Lab 3: Analisi de Correspondencies i Normal multivariant

Álvaro Ribot & Jose Pérez Cano

Exercici 1: Anticonceptius a indonesia

```
library("ca")
library("MASS")
```

a)

```
X <- read.table("http://www-eio.upc.es/~jan/Data/MVA/cmc.dat", header = T, sep = ",")
head(X)
```

```
##   Age WifeEduc HusbEduc Childs Rel Work Occup Standard Media Method
## 1  24        2        3      3   1    1     2        3     0      1
## 2  45        1        3     10   1    1     3        4     0      1
## 3  43        2        3      7   1    1     3        4     0      1
## 4  42        3        2      9   1    1     3        3     0      1
## 5  36        3        3      8   1    1     3        2     0      1
## 6  19        4        4      0   1    1     3        3     0      1
```

```
for (i in 1:ncol(X)) {
  print(range(X[,i]))
}
```

```
## [1] 16 49
## [1] 1 4
## [1] 1 4
## [1]  0 16
## [1] 0 1
## [1] 0 1
## [1] 1 4
## [1] 1 4
## [1] 0 1
## [1] 1 3
```

```
X[is.na(X)]
```

```
## integer(0)
```

Podem observar que el rang dels valors que prenen les variables categoriques coinicideix amb el de la seva codificacio. Per tant, podem concloure que no hi ha dades mancants. A mes podem veure que no hi ha cap dada faltant.

b)

```
X$Method <- factor(X$Method, levels = c(1, 2, 3), labels = c("none", "long-term", "mid-term"))
X$WifeEduc <- factor(X$WifeEduc, levels = c(1, 2, 3, 4), labels = c("1low",  "2medium-low", "3medium-hig
```

c)

```r
(taula <- table(X$Method, X$WifeEduc))
```

```
##
##             1low 2medium-low 3medium-high 4high
##   none       103         176          175   175
##   long-term    9          37           80   207
##   mid-term    40         121          155   195
```

d)

```r
chi.test <- chisq.test(taula)
chi.test$statistic
```

```
## X-squared
##  140.4589
```

```r
chi.test$p.value
```

```
## [1] 8.01877e-28
```

L'estadistic de prova pren valor 140.46 i el p-valor es $8.01877e-28$. Per tant, podem concluir que hi ha associacio entre les variables (rebutgem la hipotessi nul · la que assumeix independència)

e)

```r
# Columna
(c <- apply(taula, 2, sum))
```

```
##         1low  2medium-low 3medium-high        4high
##          152          334          410          577
```

```r
# Fila
(r <- apply(taula, 1, sum))
```

```
##      none long-term  mid-term
##       629       333       511
```

Podem veure els pesos de *Method* i de *WifeEduc*, respectivament. Per tant, el métode conceptiu més utilitzat es "none".

f)

Es mostren les matrius de perfils fila i columna, respectivament.

```r
Dr <- diag(r)
Dc <- diag(c)

(R <- solve(Dr)%*%taula)
```

```
##
##              1low 2medium-low 3medium-high     4high
##   [1,] 0.16375199   0.2798092    0.2782194 0.2782194
##   [2,] 0.02702703   0.1111111    0.2402402 0.6216216
##   [3,] 0.07827789   0.2367906    0.3033268 0.3816047
```

```r
(C <- t(solve(Dc)%*%t(taula)))
```

```
##
##                [,1]      [,2]      [,3]      [,4]
##   none      0.67763158 0.5269461 0.4268293 0.3032929
##   long-term 0.05921053 0.1107784 0.1951220 0.3587522
##   mid-term  0.26315789 0.3622754 0.3780488 0.3379549
```

g) Es mostra el perfil marginal de la taula per perfils fila

```
P <- as.matrix(taula)
P <- P/sum(P)
(marg.r <- apply(P, 2, sum))
```

```
##        1low  2medium-low 3medium-high        4high
##   0.1031908    0.2267481    0.2783435    0.3917176
```

h)

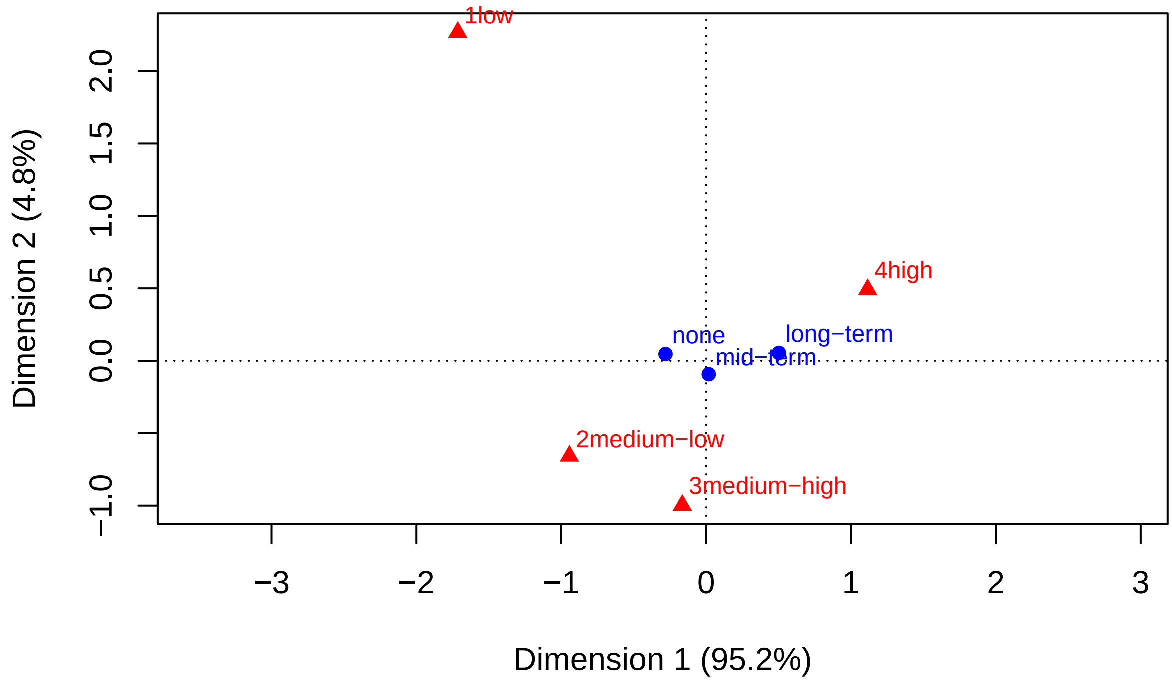
```
resultats <- ca(taula)
summary(resultats)
```

```
##
## Principal inertias (eigenvalues):
##
##  dim    value      %   cum%   scree plot
##  1      0.090755  95.2  95.2  ************************
##  2      0.004600   4.8 100.0  *
##         -------- -----
##  Total: 0.095356 100.0
##
##
## Rows:
##     name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr  
## 1 | none |  427 1000  362 | -280 973 370 |   47  27 203 |
## 2 | lngt |  226 1000  605 |  502 988 629 |   54  12 145 |
## 3 | mdtr |  347 1000   33 |   18  36   1 |  -93 964 652 |
##
## Columns:
##     name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr  
## 1 | 1low |  103 1000  314 | -517 919 303 |  154  81 531 |
## 2 | 2mdm |  227 1000  197 | -284 976 202 |  -45  24  98 |
## 3 | 3mdm |  278 1000   21 |  -50 350   8 |  -68 650 277 |
## 4 | 4hgh |  392 1000  468 |  336 990 487 |   33  10  95 |
```

Existeixen dues dimensions: la taula es pot representar de manera exacta amb min(I-1, J-1), on I i J son el nombre de files i columnes, respectivament. Però observem que amb una dimensió ja obtenim el 95.2%. Per tant, podem representar adequadament la taula de contingencia amb una dimensio.

i)

```
plot(resultats, map = "rowprincipal")
```



j)

```r
dist(rbind(marg.r,P[1,]))
```

```
##      marg.r
##   0.3354785
```

```r
dist(rbind(marg.r,P[2,]))
```

```
##      marg.r
##   0.4041845
```

```r
dist(rbind(marg.r,P[3,]))
```

```
##      marg.r
##   0.3520156
```

Per tant, el perfil fila que s'assembla mes al perfil marginal es el corresponent a la fila 1, és a dir, a la categoria "none".

