Compte rendu TP 1 SY09

ARTCHOUNIN Daniel / VALLOIS Célestin

30 mars 2016

Résumé

Dans le cadre du premier sujet des séances de Travaux Pratiques (TP) de l'Unité de Valeur (UV) SY09 enseignée à l'Université de Technologie de Compiègne (UTC), nous avons mené des analyses descriptives et des analyses en composantes principales (`ACP`) sur plusieurs jeux de données.

Dans un premier temps, nous avons mené une analyse descriptive sur une série de données portant sur des matchs de tennis dont le déroulement ou l'issue auraient été arrangés. Par ailleurs, nous avons également effectué une analyse descriptive de données portant sur 200 crabes.

Dans un second temps, nous avons réalisé une analyse en composantes principale sur une population constituée de 4 individus. De plus, nous avons effectué la même analyse sur un jeu de données portant sur des notes en utilisant certaines des fonctions directement disponibles sous `R`. Enfin, nous avons mené une ACP sur le jeu de données précédemment mentionné portant sur des crabes.

Le dossier `code_source` associé au présent rapport et contenant le code source `R` écrit afin de répondre aux différentes questions présentes dans le sujet s'organise ainsi :

— `1_dot_1.R` : le script `R` associé à la section `1.1` du sujet
— `1_dot_2.R` : le script `R` associé à la section `1.2` du sujet
— `2_dot_1.R` : le script `R` associé à la section `2.1` du sujet
— `2_dot_2.R` : le script `R` associée à la section `2.2` du sujet
— `2_dot_3.R` : le script `R` associée à la section `2.3` du sujet

1 Statistique descriptive

1.1 Le racket du tennis

Début 2016, une série d'articles paraît dans la presse à propos de matchs de tennis dont le déroulement ou l'issue auraient été arrangés. Les journalistes se basent sur une série de données agrégées dont l'étude met en évidence un certain nombre de matchs suspects.

Nous avons étudié un jeu de données pré traité issu du fichier `anonymous-betting-data.csv` caractérisant 129271 prises de position décrites par 16 variables. Le jeu de données pré traité ne contient que 126461 prises de position portant sur 25993 matchs. 1523 joueurs sont impliqués dans ces matchs qui ont eu lieu entre 2009 et 2015, soit, sur une période de 6 ans. Ces matchs ont fait l'objet de prises de position par 7 bookmakers.

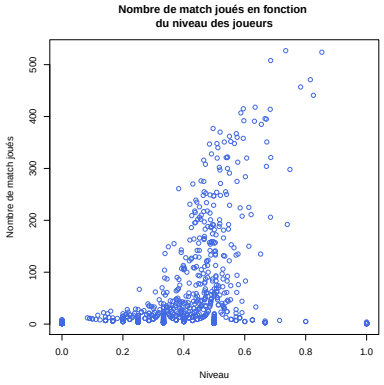

FIGURE 1 – Catégorisation du nombre de matchs joués en fonction du niveau des joueurs

Grâce aux données des 26000 matchs joués, nous avons pu calculer la propension de chaque

joueur à gagner un match. Parmi ces joueurs, seulement 899 ont au moins gagné un match et 1502 ont au moins perdu un match ce qui fait que 603 joueurs n'ont pas gagné de matchs mais ont seulement perdu. Enfin, on a pu remarquer que 66798 paris, soit 52,82% de l'ensemble, ont évolué vers le joueur gagnant

De plus, si l'on affiche le nombre de matchs de matchs joués par les joueurs en fonction de leur niveau, on peut observer les points suivants (figure 1) :

- Les joueurs qui possèdent un ratio extrême (proche de 0 ou de 1) sont surtout ceux ayant joué peu de matchs, on peut donc ne pas les considérer.
- La majorité des joueurs a un ratio compris entre 0.4 et 0.6 dès qu'un joueur a joué plus de de 100 matchs.
- Cependant, après le seuil des 100 matchs, on observe que les points semblent suivre une tendance linéaire, plus les joueurs ont joués de matchs, plus leur ratio est élevé. On peut expliquer cette tendance par le fait que les joueurs qui on joué un grand nombre de matchs sont sûrement ceux étant restés longtemps dans les tournois (et le circuit) : ils ont certainement un meilleur niveau de part leur longévité au sein du circuit professionnel et du fait qu'ils gagnent des matchs, ce qui semble cohérent.

Finalement on pourra noter que pour calculer le niveau nous avons effectuer un ratio entre le nombre de matchs gagnés par le nombre de matchs joués. Nous avons également pensé à faire un niveau en fonction de la moyenne des côtes d'entrées du joueur par match. Plus sa moyenne de côtes est faible, meilleur il est. Mais cela n'aurait pas résolu le problème lié aux joueurs ayant effectué très peu de matchs et répartis avec un niveau proche de 0 ou 1. Certes, ils auraient pu avoir un niveau plus réel mais sur un échantillon de 1 ou 2 matchs perdus, ce qui est très peu représentatif.

