Rzeszów, 24.05.2023

WNIOSKOWANIE W WARUNKACH NIEPEWNOŚCI

PROJEKT

Analiza wpływu wybranych czynników na udar

Jakub Witkowski

1. Zbiór danych

Dane zostały pobrane ze strony https://www.kaggle.com. Zawierają informacje o pacjentach z ryzykiem udaru. Dane zawierają 11 kolumn: płeć, wiek, nadciśnienie, choroby serca, stan cywilny, rodzaj pracy, rodzaj zamieszkania, poziom glukozy, bmi, palący lub nie palący oraz czy pacjent miał już wcześniej udar.

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
1	Male	67.00	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
2	Female	61.00	0	0	Yes	Self-employed	Rural	202.21	N/A	never smoked	1
3	Male	80.00	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
4	Female	49.00	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
5	Female	79.00	1	0	Yes	Self-employed	Rural	174.12	24	never smoked	1
6	Male	81.00	0	0	Yes	Private	Urban	186.21	29	formerly smoked	1
7	Male	74.00	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
8	Female	69.00	0	0	No	Private	Urban	94.39	22.8	never smoked	1
9	Female	59.00	0	0	Yes	Private	Rural	76.15	N/A	Unknown	1
10	Female	78.00	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1
11	Female	81.00	1	0	Yes	Private	Rural	80.43	29.7	never smoked	1
12	Female	61.00	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1
13	Female	54.00	0	0	Yes	Private	Urban	104.51	27.3	smokes	1
14	Male	78.00	0	1	Yes	Private	Urban	219.84	N/A	Unknown	1
15	Female	79.00	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1
16	Female	50.00	1	0	Yes	Self-employed	Rural	167.41	30.9	never smoked	1
17	Male	64.00	0	1	Yes	Private	Urban	191.61	37.5	smokes	1
18	Male	75.00	1	0	Yes	Private	Urban	221.29	25.8	smokes	1
19	Female	60.00	0	0	No	Private	Urban	89.22	37.8	never smoked	1
20	Male	57.00	0	1	No	Govt_job	Urban	217.08	N/A	Unknown	1
21	Female	71.00	0	0	Yes	Govt_job	Rural	193.94	22.4	smokes	1
22	Female	52.00	1	0	Yes	Self-employed	Urban	233.29	48.9	never smoked	1
23	Female	79.00	0	0	Yes	Self-employed	Urban	228.70	26.6	never smoked	1
24	Male	82.00	0	1	Yes	Private	Rural	208.30	32.5	Unknown	1

2. Wykonanie projektu

Pierwszym krokiem była obróbka danych. Usunięte zostały niepotrzebne kolumny oraz zostały zmienione nazwy dla pozostałych kolumn.

```
#usuwanie kolumn
dane <- dane[,-c(1,5,8,9,10,12)]
#zmiana nazw kolumn
colnames(dane) <- c("P","W","N","C","RP","S")
```

Następnie został zmieniony typ danych na factor.

```
#zmiana typu na factor
dane$P <- as.factor(dane$P)
dane$W <- as.factor(dane$W)
dane$N <- as.factor(dane$N)
dane$C <- as.factor(dane$C)
dane$RP <- as.factor(dane$RP)
dane$S <- as.factor(dane$S)
```

Po obróbce danych zostało 6 kolumn, z których zostały stworzone cechy:

- P – płeć
- W – wiek
- N – nadciśnienie
- C – stan cywilny
- RP – rodzaj pracy
- S – palacz

	P	W	N	C	RP	S
1	Male	67	0	Yes	Private	formerly smoked
2	Female	61	0	Yes	Self-employed	never smoked
3	Male	80	0	Yes	Private	never smoked
4	Female	49	0	Yes	Private	smokes
5	Female	79	1	Yes	Self-employed	never smoked
6	Male	81	0	Yes	Private	formerly smoked
7	Male	74	1	Yes	Private	never smoked
8	Female	69	0	No	Private	never smoked
9	Female	59	0	Yes	Private	Unknown
10	Female	78	0	Yes	Private	Unknown
11	Female	81	1	Yes	Private	never smoked
12	Female	61	0	Yes	Govt_job	smokes
13	Female	54	0	Yes	Private	smokes
14	Male	78	0	Yes	Private	Unknown
15	Female	79	0	Yes	Private	never smoked
16	Female	50	1	Yes	Self-employed	never smoked
17	Male	64	0	Yes	Private	smokes
18	Male	75	1	Yes	Private	smokes
19	Female	60	0	No	Private	never smoked
20	Male	57	0	No	Govt_job	Unknown
21	Female	71	0	Yes	Govt_job	smokes

