



**HAL**  
open science

# R++, logiciel d'analyse statistique ergonomique pour la médecine : l'ACP3D

Christophe Genolini, Marc Heinrich

## ► To cite this version:

Christophe Genolini, Marc Heinrich. R++, logiciel d'analyse statistique ergonomique pour la médecine : l'ACP3D. IHM'24 - 35e Conférence Internationale Francophone sur l'Interaction Humain-Machine, AFIHM; Sorbonne Université, Mar 2024, Paris, France. hal-04493682

**HAL Id: hal-04493682**

**<https://hal.science/hal-04493682>**

Submitted on 7 Mar 2024

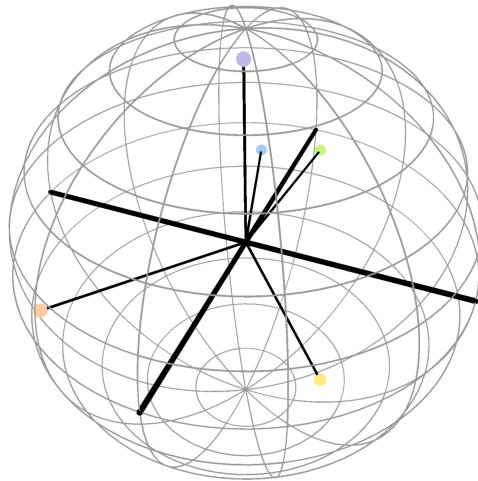
**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# R++, logiciel d'analyse statistique ergonomique pour la médecine : l'ACP3D

## R++, Ergonomic Statistical Analysis Software for Medicine: 3D PCA

CHRISTOPHE GENOLINI and MARC HEINRICH, zebrys, France



Cliquez sur la sphère  
pour la faire tourner.  
(ne fonctionne qu'avec le lecteur pdf  
«Acrobate Reader»)

Fig. 1. Représentation graphique d'une ACP3D offrant à l'utilisateur la possibilité de changer de point de vue grâce à la souris.

R++ est un logiciel d'analyse statistique particulièrement ergonomique à destination de la recherche médicale. Conçu en collaboration avec des médecins en utilisant les méthodes d'IHM (prototypage vidéo), le résultat est une grande réussite en terme de satisfaction client.

La prochaine version de R++ permettra aux utilisateurs de faire des ACP et des ACP3D, puis d'exporter les sphères 3D ainsi obtenues dans des PDF sous forme de graphiques 3D dynamiques. L'intérêt de ce genre de graphiques est que le lecteur peut lui-même manipuler le graphique, changer de point de vue et ainsi mieux visualiser les données. Un exemple d'un tel graphique est présenté ci-dessus.

R++ is a very ergonomic statistical analysis software designed for medical research. Developed in collaboration with physicians using IHM (video prototyping) methods, the result is a great success in terms of customer satisfaction.

The next version of R++ will allow users to perform PCA and 3D PCA, and then export the obtained 3D spheres into PDFs as dynamic 3D graphs. The advantage of such graphs is that the reader can manipulate the graph themselves, change the viewpoint, and thus better visualize the data. An example of such a graph is presented above.

CCS Concepts: • **Software and its engineering**;

Mots Clés et Phrases Supplémentaires: Techniques d'interaction, Modèles, ACP, statistiques

Additional Key Words and Phrases: Interaction technique, Models, statistics, PCA

### Reference:

Christophe Genolini and Marc Heinrich. 2024. R++, logiciel d'analyse statistique ergonomique pour la médecine : l'ACP3D. *IHM '24 : Articles Industriels de la 35<sup>e</sup> conférence Francophone sur l'Interaction Humain-Machine, March 25–29, 2024, Paris, France*

## 1 INTRODUCTION

L'analyse statistique devient progressivement le pivot incontournable de nombre de domaines. Dans le monde académique, on utilise des comparaisons de groupes, les estimations d'un modèle de régression, ou encore le machine learning. Dans le privé, les assureurs l'utilisent pour définir leurs tarifs, les banquiers s'en servent pour décider de l'octroi d'un prêt, l'industrie pharmaceutique valide ses essais cliniques et, de manière plus générale, tous les grands comptes rêvent d'un modèle qui permettrait de savoir qu'un client va partir avant même que le client ne prenne sa décision...