Dans le cadre de notre étude, on considérera comme suspects les matchs dont au moins un des paris présente une évolution de probabilité en valeur absolue strictement supérieure à 0.10. Ainsi, on remarque qu'il y a 2798 matchs suspects sur les 25993 étudiés, ce qui représente environs 10.76 % des matchs.

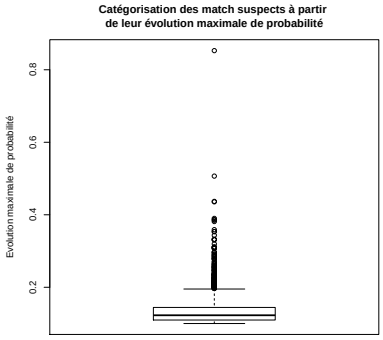

FIGURE 2 – Catégorisation des match suspects à partir de leur évolution maximale de probabilité

On remarque trivialement que plus de 75 % des matchs considérés comme suspects ont une évolution maximale de probabilité strictement inférieure à 0.2 (figure 2). Tous les outliers ont une évolution maximale de probabilité supérieure à 0.2. Ainsi, il serait probablement pertinent de s'intéresser particulièrement aux matchs présentant une évolution maximale de probabilité supérieure à 0.2. Par ailleurs, on notera également que tous les matchs, à l'exception d'un, présentent une évolution maximale de probabilité strictement inférieure à 0.6. Le match exceptionnel présente une évolution maximale de probabilité de 0.8, ce qui est particulièrement élevé et donc suspect.

Il s'avère que les 7 bookmakers ayant effectué des prises de positions sur les 25993 sont tous impliqués dans des matchs suspects. Dans le diagramme en bâtons représentant le nombre de paris suspects dans lesquels sont impliqués chaque bookmaker (figure 3), on peut noter que les bookmakers D, E, F et G sont tous impliqués dans moins de 300 paris suspects. Cependant, les bookmakers A, B et C sont tous impliqués dans plus de 1000 paris suspects. Il serait dont probablement pertinent de mener une étude plus approfondie sur ces trois derniers bookmakers.

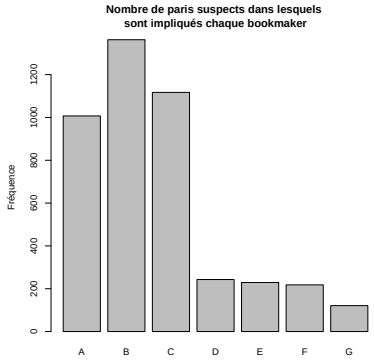

FIGURE 3 – Nombre de paris suspects dans lesquels sont impliqués chaque bookmaker

Dans notre analyse, on considère que les gagnants suspects sont ceux ayant gagné un match considéré comme suspect. De même, les perdants considérés comme suspects sont ceux ayant perdu un match considéré comme suspect. L'union des gagnants et des perdants suspects forme l'ensemble des joueurs suspects. Il est particulièrement intéressant de constater qu'il y a 559 perdants suspects, 455 gagnants suspects et 655 joueurs suspects. Ainsi, on peut facilement remarquer que 359 ($359 = (559 + 455) - 655$) des joueurs suspects sont à la fois des gagnants et des perdants suspects. En considérant qu'il est plus simple d'influencer l'issue d'un match en le perdant contre toute attente qu'en le gagnant contre toute attente, nous nous intéresserons plus particulièrement aux perdants suspects.

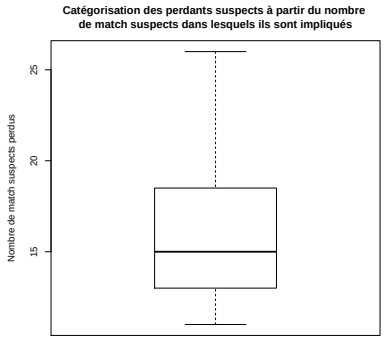

FIGURE 4 – Catégorisation des perdants suspects à partir du nombre de matchs suspects dans lesquels ils sont impliqués

Parmi ces derniers, seuls 87 ont perdu plus de 10 matchs considérés comme suspects (on pourra noter qu'il y en a 17 qui ont perdu exactement 10 matchs suspects). Plus de 75 % de ces 87 joueurs ont perdu moins de 20 matchs (figure 4). Il serait donc particulièrement pertinent de mener une étude plus approfondie sur les autres joueurs (ceux ayant perdu plus de 20 matchs suspects).

