

Exercices de Statistiques - Solutions

1. Ajustement d'une droite

- (a) D'après la fig. 1, les points sont sensiblement alignés.
(b) En appliquant la méthode matricielle, on obtient successivement :

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \quad \mathbf{X}^\top = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$$

$$\mathbf{A} = \mathbf{X}^\top \mathbf{X} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \quad \det(\mathbf{A}) = 6 \quad \mathbf{A}^{-1} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix}$$

$$\mathbf{c} = \mathbf{X}^\top \mathbf{y} = \begin{bmatrix} 11 \\ 25 \end{bmatrix} \quad \mathbf{b} = \mathbf{A}^{-1} \mathbf{c} = \begin{bmatrix} 2/3 \\ 3/2 \end{bmatrix}$$

D'où l'équation de la droite :

$$y = \frac{2}{3} + \frac{3}{2}x$$

- (c) La moyenne \bar{y} est égale à $11/3$, d'où :

$$SS_t = \left(2 - \frac{11}{3}\right)^2 + \left(4 - \frac{11}{3}\right)^2 + \left(5 - \frac{11}{3}\right)^2 = \frac{14}{3}$$

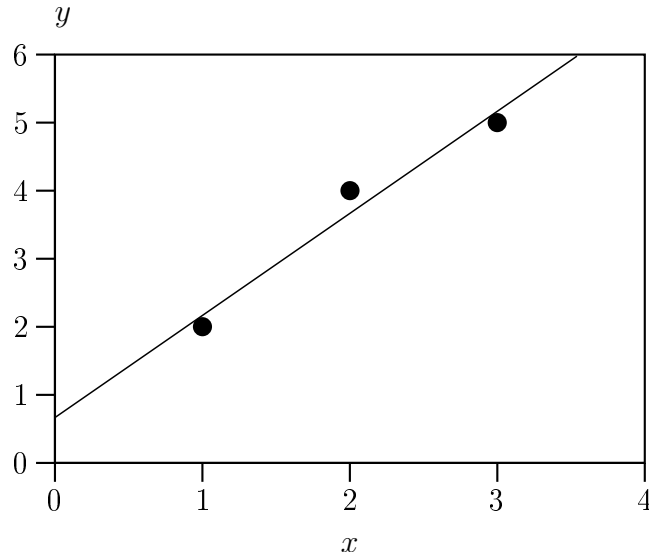
Calculons les valeurs estimées de y :

$$\hat{y}_1 = \frac{2}{3} + \frac{3}{2} \times 1 = \frac{13}{6}$$

On obtient de même :

$$\hat{y}_2 = \frac{11}{3} \quad \hat{y}_3 = \frac{31}{6}$$

FIGURE 1 – Ajustement d'une droite



D'où :

$$SS_e = \left(\frac{13}{6} - \frac{11}{3}\right)^2 + \left(\frac{11}{3} - \frac{11}{3}\right)^2 + \left(\frac{31}{6} - \frac{11}{3}\right)^2 = \frac{9}{2}$$

$$SS_r = \left(2 - \frac{13}{6}\right)^2 + \left(4 - \frac{11}{3}\right)^2 + \left(5 - \frac{31}{6}\right)^2 = \frac{1}{6}$$

On a bien :

$$SS_t = SS_e + SS_r$$

(d)

$$r^2 = \frac{SS_e}{SS_t} \approx 0,964$$

$$r = \text{sgn}(b_1) \cdot \sqrt{r^2} \approx 0,982$$

(e)

$$V_r = \frac{SS_r}{n-2} = \frac{1}{6} \quad s_r = \sqrt{V_r} \approx 0,41$$

$$V_e = \frac{SS_e}{2-1} = \frac{9}{2} \quad F = \frac{V_e}{V_r} = 27$$

(f) Calculons la matrice de variance-covariance :

$$\mathbf{V} = V_r \cdot \mathbf{A}^{-1} = \frac{1}{6} \times \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} = \begin{bmatrix} 7/18 & -1/6 \\ -1/6 & 1/12 \end{bmatrix}$$

D'où les écart-types des paramètres :

$$s_0 = \sqrt{\frac{7}{18}} \approx 0,62 \quad s_1 = \sqrt{\frac{1}{12}} \approx 0,29$$

(g) Les valeurs des paramètres sont arrondies au même nombre de décimales que les écart-types :

$$y = (0,67 \pm 0,62) + (1,50 \pm 0,29)x$$

$$n = 3 \quad r^2 = 0,964 \quad r = 0,982 \quad s_r = 0,41 \quad F = 27$$

(h)

$$r_{01} = \frac{-1/6}{\sqrt{(7/18) \times (1/12)}} \approx -0,926$$

On a une forte corrélation négative entre les 2 paramètres.