k)

```r
X$Age <- as.factor(cut(X$Age, breaks = 4))
table(X$Age)
```

```
##
##   (16,24.2] (24.2,32.5] (32.5,40.8]   (40.8,49]
##         276         504         387         306
```

l)

```r
X$AgeRel <- with(X, interaction(Age, Rel))
table(X$AgeRel)/nrow(X)
```

```
##
##   (16,24.2].0 (24.2,32.5].0 (32.5,40.8].0   (40.8,49].0   (16,24.2].1
##   0.009504413   0.042090971   0.060420910   0.037338764   0.177868296
## (24.2,32.5].1 (32.5,40.8].1   (40.8,49].1
##   0.300067889   0.202308215   0.170400543
```

m)

```r
(taula2 <- table(X$Method, X$AgeRel))
```

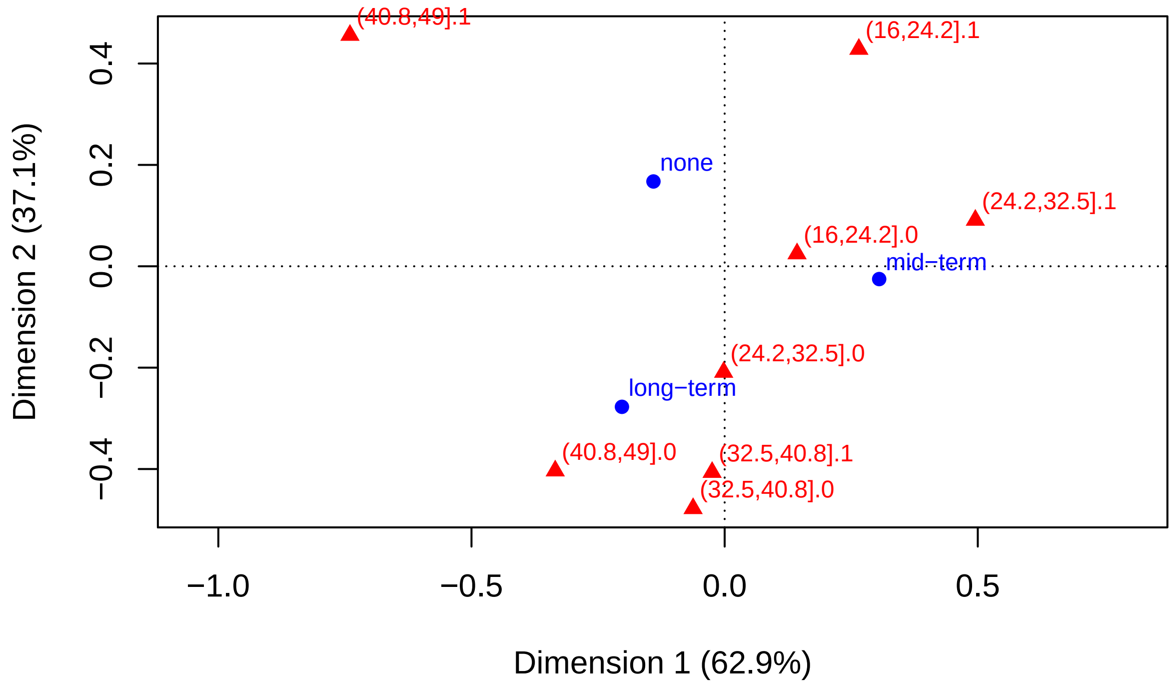
```
##
##             (16,24.2].0 (24.2,32.5].0 (32.5,40.8].0 (40.8,49].0
##   none                5            22            27          21
##   long-term           2            18            32          24
##   mid-term            7            22            30          10
##
##             (16,24.2].1 (24.2,32.5].1 (32.5,40.8].1 (40.8,49].1
##   none              121           170           109         154
##   long-term          35            77            85          60
##   mid-term          106           195           104          37
```

n)

```r
summary(resultats2 <- ca(taula2))
```

```
##
## Principal inertias (eigenvalues):
##
##  dim    value      %   cum%   scree plot
##  1      0.050065  62.9  62.9  ****************
##  2      0.029557  37.1 100.0  *********
##         -------- -----
##  Total: 0.079622 100.0
##
##
## Rows:
##     name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr
## 1 | none |  427 1000  256 | -141 414 169 |  167 586 404 |
## 2 | lngt |  226 1000  335 | -203 349 186 | -277 651 588 |
## 3 | mdtr |  347 1000  409 |  305 993 646 |  -25   7   7 |
##
## Columns:
##        name   mass  qlt  inr    k=1 cor ctr    k=2  cor ctr
## 1 |  162420 |   10 1000   13 |  328 983  20 |   44   17   1 |
## 2 | 2423250 |   42 1000   16 |   -2   0   0 | -175 1000  44 |
## 3 | 3254080 |   60 1000   87 |  -57  28   4 | -334  972 228 |
## 4 |  408490 |   37 1000  131 | -387 538 112 | -359  462 163 |
## 5 |  162421 |  178 1000  112 |  141 394  70 |  175  606 183 |
## 6 | 2423251 |  300 1000  157 |  202 980 245 |   29   20   8 |
## 7 | 3254081 |  202 1000   62 |  -12   6   1 | -155  994 165 |
## 8 |  408491 |  170 1000  421 | -401 817 548 |  190  183 208 |
```

```r
plot(resultats2, map = "rowgreen")
```



En aquest cas observem que son necessaries dues dimensions per fer una bona representació de la taula de contingencia, ja que amb una dimensio nomes representariem be un 62.9%. En general, les dones que practiquen la religio islamica tendeixen a utilitzar menys metodes anticonceptius que les no creients. Tambe podem notar que les dones joves tendeixen a utilitzar metodes anticonceptius de curt termini i a mesura que es fan gran utilitzen els de llarg termini.

o)

```r
out <- mjca(X[,c("WifeEduc", "Method", "AgeRel")], lambda="indicator")
summary(out)
```

```
##
## Principal inertias (eigenvalues):
##
##  dim    value      %   cum%   scree plot
##  1      0.490556  12.3  12.3  ***
##  2      0.456773  11.4  23.7  ***
##  3      0.357164   8.9  32.6  **
##  4      0.345038   8.6  41.2  **
##  5      0.334913   8.4  49.6  **
##  6      0.333333   8.3  57.9  **
##  7      0.333333   8.3  66.3  **
##  8      0.321610   8.0  74.3  **
##  9      0.299496   7.5  81.8  **
##  10     0.279349   7.0  88.8  **
##  11     0.232542   5.8  94.6  *
##  12     0.215894   5.4 100.0  *
##         -------- -----
##  Total: 4.000000 100.0
##
##
## Columns:
```

```
##                         name   mass  qlt  inr     k=1 cor ctr     k=2 cor
## 1  |          WifeEduc:1low |   34  487   79 |  1521 266 162 | -1387 221
## 2  |   WifeEduc:2medium-low |   76  224   65 |   599 105  55 |   637 119
## 3  | WifeEduc:3medium-high |   93  106   58 |   114   5   2 |   510 101
## 4  |         WifeEduc:4high |  131  528   54 |  -828 442 183 |  -366  86
## 5  |            Method:none |  142  418   50 |   682 346 135 |  -310  71
## 6  |       Method:long-term |   75  473   68 | -1110 360 189 |  -621 113
## 7  |        Method:mid-term |  116  335   55 |  -116   7   3 |   786 328
## 8  |     AgeRel:(16,24.2].0 |    3    8   79 |  -696   5   3 |   598   3
## 9  |   AgeRel:(24.2,32.5].0 |   14   90   78 | -1369  82  54 |  -409   7
## 10 |   AgeRel:(32.5,40.8].0 |   20  135   77 | -1424 130  83 |  -263   4
## 11 |     AgeRel:(40.8,49].0 |   12   85   78 | -1219  58  38 |  -843  28
## 12 |     AgeRel:(16,24.2].1 |   59  116   67 |   304  20  11 |   667  96
## 13 |   AgeRel:(24.2,32.5].1 |  100  205   58 |   167  12   6 |   671 193
## 14 |   AgeRel:(32.5,40.8].1 |   67   12   64 |  -202  10   6 |   -73   1
## 15 |     AgeRel:(40.8,49].1 |   57  553   73 |   777 124  70 | -1446 429
##    ctr
## 1  145 |
## 2   67 |
## 3   53 |
## 4   38 |
## 5   30 |
## 6   64 |
## 7  156 |
## 8    2 |
## 9    5 |
## 10   3 |
## 11  19 |
## 12  58 |
## 13  99 |
## 14   1 |
## 15 260 |
```