1.2 Données crabs

Le jeu de données considéré est disponible dans la bibliothèque de fonctions `MASS`. Il est constitué de 200 crabes de l'espèce *Leptograpsus variegatus* collectés à Fremantle, en Australie. Ces crabes sont décrits par 8 variables : 3 variables qualitatives et 5 quantitatives.

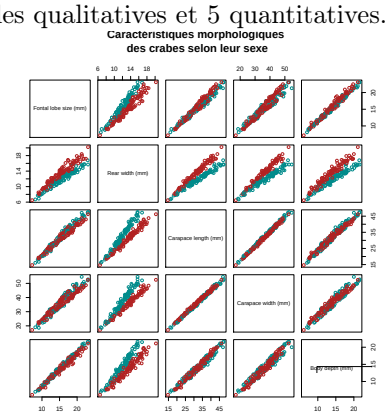

FIGURE 5 – Comparaison des caractéristiques morphologiques des crabes par sexe

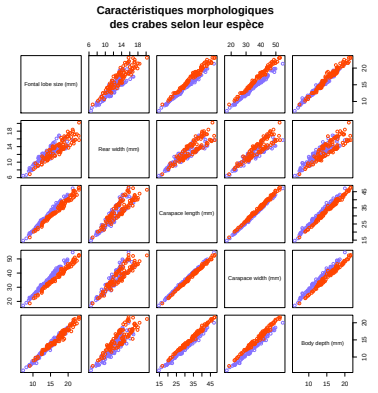

FIGURE 6 – Comparaison des caractéristiques morphologiques des crabes par espèce

Tout d'abord, il est particulièrement intéressant de constater que l'échantillon est constitué de 50 crabes mâles et bleu, 50 crabes mâles et

orange, 50 crabes femelles et bleu ainsi que de 50 crabes femelles et orange. Ainsi, chacune des quatre variétés de crabes est autant représentée que les autres.

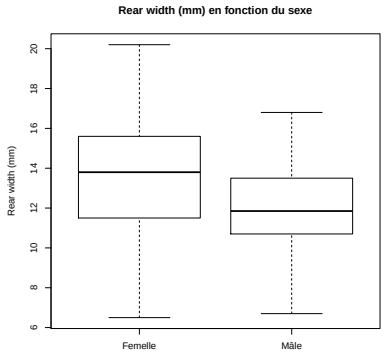

FIGURE 7 – Boxplot rear width (mm) en fonction du sexe

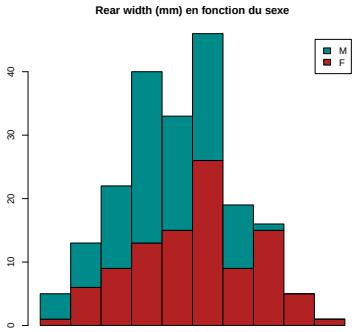

FIGURE 8 – Barplot rear width (mm) en fonction du sexe

En représentant les individus de l'échantillon selon leur sexe, tour à tour, en fonction de 2 variables quantitatives parmi les 5 à notre disposition (figure 5), il semblerait que le paramètre `Rear width` soit impacté par le sexe de l'individu. Afin de mieux observer ce phénomène, nous avons tenté de catégoriser les individus de la population selon leur sexe et leur réalisation pour le paramètre `Rear width`. Via les deux diagrammes en boîte consultables dans la figure 7, on constate que près de 50 % des femelles possèdent plus de 14 pour le paramètre `Rear width` tandis que plus de 75 % des mâles ont moins de 14 pour le même paramètre. Ainsi, notre hypothèse semble être confortée par ces deux diagrammes (figures 7 et 8).

