

# Exercices de Statistiques

## 1. Ajustement d'une droite

On cherche à ajuster, par régression linéaire, la droite d'équation  $y = \beta_0 + \beta_1 x$  aux données suivantes :

$x$	$y$
1	2
2	4
3	5

- Tracer les points et vérifier qu'ils sont approximativement alignés.
- Calculer les estimations  $b_0$  et  $b_1$  des paramètres  $\beta_0$  et  $\beta_1$  par la méthode matricielle (pour éviter les erreurs d'arrondi, on pourra exprimer les résultats sous forme de fractions).
- Calculer les sommes de carrés d'écart  $SS_t, SS_e, SS_r$  et vérifier l'équation d'analyse de la variance.
- Donner des valeurs décimales approchées (3 à 4 chiffres significatifs) pour le coefficient de détermination  $r^2$  et le coefficient de corrélation  $r$ .
- Calculer la variance résiduelle  $V_r$ , l'écart-type résiduel  $s_r$  et le rapport de variances  $F$ . Donner une valeur approchée de  $s_r$  avec 2 chiffres significatifs.
- Calculer les écart-types  $(s_0, s_1)$  des paramètres de la droite. Donner des valeurs approchées avec 2 chiffres significatifs.
- Donner la présentation des résultats sous la forme :

$$y = (b_0 \pm s_0) + (b_1 \pm s_1)x$$

$$n = \dots \quad r^2 = \dots \quad r = \dots \quad s_r = \dots \quad F = \dots$$

- Calculer le coefficient de corrélation  $r_{01}$  entre les paramètres estimés. Comment interpréter le résultat ?

## 2. Dosage chromatographique

Lors de l'étalonnage d'une méthode chromatographique, on a obtenu les données suivantes, où  $x$  désigne la quantité de produit dosé et  $y$  la réponse du détecteur (en unités arbitraires).

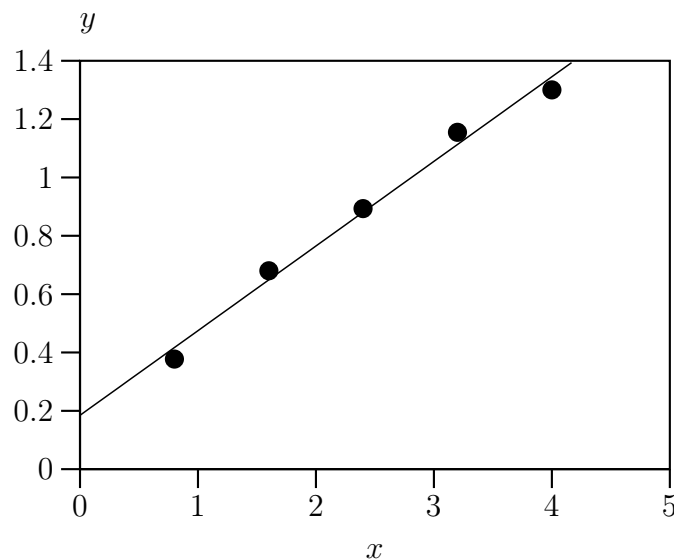
$x$	0,8	1,6	2,4	3,2	4
$y$	0,377	0,68	0,893	1,155	1,3

A l'aide du logiciel WinReg, on a modélisé la courbe d'étalonnage, soit par une droite, soit par un polynôme du deuxième degré, afin de déterminer quel est le modèle le plus approprié.

(a) Régression linéaire :  $y = \beta_0 + \beta_1 x$

Param.	Valeur	E.type	t Student	Prob(> t )
$b_0$	0.1847	0.0488	3.7851	0.0323
$b_1$	0.2901	0.0184	15.7755	0.0006

$n = 5$ ;  $s_r = 0.047$ ;  $r = 0.994$ ;  $r^2 = 0.988$ ;  $F = 249$ ;  $\text{Prob}(> F) = 0.0006$



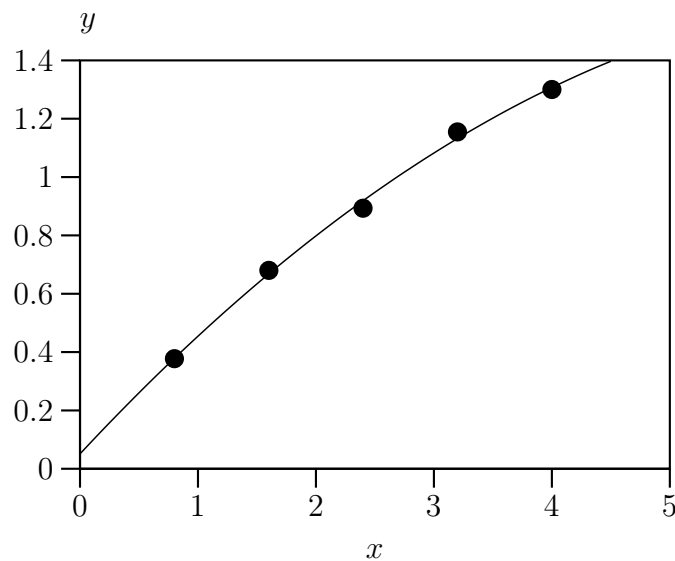
Ecrire l'équation de la droite sous la forme conseillée.