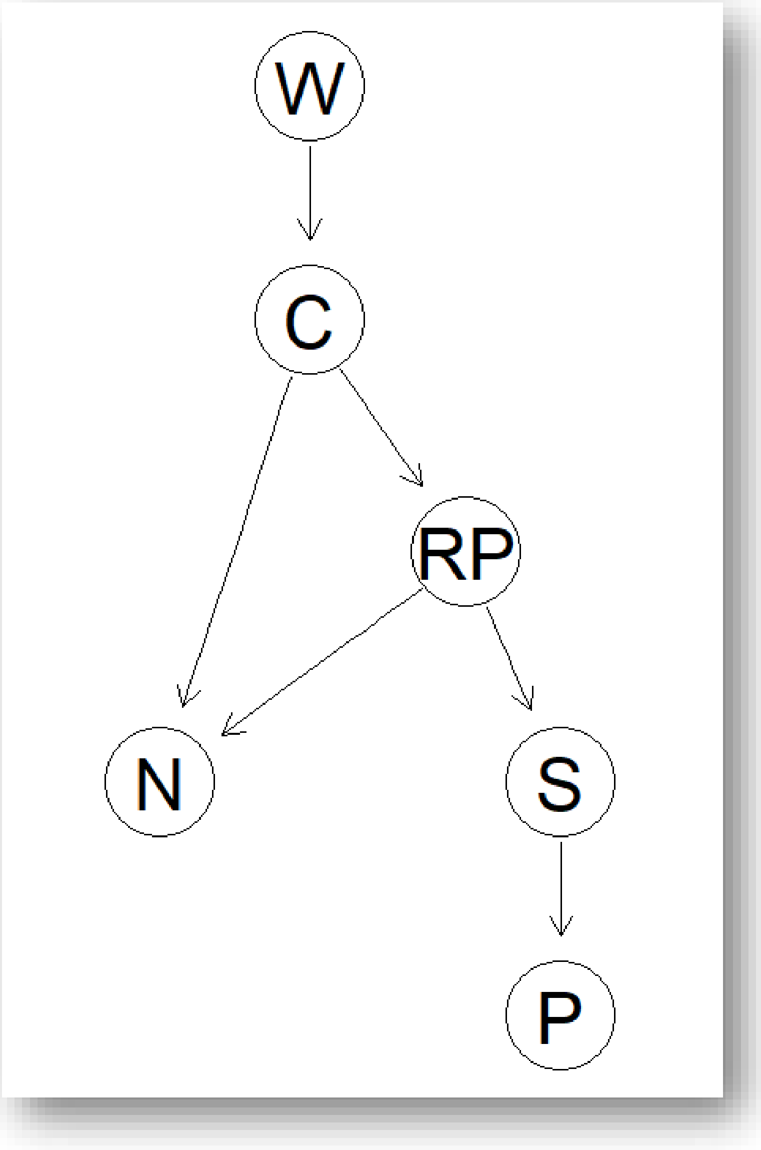
W kolejnym kroku utworzyłem ręcznie sieć Bayesa. Wybrane czynniki zostały dopasowane według moich własnych przemyśleń.

```
#tworzenie sieci
dag <- empty.graph(nodes=c('P','W','N','C','RP','S'))

dag <- set.arc(dag, from="W", to="C")
dag <- set.arc(dag, from="C", to="RP")
dag <- set.arc(dag, from="C", to="N")
dag <- set.arc(dag, from="RP", to="N")
dag <- set.arc(dag, from="RP", to="S")
dag <- set.arc(dag, from="S", to="P")

dag
score(dag,data=dane,type="bic")
graphviz.plot(dag)
```

Zwrócone wyniki:

```
  Random/Generated Bayesian network

  model:
   [W][C|W][RP|C][N|C:RP][S|RP][P|S]
  nodes:                                 6
  arcs:                                  6
    undirected arcs:                     0
    directed arcs:                       6
  average markov blanket size:           2.00
  average neighbourhood size:            2.00
  average branching factor:              1.00

  generation algorithm:                  Empty

> score(dag,data=dane,type="bic")
[1] -41338.28
```