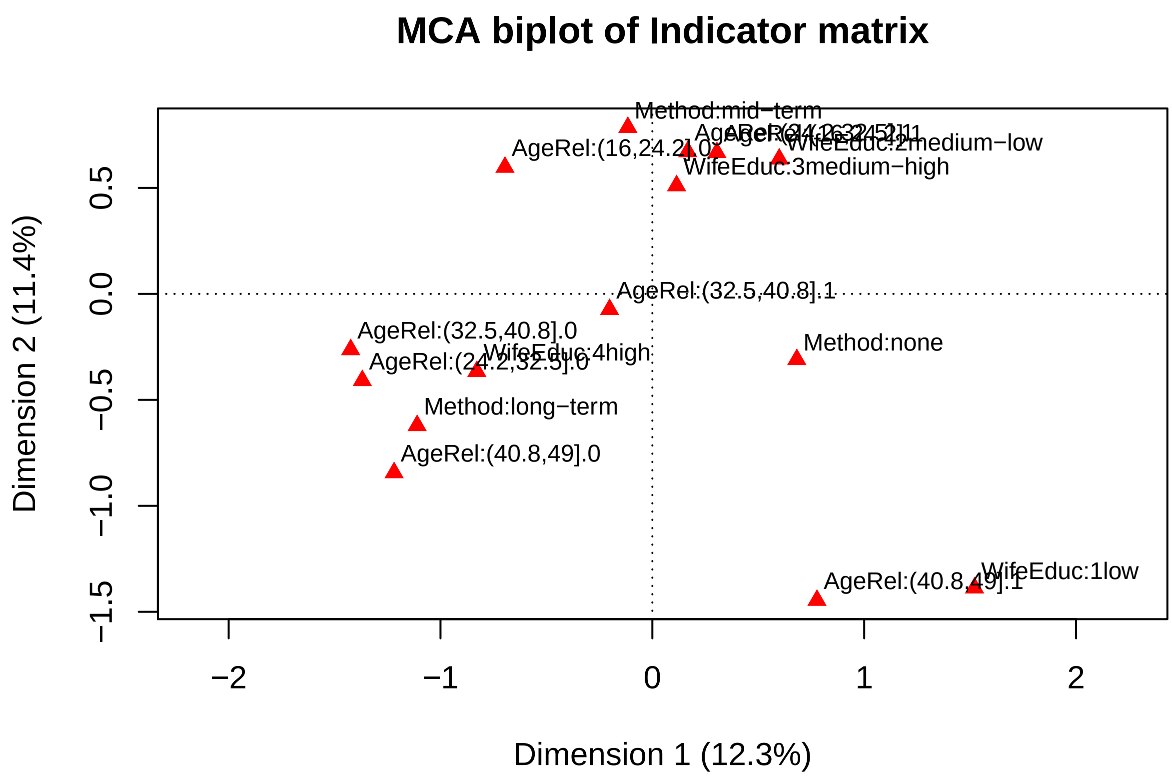
```r
out$inertia.t
```

```
## [1] 4
```

L'anàlisi té 12 dimensions, i la inercia de la matriu d'indicadors es 4. En dues dimensions la bondat de l'ajust és del 23.7%.

p)

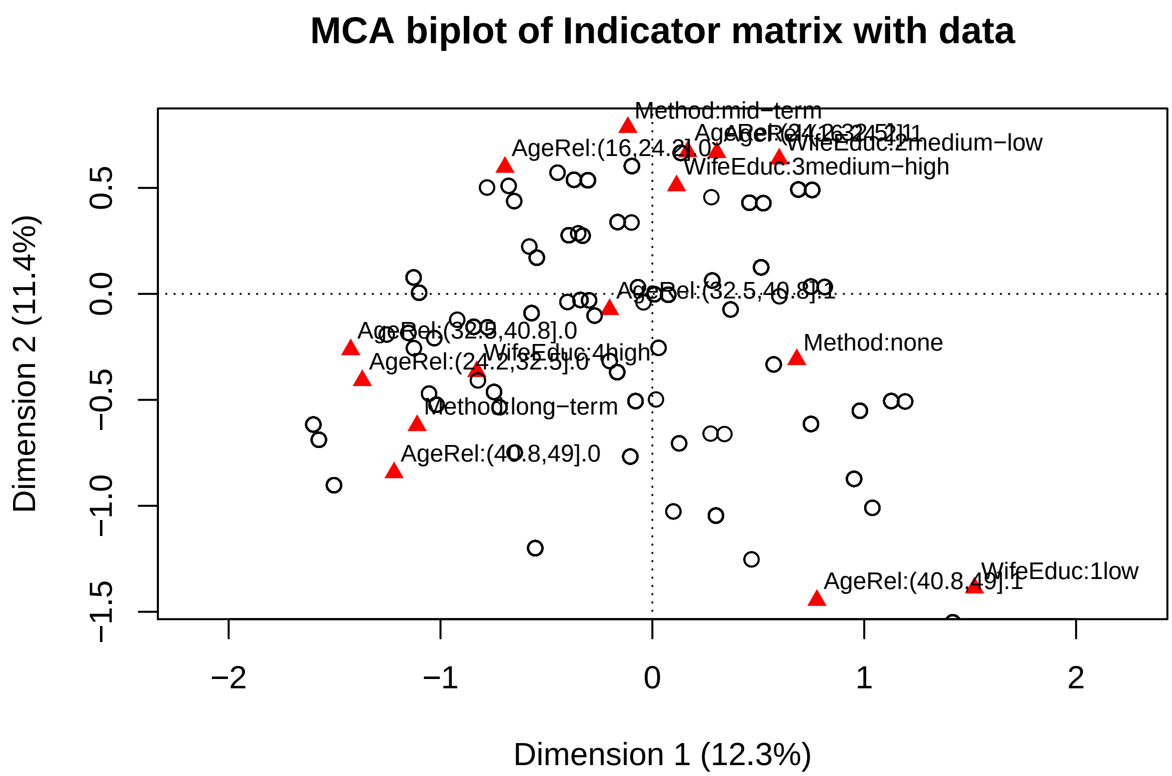
```r
plot(out, main="MCA biplot of Indicator matrix")
```



Podem veure que les dones amb poca educació tendeixen a ser grans i musulmanes, i a més són les que més tendeixen a no usar mètodes anticonceptius. Per altra banda, també es veuen dos agrupacions de punts que diuen que el nivell d'educació alt, usar mètodes a llarg termini, i no ser musulmana ni jove està correlacionat. Però ser jove o tenir educació mitjana i usar mètodes a curt termini també està correlacionat.

q)

```
plot(out, main="MCA biplot of Indicator matrix with data")
points(out$rowpcoord[,1], out$rowpcoord[,2])
```



Estàn tots els punts representats però hi ha moltes superposicions, i per això a la gràfica no es contemplen 1473 punts.

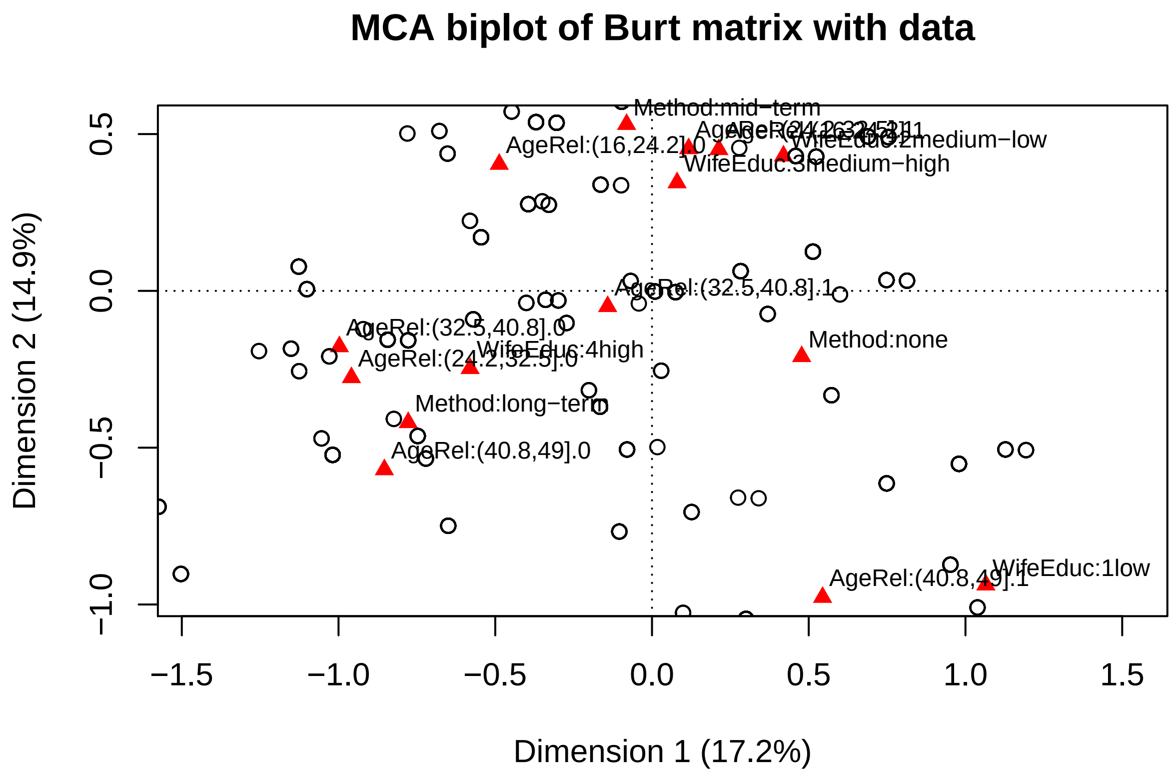
r)

```
out2 <- mjca(X[,c("WifeEduc", "Method", "AgeRel")], lambda="Burt")
summary(out2)
```

```
##
## Principal inertias (eigenvalues):
##
##  dim    value      %   cum%   scree plot
##  1      0.240645  17.2  17.2  ****
##  2      0.208641  14.9  32.0  ****
##  3      0.127566   9.1  41.1  **
##  4      0.119051   8.5  49.6  **
##  5      0.112167   8.0  57.6  **
##  6      0.111111   7.9  65.6  **
##  7      0.111111   7.9  73.5  **
##  8      0.103433   7.4  80.9  **
##  9      0.089698   6.4  87.3  **
##  10     0.078036   5.6  92.8  *
##  11     0.054076   3.9  96.7  *
##  12     0.046610   3.3 100.0  *
##         -------- -----
##  Total: 1.402145 100.0
##
##
## Columns:
```

```
##                         name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr
## 1  |          WifeEduc:1low |   34  627   79 | 1065 353 162 | -937 274 145
## 2  |   WifeEduc:2medium-low |   76  301   65 |  419 147  55 |  430 155  67
## 3  | WifeEduc:3medium-high |   93  142   58 |   80   7   2 |  345 135  53
## 4  |         WifeEduc:4high |  131  681   54 | -580 577 183 | -247 105  38
## 5  |            Method:none |  142  554   50 |  478 465 135 | -209  89  30
## 6  |       Method:long-term |   75  617   68 | -778 478 189 | -420 139  64
## 7  |        Method:mid-term |  116  436   55 |  -81  10   3 |  531 426 156
## 8  |    AgeRel:(16,24.2].0 |    3   11   79 | -487   7   3 |  404   5   2
## 9  |  AgeRel:(24.2,32.5].0 |   14  128   78 | -959 118  54 | -276  10   5
## 10 |  AgeRel:(32.5,40.8].0 |   20  192   77 | -997 186  83 | -178   6   3
## 11 |    AgeRel:(40.8,49].0 |   12  120   78 | -854  83  38 | -570  37  19
## 12 |    AgeRel:(16,24.2].1 |   59  158   67 |  213  29  11 |  451 129  58
## 13 |  AgeRel:(24.2,32.5].1 |  100  272   58 |  117  17   6 |  454 255  99
## 14 |  AgeRel:(32.5,40.8].1 |   67   17   64 | -141  15   6 |  -50   2   1
## 15 |    AgeRel:(40.8,49].1 |   57  695   73 |  544 165  70 | -977 531 260
##
## 1  |
## 2  |
## 3  |
## 4  |
## 5  |
## 6  |
## 7  |
## 8  |
## 9  |
## 10 |
## 11 |
## 12 |
## 13 |
## 14 |
## 15 |
```

```
plot(out2, main="MCA biplot of Burt matrix with data")
points(out2$rowpcoord[,1], out$rowpcoord[,2])
```



