Exercices avec le logiciel R

Épreuve Biologie & Modélisation - Contrôle terminal - 2 juin 2010

M. Bailly-Bechet & S. Mousset

Durée 1h30 (2×45 min)

Tous documents autorisés - échanges strictement interdits.

Répondre directement sur la feuille

Nom :
Prénom :
Numéro de la carte d'étudiant :

1 Statistiques et R.

1.1 Introduction

Dans un article d'Avril 2006[1], des chercheurs américains ont cherché à estimer l'impact de la prière sur les conséquences d'une opération du cœur. Plus précisément, ils ont cherché à savoir si :

- ⋆ La prière a une incidence sur les conséquences post-opératoires, après une opération de pontage coronarien (opération consistant à dévier le flux sanguin autour d'une zone malade de l'artère coronaire près du coeur).
- ⋆ Le fait, pour le malade, de savoir avec certitude que des gens ont prié pour lui, a un effet sur sa convalescence.

Les auteurs précisent bien que leur but est purement scientifique, et que leur étude "n'a pas été conçue pour, et ne pourrait répondre à un grand nombre de questions religieuses, telles que l'existence de Dieu, le fait que Dieu réponde aux prières, ou encore le fait que les prières d'un groupe religieux particulier ont le même effet que celle d'un autre groupe"[1].

L'étude est divisée en trois groupes de tailles respectives $n_1 = 604$, $n_2 = 597$, et $n_3 = 601$. Toutes les personnes dans ces trois groupes ont subi la même opération du coeur. La seule différence réside dans le fait que les patients savent ou pas que l'on prie pour eux, et que c'est effectivement le cas :

- ⋆ Les gens du premier groupe ne savent pas que l'on prie pour eux, mais un groupe de prière a effectivement prié pendant 14 jours pour eux, à partir de la veille de leur opération.

[1] En anglais dans l'article original.

- ⋆ Les gens du groupe 2 ne savent pas si l'on prie pour eux, et aucun groupe de prière n'a prié pour eux durant l'étude.
- ⋆ Les gens du groupe 3 savent que l'on prie pour eux, et le même groupe de prière que celui du groupe 1 a également prié pour eux pendant la même durée.

Il est toujours possible que des proches aient prié pour des malades durant l'étude, mais ceci étant impossible à évaluer et *a priori* similaire pour les 3 groupes, on les néglige dans l'étude.

L'étude compare les conséquences post-opératoires, et notamment les complications, dans les trois groupes, pour parvenir à une conclusion statistique. Le jeu de données présenté ici a été reconstruit à partir des paramètres statistiques de l'étude, les données originelles étant jalousement gardées par les médecins. Tous les résultats statistiques présentés dans ce document correspondent au plus près à ceux trouvés lors de l'étude originelle.

1.2 Questions

Le jeu de données se présente de la façon suivante :

```
head(data)
  Age Complications Groupe
1  68         FALSE      3
2  56         FALSE      1
3  61         FALSE      3
4  67          TRUE      3
5  83          TRUE      3
6  64          TRUE      3
```

La première colonne contient l'age des patients, la colonne "Complications" contient `TRUE` si des problèmes post-opératoires sont survenus, et la colonne "Groupe" contient le numéro du groupe.

Quelle commande donnerait le même résultat que `head()`, en utilisant les opérateurs d'indexation `[]` ?

Réponse :

Quelle commande permet de savoir combien d'individus contient chaque groupe, et donne le résultats suivant :

```
  1   2   3
604 597 601
```