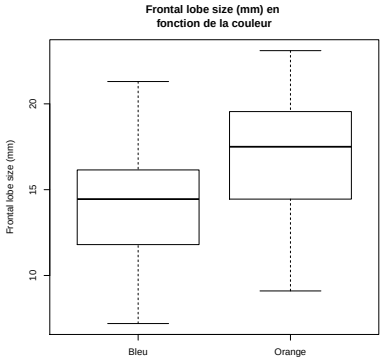

FIGURE 9 – Boxplot frontal lobe size (mm) en fonction de la couleur

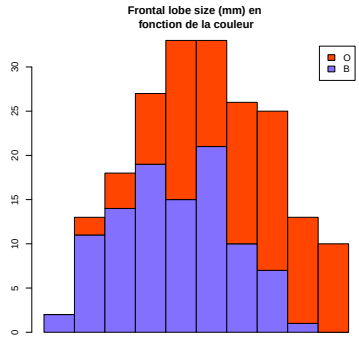

FIGURE 10 – Barplot frontal lobe size (mm) en fonction de la couleur

De la même manière, en représentant les individus de l'échantillon selon leur couleur, tour à tour, en fonction de 2 variables quantitatives parmi les 5 à notre disposition (figure 6), il semblerait que les paramètres `Frontal lobe size` et `Carapace width` soient impactés par la couleur de l'individu. Afin de mieux observer ce phénomène pour le paramètre `Frontal lobe size`, nous avons tenté de catégoriser les individus de la population selon leur couleur et leur réalisation pour le paramètre `Frontal lobe size`. Via les deux diagrammes en boîte consultables dans la figure 9, on constate que près de 75 % des crabes orange possèdent plus de 15 pour le paramètre `Fontal lobe size` tandis que plus de 50 % ont moins de 15 pour le même paramètre. Ainsi, notre hypothèse semble être confortée par ces deux diagrammes (figures 9 et 10).

En visualisant les graphiques de dispersion

de tous les couples de variables présents dans les figures 5 et 6, il semblerait que les points représentant les individus de notre échantillon se regroupent autour de droites. Ainsi, on peut logiquement penser qu'il y a des relations linéaires entre chaque couple de variables. Afin de vérifier cela, nous avons calculé la matrice des corrélations R (table 1) associée à la matrice liée au tableau individus-variables. Il s'avère que chacun des éléments de cette matrice a une valeur strictement supérieure à 0.85 : ce phénomène semble conforter notre hypothèse.

⟍	FL	RW	CL	CW	BD
FL	1.0000000	0.9069876	0.9788418	0.9649558	0.9876272
RW	0.9069876	1.0000000	0.8927430	0.9004021	0.8892054
CL	0.9788418	0.8927430	1.0000000	0.9950225	0.9832038
CW	0.9649558	0.9004021	0.9950225	1.0000000	0.9678117
BD	0.9876272	0.8892054	0.9832038	0.9678117	1.0000000

TABLE 1 – La matrice de corrélation

Afin d'illustrer nos dires, par ailleurs, nous avons mené un test de corrélation de Pearson entre les variables `Carapace length (mm)` et `Carapace width (mm)`.

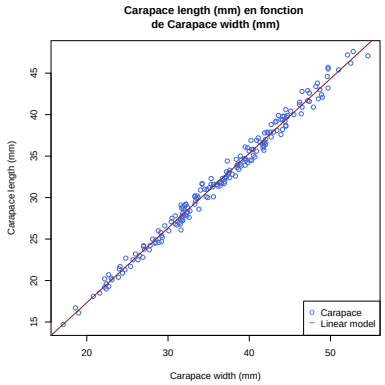

FIGURE 11 – Carapace length (mm) en fonction de Carapace width (mm)

Nous nous sommes permis d'utiliser ce test car nous jugeons que les réalisations de ces deux paramètres semblent être issues de variables aléatoires parentes de loi Gaussienne. De plus, d'après les figures 5 et 6, il y a absence de valeurs exceptionnelles pour les réalisations de ces deux variables. L'hypothèse nulle de ce test, notée H_0, est que le coefficient de corrélation entre ces deux variables est de 0, soit qu'il y ait absence de corrélation entre ces deux variables. Pour ce test, nous nous sommes fixés comme seuil de signification $\alpha^* = 0.01$. Cela signifie que l'on s'autorise à commettre une erreur de première espèce α inférieure à 0.01 (probabilité de rejeter l'hypothèse nulle H_0 sachant que cette dernière est vraie). Après avoir effectué le test, il s'avère que l'on retrouve le coefficient de corrélation 0.9950225 précédemment trouvé. Par ailleurs, l'intervalle de confiance pour ce coefficient de corrélation au niveau $1 - \alpha = 0.95$ est : $[0.9934242, 0.9962331]$. On remarque que cet intervalle est particulièrement petit pour un niveau de confiance si élevé, cela conforte à nouveau notre intuition. La valeur p (p-value) est strictement inférieure à 2.2^{-16}. Autrement dit, nous rejetons H_0 puisque cette dernière est inférieure à notre seuil de signification α^*. Ainsi, il y a donc a priori présence d'une relation linéaire entre les paramètres `Carapace length (mm)` et `Carapace width (mm)`. Afin d'exhiber ce phénomène, nous avons également réalisé une régression linéaire entre ces deux paramètres. Après calcul, on obtient une ordonnée à l'origine de $b = -0.6619$ et une pente de $a = 0.8998$. Le résultat associée à la régression linéaire est consultable dans la figure 11. Ce phénomène est particulièrement intéressant puisque en ayant recours à l'équation 1, en connaissance de `CW` (`Carapace width (mm)`), on peut prédire `CL` (`Carapace length (mm)`) et réciproquement.

