UG Reinis data

G. Marchetti

2024-05-15

Reinis Data

Data collected at the beginning of a 15 year follow-up study of probable risk factors for coronary thrombosis. Data are from all men employed in a car factory ($N = 1841$).

A table with 6 discrete variables. A: smoking, B: strenous mental work, C: strenuous physical work, D: systolic blood pressure, E: ratio of lipoproteins, F: Family anamnesis of coronary heart disease.

```r
library("readr")
data_reinis <- read_rds("data_reinis.rds")
colnames(data_reinis) <- c('A', 'B', 'C', 'D', 'E', 'F')
tab_reinis <- table(data_reinis)
head(as.data.frame(tab_reinis))
```

```
  A B C D E F Freq
1 y y y y y y   44
2 n y y y y y   40
3 y n y y y y  112
4 n n y y y y   67
5 y y n y y y  129
6 n y n y y y  145
```

Test of conditional independence

A conditional independence $F \perp\!\!\!\perp D \mid BECA$ is expected. Let's test this hypothesis

```r
ci.test("F", "D", c("B", "E", "C", "A"), data = data_reinis)
```
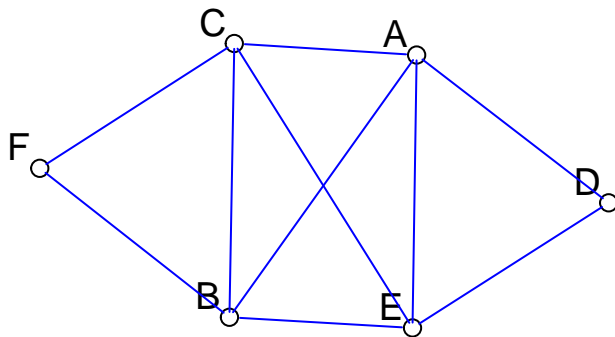
```
    Mutual Information (disc.)

data:  F ~ D | B + E + C + A
mi = 18.345, df = 16, p-value = 0.3041
alternative hypothesis: true value is greater than 0
```

The statistic `mi` is the same of the deviance. Apparently the conditional independence is not rejected.

The previous CI is implied for example by the following graph

```
G <- UG(~ F*C*B + B*E*C*A + E*A*D)
drawGraph(G, layout = layout_nicely)
```


