

$\Sigma x_i y_i = 33\,519$      
 $s_x \approx 5,23$      
 $s_y \approx 13,11$

### OBJECTIFS

Construire un tableau pour représenter les données de deux variables quantitatives.

Utiliser ce tableau pour calculer le coefficient de corrélation entre les variables.

### Mise en situation

Des amies décident de prendre chacune trois bières de 340 ml à 6,1 % en une heure et, une demi-heure plus tard, subir un test pour connaître leur taux d'alcool dans le sang. Le tableau suivant présente les valeurs obtenues

Prénom	Poids (kg)	Taux (mg/100 ml)
Anaïs	42	106
Béatrice	46	88
Claudine	47	97
Doris	49	84
Elvire	52	85
France	53	74
Guylaine	55	78
Hermance	58	64

Construire un tableau pour calculer le coefficient de corrélation.

#### TABLEAU ET CHRONOGRAMME

1. On personnalise une feuille Excel que l'on enregistre sous le nom « CorrélationLab ».
2. Dans la plage A10:C10, on écrit l'en-tête du tableau, puis on entre les données dans la plage A11:C18.
3. On sélectionne la plage A10:C18 et on indique au logiciel qu'il s'agit d'un tableau interactif.
4. Dans la plage B20:C20, on fait calculer la somme des valeurs des variables. Dans la plage B21:C21, on fait calculer la moyenne des variables et on leur donne un nom pour l'utiliser dans les calculs.
5. En D10, on écrit « Produit », Excel reconnaît que l'on souhaite ajouter une colonne au tableau. En D11, on définit le produit :

« =B11\*C11 ».

Excel incrémente automatiquement le produit dans les lignes du tableau.

6. On ajoute deux autres colonnes pour faire calculer les carrés des écarts à la moyenne,

« =(B11-moyP)^2 » et « =(C11-moyT)^2 ».

Excel incrémente automatiquement le produit dans les lignes du tableau.

#### Remarque

À l'étape 4, les noms suggérés sont « moyP » pour la moyenne des poids et « moyT » pour la moyenne des taux d'alcoolémie.

7. On fait calculer la somme dans les colonnes ajoutées.
8. En E21, on fait calculer la variance corrigée de la variable poids et en F21 celle de la variable taux.
9. Dans la plage E22: F22, on fait calculer les écart-types corrigés des variables. On fait ensuite calculer le coefficient de corrélation.

### Remarque

La valeur absolue du coefficient est plus grande que 0,9, ce qui indique une très forte corrélation. Il est à noter que le logiciel conserve plus de décimales que ce qui est affiché pour effectuer ses calculs et la valeur obtenue peut être légèrement différente de celle que l'on obtiendrait avec une calculatrice.

## EXERCICES

Dans une enquête sur les pratiques culturelles au Québec en 2014, le Bulletin de la recherche et de la statistique donne les tableaux suivants. À l'aide du logiciel Excel, construire un tableau pour calculer le coefficient de corrélation pour chacune des versions.

- a) Taux de lecture régulière (au moins une fois par semaine) de journaux quotidiens.

### Taux de lecture régulière de journaux quotidiens

Âge	Papier (%)	Numérique (%)
[15; 24[	52	46
[15; 24[	52	56
[15; 24[	48	55
[15; 24[	56	45
[15; 24[	64	40
[15; 24[	69	32
[15; 24[	77	19

Source : Bulletin de la recherche et de la statistique. Enquête sur les pratiques culturelles au Québec, 2014.

- b) Nombre moyen de visionnement de contenu télévisuel par semaine à la télévision ou sur internet.

### Heures de visionnement de contenu télévisuel

Âge	Télévision	Internet
[15; 24[	8,4	15,2
[15; 24[	11,5	15,2
[15; 24[	11,1	14,0
[15; 24[	14,2	16,1
[15; 24[	17,7	19,4
[15; 24[	21,9	23,1
[15; 24[	24,6	25,7

Source : Bulletin de la recherche et de la statistique. Enquête sur les pratiques culturelles au Québec, 2014.

- c) Taux d'écoute régulière (au moins une fois par semaine) de films, au cinéma ou ailleurs.

### Écoute de films, cinéma et internet

Âge	Cinéma (%)	Internet (%)
[15; 24[	92	51
[15; 24[	81	60
[15; 24[	79	51
[15; 24[	69	55
[15; 24[	62	59
[15; 24[	55	50
[15; 24[	38	40

Source : Bulletin de la recherche et de la statistique. Enquête sur les pratiques culturelles au Québec, 2014.

- d) Taux d'écoute régulière de musique (au moins une fois par semaine).

### Écoute de musique

Âge	Écoute (%)
[15; 24[	98
[15; 24[	97
[15; 24[	97
[15; 24[	93
[15; 24[	90
[15; 24[	84
[15; 24[	85

Source : Bulletin de la recherche et de la statistique. Enquête sur les pratiques culturelles au Québec, 2014.

- e) Nombre d'heures moyen d'utilisation d'internet à des fins personnelles.

### Écoute de musique

Âge	Écoute (%)
[15; 24[	4,1
[15; 24[	3,0
[15; 24[	2,1
[15; 24[	1,9
[15; 24[	1,7
[15; 24[	1,5
[15; 24[	1,0

Source : Bulletin de la recherche et de la statistique. Enquête sur les pratiques culturelles au Québec, 2014.