Graf:



Następnym krokiem jest badanie niezależności zmiennych. Niezależności zostały zbadane za pomocą testu χ^2:

H_0: Zmienne są niezależne

H_1: Zmienne są zależne

Wykonane testy:

- Płeć i wiek

```
> ci.test('P','W',test='x2',data=dane)

        Pearson's X^2

data:  P ~ W
x2 = 245.76, df = 206, p-value = 0.03021
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Płeć i nadciśnienie

```
> ci.test('P','N',test='x2',data=dane)

        Pearson's X^2

data:  P ~ N
x2 = 2.4096, df = 2, p-value = 0.2998
alternative hypothesis: true value is greater than 0
```

Nie ma podstaw do odrzucenia, zmienne są niezależne

- Płeć i stan cywilny

```
> ci.test('P','C',test='x2',data=dane)

        Pearson's X^2

data:  P ~ C
x2 = 6.5587, df = 2, p-value = 0.03765
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Płeć i rodzaj pracy

```
> ci.test('P','RP',test='x2',data=dane)

        Pearson's X^2

data:  P ~ RP
x2 = 43.075, df = 8, p-value = 8.503e-07
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Płeć i osoba paląca

```
> ci.test('P','S',test='x2',data=dane)

        Pearson's X^2

data:  P ~ S
x2 = 57.338, df = 6, p-value = 1.56e-10
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Wiek i nadciśnienie

```
> ci.test('W','N',test='x2',data=dane)

        Pearson's X^2

data:  W ~ N
x2 = 496.44, df = 103, p-value < 2.2e-16
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Wiek i stan cywilny

```
> ci.test('W','C',test='x2',data=dane)

        Pearson's X^2

data:  W ~ C
x2 = 3129.4, df = 103, p-value < 2.2e-16
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Wiek i rodzaj pracy

```
> ci.test('W','RP',test='x2',data=dane)

	Pearson's X^2

data:  W ~ RP
x2 = 5715.2, df = 412, p-value < 2.2e-16
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Wiek i osoba paląca

```
> ci.test('W','S',test='x2',data=dane)

	Pearson's X^2

data:  W ~ S
x2 = 1850.2, df = 309, p-value < 2.2e-16
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Nadciśnienie i stan cywilny

```
> ci.test('N','C',test='x2',data=dane)

	Pearson's X^2

data:  N ~ C
x2 = 137.85, df = 1, p-value < 2.2e-16
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Nadciśnienie i rodzaj pracy

```
> ci.test('N','RP',test='x2',data=dane)

	Pearson's X^2

data:  N ~ RP
x2 = 135.2, df = 4, p-value < 2.2e-16
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Nadciśnienie i osoba paląca

```
> ci.test('N','S',test='x2',data=dane)

	Pearson's X^2

data:  N ~ S
x2 = 103.87, df = 3, p-value < 2.2e-16
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Stan cywilny i rodzaj pracy

```
> ci.test('C','RP',test='x2',data=dane)

	Pearson's X^2

data:  C ~ RP
x2 = 1644.1, df = 4, p-value < 2.2e-16
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Stan cywilny i osoba paląca

```
> ci.test('C','S',test='x2',data=dane)

	Pearson's X^2

data:  C ~ S
x2 = 599.05, df = 3, p-value < 2.2e-16
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

- Rodzaj pracy i osoba paląca

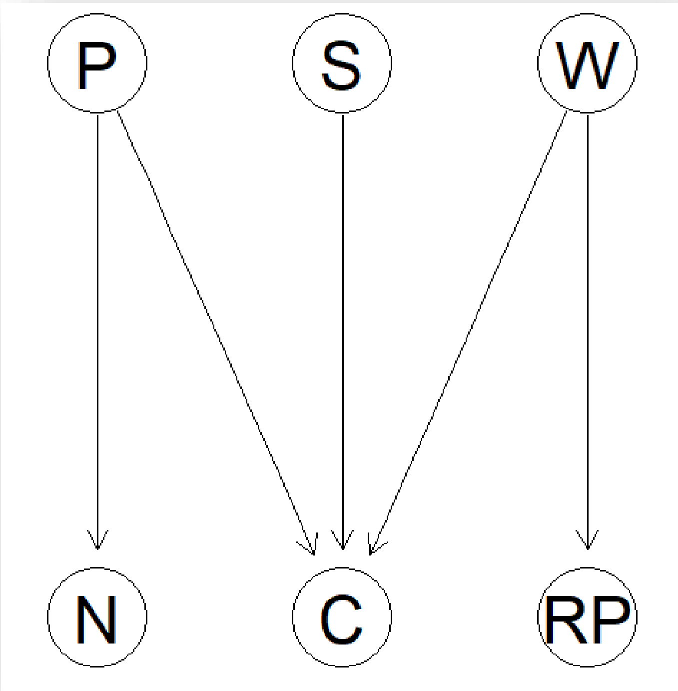
```
> ci.test('RP','S',test='x2',data=dane)