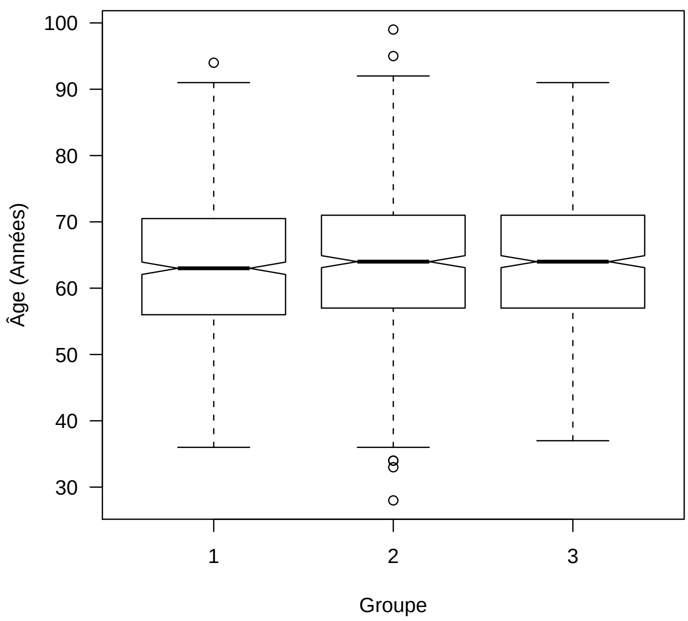
Réponse :

Que calcule la commande :

```
mean(data$Complications[data$Groupe == 1])
[1] 0.5215232
```

Réponse :

Une des premières idées à vérifier pour les auteurs, pour vérifier que leur étude est correcte, est que les trois groupes sont bien similaires en dehors des caractéristiques présentées plus haut. Notamment, il faut vérifier que l'âge des patients dans les trois groupes est globalement similaire. On peut le vérifier avec une figure comme celle-ci :



Quelle commande permet d'obtenir le graphique précédent ?

Réponse :

Sans effectuer de test, que pouvez-vous conclure, grâce à cette figure, quant à la différence d'âge moyen entre les trois groupes choisis ?

Réponse :

Les auteurs réalisent un test statistique pour savoir si la prière elle-même a un effet sur l'apparition de complications. Pour cela, ils comparent la fréquence d'apparition des complications entre les groupes 1 et 2. Le test a pour hypothèses :

⋆ H_0 : La prière n'a aucun effet sur l'apparition des complications post-opératoires.

⋆ H_1 : La prière modifie la fréquence d'apparition des complications post opératoires.

Le test est réalisé sous R, et donne le résultat suivant :

```
        2-sample test for equality of proportions with continuity correction
data:  c(sum(com[groups == 1]), sum(com[groups == 2])) out of c(604, 597)
X-squared = 0.1362, df = 1, p-value = 0.712
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.04588069  0.07050159
sample estimates:
   prop 1    prop 2
0.5215232 0.5092127
```

Si on accepte un risque de première espèce α de 5%, que pouvez-vous conclure de ce test ? Justifiez.

Réponse :

De la même manière, les auteurs veulent savoir si le fait, pour un malade, de savoir que l'on prie pour lui, a un effet sur l'apparition de complications. En comparant les données, ils réalisent que chez les patients qui savent que l'on a prié pour eux (groupe 3), des complications sont apparues dans 58.6% des cas, tandis qu'elles ne sont apparues que dans 52.2% des cas chez les patients qui ne savent pas si l'on prie pour eux ou non (groupe 1). Pour vérifier si cette différence est statistiquement significative, ils comparent donc par un test la fréquence d'apparition des complications dans les groupes 1 et 3 et obtiennent le résultat suivant :

```
        2-sample test for equality of proportions with continuity correction
data:  c(sum(com[groups == 1]), sum(com[groups == 3])) out of c(604, 601)
X-squared = 4.7627, df = 1, p-value = 0.02908
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.121845597 -0.006489077
sample estimates:
   prop 1    prop 2
0.5215232 0.5856905
```

À quoi peuvent-ils conclure ?

Réponse :

Les auteurs de l'article expliquent à la fin de celui-ci qu'ils ont conçu l'étude de manière à avoir une puissance de 95% si l'effet dépassait 10%, c'est à dire, en langage plus détaillé, que la valeur du risque de deuxième espèce β devait être

5% seulement, si la prière avait un impact sur la convalescence de plus de 10% des patients. Rappelez ce qu'est le risque de deuxième espèce.