Les utilisateurs de l'analyse statistique sont donc très variés. Les data scientists sont formés à l'utilisation des logiciels. Mais beaucoup d'autres corps de métiers ont besoin d'utiliser des statistiques sans pour autant être experts. Pour eux, l'analyse statistique n'est pas leur activité première. C'est un outil qui vient en complément de leur métier principal. Par exemple, les médecins-chercheurs font régulièrement quatre à six mois de clinique pendant lesquels ils ne font plus du tout de statistiques, puis ils utilisent les statistiques pour l'analyse de leur données pendant un à deux mois.

Autre caractéristique des utilisateurs occasionnels, leurs besoins sont assez ciblés. Ils n'utilisent que les statistiques de leur corps de métiers. Par exemple, un médecin manipule régulièrement la survie (Modèle de Cox) et ignore tout du forecast. Inversement, dans le milieu de la logistique, personne ne connaît la survie alors que le forecast est un outil quotidien.

Du point de vue soft, les logiciels métiers (conçu par les statisticiens comme SAS, R, SPSS, Stata...) comme les solutions informatiques (Oracle, C, python) ont en commun d'être intégrés dans des interfaces particulièrement difficiles à utiliser. Elles ne sont pas adaptées aux besoins des utilisateurs non-programmeurs et non-statisticiens. Autre point commun, ces logiciels sont le plus exhaustif possibles en terme de nombre de méthodes statistiques afin de pouvoir satisfaire le plus grand nombre possible d'utilisateurs.

Le projet R++ a pour objectif de développer un logiciel d'analyse statistique ergonomique à destination d'utilisateurs ayant un métier principal et pour qui l'analyse statistique est un outil secondaire. Comme différents types d'utilisateurs peuvent avoir des besoins variés et que proposer trop d'outils dans un seul logiciel complexifie son interface, R++ est consacré à un unique corps de métier : la recherche médicale.

Dans cet article, nous présentons le logiciel R++ et nous introduisons une nouvelle méthode, l'ACP3D. La Section 2 présente la méthode de prototypage vidéo qui nous a permis d'arriver à la création de R++. Section 3 présente le logiciel R++ dans sa version classique. Section 4 détaille la méthode statistique ACP, ses limites et présente l'ACP3D. Section 5 conclut l'article.

## 2 MÉTHODES

R++ est un logiciel d'analyse statistique conçu pour offrir une approche ergonomique et intuitive. L'interface de R++ a été conçue de manière collaborative lors de 19 séances de prototypage vidéo. Le résultat est une grande réussite en terme de satisfaction clients puisque 92% des médecins-chercheurs qui voient R++, veulent R++. La force de notre logiciel réside autant dans les méthodes proposées que dans les méthodes qu'il ne propose pas. En effet, R++ n'intègre que les analyses qui ont été directement demandées par des médecins. L'interface résultante est particulièrement épurée. Pas de menu déroulant ou de sous menu. Tout est directement visible. Si un utilisateur se souvient vaguement d'une fonctionnalité mais ne sait plus où elle se situe, il n'a qu'à essayer 5 ou 6 boutons avant de trouver. L'IHM de R++ est conçue pour guider l'utilisateur à travers des processus d'analyse de données de manière intuitive. Cette approche simplifie considérablement la courbe d'apprentissage. En 1h de temps, la majorité de nos utilisateurs sont autonomes.

## 2.1 Prototypage vidéo

Pour concevoir l'IHM de R++, nous avons utilisé la méthode du design exploratory et du prototypage vidéo : le principe est de réunir pendant trois heures 4 ou 5 utilisateurs (des chercheurs en santé ayant besoin de statistiques dans le cas présent) et 4 ou 5 spécialistes de l'IHM. La séance est ensuite découpée en trois phases.

(1) Pendant une heure, nous lançons un brainstorming sur ce qu'il faut améliorer dans les logiciels actuels : « Quelles sont avec vos outils actuels les tâches compliquées, chronophages, pénibles, fastidieuses ou à fort risque d'erreur ? » Ce tour de table est fait sur le mode brainstorming ouvert : les participants interviennent et un modérateur note leurs remarques au tableau. À la fin de cette première étape, le groupe sélectionne collectivement deux ou trois thèmes.