S'assembla molt a l'analisi utilitzant la matriu d'indicadors. El plot sembla el mateix pero escalat, i els punts estàn una mica desplcaçats cap a l'esquerra. Hi ha diferèncie en les inèrcies del ordre de 10^{-3} o 10^{-2}, tot i que al final la diferència total es de 2.6. Tot i amb això, amb dues dimension s'obté una bondat d'ajust del 32.0%, mentre que amb la matriu d'indicadors aconseguiem un 23.7%. Podem veure també aquesta diferència comparant els screeplots dels summaries dels dos anàlisis.

Exercici 2: Distribució de característiques esquelètiques

```r
library("ellipse")
```

```
##
## Attaching package: 'ellipse'

## The following object is masked from 'package:graphics':
##
##     pairs
```

```r
library("ICSNP")
```

```
## Loading required package: mvtnorm

## Loading required package: ICS
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

a)

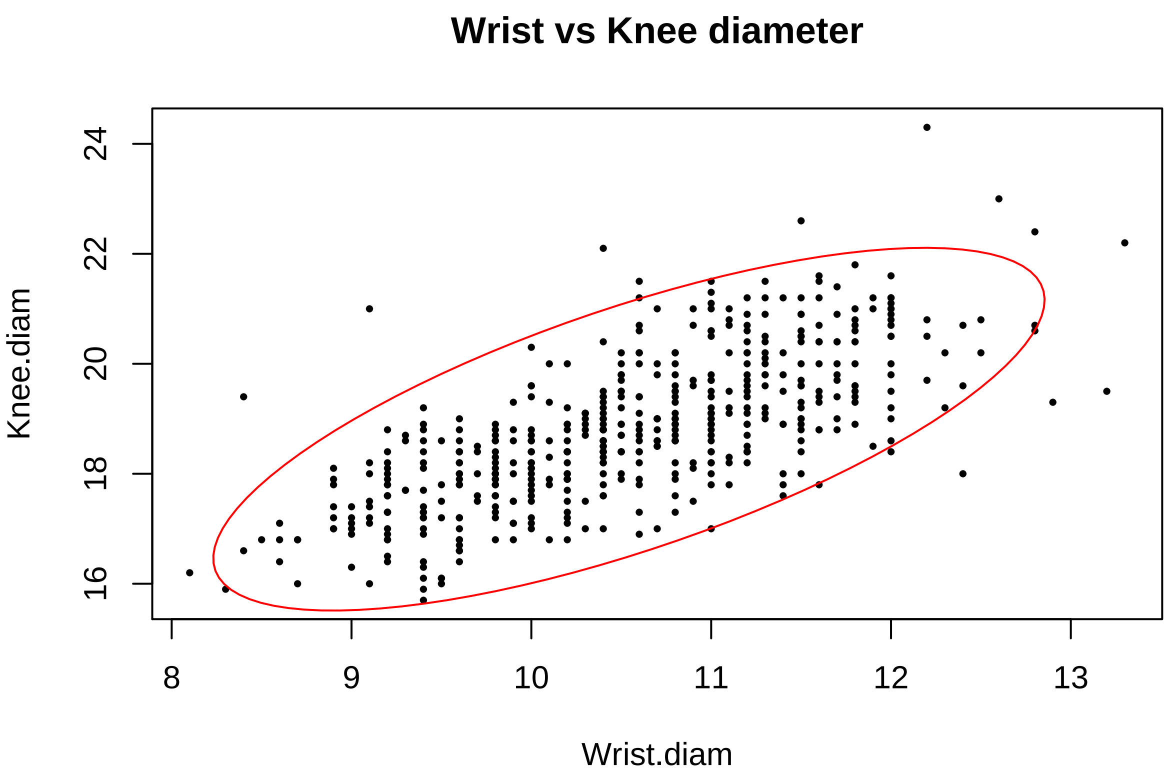
```r
esq <- read.table("http://www-eio.upc.es/%7Ejan/Data/MVA/body.dat")
esq <- esq[,c(1:9,25)]
colnames(esq) <- c("Biacromial.diam", "Biiliac.diam", "Bitrochanteric.diam", "Chest.depth", "Chest.diam
esq$Gender <- factor(esq$Gender, levels = c(0, 1), labels = c("female", "male"))
head(esq)
```

```
##   Biacromial.diam Biiliac.diam Bitrochanteric.diam Chest.depth Chest.diam
## 1            42.9         26.0                31.5        17.7       28.0
## 2            43.7         28.5                33.5        16.9       30.8
## 3            40.1         28.2                33.3        20.9       31.7
## 4            44.3         29.9                34.0        18.4       28.2
## 5            42.5         29.9                34.0        21.5       29.4
## 6            43.3         27.0                31.5        19.6       31.3
##   Elbow.diam Wrist.diam Knee.diam Ankle.diam Gender
## 1       13.1       10.4      18.8       14.1   male
## 2       14.0       11.8      20.6       15.1   male
## 3       13.9       10.9      19.7       14.1   male
## 4       13.9       11.2      20.9       15.0   male
## 5       15.2       11.6      20.7       14.9   male
## 6       14.0       11.5      18.8       13.9   male
```

```r
attach(esq)
```

b)

```r
plot(Wrist.diam, Knee.diam, main="Wrist vs Knee diameter", pch=16, cex=0.5)
mean.WK <- c(mean(Wrist.diam), mean(Knee.diam))
WK <- data.frame(Wrist.diam, Knee.diam)
cov.WK <- cov(WK)
contour.WK <- ellipse(cov.WK, centre=mean.WK)
lines(contour.WK[,1], contour.WK[,2], col = "red")
```



c) Mirant la grafica podem observar que hi ha 19 punts fora de l'elipse, y el nombre que esperariem es 25.

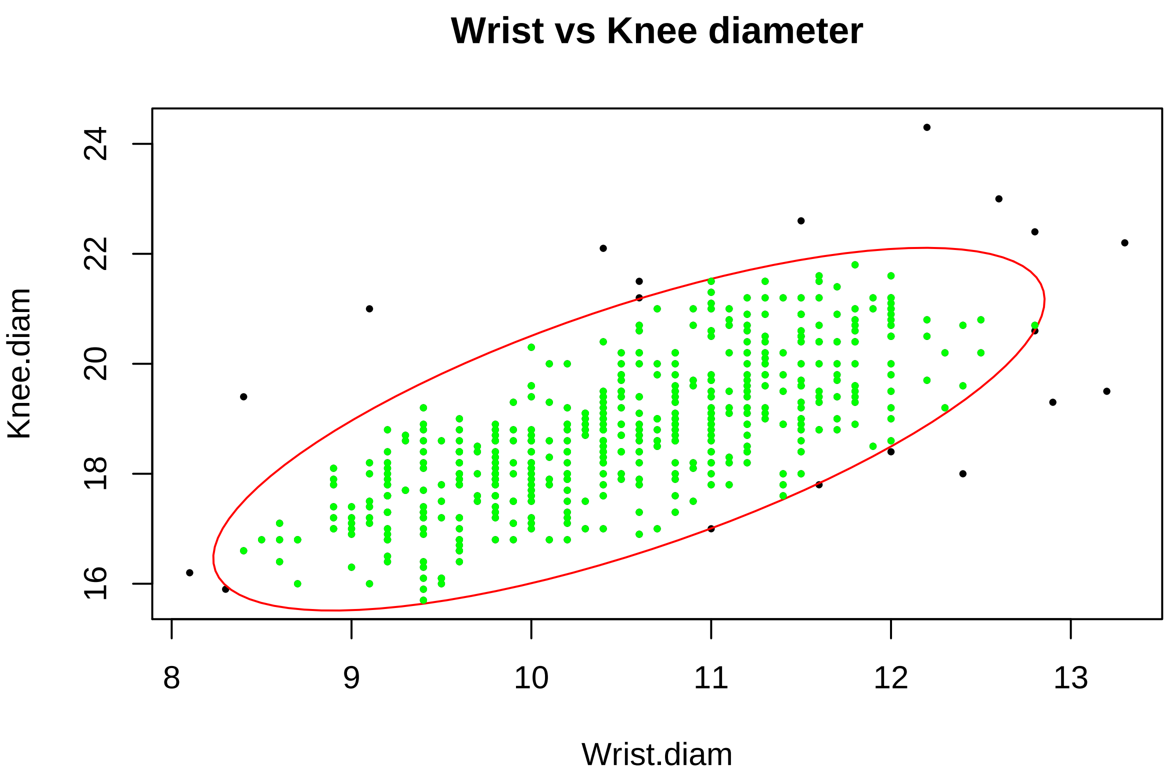
```r
print(exp <- nrow(esq)*5/100)
```

