Exercice 1 On veut étudier la relation entre la distance de freinage (D en m) nécessaire à une automobile circulant sur une route humide pour s'arrêter et sa vitesse (V en km/h). On a relevé pour différentes vitesses v_i la distance d_i associée et les résultats sont synthétisés comme suit :

											Total
v_i (en km/h)	40	50	60	70	80	90	100	110	120	130	**850**
d_i (en mètres)	18	35	60	80	100	120	148	173	193	200	**1127**
v_i^2	1600	2500	3600	4900	6400	8100	10000	12100	14400	16900	**80500**
d_i^2	324	1225	3600	6400	10000	14400	21904	29929	37249	40000	**165031**
$v_i \cdot d_i$	720	1750	3600	5600	8000	10800	14800	19030	23160	26000	**113460**

1. Calculer $\overline{V}$ et $\overline{D}$ les moyennes observées ainsi que les valeurs des variances σ_V^2 et σ_D^2.
2. Donner la valeur du coefficient de corrélation linéaire $\rho(V, D)$. Est-ce qu'on peut écrire D de la forme $D = aV + b + \epsilon$?
3. Déterminer l'équation de la droite de régression linéaire.
4. Vérifier si les hypothèses formulées sur les résidus sont vérifiées et calculer le coefficient de détermination.
5. Estimer, la distance de freinage nécessaire pour s'arrêter pour une voiture roulant à 140 km/h.

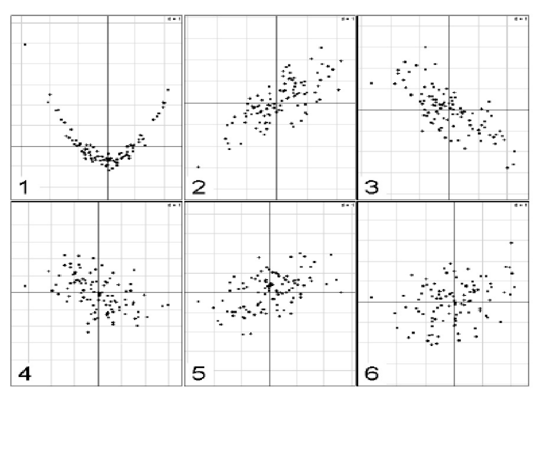
Exercice 2 Les données suivantes, sont les poids secs (moyens) d'un légume à différentes périodes de son développement (temps en semaine). Le but est de modéliser le poids secs en fonction de la durée de vie.

Poids secs (Y)	Durée de vie (X)
16.08	1
33.83	2
65.8	3
97.2	4
191.55	5
326.2	6
386.87	7
520.53	8
590.03	9
651.92	10

1. Effectuer le nuage de points de ces données. Peut-on penser, au vu de celui-ci, qu'une régression linéaire est bien adaptée à la question ?
2. Calculer le coefficient de corrélation linéaire $\rho(X, Y)$.
3. Déterminer la valeur da variable $Z = \ln\left(\left(\frac{Y}{700}\right)^{-1.28} - 1\right)$ puis calculer $\rho(X, Z)$. Interpréter.
4. Déterminer l'équation de la droite de régression de Z en X. A quelle courbe de régression cette droite correspond-elle pour les variables Y et X ? Tracer cette courbe sur le nuage de points de la première question.
5. Evaluer le poids sec de ce légume cueilli à 10.5 semaines.

Exercice 3

Six simulations de variables donnent les nuages de points suivants : Attribuer à chacune d'entre elles son coefficient de corrélation sachant que ces valeurs figurent parmi l'ensemble : $\{-1; -0.73; -0.49; -0.04; 0.33; 0.50; 0.74; 1\}$

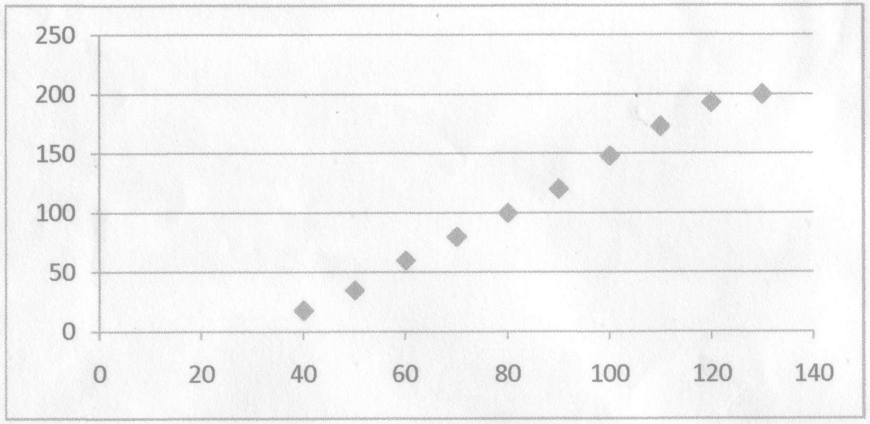


• Contact : imen.kammoun@gmail.com

Ex 1 :

