TARBES
IUT Occitanie
UNIVERSITÉ TOULOUSE III

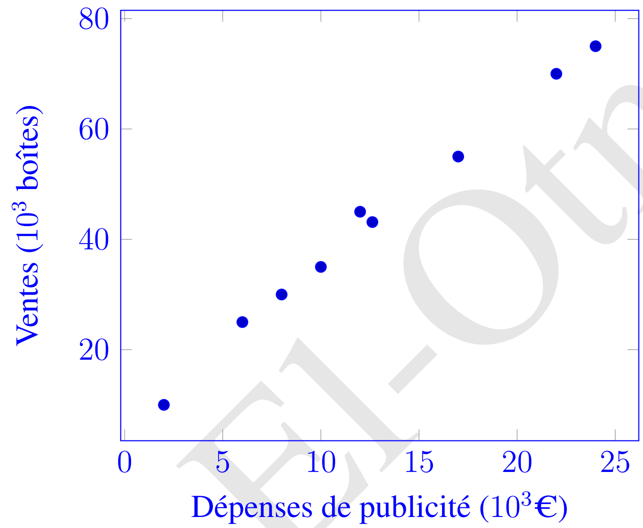
TECHNIQUES QUANTITATIVES ET REPRÉSENTATION
FEUILLE DE TRAVAUX DIRIGÉS N° 3
———-
Ajustement linéaire

BUT-Tech. de Co.
Semestre : 2
A.U. : 2021-2022
Prof. H. El-Otmany

Exercice n°1 Un chercheur universitaire souhaite mesurer l'influence des dépenses de publicité sur la vente des boîtes de conserve. Les informations sont récoltées dans le tableau ci-dessus :

Dépenses de publicité (10^3€)	2	6	8	10	12	17	22	24
Ventes (10^3 boîtes)	10	25	30	35	45	55	70	75

1. Réaliser le nuage de points en plaçant le point moyen.
   Soit X la variable désignant les dépenses de publicité (10^3€) et soit Y la variable représentant les ventes des boîtes de conserve (10^3 boîtes).



   Le point moyen de la série statistique (X, Y) est $(\overline{x}, \overline{y})$ tel que

$$\overline{x} = \frac{2+6+8+10+12+17+22+24}{8} = 12.625 \quad (\times 10^3),$$
$$\overline{y} = \frac{10+25+30+35+45+55+70+75}{8} = 43.125 \quad (\times 10^3).$$

2. Calculer le coefficient de corrélation linéaire. En déduire que l'ajustement linéaire se justifie dans cette situation.
   Le coefficient de corrélation est défini par $r(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$. Pour se faire, on calcule
   — la variance de X :

$$V(X) = \overline{X^2} - \left(\overline{X}\right)^2 = \frac{2^2+6^2+8^2+10^2+12^2+17^2+22^2+24^2}{8} - (12.625)^2$$
$$= 212.125 - (12.625)^2 \approx 52.734 \quad (\times 10^6).$$

   — la variance de Y :

$$V(Y) = \overline{Y^2} - \left(\overline{Y}\right)^2 = \frac{10^2+25^2+30^2+35^2+45^2+55^2+70^2+75^2}{8} - (43.125)^2$$
$$= 2303.125 - (43.125)^2 \approx 443.359 \quad (\times 10^6)$$

— la covariance de X et Y : $Cov(X,Y) = \overline{XY} - \overline{X} \times \overline{Y}$ telle que

$$\overline{XY} = \frac{(2\times 10)+(6\times 25)+(8\times 30)+(10\times 35)}{8}$$
$$\ldots + \frac{(12\times 45)+(17\times 55)+(22\times 70)+(24\times 75)}{8} = 696.875.$$
$$Cov(X,Y) = 696.875 - (12.625 \times 43.125) = 152.422 \quad (\times 10^6).$$