```
## [1] 25.35
```

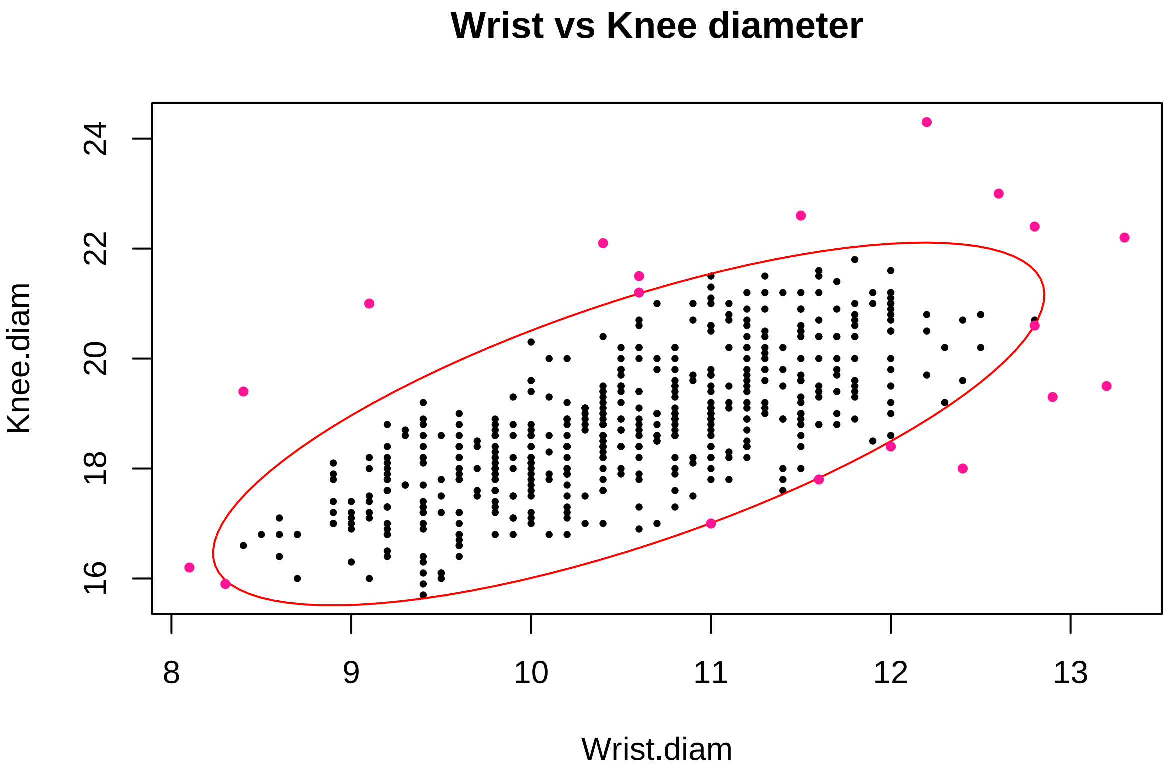
d)

Si la funció de densitat de un punt dona un valor menor que el valor que donen els punts del contorn, llavors està a fora de l' elipse. Tot i que també es pot calcular geomètricament sabent els eixos de l'elipse y el seu centre.

```r
# Método 1 (Geométrico)
plot(Wrist.diam, Knee.diam, main="Wrist vs Knee diameter", pch=16, cex=0.5)
lines(contour.WK[,1], contour.WK[,2], col = "red")
library(SIBER)
Z <- pointsToEllipsoid(WK, cov.WK, mean.WK)
inside <- ellipseInOut(Z)
points(WK[inside,], pch=16, col="green", cex=0.5)
```




```r
# Método 2 (Estadístico)
library(mvtnorm)
limit <- dmvnorm(contour.WK[1,], mean=mean.WK, sigma=cov.WK)
outside <- dmvnorm(WK, mean=mean.WK, sigma=cov.WK) < limit
plot(Wrist.diam, Knee.diam, main="Wrist vs Knee diameter", pch=16, cex=0.5)
lines(contour.WK[,1], contour.WK[,2], col = "red")
points(WK[outside,], pch=16, col="deeppink", cex=0.7)
```

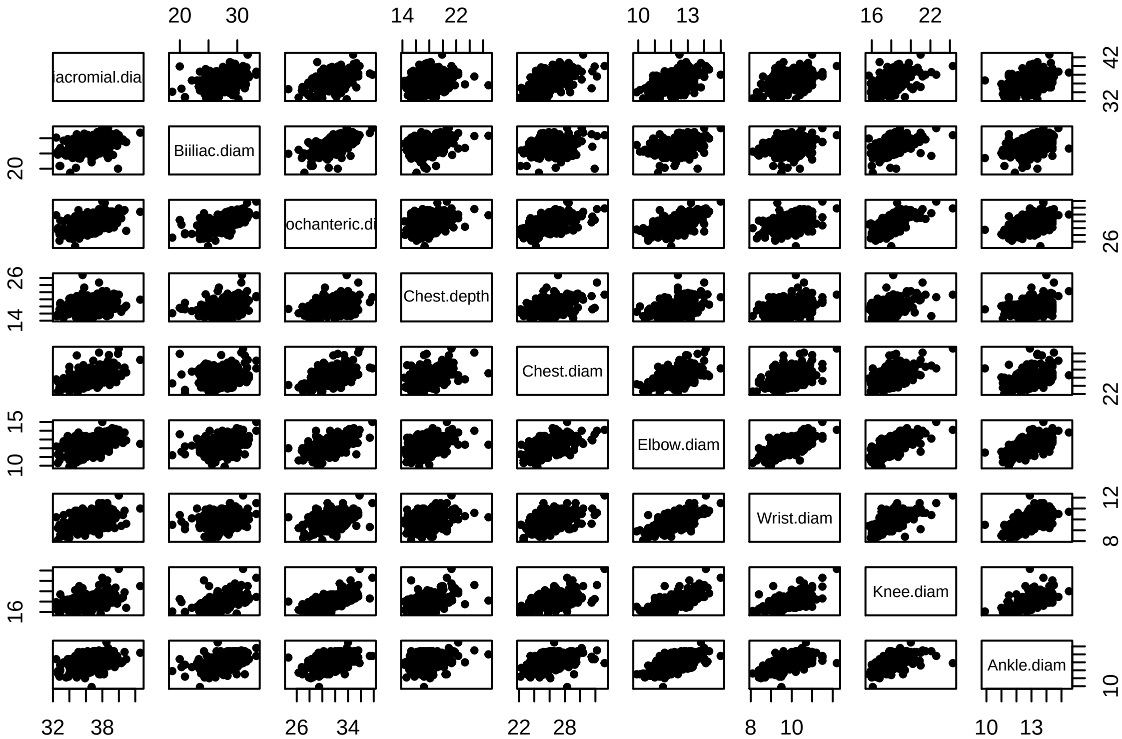


e)

```
detach(esq)
esq.f <- filter(esq, Gender=="female")
esq.f <- esq.f[,-10]
```

f)

```
pairs(esq.f, pch = 20)
```



g)

```r
chisq_plot <- function(X){
  M <- as.matrix(scale(X, center = T, scale = F))
  S <- solve(cov(X))
  v <- c()
  for (i in 1:nrow(M)) {
    aux <- t(M[i,])%*%S%*%M[i,]
    v <- c(v, aux)
  }
  sort(v)
}
(esq.f.dist.sq.ord <- chisq_plot(esq.f))
```