	Pearson's X^2

data:  RP ~ S
x2 = 1389.1, df = 12, p-value < 2.2e-16
alternative hypothesis: true value is greater than 0
```

p-value < 0.05, więc odrzucamy, mogą być zależne

3. Algorytmy

- GS

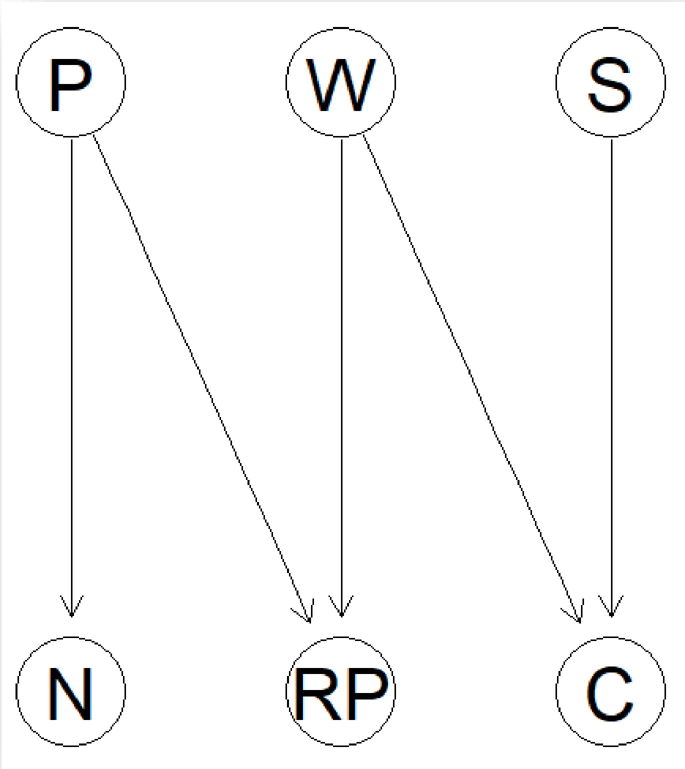
```
#GS
siec <- gs(dane,
           whitelist=matrix(c("P","N","S","C","P","C"),
                            ncol=2, byrow=TRUE))
siec <- set.arc(siec,"W","RP")
graphviz.plot(siec)
score(siec,data=dane,type="bic")
```



Score: -47120.47

- Pc.stable

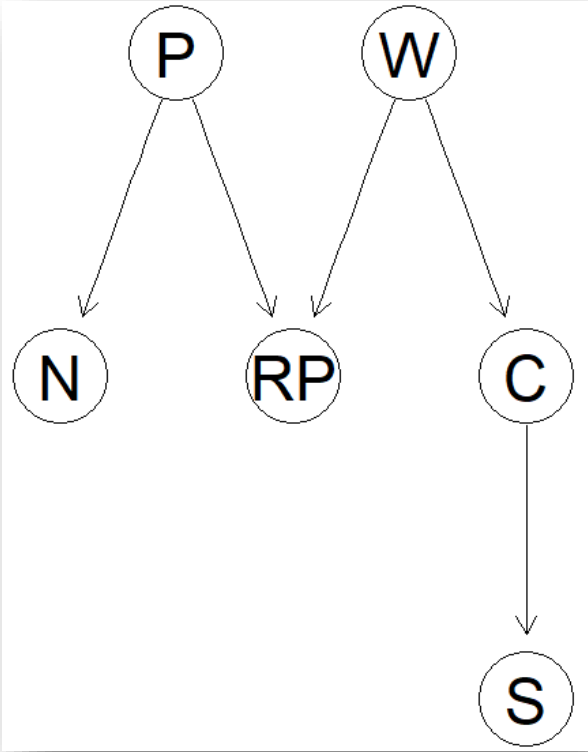
```
#pc.stable
siec <- pc.stable(dane,
                  whitelist=matrix(c("P","N","S","C","P","RP"),
                                   ncol=2, byrow=TRUE))