d*v	d^2	v^2
720	324	1600
1750	1225	2500
3600	3600	3600
5600	6400	4900
8000	10000	6400
10800	14400	8100
14800	21904	10000
19030	29929	12100
23160	37249	14400
26000	40000	16900
113460	165031	80500

vi	di	(vi-moyv)^2	(di-moyd)^2	cov
40	18	2025	8968,09	4261,5
50	35	1225	6037,29	2719,5
60	60	625	2777,29	1317,5
70	80	225	1069,29	490,5
80	100	25	161,29	63,5
90	120	25	53,29	36,5
100	148	225	1246,09	529,5
110	173	625	3636,09	1507,5
120	193	1225	6448,09	2810,5
130	200	2025	7621,29	3928,5
850	1127	8250	38018,1	17665



$\boxed{\overline{V} = 85}$ $\boxed{\overline{D} = 112,7}$ $\sigma_V^2 = 825$ $\sigma_D^2 = 3801,81$ $\boxed{1766,5 = \mathrm{cov}(V, D)}$

$$\sigma_V^2 = \frac{1}{10}\sum_{i=1}^{10}(v_i - \overline{V})^2 = \frac{1}{10}\sum v_i^2 - \overline{V}^2$$

$$= 8050 - (85)^2$$

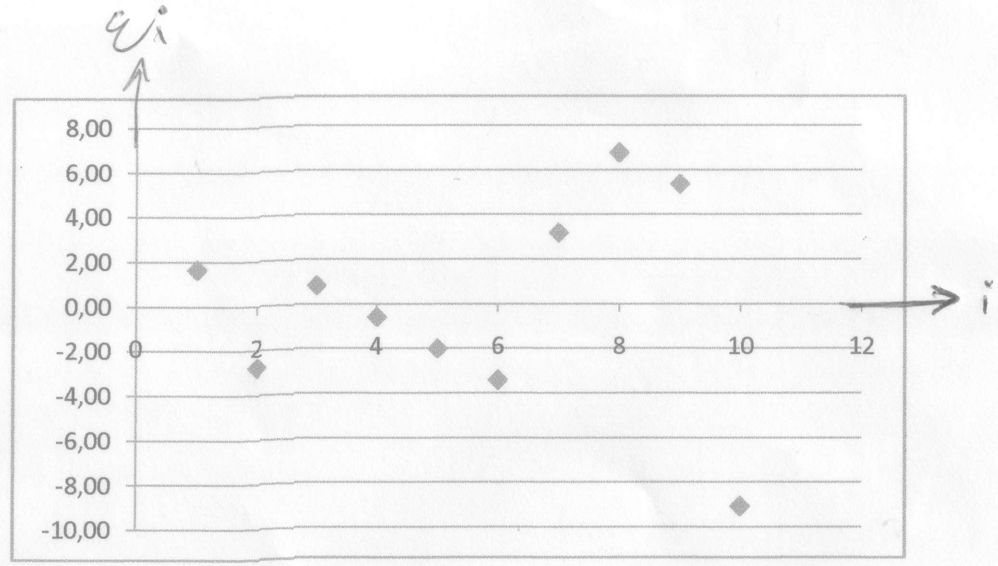
$$\sigma_D^2 = 16503,1 - (112,7)^2$$

$$\mathrm{Cov}(V, D) = \frac{1}{10}\sum(v_i - \overline{V})(d_i - \overline{d})$$