Par conséquent, le coefficient de corrélation linéaire vaut

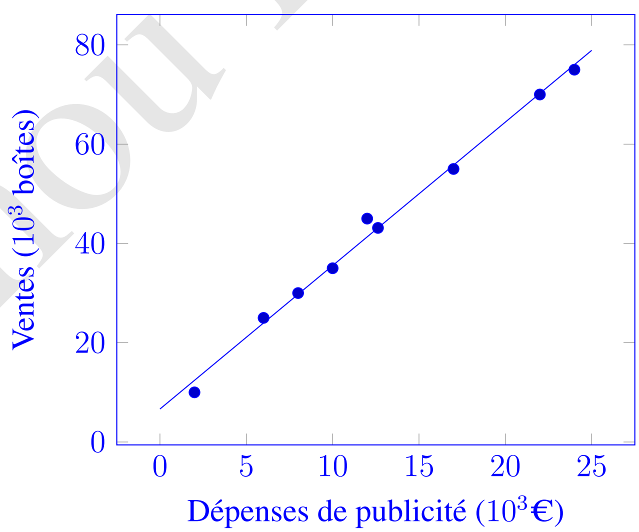
$$r(X,Y) = \frac{152.422\times 10^6}{\sqrt{52.734\times 10^6 \times 443.359 \times 10^6}} \approx 0.9968 = 99.68\%$$

Comme $r(X,Y) \approx 1$, alors il existe une relation linéaire forte positive entre la vente des boîtes de conserves et les dépenses de publicité (même sens de variation ; si les dépenses de publicité augmentent alors les ventes augmentent et vice versa). Autrement dit, les dépenses de publicité expliquent les ventes des boîtes de conserves à 99.68%.

3. Déterminer l'équation de la droite des moindres carrées.
   L'équation de la droite des moindres carrées (ou encore la droite de régression ou la droite d'ajustement linéaire) est donnée par la formule $y = ax + b$ où

$$a = \frac{Cov(X,Y)}{V(X)} = \frac{152.422\times 10^6}{52.734\times 10^6} \approx 2.89,$$
$$b = \overline{Y} - a\overline{X} = 43.125 - 2.89 \times 12.625 = 6.64.$$

   donc, la droite de régression est $y = 2.89x + 6.64 \quad (\times 10^3)$. Nous pouvons tracer cette droite sur le même graphique du nuage de points.

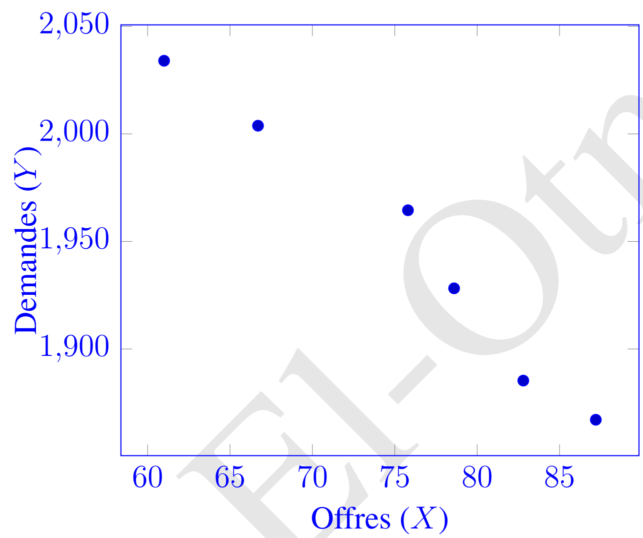


4. Grâce à cette équation, déterminer une estimation de ventes prévisible pour des dépenses de publicité de 20000€.
   En utilisant la droite des moindres carrées (ou encore la droite de régression ou la droite d'ajustement linéaire) $y = 2.89x + 6.64$, on pourra estimer (ou encore prévoir) la vente des boîtes de conserve pour une dépense de publicité $x = 20000$€. Comme les valeurs sont données à l'ordre de 10^3, soit ainsi $x = 20 \times 10^3$€. La vente prévisible des boîtes de conserves s'écrit ainsi $y_{|x=20} = 2.89 \times 20 + 6.64 = 64.44 \quad (10^3)$. Par conséquent, pour 20000€ de dépenses de publicité, on pourra vendre 64440 boîtes de conserves.