Réponse :

Pourquoi les auteurs ont-ils prêté tant d'attention au risque de deuxième espèce (le risque α de première espèce était également fixé à 5%, comme souvent) ? Plus précisément, si on avait une valeur de β de 50% et que létude concluait que la prière n'a aucun effet sur la convalescence, quel argument pourrait-on faire ?

Réponse :

Finalement, les auteurs dressent un tableau résumé des facteurs de risque augmentant la probabilité d'avoir des complications suite à une opération de pontage coronarien. Il a été montré dans l'article que toutes les variables présentées ici étaient significativement impliquées dans l'apparition des complications, indépendamment les unes des autres. La table (simplifiée) représente l'augmentation des chances d'avoir des complications :

Facteur de risque	Probabilité de complications
Âge avancé	+4%
Avoir connaissance de prières faites pour soi	+27%
Antécédents d'infarctus	+45%
Antécédents de maladies pulmonaires	+61%
Antécédents d'hypertension	+39%

Table 1 –

Commentez de manière critique ces résultats. Avez-vous une explication scientifique pour l'augmentation du risque chez les patients sachant que l'on prie pour eux ?

Réponse :

2 Modélisation.

2.1 Maladie cardiaque et vieillissement.

Les opérations du cœur concernent environ 800 000 patients par an, dont environ 350 000 aux États-Unis d'Amérique. Partout dans le monde, le risque de maladie cardiaque augmente avec l'âge des individus. Il en est de même pour de nombreux dysfonctionnements (diabète, cancer, maladie d'Alzheimer...) qui contribuent au vieillissement des individus. On sait par ailleurs qu'il existe dans les populations des lignées d'individus prédisposés à l'une ou l'autre de ces maladies. À l'aide de modèles simples, nous nous intéresserons à l'évolution du vieillissement.

2.2 Un modèle simple de dynamique d'une population structurée en âge :

On propose de s'intéresser à une population structurée en 6 classes d'âge. Ce modèle pourrait par exemple être appliqué à l'homme en considérant les classes d'âge suivantes :

1. Jeunes (chez l'homme : $[0; 15[$ ans).
2. Adultes 1 (chez l'homme : $[15; 30[$ ans).
3. Adultes 2 (chez l'homme : $[30; 45[$ ans).
4. Adultes 3 (chez l'homme : $[45; 60[$ ans).
5. Adultes 4 (chez l'homme : $[60; 75[$ ans).
6. Adultes 5 (chez l'homme : $[75; 90[$ ans).

La dynamique de cette population est modélisée en utilisant la matrice $\mathbf{A}$ suivante :

$$\mathbf{A} = \begin{pmatrix} 0.000 & 0.750 & 0.220 & 0.050 & 0.000 & 0.000 \\ 0.991 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.987 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.976 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.923 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.761 & 0.000 \end{pmatrix}$$