$$= \frac{1}{10}\sum v_i d_i - \overline{V}.\overline{d} = 11346 - 85.1127$$

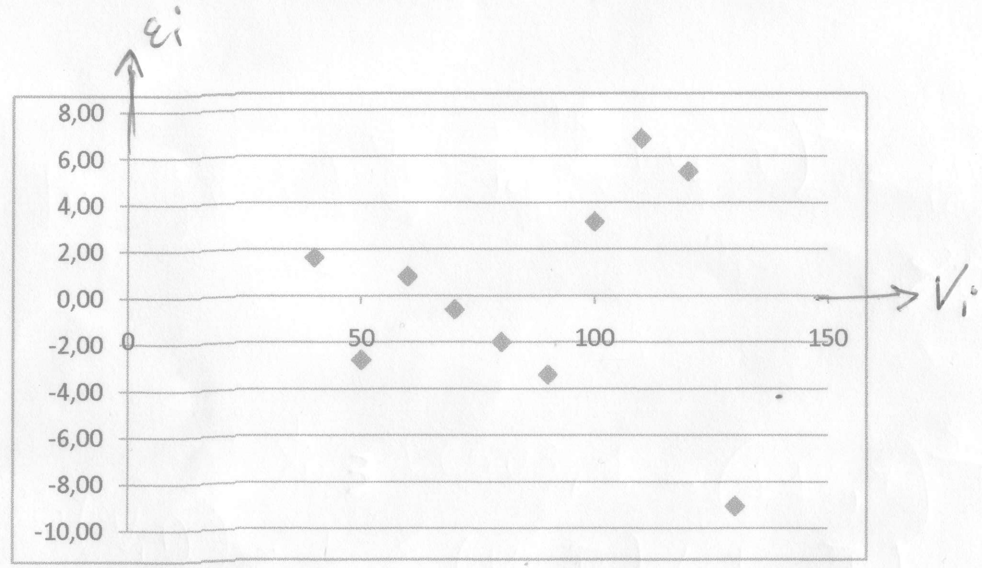
$\hat{D} = aV + b$ $\quad \varepsilon_i = D_i - \hat{D}_i$

dest	err	err^2
16,35	1,65	2,74
37,76	-2,76	7,60
59,17	0,83	0,69
80,58	-0,58	0,34
101,99	-1,99	3,98
123,41	-3,41	11,60
144,82	3,18	10,12
166,23	6,77	45,83
187,64	5,36	28,70
209,05	-9,05	81,98
	0,00	193,587879



2,14 a	R2	rho
-69,30 b	0,99490801	0,99745075

$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{193,58}{38018,1}$



$a = \frac{cov(D,V)}{\sigma_V^2}$

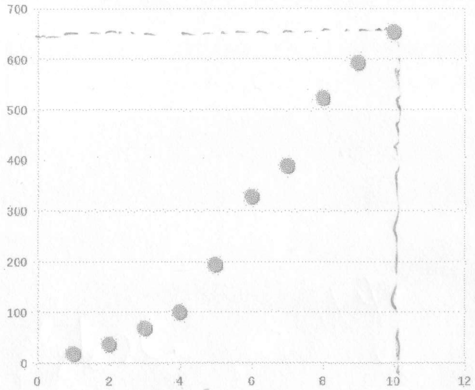
$\rho(V,D) = \frac{cov(V,D)}{\sigma_V \cdot \sigma_D} = \frac{1766,5}{\sqrt{325 \cdot 3801,81}}$

$b = \bar{D} - a\bar{V}$

$= 112,7 - 2,14 \cdot 85$

Ex2 :

1. Effectuer le nuage de points de ces données. Peut-on penser, au vu de celui-ci, qu'une régression linéaire est bien adaptée à la question ?
Réponse:



La corrélation linéaire entre X et Y semble bien adaptée à la question de la prédiction du poids sec en fonction de la durée de vie.

2. Calculer $r(x, y)$.
Réponse:

$1^{\text{ère}}$ étape : Calcul des moyennes $\bar{x}$ et $\bar{y}$
Le poids sec moyen est

$$\bar{y} = \frac{16,08 + 33,83 + \ldots + 651,92}{10} = \frac{2880}{10} = \boxed{288. = \bar{Y}}$$

La durée de vie moyenne est

$$\bar{x} = \frac{1 + 2 + \ldots + 10}{10} = \frac{55}{10} = \boxed{5,5 \text{ semaines.} = \bar{X}}$$

$2^{\text{ème}}$ étape : Calcul des écarts types $\sigma(x)$ et $\sigma(y)$
La variance du poids sec est

$$\text{Var}(y) = \frac{16,08^2 + 33,83^2 + \ldots + 651,92^2}{10} - 288^2 \simeq \boxed{52\,259 = \sigma_Y^2}$$

et donc $\sigma(y) = \sqrt{52\,259} \simeq 228,6$. La durée de vie moyenne est

$$\text{Var}(x) = \frac{1^2 + 2^2 + \ldots + 10^2}{10} - 5,5^2 = \frac{55}{10} = 8,25 \boxed{= \sigma_X^2}$$

et donc $\sigma(x) = \sqrt{8,25} \simeq 2,87$.
$3^{\text{ème}}$ étape : Calcul de la covariance Cov(x, y)

$$\text{Cov}(x, y) = \frac{16,08 \times 1 + 33,83 \times 2 + \ldots + 651,92 \times 10}{10} - 288 \times 5,5 = \frac{22\,286}{10} - 288 \times 5,5 = \boxed{644,66. = \text{COV}}$$