```
##   [1] 1.614854 1.876730 2.135547 2.293399 2.361022 2.461525 2.519774
##   [8] 2.546432 2.661758 2.782444 3.011668 3.096715 3.107248 3.116671
##  [15] 3.121416 3.244018 3.283014 3.290570 3.309792 3.445509 3.485289
##  [22] 3.490559 3.492904 3.500346 3.553546 3.589107 3.840341 3.879719
##  [29] 3.971242 3.985971 4.005135 4.017084 4.051308 4.076136 4.096903
##  [36] 4.112500 4.129188 4.137548 4.140787 4.175190 4.197765 4.207455
##  [43] 4.234116 4.274493 4.330855 4.333695 4.346814 4.398121 4.439471
##  [50] 4.455623 4.456969 4.467459 4.471037 4.486168 4.493606 4.494330
##  [57] 4.513944 4.523815 4.592725 4.642134 4.650526 4.666414 4.670324
##  [64] 4.744588 4.916874 4.928640 5.010485 5.094324 5.122366 5.130664
##  [71] 5.136954 5.237419 5.260757 5.261181 5.301205 5.331515 5.369237
##  [78] 5.383553 5.389557 5.419807 5.481445 5.506699 5.507348 5.522246
##  [85] 5.684349 5.719844 5.725653 5.803130 5.827665 5.835961 5.838214
##  [92] 5.866001 5.932005 5.943784 6.066505 6.163923 6.207472 6.284313
##  [99] 6.318578 6.336879 6.356863 6.380743 6.466149 6.490106 6.495413
## [106] 6.519300 6.544996 6.598827 6.602572 6.632325 6.754529 6.758676
```

```
## [113]  6.816060  6.838044  6.846797  6.882979  6.901981  7.013221  7.059718
## [120]  7.068515  7.073742  7.078313  7.084657  7.225861  7.249228  7.268647
## [127]  7.341242  7.389092  7.408357  7.415898  7.447574  7.477442  7.493422
## [134]  7.560681  7.561957  7.597293  7.655623  7.744225  7.765662  7.891551
## [141]  7.911550  7.941179  7.961948  8.001943  8.023236  8.029642  8.046205
## [148]  8.074883  8.136608  8.147027  8.153852  8.203672  8.230997  8.266085
## [155]  8.266092  8.400502  8.401860  8.433514  8.472850  8.474558  8.480122
## [162]  8.511126  8.521490  8.588312  8.627507  8.664994  8.704224  8.710790
## [169]  8.767446  8.892980  8.918246  8.943260  8.952328  9.054727  9.082765
## [176]  9.183776  9.314717  9.428410  9.500931  9.545189  9.622885  9.623207
## [183]  9.660806  9.663457  9.694757  9.702842  9.894264  9.915855  9.977803
## [190]  9.989690  9.995306 10.278050 10.437873 10.565243 10.694855 10.768040
## [197] 10.820655 10.856596 10.934380 11.034163 11.095348 11.100915 11.148988
## [204] 11.197680 11.289975 11.330933 11.624327 11.660842 11.774664 12.241728
## [211] 12.249628 12.428750 12.468999 12.483368 12.743209 12.947631 12.966730
## [218] 13.175455 13.409371 13.446852 13.584005 14.020391 14.188699 14.450502
## [225] 14.880239 14.930638 15.095489 15.357331 15.516899 15.592214 15.606884
## [232] 15.880392 16.056696 16.132222 16.208454 16.455393 16.463991 16.476143
## [239] 16.615482 16.697783 16.786825 17.375693 17.898106 18.181830 18.933329
## [246] 20.307927 20.516087 20.523330 21.592733 21.948983 22.349417 23.441132
## [253] 24.451724 25.848819 28.335896 28.819315 34.233823 34.372837 36.271954
## [260] 55.984541
```

h) Cal aplicar 9 graus de llibertat, que es el nombre de variables que tenim.

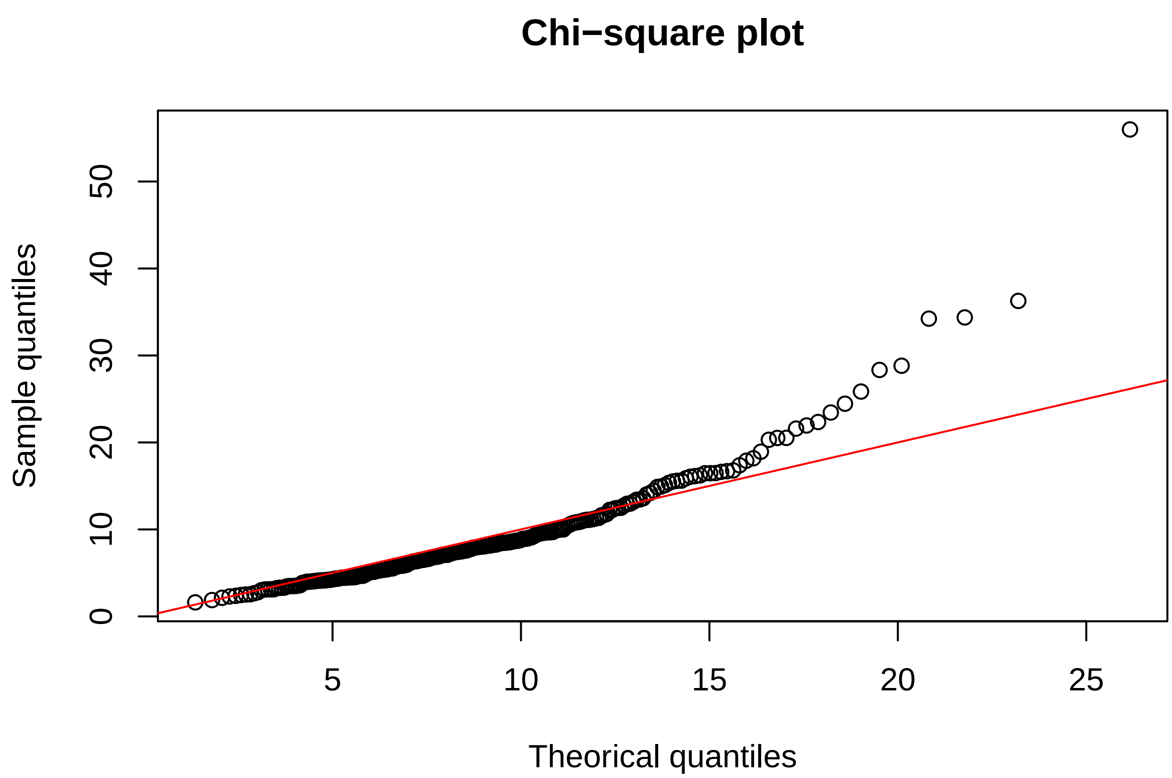
```r
n <- length(esq.f.dist.sq.ord)
p <- dim(esq)[2]-1
rang <- ((1:n)-0.5)/n
(quantils <- qchisq(rang, df=p))
```

```
##   [1]  1.356680  1.801585  2.065751  2.265573  2.430587  2.573321  2.700389
##   [8]  2.815752  2.921986  3.020876  3.113711  3.201456  3.284855  3.364494
##  [15]  3.440846  3.514296  3.585162  3.653714  3.720178  3.784749  3.847596
##  [22]  3.908865  3.968684  4.027167  4.084413  4.140514  4.195547  4.249587
##  [29]  4.302697  4.354938  4.406363  4.457020  4.506956  4.556212  4.604825
##  [36]  4.652832  4.700264  4.747151  4.793523  4.839405  4.884822  4.929796
##  [43]  4.974350  5.018503  5.062275  5.105683  5.148744  5.191474  5.233888
##  [50]  5.276002  5.317827  5.359377  5.400665  5.441702  5.482499  5.523067
##  [57]  5.563416  5.603556  5.643497  5.683246  5.722814  5.762207  5.801434
##  [64]  5.840503  5.879421  5.918195  5.956833  5.995340  6.033723  6.071989
##  [71]  6.110143  6.148191  6.186139  6.223993  6.261757  6.299438  6.337040
##  [78]  6.374568  6.412027  6.449422  6.486757  6.524038  6.561267  6.598451
##  [85]  6.635592  6.672695  6.709765  6.746805  6.783819  6.820812  6.857786
##  [92]  6.894746  6.931696  6.968640  7.005580  7.042520  7.079465  7.116417
##  [99]  7.153381  7.190359  7.227355  7.264373  7.301416  7.338487  7.375590
## [106]  7.412728  7.449904  7.487123  7.524386  7.561698  7.599061  7.636480
## [113]  7.673957  7.711496  7.749100  7.786772  7.824516  7.862336  7.900234
## [120]  7.938214  7.976279  8.014434  8.052680  8.091023  8.129465  8.168010
## [127]  8.206661  8.245423  8.284298  8.323291  8.362405  8.401644  8.441012
## [134]  8.480513  8.520150  8.559928  8.599851  8.639923  8.680147  8.720529
## [141]  8.761072  8.801781  8.842661  8.883715  8.924949  8.966367  9.007974
## [148]  9.049775  9.091774  9.133978  9.176390  9.219017  9.261864  9.304936
## [155]  9.348239  9.391779  9.435562  9.479594  9.523880  9.568428  9.613244
## [162]  9.658335  9.703707  9.749368  9.795324  9.841584  9.888156  9.935045
```