Cela signifie que si l'on surestime un paramètre, on a de fortes chances de sous-estimer l'autre, et réciproquement.

2. Dosage chromatographique

(a) Prenons 2 chiffres significatifs pour les écart-types :

$$y = (0,185 \pm 0,049) + (0,290 \pm 0,018)x$$

Soit l'hypothèse nulle : $\beta_0 = 0$. Sous cette hypothèse, la probabilité d'observer $|t| > 3,79$ est d'environ 3%. On peut donc rejeter cette hypothèse au risque 3%

Ce résultat est peu compatible avec ce que l'on attend d'une courbe d'étalonnage, à savoir qu'elle passe par l'origine.

(b) Avec les mêmes conventions que précédemment :

$$y = (0,051 \pm 0,057) + (0,433 \pm 0,054)x - (0,030 \pm 0,011)x^2$$

Soit l'hypothèse nulle : $\beta_0 = 0$. Sous cette hypothèse, la probabilité d'observer $|t| > 0,90$ est 0,46. On ne rejette donc pas cette hypothèse. Cela équivaut à admettre que la courbe théorique passe par l'origine.

Soit l'hypothèse nulle : $\beta_2 = 0$. Sous cette hypothèse, la probabilité d'observer $|t| > 2,70$ est 0,11. On peut rejeter cette hypothèse mais seulement au risque de 11%

- (c) Les 3 critères (s_r, r^2, F) sont en faveur du modèle polynomial. On peut donc préférer ce modèle, bien que le coefficient de x^2 ne soit pas très significatif.

3. Ajustement d'un modèle exponentiel

- (a) Ecrivons la forme linéarisée du modèle :

$$\ln y = \ln a - bx$$

d'où par identification :

$$\begin{aligned} \ln a^0 = 3,1445 &\quad \Rightarrow \quad a^0 = \exp(3,1445) \approx 23,208 \\ b^0 = 0,8047 &\end{aligned}$$

- (b) On a par dérivation :

$$y'_a = \exp(-bx) \quad y'_b = -ax \exp(-bx)$$

- (c) En appliquant les définitions du cours on obtient :

$$\mathbf{z} = \begin{bmatrix} y_1 - y_1^0 \\ y_2 - y_2^0 \\ y_3 - y_3^0 \end{bmatrix} = \begin{bmatrix} 10 - 23,208 \exp(-0,8047 \times 1) \\ 5 - 23,208 \exp(-0,8047 \times 2) \\ 2 - 23,208 \exp(-0,8047 \times 3) \end{bmatrix} \approx \begin{bmatrix} -0,3792 \\ 0,3582 \\ -0,07591 \end{bmatrix}$$

$$\mathbf{J} = \begin{bmatrix} y'_{a1} & y'_{b1} \\ y'_{a2} & y'_{b2} \\ y'_{a3} & y'_{b3} \end{bmatrix} = \begin{bmatrix} \exp(-0,8047 \times 1) & -1 \times 23,208 \exp(-0,8047 \times 1) \\ \exp(-0,8047 \times 2) & -2 \times 23,208 \exp(-0,8047 \times 2) \\ \exp(-0,8047 \times 3) & -3 \times 23,208 \exp(-0,8047 \times 3) \end{bmatrix}$$

$$\mathbf{J} \approx \begin{bmatrix} 0,4472 & -10,38 \\ 0,2000 & -9,284 \\ 0,08945 & -6,228 \end{bmatrix}$$

- (d) Les nouvelles estimations des paramètres sont telles que :

$$a = 23,208 - 2,11 = 21,1 \quad b = 0,8047 - 0,05931 = 0,7454$$

- (e) Il faudrait utiliser les nouvelles valeurs des paramètres comme estimations initiales et répéter les étapes (c) et (d) jusqu'à convergence des paramètres (à un niveau de précision donné).