Soit l'hypothèse nulle :  $\beta_0 = 0$ . Peut-on rejeter cette hypothèse ? A quel risque ? Ce résultat est-il compatible avec ce que l'on attend d'une courbe d'étalonnage ?

(b) Régression polynomiale :  $y = \beta_0 + \beta_1x + \beta_2x^2$

Param.	Valeur	E.type	t Student	Prob(> t )
$b_0$	0.0512	0.0568	0.9017	0.4624
$b_1$	0.4332	0.0541	8.0082	0.0152
$b_2$	-0.0298	0.0111	-2.6953	0.1145

$n = 5$ ;  $s_r = 0.027$ ;  $r = 0.999$ ;  $r^2 = 0.997$ ;  $F = 388$ ;  $\text{Prob>(> } F) = 0.0026$



Ecrire l'équation de la courbe sous la forme conseillée.

Tester à nouveau l'hypothèse nulle :  $\beta_0 = 0$

Soit l'hypothèse nulle :  $\beta_2 = 0$  Peut-on rejeter cette hypothèse?

A quel risque?

(c) Comparaison des deux modèles

Comparer les critères statistiques des 2 modèles (coefficients de détermination, écart-type résiduel, rapport de variance).

Au vu des résultats et de ceux des questions précédentes, quel modèle choisiriez-vous pour représenter la courbe d'étalonnage?

### 3. Ajustement d'un modèle exponentiel

On cherche à ajuster, par régression non linéaire, le modèle exponentiel  $y = a \exp(-bx)$  aux données suivantes :

$x$	$y$
1	10
2	5
3	2

(a) Après linéarisation on a obtenu l'équation :

$$\ln y = 3,1445 - 0,8047x$$

En déduire les estimations initiales  $a^0$  et  $b^0$  du modèle exponentiel

(b) Calculer les dérivées partielles :

$$y'_a = \frac{\partial f}{\partial a}(x; a, b) \quad y'_b = \frac{\partial f}{\partial b}(x; a, b)$$

(c) Montrer que le vecteur  $\mathbf{z}$  de la régression non linéaire et la matrice Jacobienne  $\mathbf{J}$  sont donnés approximativement par :

$$\mathbf{z} \approx \begin{bmatrix} -0,3792 \\ 0,3582 \\ -0,07591 \end{bmatrix} \quad \mathbf{J} \approx \begin{bmatrix} 0,4472 & -10,38 \\ 0,2000 & -9,284 \\ 0,08945 & -6,228 \end{bmatrix}$$

(d) On donne :

$$\delta = (\mathbf{J}^\top \mathbf{J})^{-1}(\mathbf{J}^\top \mathbf{z}) \approx \begin{bmatrix} -2,110 \\ -0,05931 \end{bmatrix}$$

En déduire les nouvelles estimations  $a$  et  $b$  des paramètres du modèle exponentiel.

(e) Que faudrait-il faire pour terminer l'estimation des paramètres du modèle par régression non linéaire ?

### 4. Equation de Michaelis

L'équation de Michaelis est utilisée en Biochimie pour représenter les variations de la vitesse initiale  $v_0$  d'une réaction catalysée par une enzyme, en fonction de la concentration initiale  $s_0$  du substrat :

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

où  $V_{max}$  désigne la vitesse maximale et  $K_m$  la constante de Michaelis.

- (a) Quelle est l'allure du graphe de  $v_0$  en fonction de  $s_0$  ?
- (b) Montrer que  $K_m$  représente la concentration de substrat correspondant à la moitié de la vitesse maximale.
- (c) Montrer que l'équation de Michaelis peut être linéarisée en exprimant  $1/v_0$  en fonction de  $1/s_0$  (transformation en double inverse dite de Lineweaver et Burk)
- (d) Quelle condition devrait vérifier l'écart-type de  $v_0$  pour que la linéarisation précédente donne une estimation fiable des paramètres  $K_m$  et  $V_{max}$  ?

