Analysis of data on Canadian women Labour Force participation

G. Marchetti

2024-05-08

Data on Women's Labour-Force Participation

The original data are in R package **carData**. The first 3 variables are transformed to binary data.

```r
wlfdata <- read.table("wlfdata.txt", header = TRUE, colClasses = "character")
wlfdata <- as.data.frame(lapply(wlfdata, factor))
```

Below is the code for the transformation

```
library(carData)
data(Womenlf)

wlfdata <- Womenlf
wlfdata$hincome <- factor(0 + (Womenlf$hincome > 14))
wlfdata$partic <-  factor(0 + (Womenlf$partic == "fulltime"))
wlfdata$children <-factor(0 + (Womenlf[,3]=="present"))
wlfdata$region <- Womenlf[,4]
colnames(wlfdata) <- c("L", "H", "C", "R")
head(wlfdata)
```

Contingency table

```r
ftable(R ~ H + C + L, data = wlfdata)
```

```
      R Atlantic BC Ontario Prairie Quebec
H C L
0 0 0          1  4       6       0      5
    1          1  4      11       3     10
  1 0         11  6      25      15     18
    1          2  0       4       4      4
1 0 0          1  4       6       1      5
    1          1  2      10       1      3
  1 0         11  8      44       7     19
    1          2  1       2       0      1
```

Test conditional independency

You need package **bnlearn** for this.

```r
ci.test( "H", "C", "R", data = wlfdata, test = "mi")
```

```
    Mutual Information (disc.)

data:  H ~ C | R
mi = 3.7358, df = 5, p-value = 0.588
alternative hypothesis: true value is greater than 0
```

```r
m_full0 <- glm(L ~ C *H * R, family = binomial, data = wlfdata)
round(summary(m_full0)$coef, 3)
```

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.000      1.414   0.000    1.000
C1              -1.705      1.610  -1.059    0.290
H1               0.000      2.000   0.000    1.000
RBC              0.000      1.581   0.000    1.000
ROntario         0.606      1.503   0.403    0.687
RPrairie        17.566   2284.102   0.008    0.994
RQuebec          0.693      1.517   0.457    0.648
C1:H1            0.000      2.276   0.000    1.000
C1:RBC         -15.861   1615.105  -0.010    0.992
C1:ROntario     -0.734      1.772  -0.414    0.679
C1:RPrairie    -17.183   2284.102  -0.008    0.994
C1:RQuebec      -0.492      1.788  -0.275    0.783
H1:RBC          -0.693      2.291  -0.303    0.762
```

```
H1:ROntario      -0.095      2.127  -0.045    0.964
H1:RPrairie     -17.566   2284.103  -0.008    0.994
H1:RQuebec       -1.204      2.198  -0.548    0.584
C1:H1:RBC        16.180   1615.106   0.010    0.992
C1:H1:ROntario   -1.163      2.553  -0.456    0.649
C1:H1:RPrairie    1.322   2730.025   0.000    1.000
C1:H1:RQuebec    -0.236      2.715  -0.087    0.931
```

```r
m_red0 <- glm(L ~ C *  H, family = binomial, data = wlfdata)
anova(m_red0, m_full0, test = "Chisq")
```

```
Analysis of Deviance Table

Model 1: L ~ C * H
Model 2: L ~ C * H * R
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       259     227.93
2       243     216.66 16   11.269   0.7926
```

Fit equation 1

```r
m_full1 <- glm(L ~ C + H + R, family = binomial, data = wlfdata)
round(summary(m_full1)$coef, 3)
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.041      0.607   1.715    0.086
C1            -2.609      0.361  -7.234    0.000
H1            -0.768      0.348  -2.210    0.027
RBC           -0.944      0.745  -1.266    0.206
ROntario      -0.254      0.590  -0.430    0.667
RPrairie       0.168      0.695   0.241    0.809
RQuebec       -0.342      0.627  -0.545    0.586
```

```r
m_red1 <- glm(L ~ C + H, family = binomial, data = wlfdata)
anova(m_red1, m_full1, test = "Chisq")
```

```
Analysis of Deviance Table

Model 1: L ~ C + H
```

```
Model 2: L ~ C + H + R
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       260     228.31
2       256     225.50  4   2.8117   0.5898
```

```r
`LRtest` <- function(m_red, m_full) {
l0 <- logLik(m_red)[1]
lsat <- logLik(m_full)[1]
w <- 2 * (lsat - l0)
df <- m_red$df.residual - m_full$df.residual
p <- 1 - pchisq(w, df)
c(w = w, df = df, p = p)
}
```

```r
LRtest(m_red1, m_full1)
```

```
        w        df         p
2.8117211 4.0000000 0.5898111
```

Fit equation 2

```r
m_full2 <- glm(C~ H+R, family = binomial, data = wlfdata)
round(summary(m_full2)$coef, 3)
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.671      0.551   3.030    0.002
H1             0.437      0.282   1.546    0.122
RBC           -1.827      0.656  -2.785    0.005
ROntario      -1.092      0.579  -1.886    0.059
RPrairie      -0.136      0.730  -0.186    0.852
RQuebec       -1.250      0.598  -2.089    0.037
```

```r
m_red2 <- glm(C~ R, family = binomial, data = wlfdata)
LRtest(m_red2, m_full2)
```

```
        w        df         p
2.4134751 1.0000000 0.1202951
```

Fit equation 3

```
m_full3 <- glm(H ~ R, family = binomial, data = wlfdata)
round(summary(m_full3)$coef, 3)
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.000      0.365   0.000    1.000
RBC            0.069      0.521   0.132    0.895
ROntario       0.298      0.414   0.721    0.471
RPrairie      -0.894      0.538  -1.660    0.097
RQuebec       -0.279      0.443  -0.629    0.529
```

```
m_red3 <- glm(H ~ 1, family = binomial, data = wlfdata)
LRtest(m_red3, m_full3)
```

```
         w         df          p
9.19254035 4.00000000 0.05646299
```