4. Equation de Michaelis

- (a) Le graphe de v_0 en fonction de s_0 est une hyperbole présentant une asymptote horizontale à V_{max} (voir fig. 6).
(b) Quand $s_0 = K_m$ on a :

$$v_0 = \frac{V_{max}K_m}{2K_m} = \frac{V_{max}}{2}$$

- (c) On a :

$$\frac{1}{v_0} = \frac{K_m + s_0}{V_{max}s_0} = \frac{K_m}{V_{max}} \times \frac{1}{s_0} + \frac{1}{V_{max}}$$

L'équation est de la forme $y = b_0 + b_1x$ avec :

$$y = \frac{1}{v_0} \quad x = \frac{1}{s_0} \quad b_0 = \frac{1}{V_{max}} \quad b_1 = \frac{K_m}{V_{max}}$$

- (d) Exprimons l'écart-type de $1/v_0$ en fonction de celui de v_0 :

$$\sigma\left(\frac{1}{v_0}\right) \approx \left|\frac{d(1/v_0)}{dv_0}\right| \sigma(v_0) = \left|-\frac{1}{v_0^2}\right| \sigma(v_0) = \frac{\sigma(v_0)}{v_0^2}$$

La régression linéaire donnera des estimations fiables si :

$$\sigma\left(\frac{1}{v_0}\right) = k \quad (\text{constante}) \quad \Rightarrow \quad \sigma(v_0) = kv_0^2$$

c'est-à-dire si l'erreur de mesure de la vitesse est proportionnelle au carré de la vitesse.

Cette condition ayant peu de chances d'être remplie, dans la plupart des cas l'estimation sera peu fiable ; on sera donc amené à affiner les paramètres par régression non linéaire.

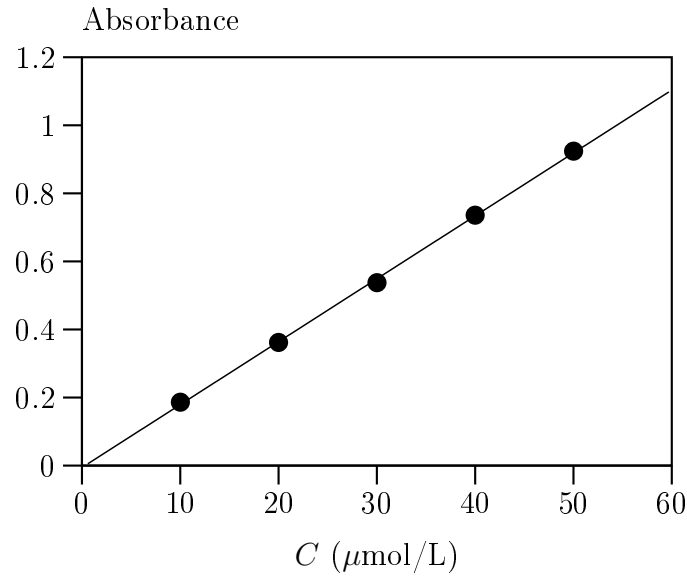
5. Dosage spectrophotométrique du 4-nitrophénol

- (a) Voir fig. 2
(b) En désignant l'absorbance par A , l'équation de la droite de régression est :

$$A = (-0,0057 \pm 0,0092) + (18489 \pm 278)C$$

$n = 5 \quad s_r = 0,0088 \quad r^2 = 0,9993 \quad r = 0,9997 \quad F = 4410$

FIGURE 2 – Dosage spectrophotométrique du 4-nitrophénol



Pour $n - 2 = 3$ degrés de liberté, $t_{0,975} = 3,18$. On en déduit les intervalles de confiance au risque 5% :

$$b_0 \in [-0,0057 - 3,18 \times 0,0092; -0,0057 + 3,18 \times 0,0092] = [-0,0350; 0,0236]$$

$$b_1 \in [18489 - 3,18 \times 278; 18489 + 3,18 \times 278] = [17604; 19373]$$

(c) La loi de Beer-Lambert s'écrit :

$$A = \epsilon l C$$

où ϵ désigne le coefficient d'extinction molaire (en $\text{L mol}^{-1} \text{cm}^{-1}$) et l la longueur du trajet optique (1 cm).

Le coefficient d'extinction correspond à la pente de la droite, soit $18489 \text{ L mol}^{-1} \text{cm}^{-1}$

