

Examen du mercredi 19 décembre 2018

Durée 2 heures. Documents, calculatrices et téléphones portables interdits.

Prière de rendre **deux copies**, l'une portant la mention **STATS** pour les exercices 1 et 2 de la partie statistique du sujet et l'autre portant la mention **MATHS** pour les exercices 3, 4 et 5 de la partie mathématique du sujet.

A) Sujet de statistiques : Analyse de données

Exercice 1. (Question de cours)

Expliquez succinctement le cadre d'utilisation, le principe et le fonctionnement d'une analyse en composantes principales.

Exercice 2. (Étude statistique d'un jeu de données)

On se place dans le contexte de l'acclimatation végétale au froid, un ensemble de mécanismes physiologiques permettant aux plantes de résister au froid et au gel. Dans la nature, les êtres vivants sont parfois confrontés à des conditions défavorables telles que la sécheresse, la salinité, le froid ou encore les inondations, qui sont des stress abiotiques. Les conséquences vont du simple ralentissement de la croissance à la mort. Ces individus ont développé diverses stratégies pour faire face à ce type de stress.

Les protéines LEA (Late Embryogenesis Abundant proteins) sont des protéines qui contribuent, principalement chez les végétaux, à l'acquisition de la tolérance à la dessiccation, en particulier dans le cas de déshydratation ou de stress induit par le froid. Elles sont assez mal connues et encore moins bien classifiées, même si des groupes de LEA sont souvent utilisés. À partir des données disponibles au NCBI plusieurs centaines de protéines annotées LEA ont été sélectionnées et complétées pour en faire la base de données. Les données qui suivent ne sont qu'un extrait de cette base.

L'objectif de l'étude consiste à analyser et décrire les caractéristiques de ces protéines.

Présentation des données : Le tableau de données à analyser contient 1295 protéines et les 6 variables suivantes :

- LENGTH : longueur de la séquence (en acides aminés) ;
- REIGN : règne associé ;
- FOLDINDEX : indice de repliement ;
- PI : point isoélectrique ou pHi ;
- MW : poids moléculaire (en g/mol et non pas en Dalton) ;
- GRAVY : hydrophatie moyenne, échelle de KD.