```
## [169]  9.982262 10.029815 10.077711 10.125961 10.174572 10.223555 10.272919
## [176] 10.322674 10.372831 10.423400 10.474393 10.525821 10.577695 10.630029
## [183] 10.682835 10.736125 10.789915 10.844217 10.899047 10.954420 11.010352
## [190] 11.066860 11.123961 11.181673 11.240014 11.299004 11.358664 11.419014
## [197] 11.480078 11.541878 11.604439 11.667786 11.731947 11.796948 11.862820
## [204] 11.929593 11.997300 12.065976 12.135657 12.206380 12.278187 12.351119
## [211] 12.425222 12.500543 12.577134 12.655048 12.734342 12.815078 12.897321
## [218] 12.981140 13.066609 13.153808 13.242823 13.333746 13.426676 13.521720
## [225] 13.618994 13.718622 13.820743 13.925503 14.033066 14.143609 14.257327
## [232] 14.374434 14.495169 14.619794 14.748604 14.881925 15.020127 15.163624
## [239] 15.312887 15.468455 15.630944 15.801067 15.979654 16.167678 16.366293
## [246] 16.576882 16.801118 17.041055 17.299255 17.578970 17.884415 18.221194
## [253] 18.596995 19.022768 19.514902 20.099582 20.822660 21.776373 23.196756
## [260] 26.160414
```

i) Sembla que al principi segueixen molt la tendencia de ser normal bivariant pero al final podem observar outliers que tenen una distancia molt mes gran a l'esperada.

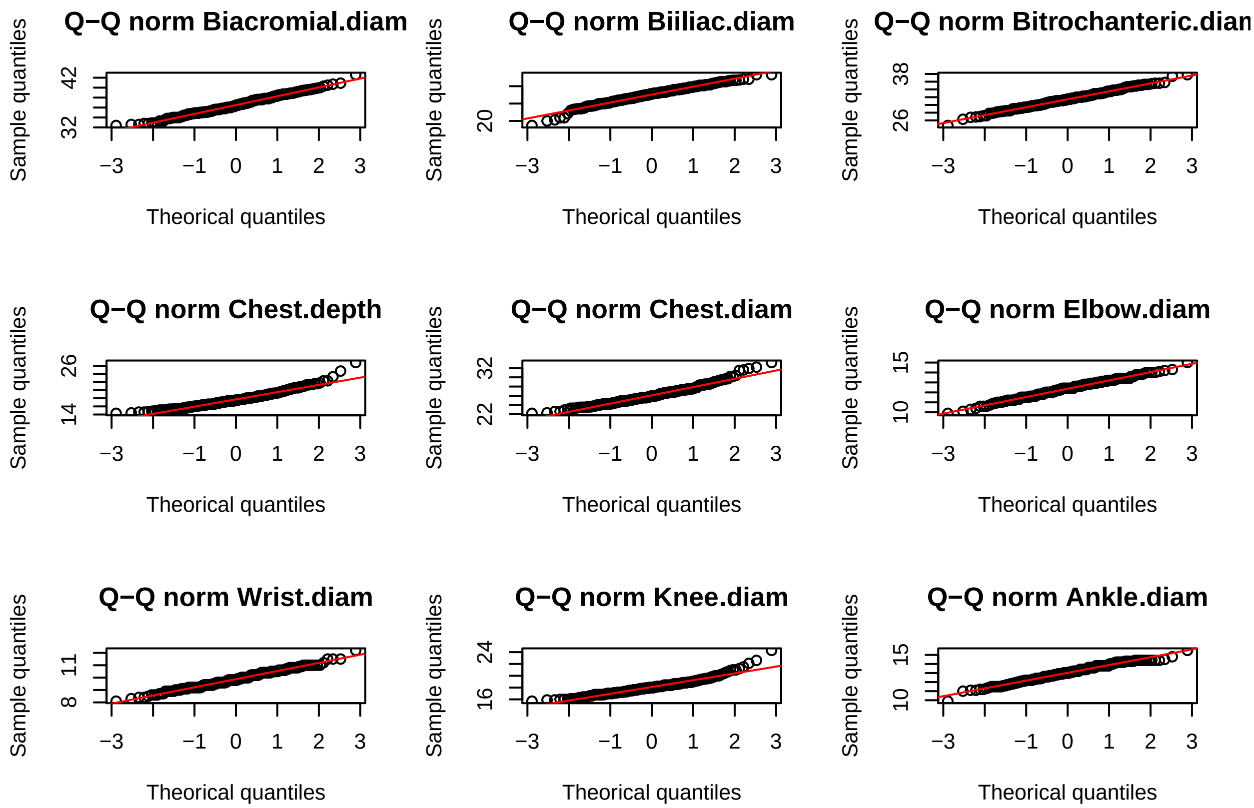
```r
plot(quantils, esq.f.dist.sq.ord, xlab="Theorical quantiles", ylab="Sample quantiles", main="Chi-square
abline(a=0, b=1, col="red")
```



j) Podem observar que es creible la normalitat marginal de les variables.

```r
sample.quantils <- apply(esq.f, 2, sort)
th.quantils <- qnorm(rang)
par(mfrow=c(3,3))
for (i in 1:9){
  plot(th.quantils, sample.quantils[,i], xlab="Theorical quantiles", ylab="Sample quantiles", main=paste
  abline(lm(sample.quantils[,i]~th.quantils), col="red")
```

```
}
```



k)

En aquest cas els punts semblen seguir una recta quasi perfecta, per tant es bastant probalbe que segueixin una distribució normal multivariada. En este caso los puntos parecen seguir la recta de forma casi perfecta, por lo que es bastante probable que sigan una distribución normal multivariada.

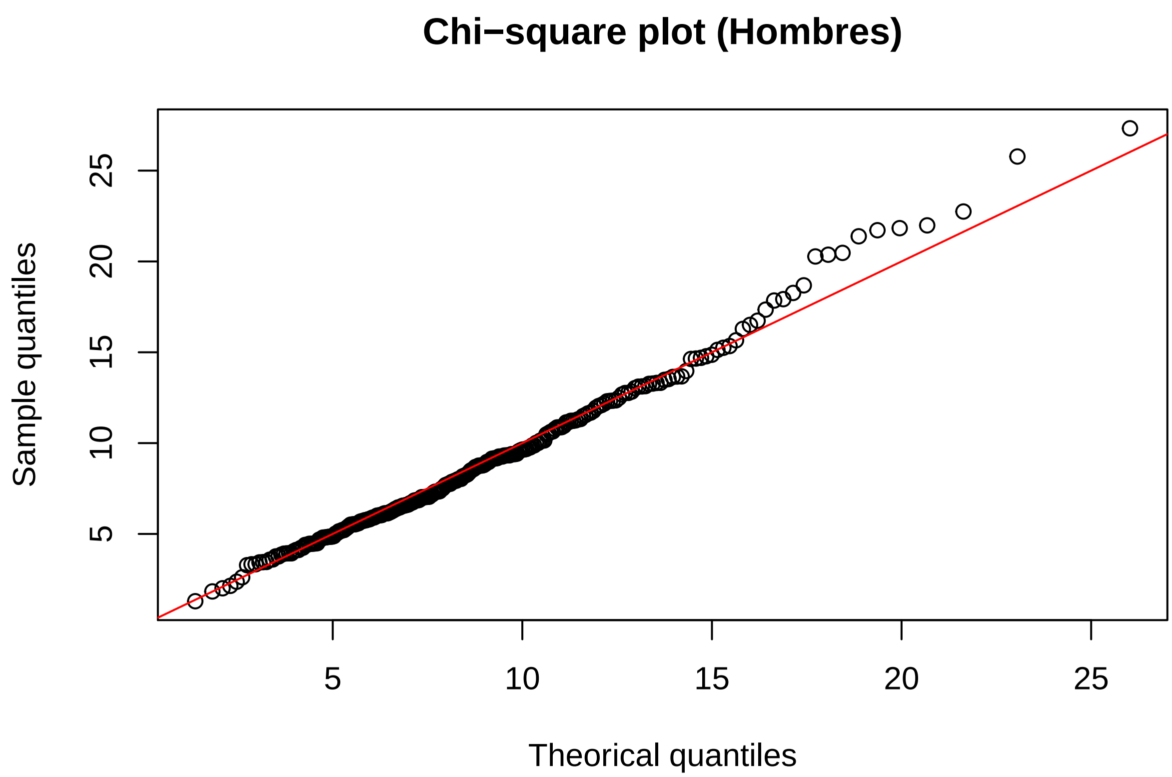
```r
esq.m <- filter(esq, Gender=="male")
esq.m <- esq.m[,-10]
esq.m.centre <- as.matrix(scale(esq.m, scale=F))
esq.m.cov <- as.matrix(cov(esq.m))
esq.m.cov.inv <- solve(esq.m.cov)
esq.m.dist.sq <- diag((esq.m.centre) %*% esq.m.cov.inv %*% t(esq.m.centre))
esq.m.dist.sq.ord <- sort(esq.m.dist.sq)

n.m <- length(esq.m.dist.sq.ord)
rang <- ((1:n.m)-0.5)/n.m
(quantils.m <- qchisq(rang, df=p))
```