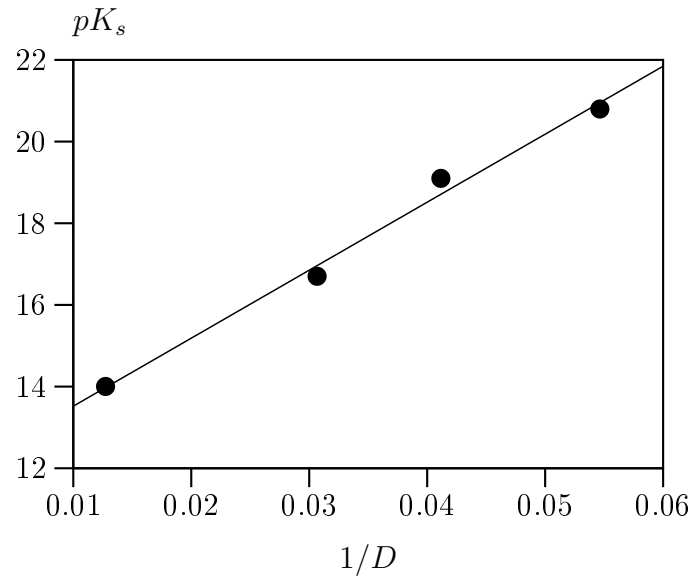
6. Produit ionique des alcools

(a) Voir fig. 3

(b) Après calcul des valeurs de $1/D$, on obtient la droite de régression suivante :

$$pK_s = (11,86 \pm 0,44) + (167 \pm 11) \left(\frac{1}{D} \right)$$

FIGURE 3 – Produit ionique des alcools



$$n = 4 \quad s_r = 0,35 \quad r^2 = 0,9906 \quad r = 0,9953 \quad F = 210$$

Pour $n - 2 = 2$ degrés de liberté, $t_{0,975} = 4,3$. On en déduit les intervalles de confiance au risque 5% :

$$b_0 \in [11,86 - 4,3 \times 0,44; 11,86 + 4,3 \times 0,44] = [9,97; 13,75]$$

$$b_1 \in [167 - 4,3 \times 11; 167 + 4,3 \times 11] = [120; 214]$$

(c) La droite de régression permet d'estimer le pK_s du n -propanol :

$$pK_s = 11,86 + 167 \times \frac{1}{20,1} \approx 20,2$$

Calculons la variance de cette estimation :

$$\text{Var}(\hat{y}_0) = [1 \quad x_0] \begin{bmatrix} \text{Var}(b_0) & \text{Cov}(b_0, b_1) \\ \text{Cov}(b_0, b_1) & \text{Var}(b_1) \end{bmatrix} \begin{bmatrix} 1 \\ x_0 \end{bmatrix}$$

$$= \text{Var}(b_0) + x_0^2 \text{Var}(b_1) + 2x_0 \text{Cov}(b_0, b_1)$$

$$= s_0^2 + x_0^2 s_1^2 + 2x_0 r_{01} s_0 s_1$$

$$\approx (0,44)^2 + \left(\frac{1}{20,1}\right)^2 \times 11^2 + 2 \times \frac{1}{20,1} \times (-0,915) \times 0,44 \times 11$$

$$\text{Var}(\hat{y}_0) \approx 0,052$$

D'où l'écart-type :

$$\sigma(\hat{y}_0) \approx \sqrt{0,052} \approx 0,23$$

puis l'intervalle de confiance au risque 5% :

$$pK_s \in [20,2 - 4,3 \times 0,23; 20,2 + 4,3 \times 0,23] = [19,2; 21,2]$$

(d) La variance $\text{Var}(\epsilon)$ est estimée par la variance résiduelle, soit $V_r = s_r^2 \approx 0,123$, d'où :

$$\text{Var}(y_0) = 0,052 + 0,123 = 0,175$$

D'où l'écart-type :

$$\sigma(y_0) \approx \sqrt{0,175} \approx 0,42$$

puis l'intervalle de confiance au risque 5% :

$$pK_s \in [20,2 - 4,3 \times 0,42; 20,2 + 4,3 \times 0,42] = [18,4; 22,0]$$

Cet intervalle est plus large que le précédent car il englobe l'erreur d'estimation et l'erreur de mesure.

7. Cinétique du premier ordre

Avec le logiciel WinReg, on obtient :

– Par linéarisation : $\ln Q = \ln Q_0 - kt$

$$\ln Q = (6,281 \pm 0,012) - (0,2605 \pm 0,0020)t$$

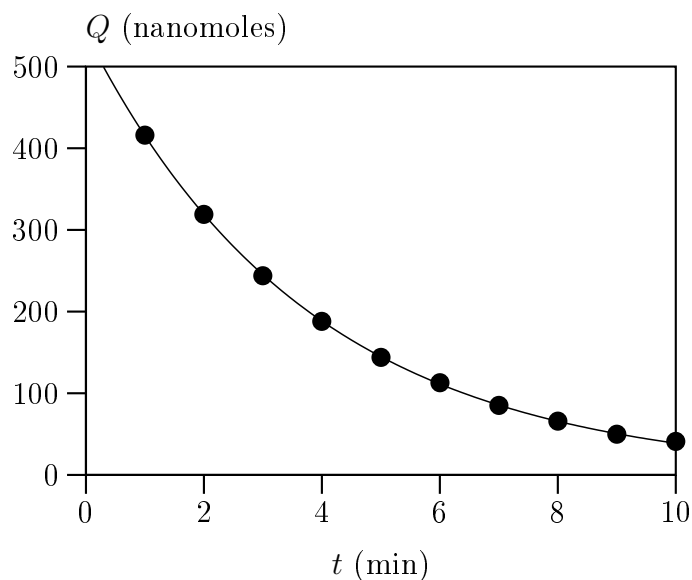
$$n = 10 \quad s_r = 0,018 \quad r^2 = 0,9995 \quad r = -0,9998 \quad F_{1,8} = 17290$$