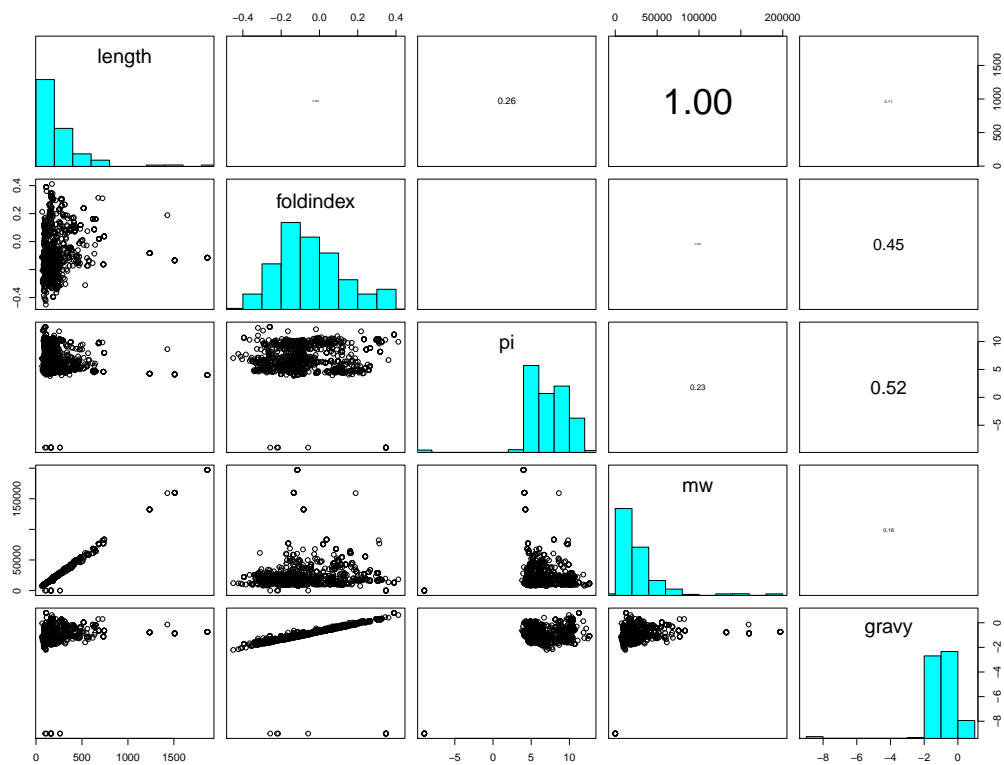


FIGURE 1 –

```

head(data)
  length      reign  foldindex  pi      mw      gravity
ABS44867    200 Viridiplantae -0.0835869 6.14 20295.80 -1.018091
2211386C    153 Viridiplantae -0.1659534 9.31 15909.32 -1.274510
ABB55135    132 Viridiplantae -0.2853720 9.44 14103.76 -1.580303
AAW59164    143 Viridiplantae -0.1855139 8.70 15209.61 -1.357343
AAW59168    144 Viridiplantae -0.2060405 9.31 15422.86 -1.401389
AAW59174    143 Viridiplantae -0.1803197 8.70 15237.67 -1.340559
...

summary(data)
  length      reign  foldindex  pi      mw
gravity
Min.   : 68.0      Viridiplantae:698  Min.   :-0.45034  Min.   :-9.000  Min.   : -9
1st Qu.: 137.0     Bacteria      :304  1st Qu.: -0.16615  1st Qu.: 5.170  1st Qu.: 14608
1st Qu.: -1.2225   Fungi         :132  Median : -0.08624  Median : 6.750  Median : 18999
Median : 182.0     Metazoa      :161  Mean   : -0.04703  Mean   : 7.167  Mean   : 28052
Mean   : 264.8
Mean   : -0.9104
3rd Qu.: 291.0
3rd Qu.: -0.4612
Max.   : 1864.0
Max.   : 0.7991
Max.   : 0.41210  Max.   : 12.610  Max.   : 197129

```

Script 1

Analyse préliminaire des données

1. D'après la présentation des données, le script 1 et la figure 1, quel type d'analyse multivariée feriez-vous ? Justifiez votre choix.

2. Interprétez la figure 1.

```

inertia.dudi(analyse)$TOT
  inertia      cum      ratio
1 2.512508831 2.512509 0.3140636
2 1.902411425 4.414920 0.5518650
3 1.280828173 5.695748 0.7119686
4 0.988823958 6.684572 0.8355715
5 0.615830799 7.300403 0.9125504
6 0.444260801 7.744664 0.9680830
7 0.254008848 7.998673 0.9998341
8 0.001327166 8.000000 1.0000000

```

Script 2

```

analyse$co
  length      Comp1      Comp2      Comp3      Comp4      Comp5      Comp6
reign.Bacteria 0.18785191 -0.1810205 -0.05839744 -0.020410906 0.22044376 0.07549110
reign.Fungi    1.96148071 -0.3501940 -0.94252474 -0.406469566 -1.90476592 -0.31066278
reign.Metazoa 0.04345214 -0.4897206 0.41156921 2.567085491 -0.08712049 -0.13314603
reign.Viridiplantae -0.46277684 -0.3250778 -0.37052439 -0.386946166 0.41295466 -0.27526989
foldindex     0.28089072 0.7447262 0.38905358 -0.067635191 -0.03550546 -0.40207574
pi            -0.24908323 0.5196573 -0.72897902 0.108792331 -0.06977667 0.20817076
mw            0.94734757 -0.1646610 -0.10096240 0.004340866 0.23338811 0.07866238
gravity       0.24245520 0.7771137 -0.37940492 0.186475391 0.20649471 -0.09929461