graphviz.plot(siec)
score(siec,data=dane,type="bic")
```



Score: -47159.4

- MMPC

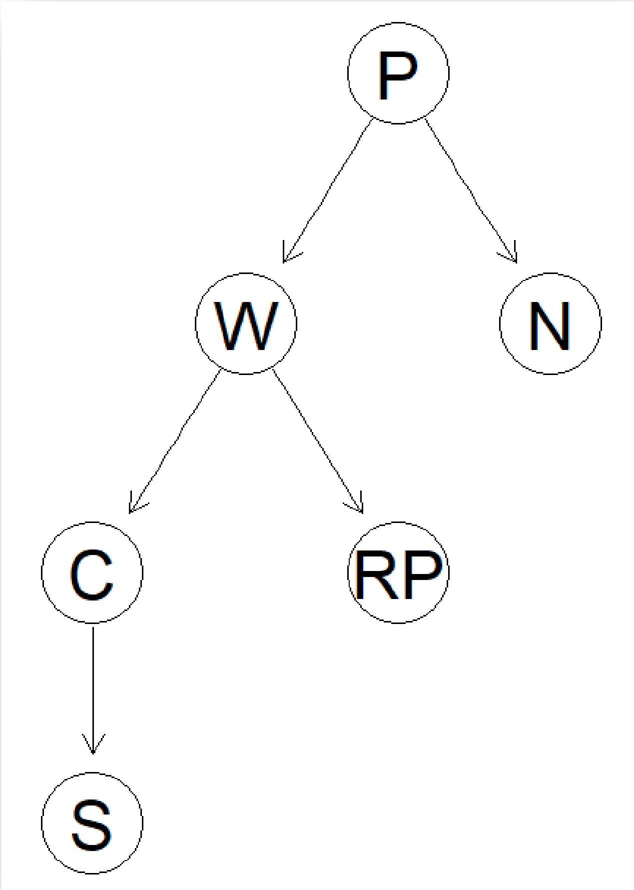
```
#MMPC
siec <- mmpc(dane,
                  whitelist=matrix(c("P","N","S","C","P","RP"),
                                   ncol=2, byrow=TRUE))
siec <- set.arc(siec, "P","N")
siec <- set.arc(siec, "P","RP")
siec <- set.arc(siec, "W","RP")
siec <- set.arc(siec, "W","C")
siec <- set.arc(siec, "C","S")
graphviz.plot(siec)
score(siec,data=dane,type="bic")
```



Score: -45637.45

- HPC

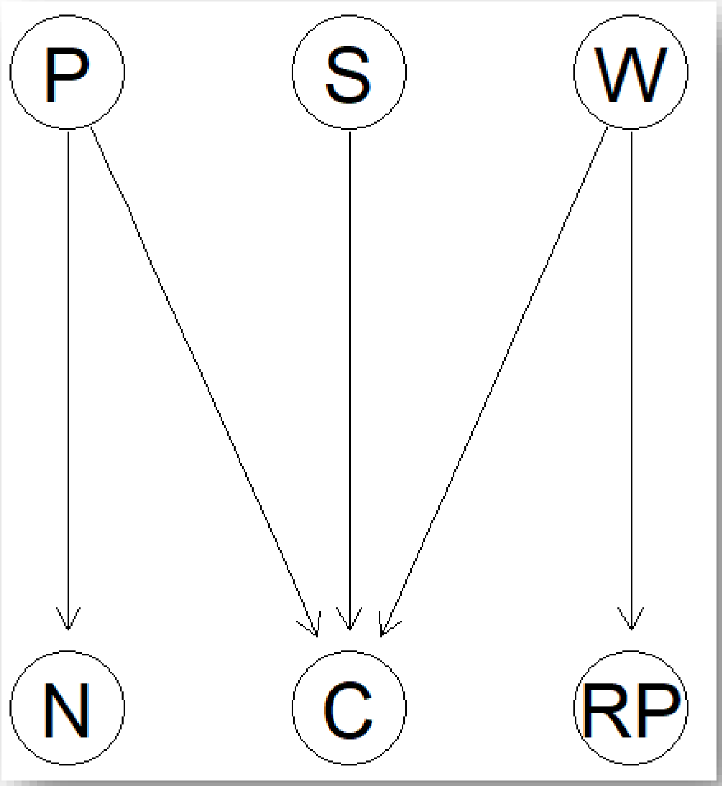
```r
#HPC
siec <- hpc(dane,
                  whitelist=matrix(c("P","N","S","C","P","W"),