$$CL = a * CW + b = 0.8998 * CW - 0.6619 \quad (1)$$

2 Analyse en composantes principales

2.1 Exercice théorique

Trois variables mesurées sur quatre individus forment la matrice Y présente dans l'équation 2.

$$Y = \begin{pmatrix} 3 & 4 & 3 \\ 1 & 4 & 3 \\ 2 & 3 & 6 \\ 4 & 1 & 2 \end{pmatrix} \quad (2)$$

Afin de calculer, les axes factoriels de l'`ACP`, nous nous sommes contentés de calculer la matrice de variance $V = X^T D_p X = \frac{1}{n} X^T I_n X$

associée à X, X étant la matrice Y ayant subi un centrage, Y étant la matrice initiale des données. A l'issue de ce calcul, nous avons calculé les valeurs propres $\lambda_1, \cdots, \lambda_p$ et leurs vecteurs propres normalisés associés $U = (U_1, \cdots, U_p)$ (ces données sont triées selon les valeurs propres dans l'ordre décroissant). Les axes factoriels correspondent aux vecteurs propres $U_1, \cdots, U_p$ de la matrice V.

Pour le calcul du pourcentage d'inertie expliquée E_{a_i} par l'axe i, où $i \in [[1, p]]$, nous nous sommes contentés d'appliquer la formule présente dans l'équation 3. Le résultat obtenue est consultable dans la figure 12.

FIGURE 12 – Scree plot des valeurs propres

$$E_{a_i} = 100 \frac{\lambda_i}{\sum_{k=1}^{p} \lambda_k} \tag{3}$$

Pour le calcul des composantes principales $C = (C_1, \cdots, C_p)$, il suffit d'appliquer la formule présente dans l'équation 4.

$$C = XMU = XI_pU \tag{4}$$

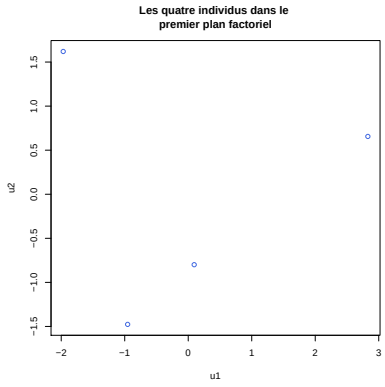

FIGURE 13 – Les quatre individus dans le premier plan factoriel

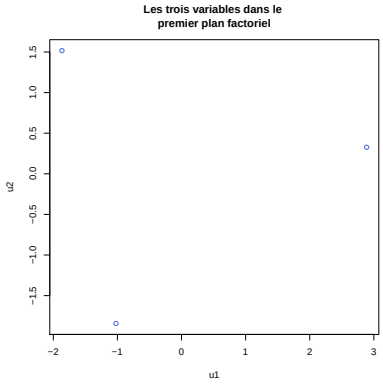

FIGURE 14 – Les trois variables dans le premier plan factoriel

Pour la représentation des quatre individus dans le premier plan factoriel, il suffit de ne représenter que les données associées aux deux premières colonnes de la matrice C (C_1 et C_2). Le résultat obtenu est consultable dans la figure 13.

Pour la représentation des trois variables dans le premier plan factoriel, il faut effectuer les même calculs que pour la représentation des quatre individus en n'oubliant pas d'effectuer ces derniers sur la transposée de Y. Le résultat obtenu est consultable dans la figure 14.

Nous savons que $XM = CU^T$. Ainsi, on peut en déduire les résultats figurant dans l'équation 5.

$$XM = X = CU^T = \sum_{\alpha=1}^{p} c_\alpha u_\alpha^T \tag{5}$$

Ainsi, on peut se permettre de calculer l'expression $\sum_{\alpha=1}^{k} c_\alpha u_\alpha^T$, pour $k = 1, 2$ et 3. Lorsque $k = 1$, on retrouve les coordonnées dans la base initiale des individus projetés dans le sous espace vectoriel E_1 de dimension 1. De même, lorsque $k = 2$, on retrouve les coordonnées dans la base initiale des individus projetés dans le sous espace vectoriel E_2 de dimension 2. Enfin, lorsque $k = 3$, on retrouve les coordonnées dans la base initiale des individus projetés dans le sous espace vectoriel E_3 de dimension 3, soit on retrouve la matrice centrée X.

2.2 Utilisation des outils R

Dans cette sous-section, on utilise les fonctions `R` afin d'effectuer l'`ACP` d'un jeu de données portant sur des notes.

Pour ce faire, comme dans l'exercice précédent, on commence par calculer la matrice des données centrées. Puis, on calcule la matrice V de covariance ainsi que les vecteurs propres et les valeurs propres lui étant associés. Ensuite, on calcule la matrice des composantes principales C. On cherche également à obtenir les contributions des axes aux individus et des individus aux axes. Enfin, on représente les variables dans la base des composantes principales.