### 5. Dosage spectrophotométrique du 4-nitrophénol

On a mesuré l'absorption de la lumière par des solutions alcalines de 4-nitrophénol, de concentrations croissantes. On a obtenu les résultats suivants (pour une lumière de longueur d'onde 400 nm et un trajet optique de 1 cm) :

concentration C (en mol/l)	$1 \times 10^{-5}$	$2 \times 10^{-5}$	$3 \times 10^{-5}$	$4 \times 10^{-5}$	$5 \times 10^{-5}$
absorbance A	0,1865	0,3616	0,5370	0,7359	0,9238

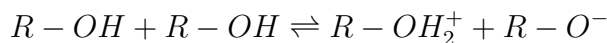
- (a) Vérifier graphiquement qu'on peut admettre l'existence d'une relation linéaire entre l'absorbance et la concentration (loi de Beer-Lambert).
- (b) En supposant que les hypothèses du cours sont satisfaites, estimer les paramètres de la droite de régression de A par rapport à C :
  - ponctuellement,
  - par des intervalles de confiance au risque 5 %.
 En déduire une estimation du coefficient d'extinction du 4-nitrophénol. Préciser son unité.

**Extrait de la table de Student pour  $\alpha = 0,05$  :**

d.d.l.	1	2	3	4	5
$t_{0,975}$	12,7	4,30	3,18	2,78	2,57

## 6. Produit ionique des alcools

Le produit ionique des alcools est défini par l'équilibre ;



$$K_s = [R-OH_2^+][R-O^-]$$

$$pK_s = -\log K_s$$

La valeur de  $pK_s$  est liée à la constante diélectrique  $D$  par une relation du type :

$$pK_s = \alpha + \frac{\beta}{D} \quad (1)$$

On connaît les résultats suivants :

solvants	$D$	$pK_s$
eau	78,5	14
éthanol	24,3	19,1
isopropanol	18,3	20,8
méthanol	32,6	16,7

- Vérifier graphiquement la validité de la relation (1) pour ces solvants.
- Estimer les valeurs de  $\alpha$  et de  $\beta$  :
  - ponctuellement,
  - par des intervalles de confiance au coefficient de sécurité 0,95.
- Pour le  $n$ -propanol, on a :  $D = 20,1$ . Estimer son  $pK_s$  :
  - ponctuellement,
  - par un intervalle de confiance au risque 0,05.

Note : Si  $\hat{y}_0$  désigne une valeur estimée à l'aide d'une droite de régression, pour une valeur  $x_0$  de la variable indépendante ( $\hat{y}_0 = b_0 + b_1x_0$ ), la variance de cette estimation peut être calculée par :

$$\text{Var}(\hat{y}_0) = \mathbf{x}^\top \mathbf{V} \mathbf{x}$$

où  $\mathbf{x}$  est le vecteur  $[1 \quad x_0]^\top$  et  $\mathbf{V}$  représente la matrice de variance-covariance des paramètres de la droite.

- Donner un intervalle de confiance du  $pK_s$  mesuré pour le  $n$ -propanol. On utilisera les relations :

$$y_0 = \hat{y}_0 + \epsilon \quad \Rightarrow \quad \text{Var}(y_0) = \text{Var}(\hat{y}_0) + \text{Var}(\epsilon)$$

## 7. Cinétique du premier ordre

Un corps chimique se décompose selon une cinétique du premier ordre caractérisée par l'équation :  $Q = Q_0 e^{-kt}$  où :

$Q$  désigne la quantité de corps restant à l'instant  $t$  ;

$Q_0$  la quantité initiale ;

$k$  la constante de vitesse de la décomposition.

On dispose des données expérimentales suivantes :

t (min)	1	2	3	4	5	6	7	8	9	10
Q (nanomoles)	416	319	244	188	144	113	85	66	50	41

Estimer la constante de vitesse  $k$  par linéarisation, puis par régression non linéaire. Quelle est la méthode la plus précise ?

Donner un intervalle de confiance de  $k$  au risque 0,05 (on donne  $t_{0,975} = 2,306$  à 8 d.d.l.)

Préciser l'unité de  $k$

## 8. Ajustement de l'équation de Michaelis

Le tableau suivant représente les variations de la vitesse initiale  $v_0$  d'une réaction enzymatique en fonction de la concentration initiale  $s_0$  du substrat :

$s_0$ (mmol L <sup>-1</sup> )	$v_0$ (μmol L <sup>-1</sup> s <sup>-1</sup> )
0,10	0,71
0,20	1,19
0,40	1,62
0,80	2,37
1,00	2,87
2,00	4,26
4,00	4,81
6,00	5,50
8,00	5,77
10,00	6,31

- Estimer  $K_m$  et  $V_{max}$  par linéarisation, puis par régression non linéaire (dans WinReg : F5 puis choisir « Michaelis »).
- Pour chaque méthode de régression, tracer les résidus normalisés en fonction de  $y_{calc}$  (dans WinReg : menu « Graphique / Axes et courbes »). Peut-on admettre que la répartition des résidus est conforme aux hypothèses du cours ? Les résultats permettent-ils de privilégier une méthode de régression ?

- (c) Donner les valeurs de  $K_m$  et  $V_{max}$  sous forme : valeur estimée  $\pm$  écart-type. Préciser les unités.