                                   ncol=2, byrow=TRUE))
siec <- set.arc(siec, "P","N")
siec <- set.arc(siec, "P","W")
siec <- set.arc(siec, "W","RP")
siec <- set.arc(siec, "W","C")
siec <- set.arc(siec, "C","S")
graphviz.plot(siec)
score(siec,data=dane,type="bic")
```



Score: -42963.51

- Si.hiton

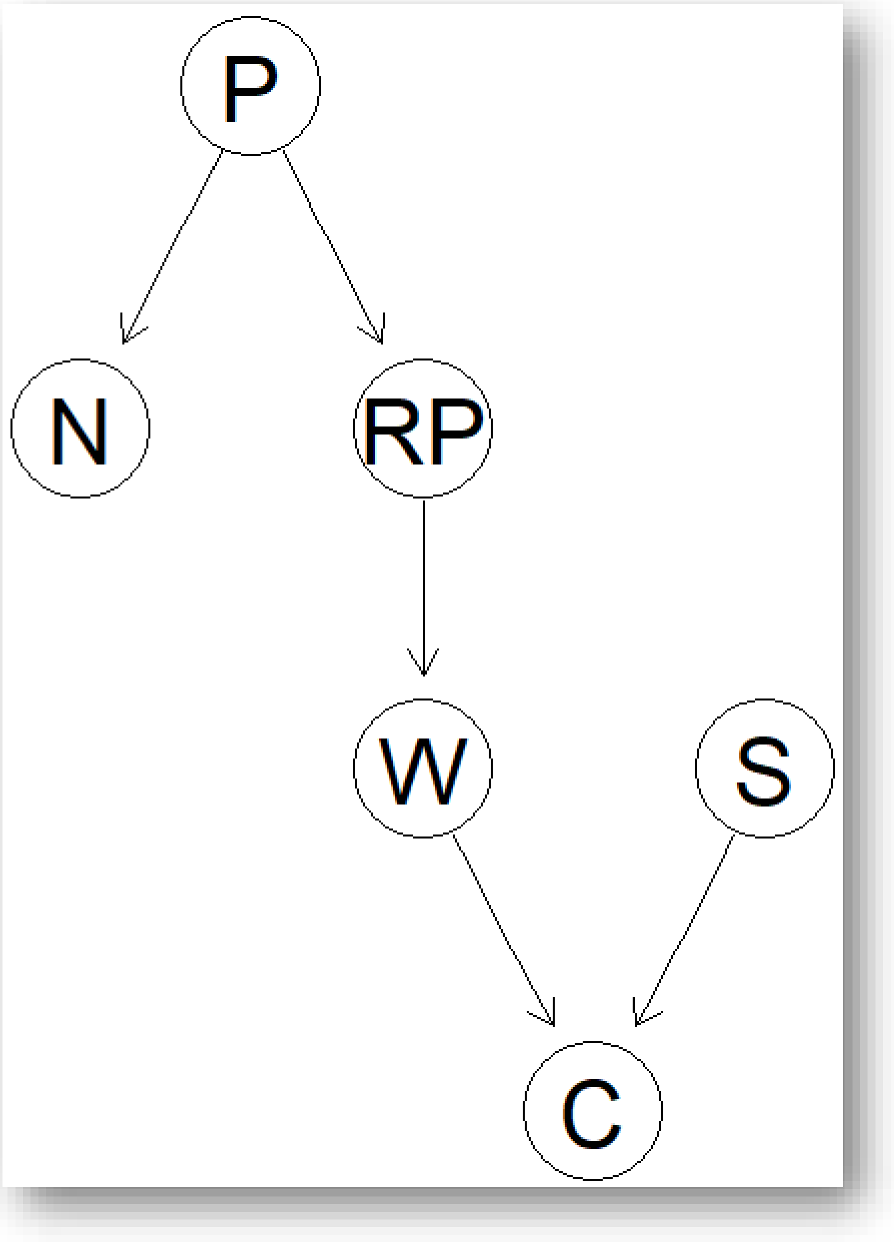
```
#si.hiton
siec <- pc.stable(dane,
                  whitelist=matrix(c("P","N","S","C","P","C"),
                                   ncol=2, byrow=TRUE))
siec <- set.arc(siec, "W","RP")
graphviz.plot(siec)
score(siec,data=dane,type="bic")
```



Score: -47120.47

- HC

