# INTRODUCTION AUX LOGICIELS EXCEL & R

# *Veillez à avoir lu les sections relatives à cette séance dans le « Guide des séances de TP de biostatistique » avant de vous présenter au TP.*

#### Exercice 1

Nous allons faire une représentation graphique d'un polynôme à une variable dont voici l'équation :  $P(x) = x^4 - 10x^3 + 35x^2 - 50x + 24$ 

Notre représentation graphique va s'intéresser aux valeurs du polynôme quand x est compris entre 0,5 et 4,5. La première étape consiste donc à construire une liste de valeurs de x pour lesquelles on calculera les valeurs de P(x).

Pour cela, nous allons espacer les valeurs successives de x de 0,1 (soient 0,5 ; 0,6 ; 0,7 ;... ; 4,4 ; 4,5). Dans la première feuille du fichier Excel, créez un tableau dont la <u>première colonne</u> sera les valeurs de x et la <u>seconde</u> les valeurs de P(x).

#### Pour créer les valeurs de x, utilisez :

- soit la technique du copier-coller (<u>TRUC</u> : mettre 0,5 dans la cellule A2 et 0,6 dans la cellule A3, sélectionner les 2 cellules et étendre la copie à partir du coin inférieur droit (curseur =
- fine croix) de la sélection vers le bas)
- soit une fonction à référence relative (<u>TRUC</u> : mettre 0,5 dans la cellule A2 ensuite mettre la formule « = A2 + 0,1 » dans la cellule A3, copier ensuite la cellule A3 vers le bas).

Pour calculer P(x) dans la colonne B, inscrivez la formule en remplaçant x par la référence de la cellule qui contient la valeur de x. Ainsi pour la première valeur de x dans la cellule A2, P(x) s'écrira dans la cellule B2 : « = $A2^{4}-10^{4}A2^{3}+35^{4}A2^{2}-50^{4}A2+24$  »

Générez ensuite le graphique à partir des valeurs du tableau créé (graphique en XY type nuage de points dans le menu insertion d'Excel). Appliquez-vous à mettre en forme correctement le tableau et le graphique.

# Exercice 2

**Dans la console de R**, générez un vecteur, appelé x, contenant des valeurs similaires aux valeurs de x utilisées dans excel grâce à la fonction *seq(...)*. Créez un vecteur y contenant les valeurs P(x) correspondantes au vecteur x que vous venez de générer.

Dans R, un vecteur est un objet virtuel qui contient un ensemble de valeurs numériques ou alphanumériques et à qui on a donné un nom. Il existe plusieurs manières pour créer ce vecteur. Elles sont abordées dans le guide des TP, l'ABC d'R ou les autres documents relatifs à R sur ce site.

Générez ensuite un graphique en nuage de points à partir de ces 2 vecteurs (utilisez la fonction graphique 2D *plot(...)*. *Pour connaître les arguments nécessaire pour paramétrer une fonction, utilisez l'aide de R en tapant par exemple ?plot ou help(plot) dans votre console*.

Pour vous familiariser avec R, nous allons modifier les vecteurs x et y :

- Comptez le nombre de valeurs présentes dans x (length(...))
- Créez le vecteur x1 contenant les valeurs (5,5 ; 5,6 ; 5,7 ; 5,8 ; 5,9 ; 6) grâce à la fonction c(...)
- Fusionnez les vecteurs x et x1 tels que les valeurs de x1 suivent celles de x, grâce à cette même fonction et nommez ce nouveau vecteur x2.
- Ajoutez 1 aux 41es premières valeurs du vecteur x2 (accédez aux indices du vecteur grâce aux crochets []) et nommez ce nouveau vecteur x3.

- Supprimez la 41<sup>e</sup> valeur de x3 qui est un doublon (valeur 5,5) et nommez ce nouveau vecteur x4.
- Supprimez les 6 premières valeurs du vecteur x4 afin d'avoir 40 valeurs dans un nouveau vecteur x5.
- Générez le vecteur y2 = P(x5) à partir de ce vecteur x5 et comptez combien de valeurs de P(x5) sont plus petites que zéro.
- Retracez le nouveau graphique, avec une courbe plutôt qu'un nuage de point (argument type= «l» pour line (<u>mettre un L minuscule</u>) dans la fonction plot(...))