– Par régression non linéaire : $Q = Q_0 \exp(-kt)$ (Fig. 4)

$$Q = (540,3 \pm 1,7) \exp[(-0,2633 \pm 0,0011)t]$$

$$n = 10 \quad s_r = 1,19 \quad F_{1,8} = 100676$$

FIGURE 4 – Cinétique du premier ordre



L'estimation de k est plus précise par régression non linéaire. On retient donc $k = 0,2633 \text{ min}^{-1}$.

Pour $n-2 = 8$ degrés de liberté, $t_{0,975} = 2,306$. On en déduit l'intervalle de confiance de k au risque 5% :

$$k \in [0,2633 - 2,306 \times 0,0011; 0,2633 + 2,306 \times 0,0011] = [0,2608; 0,2658]$$

8. Ajustement de l'équation de Michaelis

- (a) Après calcul de $1/s_0$ et $1/v_0$, on obtient avec le logiciel WinReg :
 – Par linéarisation : $1/v_0 = (1/V_{max}) + (K_m/V_{max})(1/s_0)$ (Fig. 5)

$$1/v_0 = (0,197 \pm 0,022) - (0,1253 \pm 0,0059)(1/s_0)$$

$$n = 10 \quad s_r = 0,057 \quad r^2 = 0,983 \quad F_{1,8} = 445$$

On en déduit :

$$V_{max} = \frac{1}{0,197} = 5,07 \text{ } \mu\text{mol L}^{-1} \text{ s}^{-1}$$