```
#HC
siec <- gs(dane,
           whitelist=matrix(c("P","N","S","C","P","RP"),
                            ncol=2, byrow=TRUE))
graphviz.plot(siec)
score(siec,data=dane,type="bic")
```



Score: -43708.22

Tabela wyników dla przedstawionych algorytmów:

Algorytm	Wynik
GS	-47120.47
Pc.stable	-47159.4
MMPC	-45637.45
HPC	-42963.51
Si.hiton	-47120.47
HC	-43708.22
Ręcznie	-41338.28

Jak można zauważyć siecią z najlepszym wynikiem okazała się ta zrobiona ręcznie.

4. Prawdopodobieństwa dla najlepszego wyniku

```r
siec<- bn.fit(dag, dane)

siec
```

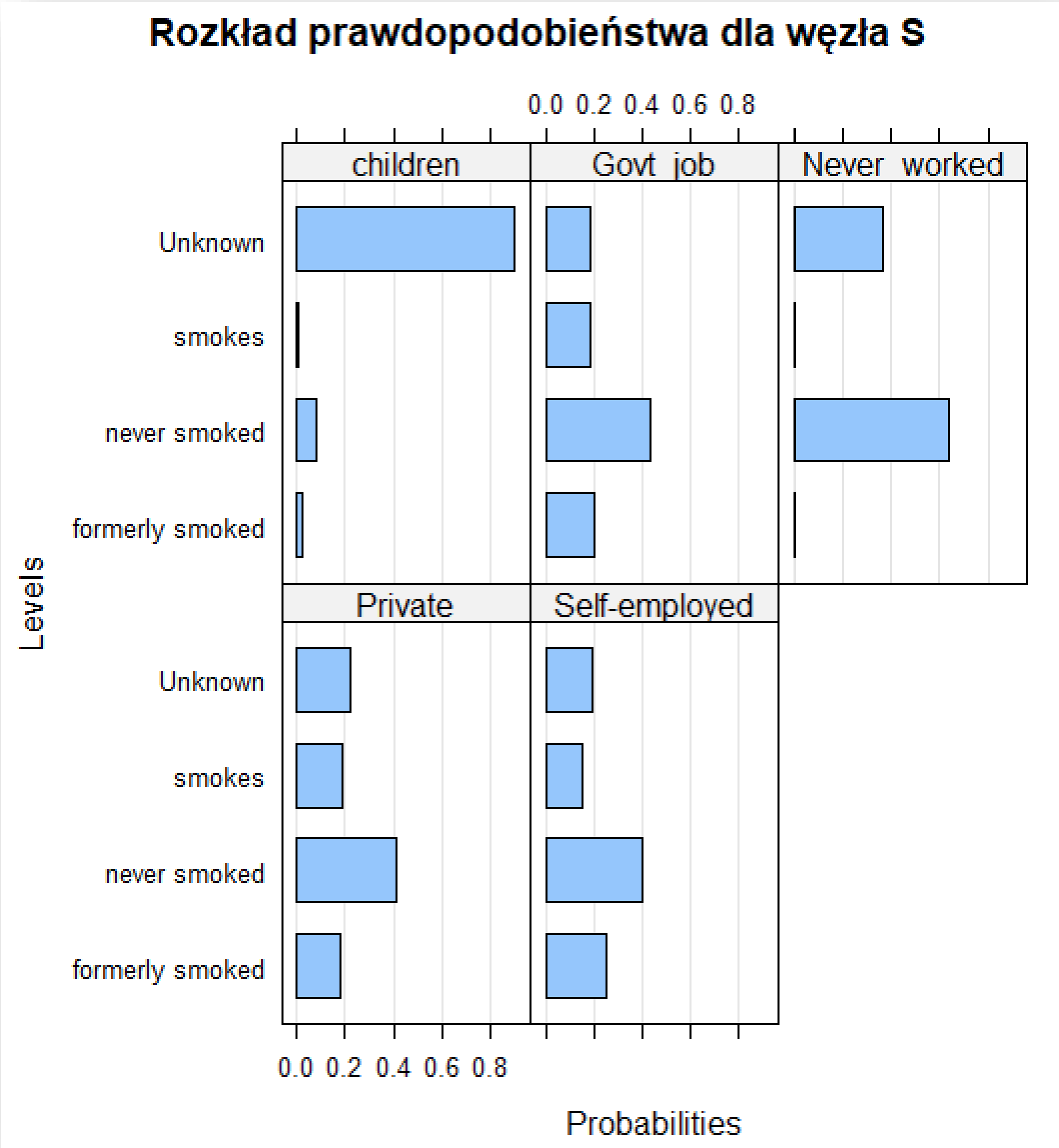
```
  Bayesian network parameters

