Dans tous les exercices on considérera un risque de première espèce de 5%.

Exercice 1 — Qui des deux à raison ? — 4 points

Al Blop a acheté à Bill le vendeur 800 oignons de tulipe qu'il a planté. Bill a déclaré qu'il garantissait plus de 80 % de floraison. Sur les 800 oignons, seuls 652 ont germé et ont donné des fleurs.

1. Peut-on en conclure que le vendeur a raison ?
2. Le vendeur des oignons a assuré à Blop que les tulipes vendues étaient de quatre types, deux couleurs (rouge ou jaunes) et deux formes de pétales (lisses ou rugueux) et qu'il devait attendre sur le lot 9/16 ème de tulipes rouges et lisses, 3/16 ème de tulipes rouges et rugueuses 3/16 ème de tulipes jaunes et lisses et 1/16 ème du tulipes jaunes et rugueuses. Blop a observé les résultats suivants : 440 de tulipes rouges et lisses ; 50 tulipes rouges de rugueuses ; 147 tulipes jaunes et lisses ; 15 tulipes jaunes et rugueses
   a) Peut-on en conclure que le vendeur a raison ?
   b) Peut-on en conclure à l'indépendance entre la couleur et le caractère lisse ou rugueux de la fleur ?

Exercice 2 — Le bol d'air de l'étudiante — 4 points

La capacité respiratoire d'une personne est une variable aléatoire X, distribuée suivant une loi normale de paramètres inconnus. On tire deux échantillons i.i.d. de taille 10 de cette loi. Le premier groupe de mesures a été effectué avant d'avoir soumis 10 individus à un traitement dont on cherche à démontrer l'efficacité. Le second groupe de mesures a été pris après avoir effectué le traitement. On cherche à savoir si ce traitement est efficace ou non.

1. En supposant que les individus du groupe 1 et du groupe 2 sont différents, proposez une stratégie de décision permettant de décider si le traitement est efficace où non.
2. après avoir observé les résultats suivants, et pour un risque de première espèce de 5%, que décidez vous ?

individu	1	2	3	4	5	6	7	8	9	10
groupe 1	102	84	91	72	93	91	135	115	101	94
groupe 2	102	115	145	107	117	114	111	120	125	101

$\sum x_i$ groupe 1 : 978 et groupe 2 : 1157 et $\sum x_i^2$ groupe 1 : 98342 et groupe 2 : 135335

Exercice 3 — Le chant du cygne — 4 points

On reprend l'exercice précédent en supposant maintenant que les individus du groupe 1 et du groupe 2 sont les mêmes (on a testé le même individu avant et après le traitement)) et sans faire l'hypothèse de normalité. On note N le nombre de fois où le traitement a augmenté la capacité respiratoire d'un individu. N est une variable aléatoire suivant une loi binômiale de paramètres p et $n = 10$.

1. Si le traitement n'a pas d'effet, le nombre de fois où le traitement a augmenté la capacité respiratoire d'un individu doit être égale au nombre de fois où le traitement a diminué la capacité respiratoire d'un individu, et donc la proportion du nombre de fois où le traitement a augmenté la capacité respiratoire $p = \frac{1}{2}$.
   Proposez une stratégie de décision permettant de vérifier cette hypothèse.
2. A la vue des données disponibles, que décidez vous ?
3. Avec la règle de décision que vous avez proposé, quelle serait la probabilité de se tromper si en fait le traitement augmentait la capacité respiratoire de 7 patients sur 10.

