Notions de statistiques, 3ème édition - Christiane Simard
Section 7.1 Mise en situation p.300

Femmes

Taux d'alcool dans le sang en fonction du poids après la consommation de trois bières

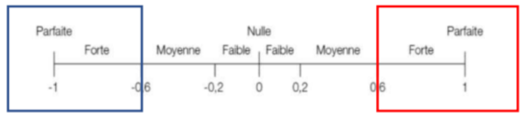
X: Poids	45 kg (100 lb)	52 kg (115 lb)	57 kg (125 lb)	68 kg (150 lb)	73 kg (161 lb)	80 kg (176 lb)	91 kg (200 lb)
Y: Taux d'alcool (en mg/100 ml)	152	133	120	101	99	87	76

Source: Éduc'alcool. *Boire, conduire, choisir. L'alcool au volant. L'alcool et la loi (0.08)*, 2014.

1. Tracez le diagramme de dispersion
2. Donner une interprétation de la relation entre le poids des femmes (X) et le taux d'alcool dans le sang (Y) à partir du diagramme de dispersion.
3. Calculer et interpréter la valeur du coefficient de corrélation (r).
4. Représenter graphiquement la droite de régression.
5. Déterminer les coefficients de la droite de régression.
6. Calculer et interpréter la valeur du coefficient de détermination r^2

Propriétés du coefficient de corrélation linéaire

- Sans unités
- Compris entre -1 et 1.
- Si la valeur de r proche de 1 (>0,6) : corrélation parfaite positive forte
- Si la valeur de r proche de -1 (<-0,6): corrélation parfaite négative forte



Coefficient de corrélation linéaire

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

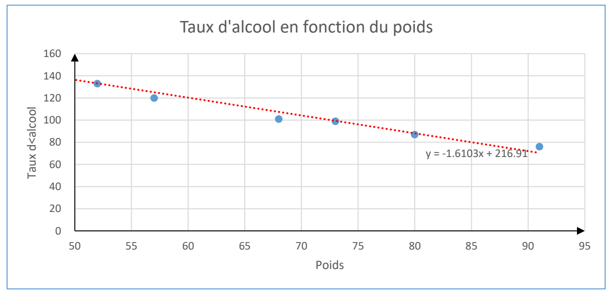
$\sum xy$ représente la somme des produits de chaque valeur de la variable X par la valeur correspondante de la variable Y

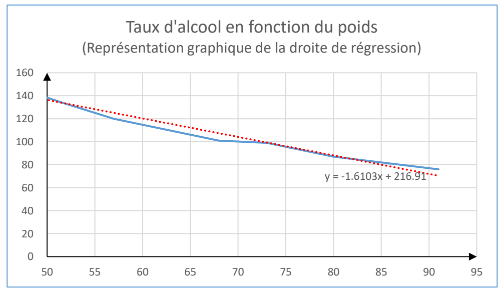
n correspond au nombre de couples (x, y)

- $\bar{x}$ représente la moyenne des valeurs de la variable X
- $\bar{y}$ représente la moyenne des valeurs de la variable Y
- s_x est l'écart type corrigé de la variable X
- s_y est l'écart type corrigé de la variable Y

Poids des femmes en kg (X) Variable indépendante	Taux d'alcool (Y) Variable dépendante	x * y	x̄	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	ȳ	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
45	152	6 840	66.6	-21.60	466.56	109.70	42.30	1789.29
52	133	6 916	66.6	-14.60	213.16	109.70	23.30	542.89
57	120	6 840	66.6	-9.60	92.16	109.70	10.30	106.09
68	101	6 868	66.6	1.40	1.96	109.70	-8.70	75.69
73	99	7 227	66.6	6.40	40.96	109.70	-10.70	114.49
80	87	6 960	66.6	13.40	179.56	109.70	-22.70	515.29
91	76	6 916	66.6	24.40	595.36	109.70	-33.70	1135.69
	∑xy =	48 567			1589.7			4279.4

1. Diagramme de dispersion





2. Interprétation Lorsque le poids des femmes augmente, le taux d'alcool diminue.

3. Coefficient de corrélation (r)

n	7	
x̄	66.57	66.57 Utiliser la fonction MOYENNE (AVERAGE en anglais) pour le calcul de la moyenne
ȳ	109.71	109.71 Utiliser la fonction MOYENNE (AVERAGE en anglais) pour le calcul de la moyenne

s_x : $s_x = \sqrt{\frac{\sum(xi - x)^2}{n-1}} = \sqrt{\frac{1\,589.7}{6}}$ $s_x =$ 16.28 16.28 Utiliser la fonction ECART.TYPE STANDARD (STDEV.S en anglais) pour le calcul de l'écart type.

s_y : $s_y = \sqrt{\frac{\sum(yi - y)^2}{n-1}} = \sqrt{\frac{4\,279.4}{6}}$ $s_y =$ 26.71 26.71

∑xy - nx̄ȳ -2,559.86
(n - 1)S_xS_y 2608.27

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{(n-1)S_xS_y} = \frac{-2\,559.86}{2\,608.27} =$$

r = -0.981

Garder 2 décimales pour la valeur du coefficient de corrélation.

r = -0.981 Utiliser la fonction COEFFICIENT.CORRELATION (CORREL en anglais) pour le calcul de l'écart type.

$r = \frac{\sum xy - n\bar{x}\bar{y}}{(n-1)s_xs_y}$

r = -0,981

<u>Interprétation:</u>
Il y a une très faible corrélation linéaire négative entre le poids d'une femme son taux d'alcool dans le sang.

Cela indique que plus le poids de la femme est bas, plus le taux d'alcool dans le sang est élevé.

Équation de la droite

(n - 1)S_x 1589.72

$a = \bar{y} - b\bar{x} \Leftrightarrow a = -b\bar{x} + \bar{y}$

b (pente): $b = \frac{\sum xy - n\bar{x}\bar{y}}{(n-1)s_x^2} = \frac{-2\,559.86}{1\,589.72}$ b = -1.61 -1.61 Avec la fonction PENTE (SLOPE en anglais)

a: $a = \bar{y} - b\bar{x}$ a = 216.91 217 Avec la fonction ORDONNEE (INTERCEPT en anglais)

d'où l'équation $\boldsymbol{y = -1.61x + 216.91}$