```

Script 3

```

> inertia.dudi(analyse , col=T)$col.abs/100
      length      Comp1      Comp2      Comp3      Comp4      Comp5      Comp6
reign.Bacteria  0.33 16.54 19.90 2.06 0.21 37.06
reign.Fungi     15.61 0.66 7.07 1.70 60.05 2.21
reign.Metazoa   0.01 1.57 1.64 82.85 0.15 0.50
reign.Viridiplantae 4.59 2.99 5.78 8.16 14.93 9.19
foldindex       3.14 29.15 11.82 0.46 0.20 36.39
pi              2.47 14.19 41.49 1.20 0.79 9.75
mw              35.72 1.43 0.80 0.00 8.84 1.39
gravity         2.34 31.74 11.24 3.52 6.92 2.22

> inertia.dudi(analyse , col=T)$col.rel/100
      length      Comp1      Comp2      Comp3      Comp4      Comp5      Comp6      con.tra
reign.Bacteria  89.92 -3.28 -0.34 -0.04 4.86 0.57 12.50
reign.Bacteria  1.08 41.12 33.31 -2.66 -0.17 21.51 9.57
reign.Fungi     43.67 -1.39 -10.08 -1.88 -41.18 -1.10 11.23
reign.Metazoa   0.03 -3.40 2.40 93.56 -0.11 -0.25 10.95
reign.Viridiplantae -25.04 -12.36 -16.05 -17.51 19.94 -8.86 5.76
foldindex       7.89 55.46 15.14 -0.46 -0.13 -16.17 12.50
pi              -6.20 27.00 -53.14 1.18 -0.49 4.33 12.50
mw              89.75 -2.71 -1.02 0.00 5.45 0.62 12.50
gravity         5.88 60.39 -14.39 3.48 4.26 -0.99 12.50

```

Script 4

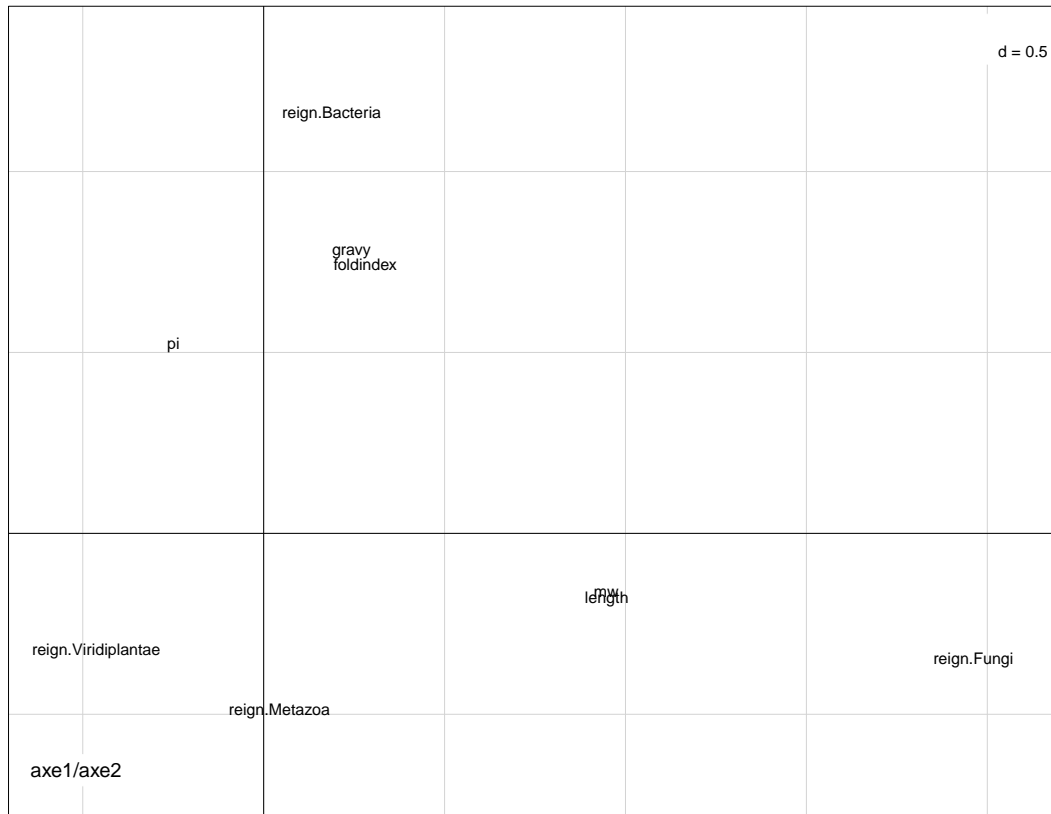


FIGURE 2 –

					reign.Bacteria	d =
		reign.Metazoa		foldindex		
		length mw				
		reign.Viridiplantae		gravy		
				pi		
		reign.Fungi				
e2/axe3						

FIGURE 3 –

Analyse du jeu de données

- Interprétez le listing ci-dessus (scripts 2 à 4, et figures 2 et 3). Précisez si vous avez des conditions à vérifier et si c'est le cas, comment vous le feriez en R. Précisez également s'il vous manque des informations dans ce listing et comment vous les obtiendriez en R.
- Concluez votre analyse.
- Que proposeriez-vous pour améliorer cette analyse ?