Exercice 4 — restons simple — 3 points

Cherchant à expliquer la température d'un four (y) en fonction de l'hygrométrie (x) un boulanger a réalisé dix huit expériences différentes. Il a exécuté un programme de régression linéaire simple ($y = ax + b + \varepsilon$) sur ses données et il a trouvé une estimation de la pente de la droite de $\widehat{a} = 1.0355$, un coefficient $\widehat{b} = 0.2353$ et un coefficient de détermination $R^2 = 0.8534$. Le logiciel a aussi donné les résultats suivants :

Point	x	y	e	r	ei	c
1	-0.20	-0.04	-0.06	-0.28	-0.08	0.01
2	0.72	1.30	0.32	1.35	0.34	0.06
3	0.08	0.08	-0.24	-1.03	-0.27	0.05
4	1.39	1.37	-0.31	-1.45	-0.40	0.30
5	0.22	0.27	-0.18	-0.78	-0.20	0.02
6	0.87	1.36	0.22	0.95	0.24	0.04
7	1.38	1.44	-0.22	-1.03	-0.28	0.15
8	0.94	1.28	0.07	0.32	0.08	0.01
9	0.36	0.62	0.01	0.03	0.01	0.00
10	1.22	1.29	-0.21	-0.96	-0.26	0.09
11	0.74	1.41	0.41	1.73	0.44	0.11
12	0.13	0.12	-0.24	-1.04	-0.27	0.05
13	0.21	0.26	-0.19	-0.82	-0.21	0.03
14	0.56	1.16	0.35	1.46	0.37	0.06
15	-0.35	-0.04	0.09	0.42	0.11	0.02
16	-0.30	-0.16	-0.09	-0.40	-0.11	0.02
17	0.48	0.93	0.20	0.84	0.21	0.02
18	0.48	0.83	0.10	0.40	0.10	0.00

où e désigne l'erreur d'estimation (les résidus), r les résidus standardisés, ei les résidus studentisés et c les distance de Cook (encore appelées contributions).

1. A propos du modèle :
   a) quelle différence faites vous entre a et $\widehat{a}$?
   b) Qu'est-ce que ε dans le modèle ?
2. Que pensez vous des résultats de cette régression linéaire ?

Exercice 5 — Passons aux choses sérieuses — 5 points

Le voisin du boulanger, lui même boulanger à la retraite, explique que le modèle linéaire n'est pas le bon. Le bon modèle est le suivant :

$$y = \begin{cases} ax + bx^2 + \varepsilon & \text{si } x \leq 0,6 \\ c + \varepsilon & \text{si } x > 0,6 \end{cases}$$

1. En supposant que le modèle est continu, posez le problème sous la forme matricielle

   $$\mathbf{y} = X\alpha + \epsilon$$

   où X est une matrice et $\mathbf{y}, \alpha, \epsilon$ des vecteurs que l'on précisera.
2. A partir des données du tableau de l'exercice précédent (les x et les y) donnez une estimation des paramètres a, b et c.

$$\text{indication :} \quad \begin{pmatrix} 3.7927 & 1.8979 \\ 1.8979 & 1.1579 \end{pmatrix}^{-1} = \begin{pmatrix} 1.4670 & -2.4046 \\ -2.4046 & 4.8052 \end{pmatrix}$$

Quelques extraits des tables des lois (choisies au hasard) .