Revenons à nos vecteurs initiaux x et y :

- Organisez les 2 vecteurs sous la forme d'un « data frame » nommé t (chaque vecteur deviendra une colonne de t) avec la fonction *data.frame()*
- Sélectionnez dans le tableau la 3<sup>ème</sup> valeur de la colonne y, soit :
  - Uniquement grâce aux indices de positionnement [*a*,*b*] <u>du tableau</u>, où *a* fait référence aux lignes et *b* aux colonnes.
  - Grâce aux noms des colonnes du tableau (sélectionnés à l'aide des \$) et aux indices du <u>vecteur</u>.

# Exercice 3

A la suite du premier exercice dans Excel, vous allez effectuer la représentation graphique d'un polynôme du second degré à 2 inconnues dont voici l'équation :  $P(x,y) = x^2+2y^2+2xy-4$ 

Pour cela, réalisez un tableau à 2 dimensions (type table de multiplication) dont la première ligne contiendra les valeurs de x et la première colonne les valeurs de y. Les valeurs de x iront de -4 à 5 par pas unitaire (-4, -3, ..., 5) sur la plage B1 à J1 et les valeurs de y iront de -2,5 à 2 par pas de 0,5 (-2.5, -2.0, -1.5, ..., 2) sur la plage A2 à A12.

A chaque intersection x,y de ce tableau, la valeur calculée de P(x,y) devra apparaître. A cette fin, rentrez la formule dans la première case du tableau en veillant à y utiliser les références cellulaires des valeurs de x et de y et à placer éventuellement les \$ aux bons endroits (ceci afin de bloquer certaines références cellulaires lors du copier-coller) et effectuez un « copier-coller » afin que la valeur de la fonction apparaisse dans chacune des cases du tableau. Comme pour l'exercice 1, les variables x et y sont remplacées par la référence cellulaire, ce qui donne pour la première valeur de P(x,y) inscrite dans la cellule B2: « =B1^2+2\*A2^2+2\*B1\*A2-4». <u>Remarque</u> : le(s) \$ bloque(nt) l'élément de la cellule devant lequel il(s) se trouve(nt). Par exemple, si on écrit « A\$2 », cela bloque la ligne 2 mais pas la colonne A.

Ensuite, à partir du menu « Insertion », affichez un graphique de type surface (vue 3D) et un graphique de type courbe de niveau (vue 2D). Il faut toujours sélectionner au préalable les données P(x,y) avant d'effectuer le graphique. Les valeurs des axes x et y peuvent ensuite être adaptées (si nécessaire) dans les paramètres du graphique.

Quel type de polynôme attendez-vous ? Cela correspond-il à ce qui est observé sur le graphique ?

# Envie d'aller plus loin : exercices facultatifs pour le brain storming (à faire à domicile)

**1.** Les valeurs du tableau de l'exercice 3 peuvent aussi être calculées dans R mais cela nécessite de passer par une fonction qui va calculer chaque valeur de P(x,y) pour chaque couple de valeurs x,y possible. Créez les vecteurs x et y grâce à la fonction *seq(...)* vue à l'exercice 1.

Ensuite, créez un vecteur nommé Pxy en tapant dans la console : « Pxy <- x^2+2\*y^2+2\*x\*y-4 ». Si vous demandez d'afficher le résultat en tapant Pxy, seulement 10 valeurs apparaissent... Pourquoi ?

R a travaillé en associant la première valeur de x avec la première de y pour donner le premier résultat de Pxy, puis a fait la même chose avec les secondes valeurs et ainsi de suite jusqu'aux 10èmes ! Cette méthode ne peut donc fonctionner que si les vecteurs x et y ont le même nombre de données (ce qui était le cas ici) et cela conduit à seulement 10 résultats de P(x,y) à la place de 10x10 = 100 qui sont l'ensemble des possibilités de couples x,y !

Pour pallier ce problème, nous allons devoir utiliser la fonction *outer(...)* qui crée une table en 2 dimensions avec chacune des valeurs de x et de y, et effectue ensuite une opération à l'intersection de chaque ligne-colonne de la table.