Exercice n°2 On donne pour les six derniers mois de l'année 2021 les nombres d'offres d'emploi (concernant les emplois durables et à plein temps) et le nombre des demandes d'emploi (déposées par des personnes sans emploi, immédiatement disponibles et à la recherche d'un emploi durable et à plein temps). Les données sont exprimées en milliers d'individus.

Offres (X)	61	66.7	75.8	78.6	82.8	87.2
Demandes (Y)	2034	2003.8	1964.5	1928.2	1885.3	1867.1

1. Représenter le nuage de points. Le nuage de points vous semble-t-il aligné le long d'une droite ?



On constate que l'ensemble des points sont alignés le long d'une droite affine d'équation $y = ax+b$ avec a et b sont des constantes à déterminer.

2. Calculer le coefficient de corrélation linéaire entre X et Y. Commenter.
Le coefficient de corrélation est défini par $r(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$. Pour se faire, on calcule

— la variance de X : $V(X) = \overline{X^2} - \left(\overline{X}\right)^2$ telle que

$$\overline{X} = \frac{61 + 66.7 + 75.8 + 78.6 + 82.8 + 87.2}{6} = 75.35,$$
$$\overline{X^2} = \frac{61^2 + 66.7^2 + 75.8^2 + 78.6^2 + 82.8^2 + 87.2^2}{6} \approx 5758.86,$$
$$V(X) = 5758.86 - (75.35)^2 \approx 81.24.$$

— la variance de Y : $V(Y) = \overline{Y^2} - \left(\overline{Y}\right)^2$ telle que

$$\overline{Y} = \frac{2034 + 2003.8 + 1964.5 + 1928.2 + 1885.3 + 1867.1}{6} = 1947.15,$$
$$\overline{Y^2} = \frac{2034^2 + 2003.8^2 + 1964.5^2 + 1928.2^2 + 1885.3^2 + 1867.1^2}{6} = 3795000.74,$$
$$V(Y) = 3795000.74 - (1947.15)^2 \approx 3607.62.$$

— la covariance de X et Y : $Cov(X,Y) = \overline{XY} - \overline{X} \times \overline{Y}$ telle que

$$\overline{XY} = \frac{(61 \times 2034) + (66.7 \times 2003.8) + (75.8 \times 1964.5)}{6}$$
$$... + \frac{(78.6 \times 1928.2) + (82.8 \times 1885.3) + (87.2 \times 1867.1)}{6} \approx 146184.51,$$
$$Cov(X,Y) = 146184.51 - (75.35 \times 1947.15) \approx -533.24.$$

Par conséquent, le coefficient de corrélation linéaire vaut

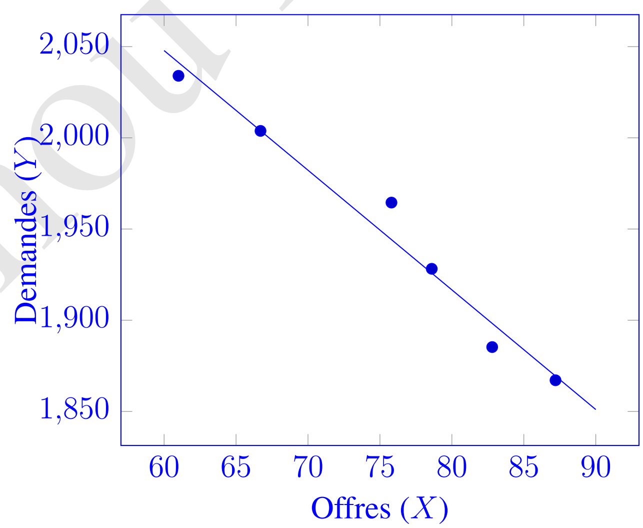
$$r(X,Y) = \frac{-533.24}{\sqrt{81.24 \times 3607.62}} \approx -0.9850 = -98.50\%$$

Comme $r(X,Y) \approx -1$, alors il existe une relation linéaire forte négative entre l'offre et la demande (différent sens de variation ; si l'offre augmente alors la demande diminue et vice versa). Autrement dit, l'offre explique la demande à 98.50% dans le sens inverse.