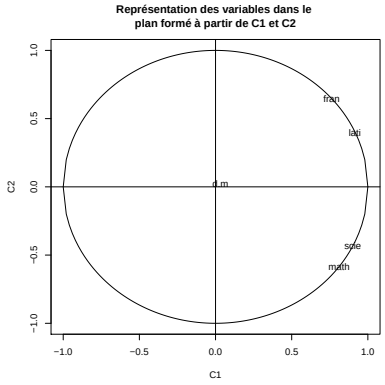

FIGURE 15 – Représentation des variables dans le plan formé à partir de C1 et C2

Nous avons représenté les 5 variables dans le plan formé à partir deux premières composantes principales (C_1 et C_2). Dans la figure 15, on constate que C_1 et C_2 résument les 4 premières variables (`fran`, `lati`, `scie` et `math`). Cependant, C_1 et C_2 ne donnent pas d'information sur la dernière variable (`d.m`).

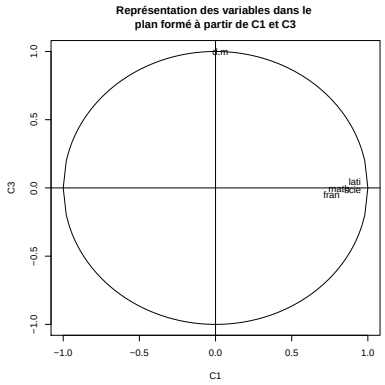

FIGURE 16 – Représentation des variables dans le plan formé à partir de C1 et C3

De plus, nous avons représenté les variables dans le plan formé à partir de la première et de la troisième composantes principales (C_1 et C_3). Dans la figure 16, on peut remarquer C_3 résume les 4 premières variables (`fran`, `lati`, `scie` et `math`) et que C_1 représente la dernière variable (`d.m`).

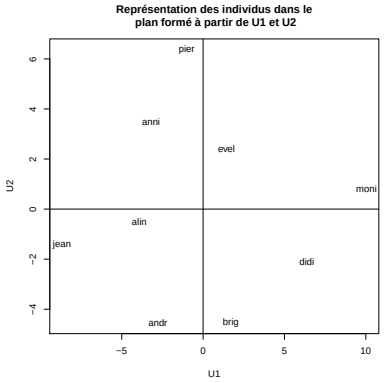

FIGURE 17 – Représentation des individus dans le plan formé à partir de U1 et U2

Par ailleurs, on a décidé de représenter dans la figure 17 les individus dans le plan formé des deux premiers axes factoriels (U_1 et U_2). Ce graphique souligne le fait que `jean`, `alin` et `andr` ont des notes peu élevées dans les 4 premières matières (`fran`, `lati`, `scie` et `math`).

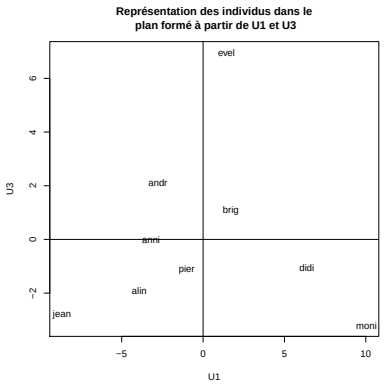

FIGURE 18 – Représentation des individus dans le plan formé à partir de U1 et U3

De plus, on a décidé de représenter les individus dans le plan formé du premier et du troisième axes factoriels (figure 18). On constate à nouveau que `jean`, `alin` et `andr` ont des notes peu élevées dans les 4 premières matières. De

plus, on note également que `anni` est dans le même cas. Enfin, on observe que `evel` a une note très élevée en (`d.m`). On remarque également que `moni` est très douée dans les 4 premières matières mais beaucoup moins en (`d.m`).

La fonction `princomp` de `R` effectue une `ACP` sur la matrice des données lui étant envoyée en paramètre et renvoie les résultats dans un objet de la classe `princomp`.

L'attribut `sdev` de l'objet retourné contient l'écart-type de chaque composante principale. En élevant au carré cet attribut (cela revient à calculer la variance de chaque composante principale), on retrouve logiquement les valeurs propres de la matrice de covariance V.

Sur l'objet retourné, la fonction `plot` affiche un diagramme en bâtons des valeurs propres (scree plot). Cela permet de trouver visuellement et simplement les composantes principales principalement responsables de la dispersion dans les données 19.