Cependant, le  $3^e$  argument de la fonction *outer(...)* doit être lui-même une fonction. Sur la Figure 1, vous voyez qu'en ayant simplement inscrit la formule de Pxy, cela ne fonctionne pas. Nous devons donc créer la fonction qui calcule P(x,y) grâce à l'instruction *function(...)*{...} comme affiché sur la Figure 1. Cette méthode permettra donc de calculer l'ensemble des 100 valeurs possibles de P(x,y) car *outer* exécute la fonction f pour chaque couple de valeur x,y.

```
R Console
> x <- -4:5
> y <- seq(-2.5,2,0.5)
> Pxy <- x^2+2*y^2+2*x*y-4
> Pxy <- outer(x,y,Pxy)</pre>
Erreur dans get(as.character(FUN), mode = "function", envir = envir) :
 objet 'Pxy' de mode 'function' introuvable
>
> f <- function(x,y) {x^2+2*y^2+2*x*y-4}</pre>
> Pxy <- outer(x,y,f)
> Pxy
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,] 44.5 36 28.5 22 16.5 12 8.5
                                      6 4.5
                                                   4
           25 18.5
                              5 2.5
 [2,] 32.5
                    13 8.5
                                        1 0.5
                                                   1
                    6 2.5
1 -1.5
 [3,] 22.5
           16 10.5
                               0 -1.5
                                       -2 -1.5
                                                   0
 [4,] 14.5
            9 4.5
                              -3 -3.5
                                       -3 -1.5
                                                  1
            4 0.5
 [5,]
      8.5
                    -2 - 3.5
                              -4 - 3.5
                                       -2 0.5
                                                  4
            1 -1.5 -3 -3.5
      4.5
                              -3 -1.5
                                        1 4.5
                                                  9
 [6,]
     2.5
           0 -1.5 -2 -1.5 0 2.5
                                        6 10.5
 [7,]
                                                  16
 [8,] 2.5
           1 0.5 1 2.5
                              5 8.5 13 18.5
                                                  25
 [9,] 4.5
           4 4.5
                    6 8.5 12 16.5 22 28.5
                                                  36
[10,] 8.5
            9 10.5 13 16.5 21 26.5 33 40.5
                                                  49
>
```

Figure 1 : console R montrant les opérations de calcul du tableau Pxy

**2.** Le logiciel R est extrêmement puissant pour réaliser notamment des graphiques et des programmes. Par cet exercice, vous allez créer de petites fonctions grâce à l'instruction *function(...)* et vous allez générer des graphiques en 3D colorés. Le but est de comprendre chacune des instructions reprises ci-dessous et de les retranscrire une par une dans la console de R.

```
tp1_fonction_3d - Bloc-notes
 Fichier Edition Format Affichage ?
# Representation 3-d de f(x,y) du second degre
x<-seq(-5,5,0.1)
y<-seq(-3,3,0.1)
 Ellipsoide, parabolide et hyperboloide
ellip<-function(x,y) {
                            x**2+2*y**2+ x*y-x+2*y-3
                         { -x**2+2*y**2+ x*y-x+2*y-3 }
{ x**2+ y**2+2*x*y-x+2*y-3 }
hyper<-function(x,y)
parab < -function(x,y)
ze<-outer(x,y,ellip)
zh<-outer(x,y,hyper)</pre>
zp<-outer(x,y,parab)
# Colorie en bleu les facettes dont la valeur < 0, et en vert celles > 0
# 0) Choisis une surface
z<-ze
                        # ou z<-zp ou z<-zh
# 1) valeur des facettes = somme des coins
nrz <- nrow(z)
ncz <- ncol(z)
nb_facettes<-(nrz-1)*(ncz-1)
valeurs<-matrix(rep(0,nb_facettes),nr=(nrz-1))
for (i in 1:(nrz-1))
for (j in 1:(ncz-1))
                         - {
  valeurs[i,j]<-z[i,j]+z[i,j+1]+z[i+1,j]+z[i+1,j+1]
 }
}
# 2) Couleurs
couleurs<-matrix(rep(0,nb_facettes),nr=(nrz-1))
content s(nz=1)) {
  for (j in 1:(nz=1)) {
    couleurs[i,j]<-"blue"
    if (valeurs[i,j]>0) {    couleurs[i,j]<-"green"}</pre>
 3
3
# 3) Diagramme 3-d
persp(x, y, z, col = couleurs, phi = 30, theta = -300)
```

Le signe # permet d'entrer des commentaires dans la console ou un programme. La ligne n'est donc pas à proprement parlé exécutée.