$4^{\text{ème}}$ étape : Calcul de $r(x, y)$
On en déduit

$$r(x, y) = \frac{\cancel{22286}\ 644,66}{2,87 \times 228,6} \simeq 0,982.$$

3. Déterminer les valeurs de la variable $Y' = \ln\left(\left(\frac{Y}{700}\right)^{-1,28} - 1\right)$ puis calculer $r(x, y')$. Conclusion ?
Réponse:

Les valeurs de la variable Y' sont données par

Poids secs	Durée de vie	Z
16,08	1	4,82
33,83	2	3,86
65,8	3	2,98
97,2	4	2,44
191,55	5	1,45
326,2	6	0,51
386,87	7	0,13
520,53	8	-0,77
590,03	9	-1,41
651,92	10	-2,35

$1^{\text{ère}}$ étape : Calcul de la moyenne $\bar{y'}$

$$\bar{y'} = \frac{4,82 + 3,86 + \ldots - 2,35}{10} = \frac{11,66}{10} = 1,166. = \bar{Z}$$

$2^{\text{ème}}$ étape : Calcul de l'écart type $\sigma(y')$

$$\text{Var}(y') = \frac{4,82^2 + 3,86^2 + \ldots + (-2,35)^2}{10} - 1,166^2 \simeq 4,99 \quad = \sigma_Z^2$$

et donc $\sigma(y) = \sqrt{4,99} \simeq 2,23$.
$3^{\text{ème}}$ étape : Calcul de la covariance Cov(x, y')

$$\text{Cov}(x, y') = \frac{4,82 \times 1 + 3,86 \times 2 + \ldots - 2,35 \times 10}{10} - 1,17 \times 5,5 = \frac{0,11}{10} - 2,35 \times 5,5 = -6,4. = \text{cov}(Z, X)$$

$4^{\text{ème}}$ étape : Calcul de $r(x, y)$
On en déduit

$$r(x, y') = \frac{-6,4}{2,87 \times 2,23} \simeq -0,998.$$

La corrélation linéaire de $\cancel{Y'}$ Z en X est encore meilleure que la corrélation linéaire de Y en X

4. Déterminer l'équation de la droite de régression de Y' en X. À quelle courbe de régression cette droite correspond-elle pour les variables Y et X ? Tracez cette courbe sur le nuage de points de la première question.
Réponse:

La droite de regression de Y' en X a pour équation $\overset{Z}{\cancel{y'}} = ax + b$ avec

$$a = \frac{-6,4}{8,25} = -0,78 \text{ et } b = 1,166 - (-0,78) \times 5,5 = 5,43.$$

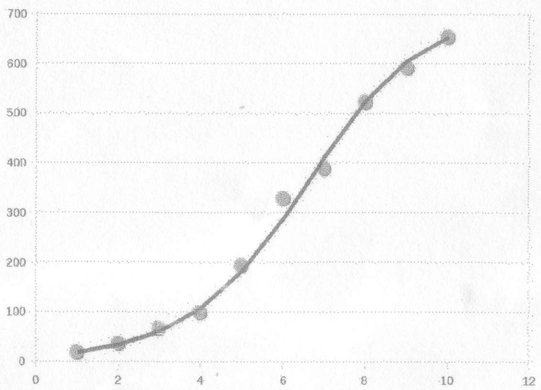
Ainsi, en remplaçant Y' par sa valeur en fonction de Y, on obtient la relation suivante entre X et Y :

$$\ln\left(\left(\frac{y}{700}\right)^{-1,28} - 1\right) = -0,78x + 5,43 \Leftrightarrow \left(\frac{y}{700}\right)^{-1,28} - 1 = e^{-0,78x+5,43}$$
$$\Leftrightarrow \left(\frac{y}{700}\right)^{-1,28} = e^{-0,78x+5,43} + 1$$
$$\Leftrightarrow \frac{y}{700} = \left(e^{-0,78x+5,43} + 1\right)^{-1/1,28}$$
$$\Leftrightarrow y = 700\left(e^{-0,78x+5,43} + 1\right)^{-1/1,28}$$

y en f° de x

On calcule pour les ≠ x la valeur de y.

ce qui donne la figure ci-dessous :



(le graphe de la courbe de tendance est obtenu, par exemple, en calculant les valeurs prédites pour Y à partir des valeurs de X : 1, 2, ..., 10.)

5. Évaluer le poids d'un oignon sec cueilli à 10,5 semaines.

Réponse:

Le poids d'un oignon sec cueilli à 10,5 semaine est estimé à :

$$y = 700\left(e^{-0,78*10,5+5,43} + 1\right)^{-1/1,28} \simeq 667,23$$

La regression de Z % à X affine plus cette estimation

Exo3 :

1 -0,04

2 0,74

3 -0,73

4 -0,49

5 0,50

6 0,33