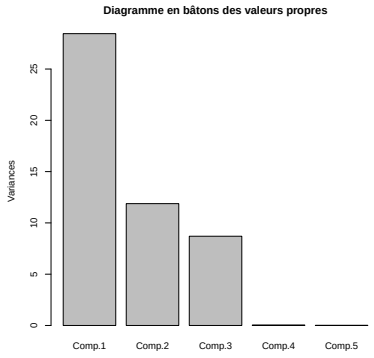

FIGURE 19 – Scree plot des valeurs propres

La fonction `biplot` permet de représenter les individus et les variables sur le même graphique. Ainsi, cela permet de projeter les variables sur un plan formé de composantes principales. Sur le même graphique, cela permet également de représenter les individus sur certains axes de la nouvelle base.

Ainsi, nous avons représenté les variables dans le plan formé à partir de C_1 et de C_2 ainsi que les individus dans le plan formé à partir de U_1 et U_2. Le résultat obtenu est consultable dans la figure 20.

De plus, nous avons représenté les variables dans le plan formé à partir de C_1 et de C_3 ainsi que les individus dans le plan formé à partir de U_1 et U_3. Le résultat obtenu est consultable dans la figure 21.

On peut effectuer les mêmes observations sur les figures 20 et 21 que celles effectuées sur les figures générées sans utiliser la fonction `princomp` de `R`.

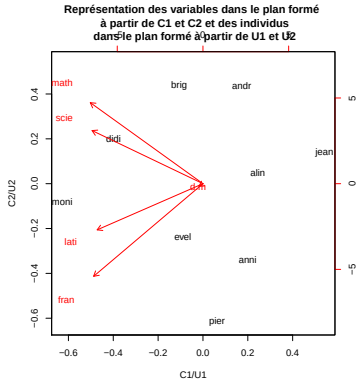

FIGURE 20 – Représentation des variables dans le plan formé à partir de C1 et C2 et des individus dans le plan formé à partir de U1 et U2

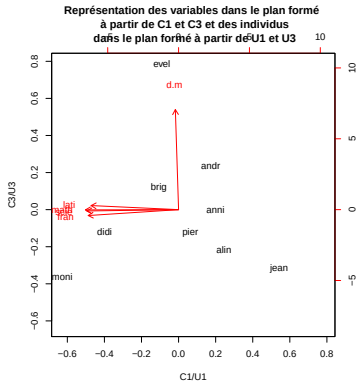

FIGURE 21 – Représentation des variables dans le plan formé à partir de C1 et C3 et des individus dans le plan formé à partir de U1 et U3

2.3 Traitement des données Crabs

On effectue l'`ACP` sur nos données `crabsquant` sans pré traitement pour le moment.

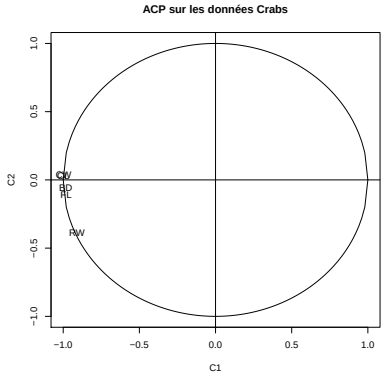

FIGURE 22 – Corrélation entre les variables des données Crabs et les composantes principales

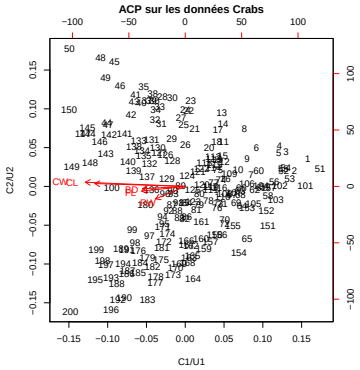

FIGURE 23 – Plan de représentation des données Crabs

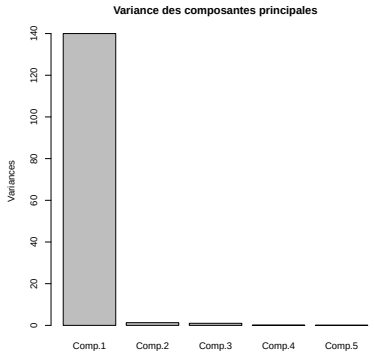

FIGURE 24 – Variance des composantes principales

On observe dans les figures 22 et 23 que les variables sont très corrélées à la première composante principale et que la variance de la première composante (la première valeur propre) est très élevée. En appelant la fonction *summary*, on constate que le pourcentage d'inertie expliquée par le premier axe est de 98%.

Ces résultats étaient particulièrement prévisibles. Effectivement, dans la sous section 1.2, nous avions stipulé qu'il y avait des relations linéaires entre chaque couple de variables quantitatives. Ainsi, le sous espace vectoriel $E_1 = \Delta u_1$ de dimension 1 maximisant l'inertie expliquée par l'axe Δu_1 capture ces relations linéaires entre chaque couple de variables quantitatives. Par conséquent, la projection des individus sur E_1 devrait limiter la perte d'informations.