(2) Dans une deuxième étape, le groupe cherche des solutions aux thèmes sélectionnés. La formule introductive est : « Imaginez : vous marchez sur une plage. Vous trouvez une lampe. Vous la frottez et une fée en sort. Chance, vous avez le droit à trois vœux. Par contre, la fée est la fée des logiciels d'analyse statistique (!), les trois vœux ne peuvent servir que pour améliorer un logiciel d'analyse statistique. Comment fonctionnerait le logiciel d'analyses statistique de vos rêves ? » Pendant 10 minutes, les participants doivent noter par écrit trois idées pour résoudre les problèmes choisis, plus une idée « farfelue ». L'objectif de l'idée farfelue est d'éviter l'autocensure : en effet, un participant peut avoir une bonne idée mais ne pas la dire par peur d'être ridicule. La consigne de l'idée farfelue permet d'éviter cela. Quand tout le monde a trouvé ses trois plus une idées, elles sont présentées au reste du groupe, puis elles sont débattues, combinées et améliorées.

(3) Enfin, dans une troisième heure, les participants créent des prototypes basses fidélités. À l'aide de papiers, feutres, post-it et découpage, ils fabriquent un scénario qu'ils filment à l'aide d'un téléphone. Ces prototypes sont ensuite présentés à l'ensemble du groupe. Cela permet d'obtenir un premier retour.

Bilan, en à peine 3 heures, le groupe a identifié de vrais problèmes utilisateurs, puis a trouvé collectivement des solutions et enfin a pu avoir un premier retour sur les solutions.

## 2.2 Choix des méthodes statistiques « de bases » intégrées dans R++

Comme indiqué dans l'introduction, R++ est dédié aux analyses élémentaires du domaine médical. Dès lors, le choix des méthodes à intégrer dans le logiciel a été assez restreint : analyse univariée, les quatre tests statistiques de base ( $\chi^2$ ,  $t$  de Student,  $f$  de Fisher et corrélation de Pearson), leurs équivalents non paramétriques, les équivalents appariés, les régressions linéaires et logistiques, les courbes ROC et les modèles de Survie (Courbes de Kaplan-Meier, log-rank et modèle de Cox). Rien de plus.

Pour toutes ses méthodes, nous avons fait le choix de l'hyper-simplicité. Toutes sont accessibles en précisément 2 ou 3 clics souris. Pour obtenir un tel résultat, nous avons choisi de fortement limiter les options proposées à l'utilisateur. Par exemple, il existe deux variantes du test  $t$  de Student. Le premier considère que les variances des groupes sont égales. Le second considère que les variances des groupes sont différentes. En pratique, cela change la méthode du calcul du degré de liberté mais cela n'a qu'une incidence mineure sur la significativité. Dans R++, nous avons décidé de proposer uniquement le  $t$  de Student avec variances inégales, le plus conservateur des deux tests. Proposer les deux méthodes aurait complexifier l'interface. De plus, nos utilisateurs qui ne savent généralement pas qu'il existe deux variantes du  $t$ , n'auraient pas su comment faire pour choisir l'une au l'autre. De même, il existe de nombreuses méthodes de régressions linéaires que même les statisticiens chevronnés ne maîtrisent pas. R++ propose une unique méthode, la classique « moindres carrés ».

La décision d'intégrer des méthodes plus avancées comme l'ACP présenté dans cet article, ou encore des méthodes de classification non supervisée comme k-means se fait dans le respect de ce principe hyper-simplicité. Pour l'ACP, notre interface permet à l'utilisateur de simplement cliquer sur les variables qu'il veut inclure dans son modèle. Une ACP à 10 variables nécessite 11 clics, un sur « ACP » et un sur chacune des variables.

Pour k-means classique, l'utilisateur doit par exemple définir une distance inter-individus, une distance intergroupe, une méthode de gestion des manquantes. Rien qu'avec ces trois paramètres, il peut obtenir plus de 500 configurations différentes. Dans R++, k-means utilise la distance Euclidienne, la distance intergroupe entre les barycentres et les ajustements de Gower pour la gestion des manquantes.

Au final, même lorsque nous introduisons de nouvelles méthodes plus avancées dans R++, nous prenons garde à toujours favoriser la simplicité.

### 3 PRÉSENTATION DE L'INTERFACE

L'interface de R++ est divisée en 3. Sur la gauche, un « chemin de fer » est constitué de 5 onglets reprenant les 5 grandes étapes d'une analyse statistique en médecine : 1/ l'import des données, 2/ le nettoyage, 3/ les tests statistiques, 4/ la modélisation et enfin 5/ l'export des résultats vers l'article scientifique. En haut se trouve la barre d'outil. Et au centre de l'écran, on trouve la fenêtre principale.