normale $X \sim \mathcal{N}(0,1)$, $\mathbb{P}(X > 1,96) = 0,025$
normale $X \sim \mathcal{N}(0,1)$, $\mathbb{P}(X > 1,65) = 0,05$
student $X \sim T_9$, a 9 degrés de liberté $\mathbb{P}(X > 2.82) = 0,02$
student $X \sim T_9$, a 9 degrés de liberté $\mathbb{P}(X > 2.26) = 0,05$
student $X \sim T_{10}$, a 10 degrés de liberté $\mathbb{P}(X > 2.76) = 0,02$
student $X \sim T_{10}$, a 10 degrés de liberté $\mathbb{P}(X > 2.23) = 0,05$
student $X \sim T_{11}$, a 11 degrés de liberté $\mathbb{P}(X > 2.72) = 0,02$
student $X \sim T_{11}$, a 11 degrés de liberté $\mathbb{P}(X > 2.20) = 0,05$
student $X \sim T_{12}$, a 12 degrés de liberté $\mathbb{P}(X > 2.65) = 0,02$
student $X \sim T_{12}$, a 12 degrés de liberté $\mathbb{P}(X > 2.17) = 0,05$
student $X \sim T_{18}$, a 20 degrés de liberté $\mathbb{P}(X > 2.49) = 0,02$
student $X \sim T_{18}$, a 20 degrés de liberté $\mathbb{P}(X > 2.06) = 0,05$
student $X \sim T_{20}$, a 20 degrés de liberté $\mathbb{P}(X > 2.52) = 0,02$
student $X \sim T_{20}$, a 20 degrés de liberté $\mathbb{P}(X > 2.08) = 0,05$
student $X \sim T_{22}$, a 22 degrés de liberté $\mathbb{P}(X > 2.50) = 0,02$
student $X \sim T_{22}$, a 22 degrés de liberté $\mathbb{P}(X > 2.07) = 0,05$
chi 2 $X \sim \chi^2_1$, a 1 degrés de liberté $\mathbb{P}(X > 7,88) = 0,005$
chi 2 $X \sim \chi^2_1$, a 1 degrés de liberté $\mathbb{P}(X > 6.63) = 0,01$
chi 2 $X \sim \chi^2_1$, a 1 degrés de liberté $\mathbb{P}(X > 5.02) = 0,025$
chi 2 $X \sim \chi^2_1$, a 1 degrés de liberté $\mathbb{P}(X > 3.84) = 0,05$
chi 2 $X \sim \chi^2_2$, a 2 degrés de liberté $\mathbb{P}(X > 10,59) = 0,005$
chi 2 $X \sim \chi^2_2$, a 2 degrés de liberté $\mathbb{P}(X > 9,21) = 0,01$
chi 2 $X \sim \chi^2_2$, a 2 degrés de liberté $\mathbb{P}(X > 7,35) = 0,025$
chi 2 $X \sim \chi^2_2$, a 2 degrés de liberté $\mathbb{P}(X > 5,99) = 0,05$
chi 2 $X \sim \chi^2_3$, a 3 degrés de liberté $\mathbb{P}(X > 12,83) = 0,005$
chi 2 $X \sim \chi^2_3$, a 3 degrés de liberté $\mathbb{P}(X > 11,34) = 0,01$
chi 2 $X \sim \chi^2_3$, a 3 degrés de liberté $\mathbb{P}(X > 9,35) = 0,025$
chi 2 $X \sim \chi^2_3$, a 3 degrés de liberté $\mathbb{P}(X > 7,81) = 0,05$
$B(n,p)$ $X \sim B(n,p)$, $p = 1/2$ $\mathbb{P}(X >) = 0,05$

Soit $Y \sim \mathcal{N}(0,1)$ une variable aléatoire normale centrée réduite. Soit $Y_1, Y_2, ..., Y_n$ un échantillon de n réalisation i.i.d. de cette variable aléatoire.

La loi du χ^2 On appelle loi du χ^2 à n degrés de libertés la loi de la variable aléatoire $Z_n = \sum_{i=1}^n Y_i^2$

La loi de student On appelle loi de student à n degrés de libertés la loi de la variable aléatoire T_n

$$T_n = \frac{N}{\sqrt{\frac{X_n}{n}}} \qquad \begin{array}{l} N \sim \mathcal{N}(0,1) \\ \\ X_n \sim \chi^2_n \end{array}$$

Theorem 0.1 (Théorème du χ^2 (Pearson)) *pour N_{ij} effectif observés et pour n_{ij} effectif théorique*

$$X_{ij} = \frac{N_{ij} - n\,\widehat{p}_{ij}}{\sqrt{n\,\widehat{p}_{ij}}} \qquad \sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}^2 \longrightarrow \chi^2_{(I-1)(J-1)}$$

la variable $T_{n_x+n_y-2}$ suit une loi de student à $n_x + n_y - 2$ degrés de liberté :

$$T_{n_x+n_y-2} = \sqrt{n_x + n_y - 2}\,\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\left(\frac{1}{n_x} + \frac{1}{n_y}\right) S^2_{xy}}}$$

avec $S^2_{xy} = S^2_x + S^2_y = \sum_{i=1}^{n_x}(X_i - \overline{X})^2 + \sum_{i=1}^{n_y}(Y_i - \overline{Y})^2$