Toutefois, afin d'améliorer notre représentation en termes de visualisation des différents groupes (les crabes mâles et bleu, les crabes mâles et orange, les crabes femelles et bleu ainsi que les crabes femelles et orange), on va donc pré traiter les données.

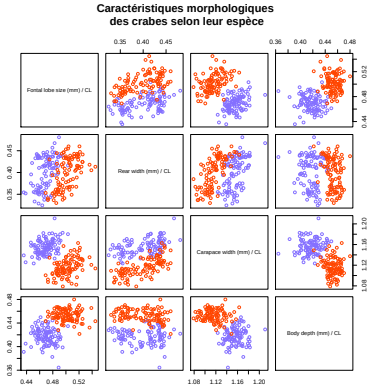

FIGURE 25 – Comparaison des caractéristiques morphologiques des crabes par espèce

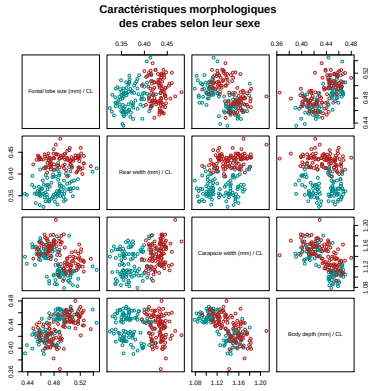

FIGURE 26 – Comparaison des caractéristiques morphologiques des crabes par sexe

Dans les figures 5 et 6, nous voyons que les lignes et les colonnes avec le paramètre `CL` nous permettent de dissocier les individus des différents groupes. Ainsi, nous suspectons que le rapport entre une variable autre que `CL` et le paramètre `CL` sont de "bons" indicateurs permettant de dissocier les individus. Dans les graphiques 25 et 26, nos hypothèses sur les 4 nouvelles variables sont confirmées.

En effectuant une `ACP` en se basant sur les quatre nouveaux indicateurs `FL/CL`, `RW/CL`, `CW/CL` et `BD/CL`, on obtient alors les résultats suivants.

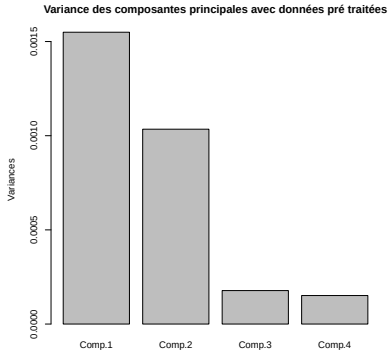

FIGURE 27 – Variance des composantes principales

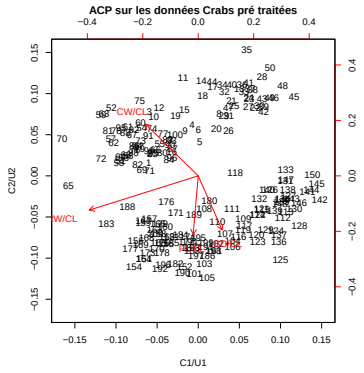

FIGURE 28 – Plan de représentation des données Crabs pré traitées (`FL/CL`, `RW/CL`, `CW/CL` et `BD/CL`)

Dans la figure 27, on constate logiquement que le pourcentage d'inertie expliquée par le premier axe factoriel est plus faible que celui dans la figure 24.

Dans la figure 28, on constate que le premier axe factoriel traduit principalement les variables `BD/CL` et `FL/CL`. Le deuxième axe factoriel semble traduire la variable `RW/CL`. Ainsi, on peut émettre les observations suivantes :

- un groupe a des rapports `BD/CL` et `FL/CL` relativement élevés mais un rapport `RW/CL` relativement bas
- un groupe a des rapports `BD/CL` et `FL/CL` relativement élevés ainsi qu'un rapport `RW/CL` relativement élevé
- un groupe a des rapports `BD/CL` et `FL/CL` relativement bas mais un rapport `RW/CL` relativement élevé
- un groupe a des rapports `BD/CL` et `FL/CL` relativement bas ainsi qu'un rapport `RW/CL` relativement bas

3 Conclusion

En conclusion, au cours de ce rapport, nous avons pu utiliser de façon concrète la puissance et les multiples possibilités qui s'offrent à nous en terme de traitement statistique de données avec `R`. Grâce à la statistique descriptive, nous avons appris à faire ressortir les données qui nous intéressent à travers un large jeu. Nous avons ensuite appliqué la méthode de l'`ACP` qui nous a considérablement aidés à obtenir des données analysables sous un plan qui ne s'offrait pas directement à nous d'un premier abord.