$$K_m = \frac{0,1253}{0,197} = 0,64 \text{ mmol L}^{-1}$$

FIGURE 5 – Equation de Michaelis : régression linéaire

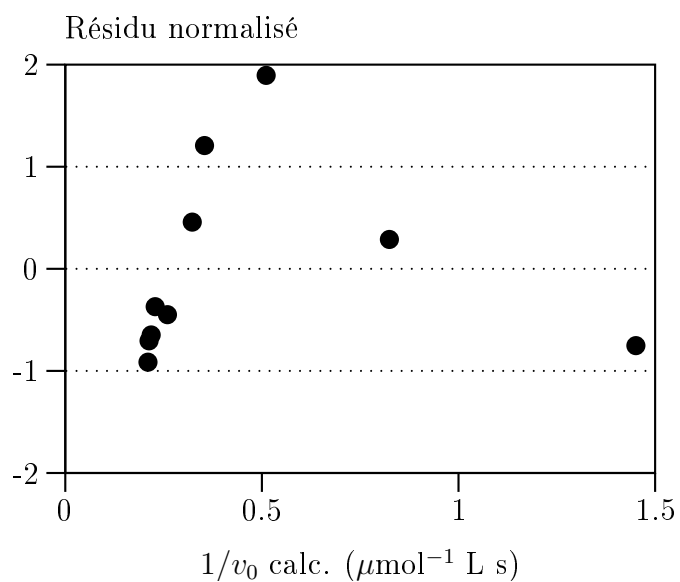
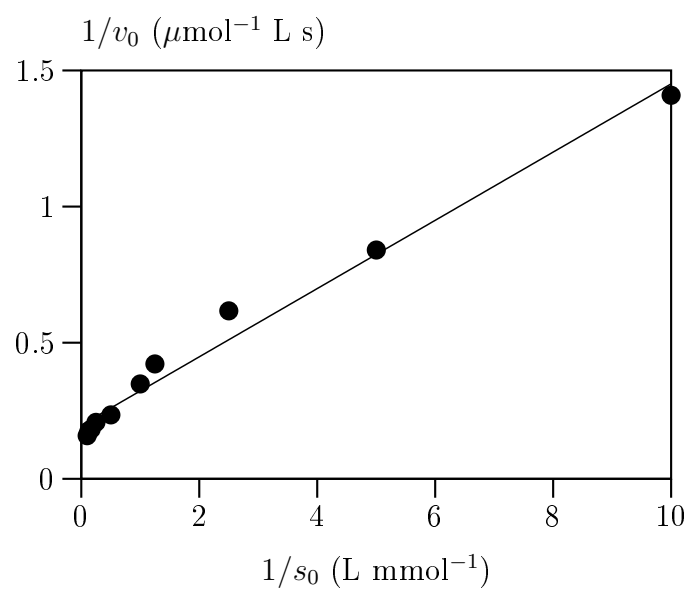
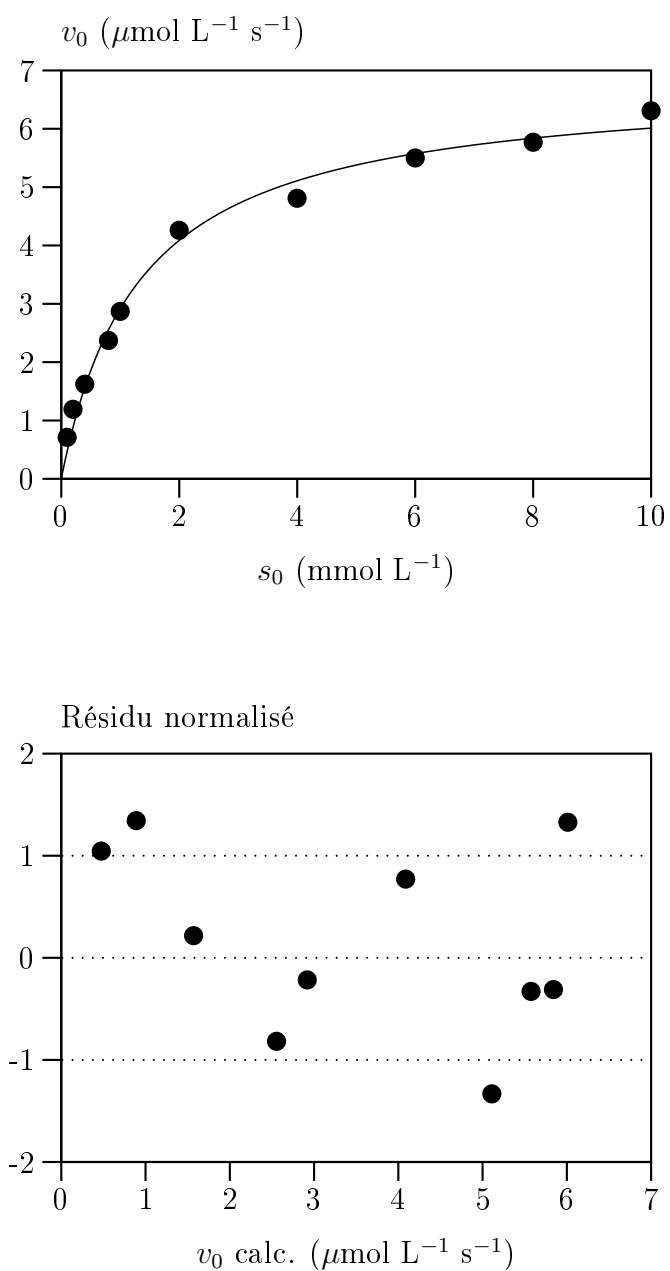


FIGURE 6 – Equation de Michaelis : régression non linéaire



- Par régression non linéaire : $v_0 = V_{max}s_0/(K_m + s_0)$ (Fig. 6)

$$v_0 = \frac{(6,81 \pm 0,21)s_0}{(1,33 \pm 0,14) + s_0}$$

$$n = 10 \quad s_r = 0,22 \quad F_{1,8} = 800$$

- (b) En comparant les diagrammes des résidus, on constate que :
- Pour la régression linéaire (fig. 5), les résidus normalisés ne sont pas répartis symétriquement par rapport à 0, et ils ont tendance à varier en fonction de la variable dépendante (augmentation puis diminution).
 - Pour la régression non linéaire (fig. 6), les résidus normalisés sont répartis symétriquement par rapport à 0 ; ils sont compris dans l'intervalle $[-3, 3]$, la plupart dans l'intervalle $[-2, 2]$, et ne montrent pas de tendance. On peut donc admettre qu'ils sont indépendamment et identiquement distribués selon une loi normale de moyenne 0 et d'écart-type estimé par s_r .

Ces résultats conduisent à privilégier la régression non linéaire.