On donne les résultat suivant :

```
A1 <- eigen(A)
sumvecA <- matrix(rep(colSums(A1$vectors), 6), ncol = 6, byrow = TRUE)
A1$values
[1]  1.003+0.000i -0.734+0.000i -0.134+0.216i -0.134-0.216i  0.000+0.000i
[6]  0.000+0.000i
A1$vectors
         [,1]      [,2]               [,3]               [,4] [,5]        [,6]
[1,] 0.446+0i  0.179+0i -0.00054-0.00141i -0.00054+0.00141i 0+0i   0.0e+00+0i
[2,] 0.441+0i -0.242+0i -0.00356+0.00467i -0.00356-0.00467i 0+0i   0.0e+00+0i
[3,] 0.434+0i  0.326+0i  0.02265+0.00219i  0.02265-0.00219i 0+0i   0.0e+00+0i
[4,] 0.423+0i -0.433+0i -0.03866-0.07817i -0.03866+0.07817i 0+0i   0.0e+00+0i
[5,] 0.389+0i  0.545+0i -0.16675+0.26850i -0.16675-0.26850i 0+0i  7.9e-292+0i
[6,] 0.295+0i -0.565+0i  0.94443+0.00000i  0.94443+0.00000i 1+0i  -1.0e+00+0i
```

```
sumvecA
```

```
        [,1]     [,2]         [,3]         [,4] [,5]  [,6]
[1,] 2.43+0i -0.19+0i 0.758+0.196i 0.758-0.196i 1+0i -1+0i
[2,] 2.43+0i -0.19+0i 0.758+0.196i 0.758-0.196i 1+0i -1+0i
[3,] 2.43+0i -0.19+0i 0.758+0.196i 0.758-0.196i 1+0i -1+0i
[4,] 2.43+0i -0.19+0i 0.758+0.196i 0.758-0.196i 1+0i -1+0i
[5,] 2.43+0i -0.19+0i 0.758+0.196i 0.758-0.196i 1+0i -1+0i
[6,] 2.43+0i -0.19+0i 0.758+0.196i 0.758-0.196i 1+0i -1+0i
```

```
A1$vectors/sumvecA
```

```
         [,1]      [,2]                [,3]                [,4] [,5]        [,6]
[1,] 0.184+0i -0.942+0i -0.00111-0.00157i -0.00111+0.00157i 0+0i    0.0e+00+0i
[2,] 0.182+0i  1.273+0i -0.00292+0.00691i -0.00292-0.00691i 0+0i    0.0e+00+0i
[3,] 0.179+0i -1.711+0i  0.02873-0.00453i  0.02873+0.00453i 0+0i    0.0e+00+0i
[4,] 0.174+0i  2.276+0i -0.07284-0.08437i -0.07284+0.08437i 0+0i    0.0e+00+0i
[5,] 0.160+0i -2.862+0i -0.12047+0.38556i -0.12047-0.38556i 0+0i -7.9e-292+0i
[6,] 0.122+0i  2.967+0i  1.16861-0.30200i  1.16861+0.30200i 1+0i    1.0e+00+0i
```

Comment appelle-t-on le type de modèle proposé, utilisant la matrice **A** ?

Réponse :

Comment nomme-t-on la catégorie de modèles pour lesquels le hasard n'intervient pas, et à laquelle appartient ce modèle que nous allons étudier ?

Réponse :

Donnez le taux de fécondité et le taux de survie des individus de la classe "Adultes 3".

Réponse :

De quelle façon peut-on mettre en évidence les effets du vieillissement des individus à l'aide de la matrice **A** ?

Réponse :

En termes d'effectif total de la population, que pouvez-vous prédire pour l'évolution de cette population ?

Réponse :

Lorsque la population aura atteint sa structuration à l'équilibre, quelle sera la proportion d'individus dans la classe d'âges $[15; 60[$ ans ?

Réponse :

2.3 Effets d'une mutation selon l'âge des individus.

Parmi les théories sur le vieillissement, on propose l'existence de mutations dont les effets dépendraient de l'âge des individus : elles seraient bénéfiques à un certain âge et néfastes à d'autres âges (en anglais, on parle de *trade-off*). On propose que dans la population survient une mutation telle que la dynamique des individus porteurs de cette mutation peut être modélisée à l'aide de la matrice **B**.

$$\mathbf{B} = \begin{pmatrix} 0.000 & 0.770 & 0.220 & 0.050 & 0.000 & 0.000 \\ 0.991 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.987 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.976 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.842 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.381 & 0.000 \end{pmatrix}$$

En comparant les matrices **A** et **B**, décrivez brièvement les effets de la mutation.

