Feuille de Travaux Dirigés n$^{\text{o}}$ 8
Mesures de liaison

Exercice VIII.1. Densité de lois normales bivariées

Le package dans R permettant d'obtenir la densité, les quantiles ou de générer des réalisations de lois normales multivariées est le package `mvtnorm`.
Que font les instructions suivantes ?

```
> library(mvtnorm)
> help(package="mvtnorm")
> dmvnorm(c(0,0), c(0,0), diag(2), log=FALSE)
```

Nous allons maintenant représenter graphiquement la fonction de densité de deux lois normales bivariées.
Éxecuter les instructions suivantes et identifier leur action. En particulier il faudra déterminer à quoi servent les fonctions `c()`, `cbind()`, `diag()`, `matrix()` ?

```
library(lattice)

# N2(c(0,0),I_2)

> g <- expand.grid(x = seq(-2,2,0.05), y = seq(-2,2,0.05))
> g$z <- dmvnorm(x=cbind(g$x,g$y),mean=c(0,0), sigma=diag(2),
+ log=FALSE)

> wireframe(z ~ x * y, data = g,colorkey = TRUE,drape=TRUE)

# N2(c(0,0),matrix(c(1,0.75,0.75,1),byrow=T,nrow=2))

> var <- matrix(c(1,0.75,0.75,1),byrow=T,nrow=2)
> g <- expand.grid(x = seq(-2,2,0.05), y = seq(-2,2,0.05))
> g$z <- dmvnorm(x=cbind(g$x,g$y),mean=c(0,0), sigma=var, log=FALSE)

> wireframe(z ~ x * y, data = g,colorkey = TRUE,drape=TRUE)
```

Les graphiques précédents sont statiques. Le package `rgl` permet de rémédier à ce défaut et de faire pivoter les graphiques à l'aide de la souris.

```
library(rgl)

# N2(c(0,0),matrix(c(1,0,0,1),byrow=T,nrow=2))

> g <- expand.grid(x = seq(-4,4,0.05), y = seq(-4,4,0.05))
> g$z <- dmvnorm(x=cbind(g$x,g$y),mean=c(0,0), sigma=diag(2),
+ log=FALSE)

> g2z <- matrix(g$z*5000,byrow=T,nrow=length(seq(-4,4,0.05)))
```

```
> g2x <- 10 * (1:nrow(g2z))
> g2y <- 10 * (1:ncol(g2z))

> zlim <- range(g2y)
> zlen <- zlim[2] - zlim[1] + 1

> colorlut <- terrain.colors(zlen) # height color lookup table

> col <- colorlut[ g2y-zlim[1]+1 ] # assign colors to heights for
each point

> open3d()
> surface3d(g2x, g2y, g2z, color=col, back="lines")


# N2(c(0,0),matrix(c(1,0.75,0.75,1),byrow=T,nrow=2))

> g <- expand.grid(x = seq(-4,4,0.05), y = seq(-4,4,0.05))
> g$z <- dmvnorm(x=cbind(g$x,g$y),mean=c(0,0),
+ sigma=matrix(c(1,0.75,0.75,1), byrow=T,nrow=2), log=FALSE)

> g2z <- matrix(g$z*5000,byrow=T,nrow=length(seq(-4,4,0.05)))

> g2x <- 10 * (1:nrow(g2z))
> g2y <- 10 * (1:ncol(g2z))

> zlim <- range(g2y)
> zlen <- zlim[2] - zlim[1] + 1

> colorlut <- terrain.colors(zlen) # height color lookup table

> col <- colorlut[ g2y-zlim[1]+1 ] # assign colors to heights for
each point

> open3d()
> surface3d(g2x, g2y, g2z, color=col, back="lines")
```

Exercice VIII.2. BMI, poids et taille.
Nous disposons de trois variables continues réelles.
- Le BMI X_1.
- Le poids[1] X_2.
- La taille X_3.

Le BMI d'un individu est un indice qui se calcule à partir de la taille et du poids

[1] D'un poids de vue physique, il s'agit de la masse.

par la formule suivante :

$$\text{BMI} = \frac{\text{Weight}}{\text{Height}^2}$$

où le poids s'exprime en kg et la taille en m.
L'échantillon étudié a un effectif $n = 38$ et est exclusivement constitué d'individus de sexe féminin. Les données sont disponibles dans le fichier BMI.CSV.

1. Représenter graphiquement les données à l'aide de la fonction `plot`.

2. Calculer les coefficients de corrélation de Pearson, Spearman et Kendall en utlisant les commandes suivantes :

```
cor(BMIdata,method="pearson")
cor(BMIdata,method="spearman")
cor(BMIdata,method="kendall")
```

3. Quelle hypothèse doit être vérifiée pour pouvoir tester la nullité des coefficients de corrélation de Pearson ? Tester cette hypothèse avec la fonction `mardia` de la bibliothèque `dprep` qu'il faudra sans doute télécharger depuis Internet :

```
library(dprep)
mardia(cbind(BMIdata,rep(1,nrow(BMIdata))))
```

4. Commenter les commandes suivantes :

```
cor.test(BMIdata$BMI,BMIdata$Weight,data=BMIdata,method="pearson")
cor.test(BMIdata$Height,BMIdata$Weight,data=BMIdata,method="pearson")
cor.test(BMIdata$BMI,BMIdata$Height,data=BMIdata,method="spearman",exact=TRUE)
cor.test(BMIdata$BMI,BMIdata$Height,data=BMIdata,method="kendall",exact=TRUE)

library(coin)
spearman_test(BMI ~ Height, data = BMIdata, distribution = approximate(B = 10000)

library(coin)
spearman_test(BMI ~ Height, distribution = approximate(B = 10000))

library(agricolae)
correlation(BMIdata,BMIdata,method="pearson")
correl(Height,Weight,method="pearson")
correl(BMI,Height,method="spearman")
correl(BMI,Height,method="kendall")
```

4. Calculer les coefficients de détermination multiples de chacune des variables par rapport aux deux autres.

5. Calculer les coefficients de détermination partiels de chaque couple de vaiables par rapport à la troisième.

Exercice VIII.3. École

Les données sont celles de l'exemple du cours. Elles sont disponibles dans le fichier `Ecole.CSV`.
Étudier la liaison entre les différentes variables du jeu de données.