La fonction *outer(...)* génère un tableau 2D de données, similaire à celui généré dans l'exemple Excel qui précède : les valeurs retournées dans le tableau 2D sont des valeurs de f(x,y), où f est une fonction fournie comme  $3^e$  argument et (x,y) correspond à chaque combinaison des valeurs de X et de Y.

Les fonctions *nrow(...)* et *ncol(...*) compte le nombre de lignes et de colonnes d'un tableau généré par *outer(...)* dans notre exercice). La fonction *matrix(...)* crée un tableau type matrice (voir l'aide R pour les détails).

L'instruction *for* crée, comme dans beaucoup d'autres langages informatiques, une boucle qui exécute le nombre de fois requis les instructions mentionnées. L'instruction *if...else* est une instruction conditionnelle similaire à la fonction SI() d'Excel.

La fonction *persp(...)* permet de réaliser des graphiques en 3 dimensions (voir l'aide pour les différents arguments de la fonction).

N'hésitez pas à expérimenter, par exemple en affichant certaines variables (x, y, z...), en changeant les paramètres de certaines fonctions, etc... pour mieux comprendre le fonctionnement du logiciel.

**3.** Pour cet exercice, ouvrez un script d'édition depuis R (menu File -> New script ...). Dans ce script, inscrivez toutes les lignes de commande R qui permettent de résoudre le problème posé. Enregistrez le script dans le répertoire de travail de R. Retournez dans la console et tapez l'instruction source(« nom de votre script et son extension ») afin de lancer son exécution. Si une erreur apparaît lors de son exécution, modifiez le script, réenregistrez le et ré-exécutez le depuis la console. Cette manière de faire vous permet de ne pas retaper toutes les instructions après chaque essai, en gardant dans le script ce qui fonctionne.

Soit une équation du 2e degré à une inconnue du type  $ax^2+bx+c = 0$ , dont vous allez devoir rechercher les racines (valeurs de x qui vérifient l'équation).

Créez 3 vecteurs a, b et c, contenant chacun 5 valeurs entières aléatoires. Créez ensuite une fonction qui permettrait de calculer les deltas de vos 5 équations en fonction des valeurs de a, b et c et qui ensuite affichera les résultats (1 ou 2 valeurs possibles si delta >=0, le message « aucune valeur possible » si delta < 0).

**4.** Le polynôme que nous allons étudier est  $P(x) = x^5+6,8 * x^4+14,4 * x^3+3,2 * x^2-18,4 * x-24$ . L'objet de l'exercice est de calculer la (seule) racine réelle de ce polynôme, c'est-à-dire trouver la valeur de x telle que P(x)=0. L'idée à exploiter est la suivante: on choisit (au hasard) deux valeurs de x, notées x<sub>1</sub> et x<sub>2</sub>, telles que P(x<sub>1</sub>) < 0 et P(x<sub>2</sub>) > 0.

Par conséquent, la valeur recherchée est nécessairement comprise entre ces deux valeurs de x. On réduit ensuite l'intervalle (de moitié) en calculant la valeur  $x_3$ , qui est au milieu de l'intervalle formé par  $x_1$  et  $x_2$ . On calcule ensuite P( $x_3$ ), et on remplace  $x_1$  par  $x_3$  si P( $x_3$ ) < 0 ou  $x_2$  par  $x_3$  si P( $x_3$ ) > 0. On obtient donc bien un intervalle, réduit de moitié, dans lequel la racine se trouve. Répétant cette opération autant de fois que nécessaire, on peut se rapprocher autant qu'on veut de la racine recherchée. Cet algorithme est appelé « méthode dichotomique ».

Résolvez ce calcul dans Excel et dans R. (Réponse : valeur proche de 1,2)