Réponse :

On donne à présent les résultats suivants :

```
eigen(B)
$values
[1]  1.011+0.000i -0.749+0.000i -0.131+0.214i -0.131-0.214i  0.000+0.000i
[6]  0.000+0.000i
```

```
$vectors
          [,1]      [,2]                [,3]                [,4] [,5]          [,6]
[1,] 0.477+0i  0.228+0i -0.00103-0.00248i -0.00103+0.00248i 0+0i   0.00e+00+0i
[2,] 0.467+0i -0.302+0i -0.00622+0.00856i -0.00622-0.00856i 0+0i   0.00e+00+0i
[3,] 0.456+0i  0.398+0i  0.04148+0.00325i  0.04148-0.00325i 0+0i   0.00e+00+0i
[4,] 0.440+0i -0.519+0i -0.07346-0.14411i -0.07346+0.14411i 0+0i   0.00e+00+0i
[5,] 0.366+0i  0.584+0i -0.28340+0.46255i -0.28340-0.46255i 0+0i  1.58e-291+0i
[6,] 0.138+0i -0.297+0i  0.82324+0.00000i  0.82324+0.00000i 1+0i  -1.00e+00+0i
```

En termes d'effectifs, que pouvez-vous dire de la vitesse de croissance de la population des individus porteurs de la mutation, par rapport à la population des individus non porteurs de la mutation ?

Réponse :

2.4 Devenir des mutations

Dans la population de départ, il existe des individus non porteurs de la mutation et des individus porteurs de la mutation. On propose un modèle matriciel à 12 classes dont la matrice **C** est simplement constituée des matrices **A** et **B** de la façon suivante :

$$\mathbf{C} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{B} \end{array}\right]$$

De cette façon, les classes 1 à 6 correspondent aux classes 1 à 6 de la population des individus non porteurs de la mutation, et les classes 7 à 12 correspondent aux mêmes classes 1 à 6 de la population des individus porteurs de la mutation.

On donne le résultat suivant :

```
C <- matrix(data = 0, ncol = 12, nrow = 12)
C[1:6, 1:6] <- A
C[7:12, 7:12] <- B
C1 <- eigen(C)
C1$values
```

```
 [1]  1.011+0.000i  1.003+0.000i -0.749+0.000i -0.734+0.000i -0.134+0.216i
 [6] -0.134-0.216i -0.131+0.214i -0.131-0.214i  0.000+0.000i  0.000+0.000i
[11]  0.000+0.000i  0.000+0.000i
```

En termes d'effectifs, que pouvez-vous dire de la vitesse de croissance à long terme de la population totale ? Comparez à vos réponses précédentes. Que remarquez-vous ?

Réponse :

On donne à présent les résultats suivant :

```
 C1$values[c(1, 2)]
[1] 1.011+0i 1.003+0i
 C1$vectors[, c(1, 2)]
             [,1]        [,2]
 [1,] 0.00000+0i 0.1838+0i
 [2,] 0.00000+0i 0.1816+0i
 [3,] 0.00000+0i 0.1788+0i
 [4,] 0.00000+0i 0.1740+0i
 [5,] 0.00000+0i 0.1602+0i
 [6,] 0.00000+0i 0.1216+0i
 [7,] 0.20333+0i 0.0000+0i
 [8,] 0.19926+0i 0.0000+0i
 [9,] 0.19449+0i 0.0000+0i
[10,] 0.18772+0i 0.0000+0i
[11,] 0.15631+0i 0.0000+0i
[12,] 0.05889+0i 0.0000+0i
```

Que pouvez-vous dire de l'évolution à long terme de la proportion d'individus non porteurs de la mutation ?

Réponse :

2.5 Une théorie du vieillissement.

Les modèles présentés précédemment sont à la base de la théorie évolutive du vieillissement. D'après ce que vous avez compris de l'exemple précédent, comment pourriez-vous expliquer l'existence dans la population de nombreuses maladies atteignant les individus âgés ?

Réponse :

Références

[1] Herbert Benson, Jeffery A Dusek, Jane B Sherwood, Peter Lam, Charles F Bethea, William Carpenter, Sidney Levitsky, Peter C Hill, Donald W Clem, Manoj K Jain, David Drumel, Stephen L Kopecky, Paul S Mueller, Dean Marek, Sue Rollins, and Patricia L Hibberd. Study of the therapeutic effects of intercessory prayer (step) in cardiac bypass patients : a multicenter randomized trial of uncertainty and certainty of receiving intercessory prayer. *Am Heart J*, 151(4) :934–942, Apr 2006.