#### 3.1 Import

L'import est assez classique, il permet d'ouvrir un fichier à plat. Les médecins lisent du csv ou de l'Excel. Notez qu'à la différence d'autres corps de métiers comme par exemple des informaticiens, jamais personne ne nous a demandé une connexion avec une base de donnée.

#### 3.2 Le nettoyage de donnée

Il propose différents outils :

- Coloration des colonnes selon leur type : à chaque type de variable<sup>1</sup> est associé une couleur (orange pour binaire, jaune pour entier, vert pour numérique, bleu pour nominale, etc). Si une variable n'est pas de la couleur qu'elle est supposée avoir, c'est qu'il y a probablement un problème de typage. Par exemple, si la variable Genre n'est pas orange, c'est probablement qu'elle a été mal codée. Un examen détaillé montrera qu'elle comporte 4 modalités qui sont Homme, Femme, homme et femme (comme l'exemple présenté Fig. 2). De même, si une variable supposée entière est verte, c'est qu'elle contient probablement des numériques. Et ainsi de suite.
- Graphique en un clic : affiche sous chaque colonne un graphique qui représente le mieux la variable, en accord avec son type. Cette option, particulièrement appréciée, donne en un coup d'œil un aperçu de toute la base de données. Elle permet de détecter les variables dont le nombre de modalités n'est pas le bon (comme GroupeSanguin sur l'exemple Fig. 2) ou encore les valeurs aberrantes des variables numériques (comme la variable taille).
- Résumé en un clic : affiche sous chaque colonne un résumé statistique en accord avec son type. Moyenne, écart type, quartiles, nombre de manquantes, non manquantes pour les numériques et les entières; effectifs, pourcentage, manquantes, non manquantes pour les nominales, ordonnées et binaires.

1. Le terme variable est utilisé ici selon la définition des statisticiens : une variable est une mesure qui a été effectuée chez tous les individus et qui, par définition, *varie* d'un individu à l'autre. Autre manière de voir les choses, une variable est une colonne (obligatoirement) dans un tableau Excel. Si on travaille sur une population de  $n$  individus, toutes les variables sont des vecteurs de taille  $n$ . De fait, dans la plupart des lignes cet article, variable et colonne sont des termes interchangeables.

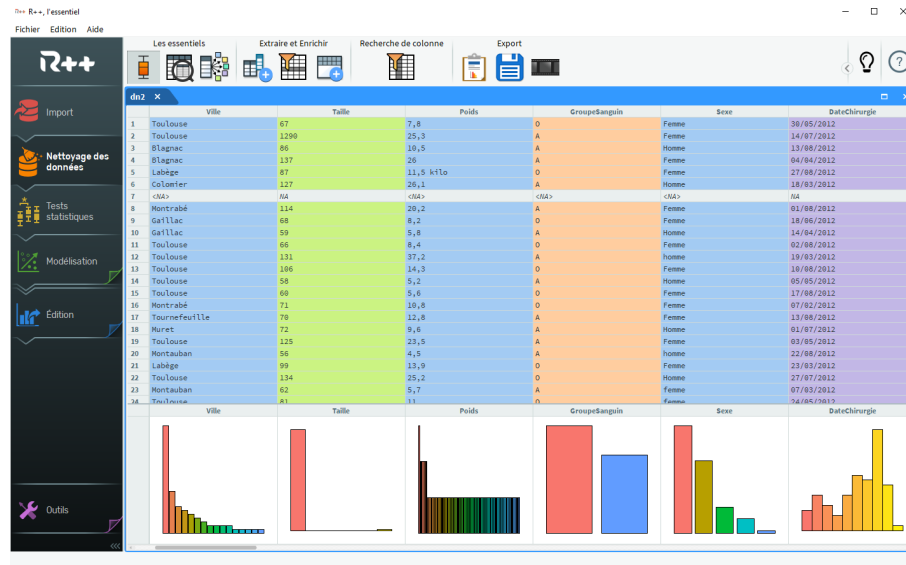


Fig. 2. Détection des variables problématique en un « clin d'œil ». L'histogramme de Taille présente une valeur aberrante; Poids est bleu (nominale) au lieu de vert (numérique), cela indique la présence d'au moins une chaîne de caractère dans la colonne (la valeur « 11.5 kilo » ligne 5); GroupeSanguin est orange (binaire) ce qui signifie que les modalités B et AB sont absentes de la colonne. Sexe est bleu (nominale) alors qu'elle devrait être binaire (orange) ce qui est lié à des codages différents pour la même modalité (« Homme / homme / H »)