- (c) D'après les résultats de la régression non linéaire :

$$V_{max} = (6,81 \pm 0,21) \mu\text{mol L}^{-1} \text{ s}^{-1}$$

$$K_m = (1,33 \pm 0,14) \text{ mmol L}^{-1}$$

9. Influence de la température sur la cinétique d'une réaction enzymatique

- (a) D'après l'exercice précédent, $V_{max} = 6,81 \mu\text{mol L}^{-1} \text{ s}^{-1}$, d'où :

$$e_{\text{tot}} = \frac{V_{max}}{k_{\text{cat}}} = \frac{6,81}{796} = 8,56 \times 10^{-3} \mu\text{M} = 8,56 \text{ nM}$$

- (b) ΔH° est défini pour la réaction $E + S \rightarrow ES$ alors que K_m est défini pour la réaction inverse : il faut donc utiliser $1/K_m$
- (c) Les équations (2) et (3) sont de la forme $y = b_0 + b_1x$:

Eq.	x	y	b_0	b_1
(2)	$1/T$	$\ln(1/K_m)$	$\Delta S^\circ/R$	$-\Delta H^\circ/R$
(3)	$1/T$	$\ln(k_{\text{cat}}/T)$	$\ln(k_B/h) + \Delta S^\ddagger/R$	$-\Delta H^\ddagger/R$

Calculons dans chaque cas l'écart-type de la variable dépendante (y) :

$$\sigma\left(\ln \frac{1}{K_m}\right) \approx \left| \frac{d}{dK_m} \ln \frac{1}{K_m} \right| \cdot \sigma(K_m) = \left| -\frac{1}{K_m} \right| \cdot \sigma(K_m) = \frac{\sigma(K_m)}{K_m}$$

Pour le calcul avec k_{cat}/T on peut traiter la température T comme une constante car elle varie peu ($\approx 11\%$) sur l'intervalle de température considéré.

$$\sigma\left(\ln \frac{k_{\text{cat}}}{T}\right) \approx \left| \frac{d}{dk_{\text{cat}}} \ln \frac{k_{\text{cat}}}{T} \right| \cdot \sigma(k_{\text{cat}}) = \left| \frac{1}{k_{\text{cat}}} \right| \cdot \sigma(k_{\text{cat}}) = \frac{\sigma(k_{\text{cat}})}{k_{\text{cat}}}$$

L'estimation par régression linéaire sera valable si l'écart-type de la variable dépendante est constant, c'est-à-dire si l'écart-type de K_m ou k_{cat} est proportionnel à la grandeur correspondante.

(d) Pour calculer les droites de régression on forme le tableau suivant :

t ($^{\circ}\text{C}$)	$1/T$ (K^{-1})	K_m (M)	k_{cat} (s^{-1})	$\ln(1/K_m)$	$\ln(k_{\text{cat}}/T)$
10	0,00353		454		0,47
15	0,00347	0,00097	503	6,94	0,56
20	0,00341	0,00105	659	6,86	0,81
25	0,00335	0,00133	796	6,62	0,98
30	0,00330	0,00180	1200	6,32	1,38
35	0,00325	0,00207	1541	6,18	1,61
40	0,00319	0,00303	1901	5,80	1,80

On en déduit les équations des droites (Fig. 7) :

$$\ln \frac{1}{K_m} = (4144 \pm 341) \left(\frac{1}{T}\right) - (7,4 \pm 1,1)$$

$$n = 6 \quad s_r = 0,079 \quad r^2 = 0,987 \quad F = 147$$

$$\ln \frac{k_{\text{cat}}}{T} = (-4234 \pm 280) \left(\frac{1}{T}\right) - (15,31 \pm 0,94)$$

$$n = 7 \quad s_r = 0,083 \quad r^2 = 0,979 \quad F = 228$$

En tenant compte que $\ln(k_B/h) \approx 23,76$ on calcule les grandeurs thermodynamiques :

$$\begin{aligned} \Delta H^{\circ} &= (-34,4 \pm 2,8) \text{ kJ mol}^{-1} \\ \Delta S^{\circ} &= (-61,5 \pm 9,1) \text{ J mol}^{-1} \text{ K}^{-1} \\ \Delta H^{\ddagger} &= (+35,1 \pm 2,3) \text{ kJ mol}^{-1} \\ \Delta S^{\ddagger} &= (-70,2 \pm 7,8) \text{ J mol}^{-1} \text{ K}^{-1} \end{aligned}$$