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

Le tableau suivant représente les variations de  $K_m$  et de la constante catalytique  $k_{cat} = V_{max}/e_{tot}$ , où  $e_{tot}$  désigne la concentration totale en enzyme, en fonction de la température, pour l'enzyme étudiée à l'exercice précédent.

Temperature (°C)	$K_m$ (mM)	$k_{cat}$ (s <sup>-1</sup> )
10		454
15	0,97	503
20	1,05	659
25	1,33	796
30	1,80	1200
35	2,07	1541
40	3,03	1901

Le mécanisme de la réaction peut être représenté sous la forme :



avec  $E$  = enzyme,  $S$  = substrat,  $ES$  = complexe enzyme-substrat,  $P$  = produit de la réaction.  $K_m$  représente la constante de dissociation du complexe  $ES$  ( $K_m = [E][S]/[ES]$ ) et  $k_{cat}$  représente la constante de vitesse de la réaction  $ES \rightarrow E + P$

La variation de  $K_m$  et  $k_{cat}$  avec la température s'exprime par les relations suivantes :

$$\ln \frac{1}{K_m} = -\frac{\Delta H^\circ}{RT} + \frac{\Delta S^\circ}{R} \quad (2)$$

$$\ln \frac{k_{cat}}{T} = \ln \frac{k_B}{h} - \frac{\Delta H^\ddagger}{RT} + \frac{\Delta S^\ddagger}{R} \quad (3)$$

$K_m$  : constante de Michaelis (mol L<sup>-1</sup>)

$k_{cat}$  : constante catalytique (s<sup>-1</sup>)

$T$  : température absolue (K)

$\Delta H^\circ$ ,  $\Delta S^\circ$  : variations d'enthalpie et d'entropie pour la formation du complexe  $ES$

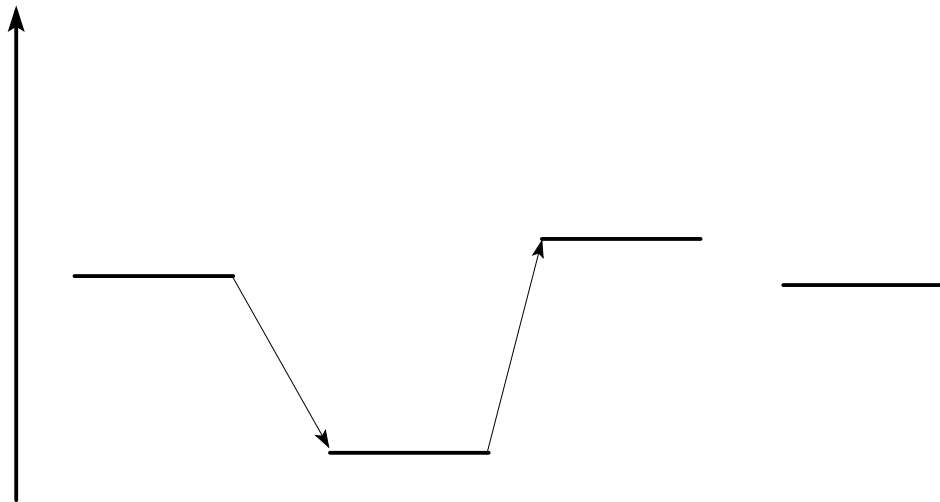
$\Delta H^\ddagger, \Delta S^\ddagger$  : enthalpie et entropie d'activation pour la réaction  $ES \rightarrow E + P$

$R$  : constante des gaz parfaits =  $8,31 \text{ J mol}^{-1} \text{ K}^{-1}$

$k_B$  : constante de Boltzmann =  $1,38 \times 10^{-23} \text{ J K}^{-1}$

$h$  : constante de Planck =  $6,62 \times 10^{-34} \text{ J s}$

- Les valeurs déterminées au TP précédent correspondent à la température de  $25^\circ\text{C}$ . Quelle était la concentration de l'enzyme?
- Pourquoi utilise-t-on  $1/K_m$  dans l'équation 2?
- Montrer que l'on peut estimer  $\Delta H^\circ, \Delta S^\circ, \Delta H^\ddagger, \Delta S^\ddagger$  par régression linéaire à partir des équations 2 et 3. Quelles conditions doivent vérifier les écart-types de  $K_m$  et  $k_{cat}$  pour que ces estimations soient valables?
- En supposant ces conditions vérifiées, estimer ces paramètres. Précisez les écart-types et les unités.
- Le schéma suivant représente l'évolution de l'enthalpie du système au cours de la réaction :



- Faire correspondre aux niveaux d'enthalpie les états du système :  $E + S, ES, ES^\ddagger$  (état de transition),  $E + P$
- A quoi correspondent les variations d'enthalpie symbolisées par les flèches ?