- Le typeur : permet de changer le type d'une variable soit de manière canonique (par exemple entier vers numérique), soit en demandant à l'utilisateur de corriger des erreurs. Par exemple, si la variable Genre est nominale et que l'utilisateur veut la transformer en Binaire, R++ ouvrira une interface permettant à l'utilisateur de fusionner les modalités jusqu'à ce qu'il n'en reste plus que deux. Le typeur permet en particulier de corriger les erreurs de typage. Si Genre est du type nominale parce qu'il est composé de 3 modalités Homme, Femme et H, forcer le type à Binaire va permettre à l'utilisateur de fusionner Homme et H.
- Ajout de colonne : l'ajout de colonne fonctionne en sélectionnant une colonne de référence. Puis l'interface propose différentes options selon le type de la colonne de référence. Pour une variable numérique, elle propose de la découper en intervalle. Pour une variable date, elle calcule une durée par différence avec une autre variable date. Pour une variable nominale à k modalités, elle la découpe en k indicatrices binaires.
- Les autres options incluent un filtre ligne, un filtre colonne, l'export des données nettoyées au format csv et un rapport automatique exportant un résumé (graphique en un clic et résumé en un clic) de toutes les variables vers Word.

### 3.3 Les tests statistiques

L'étape test statistique commence par demander à l'utilisateur de choisir une colonne de référence. Ensuite, toutes les actions de l'étape tests statistiques seront relatives à cette colonne.

- Graphiques bivariés en un clic : affiche sous chaque colonne le graphique bivarié qui représente le mieux la colonne relativement à la colonne de référence, en accord avec les types des deux colonnes. Mosaïc plot, boîtes

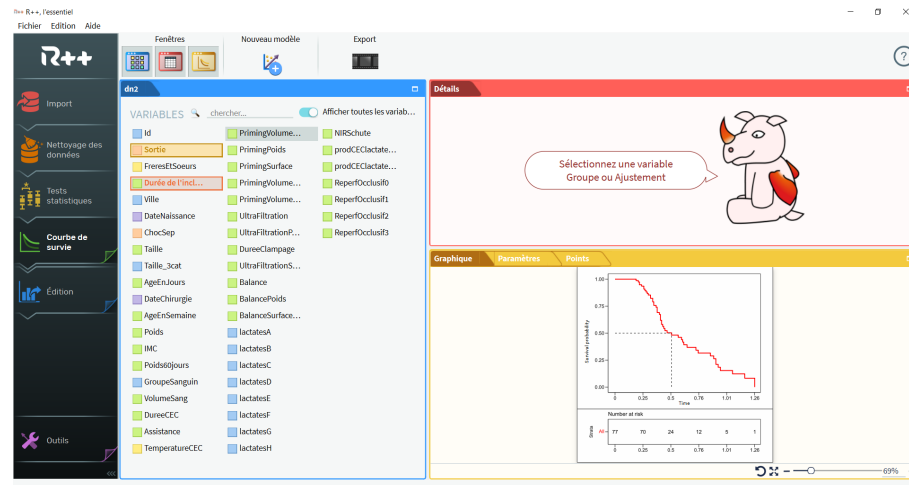


Fig. 3. Un modèle de Survie. La variable durée (en rouge) et la variable censure (en orange) ont été choisies. L'aide indique qu'il faut maintenant choisir la variable groupe.

à moustaches multiples, nuage de point ou barplot bivariés selon les cas. Cette option donne en coup d'œil un aperçu de toute les variables qui ont potentiellement un lien avec la variable de référence.

- Résumé en un clic : affiche sous chaque colonne un résumé statistique bivarié en accord avec les types des deux variables (analyse univarié par groupe ou tableau croisée selon les cas).
- Tests Statistiques en un clic : affiche sous chaque colonne les tests statistique entre la variable et la variable de référence. A chaque fois, les tests paramétriques et non paramétriques sont calculés. Une aide rappelle les règles qui permettent de choisir entre les deux. En complément, une proposition de rédaction d'article est également proposée.
- Tests appariés : permet d'apparier deux variables, ce qui transforme les tests correspondant en tests appariés.