B) Sujet de mathématiques : Systèmes dynamiques

Exercice 3. On suppose que l'évolution de la taille x_n d'une population est décrite par une dynamique de Beverton-Holt $x_{n+1} = f(x_n)$ avec

$$f(x) = \frac{Rx}{1 + (x/M)} \quad R = 1.2 \quad M = 500$$

1. Rappeler une limitation importante des modèles de Malthus simples à laquelle ce type de dynamique permet de remédier.
2. Déterminer le sens de variation de la fonction f sur $[0, +\infty)$.
3. On note $g(x) = f(x) - x$. Dresser le tableau de signe de la fonction g sur l'intervalle $[0, +\infty)$. On trouvera que g admet deux zéros sur l'intervalle $[0, +\infty)$, dont on calculera les valeurs.
4. Dédire des questions 2. et 3. le comportement de la population x_n quand $n \rightarrow \infty$, selon la population initiale $x_0 \geq 0$. On précisera si la population s'éteint ($x_n \rightarrow 0$) ou explose ($x_n \rightarrow \infty$) ou se stabilise ($x_n \rightarrow x_\infty$ avec $x_\infty > 0$ fini) et, en cas de stabilisation, on calculera la valeur de x_∞ .

Exercice 4. Des nutriments sont absorbés par une cellule à un taux de α molécules par unité de temps et en sont éliminés à un taux de $\beta c(t)$ molécules par unité de temps, proportionnel à leur concentration $c(t)$ dans la cellule au temps t . Ce processus se traduit par l'évolution

$$c'(t) = \alpha - \beta c(t)$$

pour une concentration initiale $c(0) = c_0 \geq 0$ et des paramètres d'absorption $\alpha > 0$ et d'élimination $\beta > 0$ donnés.

1. On note $x(t) = \alpha - \beta c(t)$. Montrer que $x'(t) = \lambda x(t)$ pour une valeur de λ que l'on précisera en fonction de α et β . En déduire la valeur de $x(t)$ pour tout $t \geq 0$, en fonction de $x(0) = x_0$, α et β .
2. Exprimer $c(t)$ en fonction de $x(t)$. Exprimer x_0 en fonction de la concentration initiale c_0 . En déduire la valeur de $c(t)$ pour tout $t \geq 0$, en fonction de c_0 , α et β .
3. Montrer que, pour toute concentration initiale $c_0 \geq 0$, la concentration $c(t)$ se stabilise sur un équilibre c^* , au sens où $c(t) \rightarrow c^*$ quand $t \rightarrow \infty$, et calculer la valeur de c^* en fonction de c_0 , α et β .
4. Expliquer comment on aurait pu démontrer le résultat de la question 3. sans calculer la solution $c(t)$.

Exercice 5. Une population est composée d'individus jeunes en nombre $j(t)$ et d'individus matures en nombre $m(t)$. L'évolution en temps continu de cette population est basée sur les phénomènes suivants :

- les individus jeunes deviennent matures à un taux α ,
- les individus jeunes meurent à un taux de mort μ ,
- les individus matures meurent à un taux de mort θ ,
- les individus matures donnent naissance à de nouveaux individus jeunes avec un taux de fertilité φ .

1. Expliquer brièvement pourquoi ces hypothèses suggèrent d'étudier le système dynamique

$$\begin{cases} j'(t) &= -\alpha j(t) - \mu j(t) + \varphi m(t) \\ m'(t) &= -\theta m(t) + \alpha j(t) \end{cases}$$

On considère désormais le vecteur

$$X(t) = \begin{pmatrix} j(t) \\ m(t) \end{pmatrix}$$

2. Calculer en fonction des paramètres $(\alpha, \mu, \theta, \varphi)$ du modèle, la matrice A de taille 2×2 telle que l'évolution de $j(t)$ et $m(t)$ est donnée par $X'(t) = AX(t)$.

On suppose désormais que

$$\alpha = 0.2 \quad \mu = 0.3 \quad \theta = 0.2 \quad \varphi = 0.6$$

3. Calculer la trace et le déterminant de A . Montrer que le discriminant de A est positif.

4. Montrer que A admet deux valeurs propres réelles et calculer leur signe.

5. Utiliser les questions précédentes pour déterminer si, à partir de conditions initiales $j(0) = j_0 > 0$ et $m(0) = m_0 > 0$ données, la population considérée s'éteint ou si elle explose.