FIGURE 7 – Influence de la température sur la cinétique enzymatique

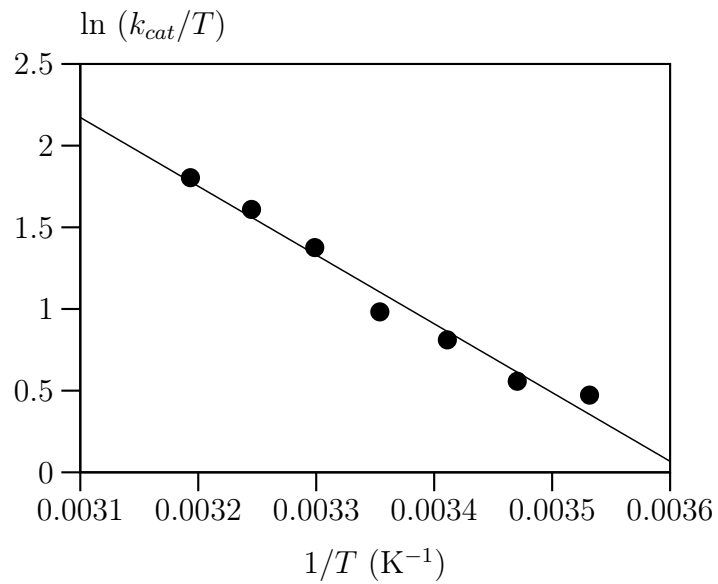
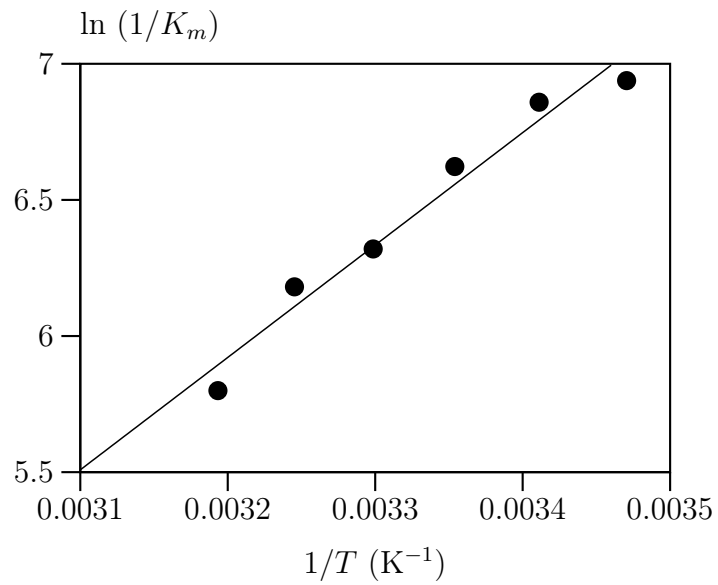
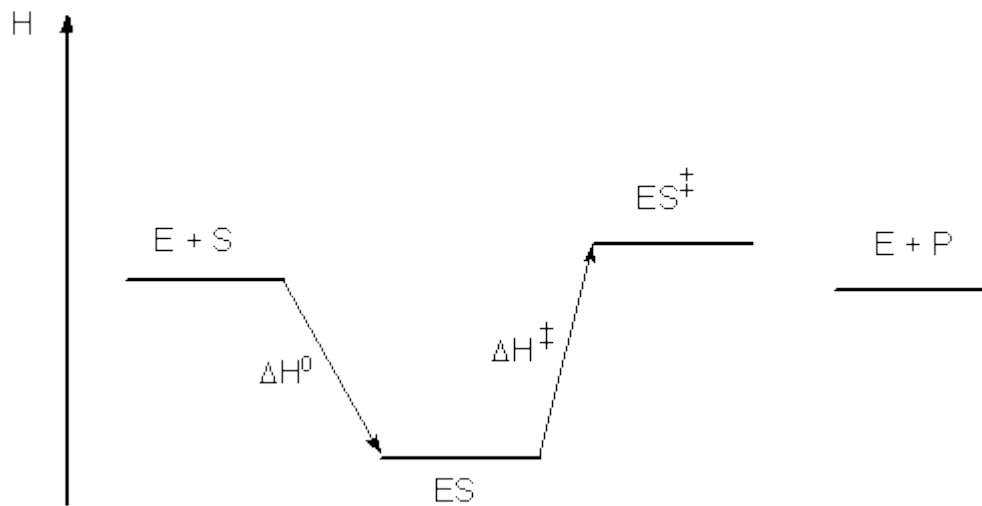


FIGURE 8 – Variations d'enthalpie au cours de la réaction enzymatique



- (e) La fig. 8 montre les différents états du système et les variations d'enthalpie correspondantes. Notons que l'enthalpie d'activation pour la formation du complexe ES ne peut pas être déterminée car le système est toujours à l'équilibre.