### 3.4 Modélisation simple

Après avoir fait des tests bivariés simples, les médecins utilisent des modèles multivariés. R++ propose 4 modèles : les régressions linéaires, les régressions logistiques, les courbes ROC et le modèle de Cox. A chaque fois, la liste des variables est affichée sur la gauche de l'écran, une aide est sur la droite. Le principe est tout simplement de choisir les variables du modèle en cliquant dessus dans un ordre précis. Cet ordre est précisé par l'aide. Par exemple, une régression linéaire nécessite (1) une variable à expliquer (numérique) et (2) des variables explicatives. L'aide affiche cet ordre, l'utilisateur doit simplement cliquer sur les variables. Un modèle de Cox nécessite une (1) variable durée (numérique), (2) une variable censure (binaire), (3) une variable groupe (nominale ou binaire) et (4) des variables d'ajustement. L'aide affiche l'ordre et l'utilisateur doit simplement cliquer sur les variables comme le montre la figure 3. A noter, toutes les variables non utilisables à une étape précise sont grisées : par exemple, quand il faut choisir une variable binaire, toutes les variables non binaire sont non cliquables.

- Régression linéaire : établie la relation linéaire entre une variable d'intérêt  $Y$  de type numérique et une combinaison linéaire de variables explicative les  $X_i$ .

- Régression logistique : établie la relation entre une variable d'intérêt  $Y$  de type binaire et une combinaison linéaire de variables explicatives. Plus précisément, le lien est construit entre le logit de  $Y$  et les explicatives, avec  $\text{logit}(Y) = \log\left(\frac{Y}{1-Y}\right)$ . Au final, le lien entre  $Y$  et un des  $X$  peut plus ou moins s'interpréter comme un Odd Ratio, concept que les médecins manipulent avec une dextérité remarquable.
- Courbes ROC : étant donnée une variable représentant un symptôme binaire (avoir le Covid oui/non) et des colonnes numériques qui sont candidates pour devenir prédictrices de ce symptôme, on cherche la meilleure prédictrice. La courbes ROC permet de montrer les performances relatives des différentes prédictrices.
- Modèle de Cox (ou courbes de Survie) : il s'utilise dans le cas où on mesure un phénomène dans lequel on s'attend à avoir beaucoup de perdus de vue. Par exemple, lorsque l'on suit des cohortes de personnes âgées sur une maladie particulière, comme l'apparition de parkinson, on sait que l'on va avoir beaucoup de perdus de vu pour des raisons autres (décès, ou des maladies invalidantes). Le modèle de Cox permet de garder dans le modèle les perdus de vue sans les exclure comme on le ferait pour des valeurs manquantes (voir exemple Fig. 3).

### 3.5 L'Export

L'export permet la mise en forme des résultats avant de les copier-coller vers un autre logiciel. Deux outils sont proposés :

- Editeur graphique : il permet de personnaliser les graphiques en modifiant tous les paramètres (styles de graphiques, couleurs, polices, etc).
- Editeur de table 1 : de nombreuses revues de médecine demandent à ce que la première table de l'article soit un tableau résumé des caractéristiques du groupe Intervention par rapport au groupe Témoin (ou la population générale). Il y a tant d'hommes dans le groupe I et tant d'hommes dans le groupe T, la moyenne d'âge est de tant dans le groupe I et de tant dans le groupe T, ainsi de suite. Un tel tableau n'est pas difficile à construire, il est juste particulièrement chronophage. L'éditeur de table 1 permet de construire ce tableau en quelques clics.

## 4 L'ANALYSE EN COMPOSANTE PRINCIPALE EN 3D

La prochaine version de R++ proposera un cinquième modèle, l'ACP. L'Analyse en Composantes Principales est une technique statistique multivariée qui permet de réduire le nombre de dimensions  $N$  des données tout en maximisant la conservation de la variance.

Au final, les données de dimensions  $N$  sont représentées en 2D. Ce choix est fait en espérant que ces deux dimensions seront suffisantes pour correctement résumer les données. Si ça n'est pas le cas, de l'information est perdue. Une alternative est alors de résumer l'information en utilisant non pas deux mais trois dimensions. Se posent alors tous les problèmes liés aux graphiques 3D : production, manipulation et surtout une visualisation qui permette de garder l'aspect 3D. Tant que l'utilisateur reste dans R++, la solution est simple : l'ACP3D est représenté sur une sphère. L'utilisateur peut la faire tourner avec la souris et visualiser les données en 3D.