3. Trouver la droite de régression des demandes d'emploi en fonction des offres d'emploi et la tracer sur le graphique précédent. L'équation de la droite de régression des demandes Y en fonction de l'offre X est $y = ax + b$ où

$$a = \frac{Cov(X,Y)}{V(X)} = \frac{-533.24}{81.24} \approx -6.56,$$
$$b = \overline{Y} - a\overline{X} = 1947.15 - (-6.56) \times 75.35 = 2441.446.$$

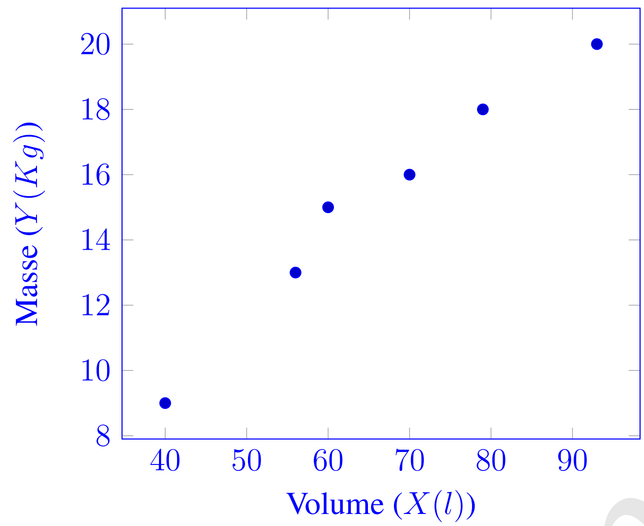
Par conséquent, la droite de régression des demandes en fonction des offres est $y = -6.56x + 2441.446$.



Exercice n°3 On souhaite étudier le lien entre le volume d'ingrédients et la masse d'un lot de fabrication.

Volume (l)	40	60	56	70	79	93
Masse (Kg)	9	15	13	16	18	20

1. Réaliser le nuage de points en plaçant le point moyen de cette série.



2. Calculer le coefficient de corrélation linéaire des variables X et Y. En déduire que l'ajustement linéaire se justifie dans cette situation.
Le coefficient de corrélation est défini par $r(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$. Pour se faire, on calcule

— la variance de X : $V(X) = \overline{X^2} - \left(\overline{X}\right)^2$ telle que

$$\begin{aligned}
\overline{X} &= \frac{40+60+56+70+79+93}{6} \approx 66.33, \\
\overline{X^2} &= \frac{40^2+60^2+56^2+70^2+79^2+93^2}{6} \approx 4687.67, \\
V(X) &= 4687.67 - (66.3)^2 \approx 287.99.
\end{aligned}$$

— la variance de Y : $V(Y) = \overline{Y^2} - \left(\overline{Y}\right)^2$ telle que

$$\begin{aligned}
\overline{Y} &= \frac{9+15+13+16+18+20}{6} \approx 15.17, \\
\overline{Y^2} &= \frac{9^2+15^2+13^2+16^2+18^2+20^2}{6} = 242.50, \\
V(Y) &= 242.50 - (15.17)^2 \approx 12.37.
\end{aligned}$$

— la covariance de X et Y : $Cov(X,Y) = \overline{XY} - \overline{X} \times \overline{Y}$ telle que

$$\begin{aligned}
\overline{XY} &= \frac{(40\times 9)+(60\times 15)+(56\times 13)+(70\times 16)+(79\times 18)+(93\times 20)}{6} = 1065, \\
Cov(X,Y) &= 1065 - (66.33 \times 15.17) \approx 58.77.
\end{aligned}$$