Fit the graph

```
m_red <- glm(Freq ~ F*C*B + B*E*C*A + E*A*D,
             family = poisson,
             data = as.data.frame(tab_reinis))
summary(m_red)
```

```
Call:
glm(formula = Freq ~ F * C * B + B * E * C * A + E * A * D, family = poisson,
    data = as.data.frame(tab_reinis))

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.76049    0.11497  32.708  < 2e-16 ***
```

```
Fn             -1.96308     0.18589 -10.560  < 2e-16 ***
Cn              1.08271     0.12610   8.586  < 2e-16 ***
Bn              0.89935     0.12894   6.975 3.06e-12 ***
En             -0.52184     0.18076  -2.887 0.003891 **
An             -0.15948     0.17283  -0.923 0.356127
Dn             -0.22841     0.08232  -2.775 0.005525 **
Fn:Cn           0.03573     0.21424   0.167 0.867559
Fn:Bn           0.25384     0.21506   1.180 0.237867
Cn:Bn          -3.34495     0.25075 -13.340  < 2e-16 ***
Bn:En           0.14105     0.19736   0.715 0.474819
Cn:En          -0.40269     0.20281  -1.986 0.047080 *
Bn:An          -0.29927     0.20126  -1.487 0.137011
En:An           0.27564     0.26177   1.053 0.292347
Cn:An           0.29101     0.18812   1.547 0.121872
En:Dn           0.22290     0.13340   1.671 0.094738 .
An:Dn          -0.49731     0.12886  -3.859 0.000114 ***
Fn:Cn:Bn        0.34909     0.32898   1.061 0.288629
Cn:Bn:En        0.45540     0.37614   1.211 0.225999
Bn:En:An        0.12678     0.29504   0.430 0.667409
Cn:Bn:An        0.77717     0.33870   2.295 0.021757 *
Cn:En:An        0.08344     0.28944   0.288 0.773116
En:An:Dn        0.32489     0.19311   1.682 0.092487 .
Cn:Bn:En:An    -0.42525     0.50350  -0.845 0.398340
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2026.740  on 63  degrees of freedom
Residual deviance:   42.298  on 40  degrees of freedom
AIC: 379.95

Number of Fisher Scoring iterations: 4
```
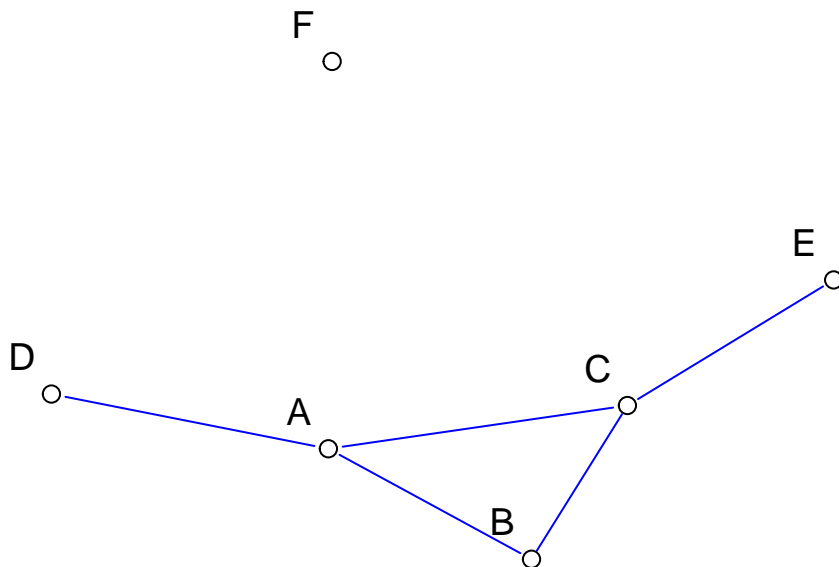
The model shows some further possible reductions

The strongest terms correspond to CB, CE, AD and CBA. Therefore a suitable graph should be

```
G0 <- UG(~ F + C*B*A + C*E + A*D)
drawGraph(G0, layout = layout_nicely)
```

Notice that surprisingly ther is a complete independence of family anamnesis of CHD from the other variables.

Fitting this graphical model we get

```
m0 <- glm(Freq ~  F + C*B*A + C*E + A*D,
              family = poisson,
              data = as.data.frame(tab_reinis))
summary(m0)
```

```
Call:
glm(formula = Freq ~ F + C * B * A + C * E + A * D, family = poisson,
    data = as.data.frame(tab_reinis))

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.57204    0.09380  38.083  < 2e-16 ***
Fn          -1.80513    0.06692 -26.973  < 2e-16 ***
Cn           1.10930    0.10549  10.516  < 2e-16 ***
Bn           0.99274    0.09689  10.246  < 2e-16 ***
An          -0.04598    0.12910  -0.356 0.721691
En          -0.11880    0.06580  -1.805 0.071016 .
Dn          -0.14385    0.06468  -2.224 0.026156 *
Cn:Bn       -3.11567    0.18512 -16.831  < 2e-16 ***
Cn:An        0.28787    0.14175   2.031 0.042270 *
```

```
Bn:An        -0.21704    0.14614  -1.485 0.137519
Cn:En        -0.38619    0.09482  -4.073 4.65e-05 ***
An:Dn        -0.31399    0.09472  -3.315 0.000917 ***
Cn:Bn:An      0.60616    0.24928   2.432 0.015030 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2026.740  on 63  degrees of freedom
Residual deviance:   96.812  on 51  degrees of freedom
AIC: 412.46

Number of Fisher Scoring iterations: 4
```

The deviance however indicates a quite poor agreement with the hypothesis so this model is rejected.