Pour l'export, depuis quelques années, il est possible d'intégrer des graphiques 3D dynamiques dans du PDF. L'opération est compliquée : elle nécessite de passer par un logiciel de production de graphiques 3D dynamiques (par exemple R), puis de convertir le graphique au format Asymptote, utiliser Asymptote pour créer un .tex, et enfin utiliser LaTeX pour créer un pdf. De plus, l'ensemble de la procédure est très mal documentée. Pour ces raisons, rares sont ceux qui utilisent cette fonctionnalité.

C'est néanmoins possible. Aussi, dans R++, nous avons décidé de simplifier la procédure et nous avons intégré une option « génération d'un graphique 3D dynamique en un clic » : quand le graphique 3D est prêt dans R++, un simple



clic permet de générer un graphique Azy, d'appeler Asympote puis LaTeX et de générer le graphique 3D dynamique dans un pdf. Un utilisateur qui reçoit le pdf peut alors le faire tourner. Il se rendre compte par lui-même de la distance qui sépare des points qui seraient proches sur une projection 2D. Un exemple d'une tel graphique 3D est présenté Fig 1.

Au final, la procédure complète de production d'une ACP3D est :

- Etape ACP : dans l'étape ACP, construction d'une ACP. Sur le diagramme des valeurs propres, on étudie le nombre de dimensions principales. S'il y en a 2, l'ACP2D est suffisantes. S'il y en a 4 ou 5, l'ACP 2D est biaisé, l'ACP 3D sera également biaisée (même si elle le sera moins). S'il y en a 3, on est dans le cas où l'ACP 3D prend toute sa justification.
- On exporte l'ACP3D dans l'éditeur de graphique 3D. On peut alors mettre le graphique en forme : changer les couleurs, les axes, augmenter ou diminuer le nombre de graduations ou encore les lignes qui permettent de visualiser la sphère.
- Quand c'est fait, on exporte le graphique 3D en pdf.

A noter, seul le lecteur « Adobe Reader » permet d'exploiter la fonctionnalité « Graphique dynamique ». Les autres lecteurs de pdf se contentent d'afficher un graphique immobile, ou pas de graphique du tout.

## 5 CONCLUSION

Dans cet article, nous vous avons présenté R++, un logiciel d'analyse statistique particulièrement ergonomique à destination de la recherche médicale. R++ répond aux besoins des médecins en proposant les outils dont ils ont besoin et uniquement ceux ils ont besoin, ce qui assure à son interface de rester simple, épurée, facile à apprendre.

Cette interface était jusqu'à présent découpée en 5 parties : import des données, nettoyage, tests statistiques, modélisation et export. Une nouvelle étape vient de faire son apparition, la modélisation avancée. Cette étape intègre le clustering, l'ACP et l'ACP3D, ACP3D dont il est possible d'extraire un graphique3D et de l'envoyer très simplement dans un PDF. Un tel graphique appelé Graphique 3D dynamique peut être manipulé par le lecteur. Cela lui permet de changer l'angle de vue et ainsi d'avoir une meilleure représentation des données.

Il reste encore de très nombreux domaines dans lesquels l'analyse statistique pourrait bénéficier des apports de l'IHM. R++ est très simple d'utilisation. Mais malgré cela, certains médecins réfractaires à l'outil informatique (de leur propre aveu) peinent à l'utiliser. L'introduction de techniques issues du serious gaming pourrait faciliter plus encore l'apprentissage.

Les langages ML ouvrent un grand nombre de possibilités. Dans le cadre d'un logiciel d'analyse statistique, un ChatBot pourrait apprendre à utiliser la documentation dans son intégralité (le manuel utilisateur, les fiches d'aides individuelles, les tutoriels). Devenu expert du logiciel, il pourrait alors lui-même enseigner à l'utilisateur comment faire pour se servir de R++. L'utilisateur poserait ses questions directement au ChatBot, par écrit ou directement à l'oral via les fonctions de reconnaissances vocale, au lieu d'avoir à lui-même avoir à parcourir l'aide.

Enfin, les graphiques 3D dynamiques ne sont pour l'instant pas imprimables. Nous travaillons actuellement avec l'équipe du Professeur Dumbledore (Poudlard) à la création d'une imprimante qui permettrait d'imprimer des graphiques 3D dynamiques, puis de les faire bouger après impression.