Par conséquent, le coefficient de corrélation linéaire vaut

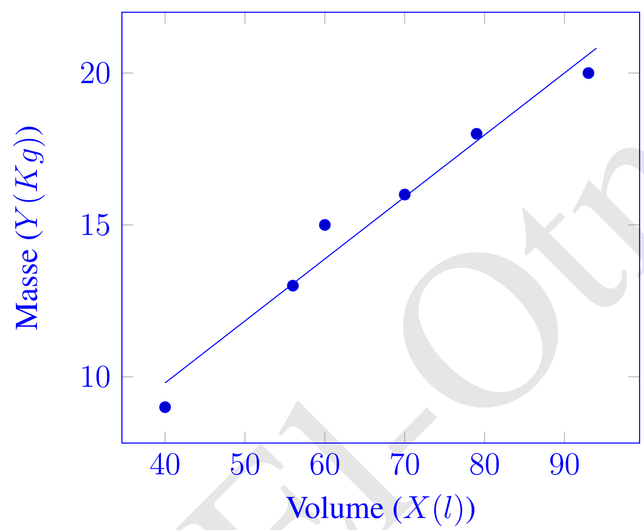
$$r(X,Y) = \frac{58.77}{\sqrt{287.99 \times 12.37}} \approx 0.9847 = 98.47\%$$

Comme $r(X,Y) \approx 1$, alors il existe une relation linéaire forte positive entre le volume d'ingrédients et la masse d'un lot de fabrication. (même sens de variation ; si le volume d'ingrédients (X) augmente alors la masse d'un lot de fabrication (Y) augment et vice versa). Autrement dit, le volume d'ingrédients (X) explique la masse d'un lot de fabrication (Y) à 98.47%.

3. Déterminer l'équation de la droite des moindres carrées. L'équation de la droite de régression de la masse d'un lot de fabrication Y en fonction du volume d'ingrédients (X) est $y = ax + b$ où

$$a = \frac{Cov(X,Y)}{V(X)} = \frac{58.77}{287.99} \approx 0.2041,$$
$$b = \overline{Y} - a\overline{X} = 15.17 - 0.2041 \times 66.33 = 1.6341.$$

Par conséquent, la droite de régression des demandes en fonction des offres est $y = 0.2041x + 1.6341$.



4. Grâce à cette équation, déterminer une estimation de la masse prévisible pour un lot faisant 120 litres de volume. En utilisant la droite des moindres carrées (ou encore la droite de régression ou la droite d'ajustement linéaire) $y = 0.2041x + 1.6341$, on pourra estimer (ou encore prévoir) la masse d'un lot de fabrication pour un volume $x = 120$ litres. On a donc $y_{|x=120} = 0.2041 \times 120 + 1.6341 \approx 26.13Kg$ de lots de fabrication.

Exercice n°4 (Travail personnel) On s'intéresse à l'étude des cours du baril de pétrole à la fin de Covid-19 et l'invasion de l'Ukraine par la Russie, voir le tableau ci-dessous.

Mois	Juin	Août	Septembre	Octobre	Novembre	Décembre	Janvier	Février	Mars
N° du mois X	1	2	3	4	5	6	7	8	9
Cours Y(€)	37.28	36	71.59	80.17	83.68	69.80	97.92	88.93	129.44

1. Représenter les nuages de points (X, Y).
2. Déterminer l'équation de la droite de régression de Y en fonction de X.
3. Quel est le coefficient corrélation entre X et Y ? Commenter le résultat.
4. On désire modéliser l'évolution du cours du baril par une fonction exponentielle de la forme $Y = A \times B^X$.
   a) Déterminer les coefficients A et B.
   b) Avec ce modèle, quel cours peut-on prévoir pour le mois d'avril 2022.