```
##   [1]  1.374400  1.826209  2.094777  2.298091  2.466097  2.611505  2.741022
##   [8]  2.858664  2.967048  3.067984  3.162779  3.252413  3.337641  3.419059
##  [15]  3.497144  3.572289  3.644817  3.715002  3.783073  3.849228  3.913639
##  [22]  3.976453  4.037802  4.097801  4.156551  4.214143  4.270659  4.326172
##  [29]  4.380749  4.434450  4.487328  4.539436  4.590817  4.641515  4.691568
##  [36]  4.741012  4.789880  4.838204  4.886011  4.933329  4.980182  5.026595
##  [43]  5.072588  5.118183  5.163398  5.208253  5.252765  5.296949  5.340822
##  [50]  5.384398  5.427691  5.470713  5.513479  5.556000  5.598286  5.640351
##  [57]  5.682203  5.723852  5.765310  5.806584  5.847683  5.888617  5.929394
```

```
## [64]  5.970021  6.010505  6.050856  6.091079  6.131182  6.171171  6.211053
## [71]  6.250834  6.290521  6.330119  6.369634  6.409072  6.448439  6.487739
## [78]  6.526979  6.566162  6.605295  6.644383  6.683429  6.722440  6.761419
## [85]  6.800371  6.839301  6.878213  6.917112  6.956002  6.994887  7.033771
## [92]  7.072659  7.111555  7.150462  7.189386  7.228329  7.267297  7.306292
## [99]  7.345319  7.384383  7.423486  7.462632  7.501826  7.541072  7.580373
## [106]  7.619733  7.659156  7.698646  7.738207  7.777843  7.817558  7.857355
## [113]  7.897239  7.937213  7.977282  8.017450  8.057720  8.098097  8.138585
## [120]  8.179187  8.219909  8.260755  8.301728  8.342833  8.384074  8.425457
## [127]  8.466984  8.508662  8.550494  8.592486  8.634641  8.676966  8.719464
## [134]  8.762141  8.805002  8.848053  8.891297  8.934742  8.978391  9.022252
## [141]  9.066329  9.110629  9.155158  9.199920  9.244924  9.290175  9.335680
## [148]  9.381446  9.427478  9.473786  9.520375  9.567253  9.614427  9.661907
## [155]  9.709698  9.757811  9.806253  9.855033  9.904161  9.953644 10.003494
## [162] 10.053720 10.104331 10.155339 10.206755 10.258589 10.310854 10.363561
## [169] 10.416723 10.470352 10.524462 10.579067 10.634181 10.689818 10.745996
## [176] 10.802729 10.860034 10.917929 10.976431 11.035560 11.095335 11.155777
## [183] 11.216908 11.278749 11.341324 11.404657 11.468775 11.533704 11.599472
## [190] 11.666109 11.733646 11.802116 11.871553 11.941994 12.013477 12.086042
## [197] 12.159732 12.234593 12.310672 12.388021 12.466693 12.546746 12.628241
## [204] 12.711243 12.795823 12.882055 12.970018 13.059799 13.151490 13.245191
## [211] 13.341008 13.439059 13.539468 13.642372 13.747921 13.856277 13.967618
## [218] 14.082140 14.200058 14.321610 14.447061 14.576704 14.710868 14.849923
## [225] 14.994284 15.144424 15.300882 15.464276 15.635320 15.814847 16.003831
## [232] 16.203431 16.415031 16.640307 16.881318 17.140632 17.421506 17.728165
## [239] 18.066224 18.443383 18.870614 19.364335 19.950776 20.675863 21.631978
## [246] 23.055497 26.024403
```

```r
plot(quantils.m, esq.m.dist.sq.ord, xlab="Theorical quantiles", ylab="Sample quantiles", main="Chi-squa
abline(a=0, b=1, col="red")
```



l) Existeixen diferencies significatives

```r
HotellingsT2(esq.f, esq.m, test = "f")
```

```
##
##  Hotelling's two sample T2-test
##
## data:  esq.f and esq.m
## T.2 = 175.74, df1 = 9, df2 = 497, p-value < 2.2e-16
## alternative hypothesis: true location difference is not equal to c(0,0,0,0,0,0,0,0,0)
```

Obtenim un estadístic $T^2 = 175.74$, i un p-valor més petit que $2.2\mathrm{e}-16$.

m) No es rellevant com podem veure a continuacio

```r
HotellingsT2(esq.f, esq.m, test = "chi")
```

```
##
##  Hotelling's two sample T2-test
##
## data:  esq.f and esq.m
## T.2 = 1607.1, df = 9, p-value < 2.2e-16
## alternative hypothesis: true location difference is not equal to c(0,0,0,0,0,0,0,0,0)
```

n)

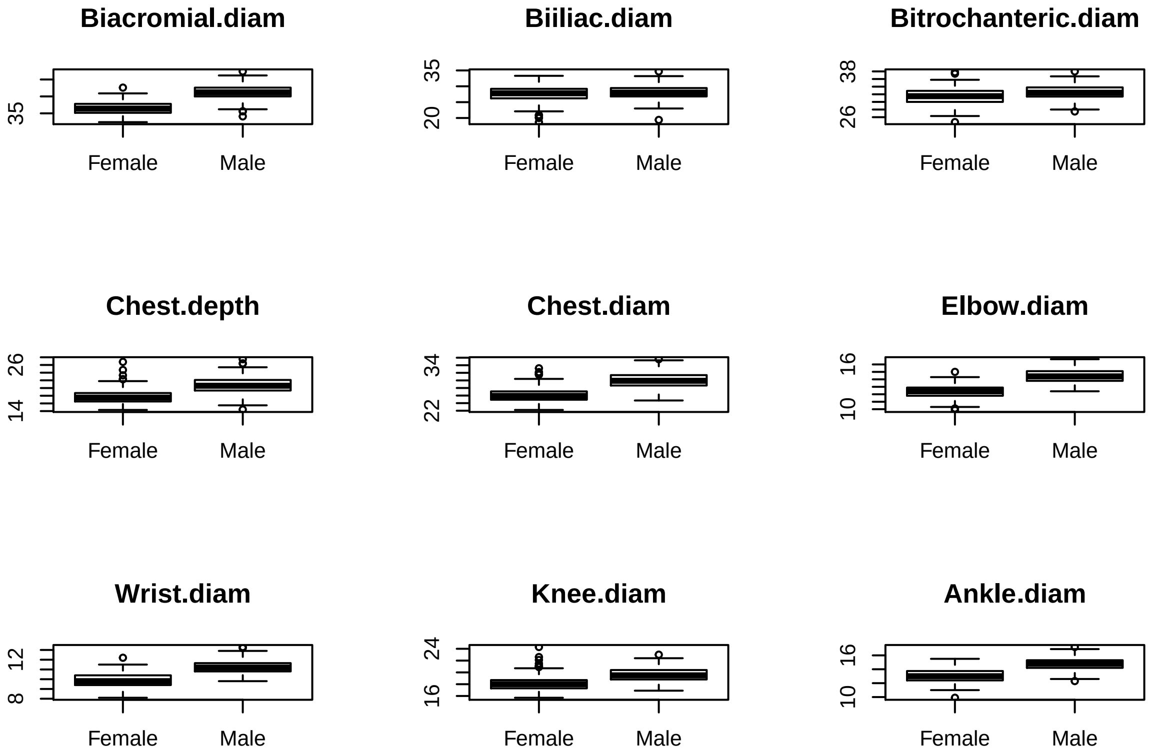
```r
for (i in 1:9){
  test.marginal <- t.test(esq.f[,i],esq.m[,i], var.equal=F)
  if (test.marginal$p.value < 0.001) {
    print(paste("Difference in", colnames(esq)[i]))
  } else {
    print(paste("No difference in", colnames(esq)[i]))
  }
}
```

```
## [1] "Difference in Biacromial.diam"
## [1] "No difference in Biiliac.diam"
## [1] "Difference in Bitrochanteric.diam"
## [1] "Difference in Chest.depth"
## [1] "Difference in Chest.diam"
## [1] "Difference in Elbow.diam"
## [1] "Difference in Wrist.diam"
## [1] "Difference in Knee.diam"
## [1] "Difference in Ankle.diam"
```

Si prenem $\alpha = 0.001$ trobem diferencies significatives en totes les variables execpte en el diàmetre biilical

o)

```r
par(mfrow=c(3,3))
for (i in 1:9){
  boxplot(esq.f[,i], esq.m[,i], names=c("Female","Male"), main=colnames(esq)[i])
}
```



Observem que en general, els homes tenen mides mes grans en les diferents parts del cos (excepte pel diàmtre biilical). Una caracterísitca potser també interessant és que s'observen més outliers en els boxplots de les dones que no pas en els dels homes.