  Parameters of node P (multinomial distribution)

Conditional probability table:

        S
P         formerly smoked never smoked      smokes     Unknown
  Female      0.538983051  0.649577167 0.572877060 0.541450777
  Male        0.459887006  0.350422833 0.427122940 0.458549223
  Other       0.001129944  0.000000000 0.000000000 0.000000000

  Parameters of node W (multinomial distribution)

Conditional probability table:
         0.08         0.16         0.24         0.32          0.4         0.48         0.56
0.0003913894 0.0005870841 0.0009784736 0.0009784736 0.0003913894 0.0005870841 0.0009784736
         0.64         0.72          0.8         0.88            1         1.08         1.16
0.0007827789 0.0009784736 0.0007827789 0.0009784736 0.0009784736 0.0015655577 0.0007827789
         1.24         1.32          1.4         1.48         1.56         1.64         1.72
0.0015655577 0.0015655577 0.0005870841 0.0011741683 0.0007827789 0.0015655577 0.0011741683
          1.8         1.88            2            3            4            5            6
0.0017612524 0.0015655577 0.0107632094 0.0090019569 0.0066536204 0.0127201566 0.0046966732
            7            8            9           10           11           12           13
0.0062622309 0.0113502935 0.0074363992 0.0068493151 0.0070450098 0.0088062622 0.0111545988
           14           15           16           17           18           19           20
0.0105675147 0.0088062622 0.0101761252 0.0117416830 0.0117416830 0.0097847358 0.0115459883
           21           22           23           24           25           26           27
0.0091976517 0.0088062622 0.0125244618 0.0107632094 0.0111545988 0.0121330724 0.0107632094
           28           29           30           31           32           33           34
0.0105675147 0.0099804305 0.0107632094 0.0154598826 0.0138943249 0.0111545988 0.0133072407
           35           36           37           38           39           40           41
0.0105675147 0.0101761252 0.0148727984 0.0140900196 0.0138943249 0.0144814090 0.0142857143
           42           43           44           45           46           47           48
0.0138943249 0.0136986301 0.0146771037 0.0166340509 0.0121330724 0.0146771037 0.0129158513
           49           50           51           52           53           54           55
0.0154598826 0.0162426614 0.0168297456 0.0176125245 0.0166340509 0.0170254403 0.0162426614
           56           57           58           59           60           61           62
0.0150684932 0.0185909980 0.0133072407 0.0156555773 0.0140900196 0.0148727984 0.0144814090
           63           64           65           66           67           68           69
0.0144814090 0.0103718200 0.0121330724 0.0117416830 0.0095890411 0.0091976517 0.0105675147
           70           71           72           73           74           75           76
0.0088062622 0.0119373777 0.0088062622 0.0090019569 0.0078277886 0.0103718200 0.0097847358
           77           78           79           80           81           82
0.0082191781 0.0199608611 0.0166340509 0.0136986301 0.0117416830 0.0109589041
```

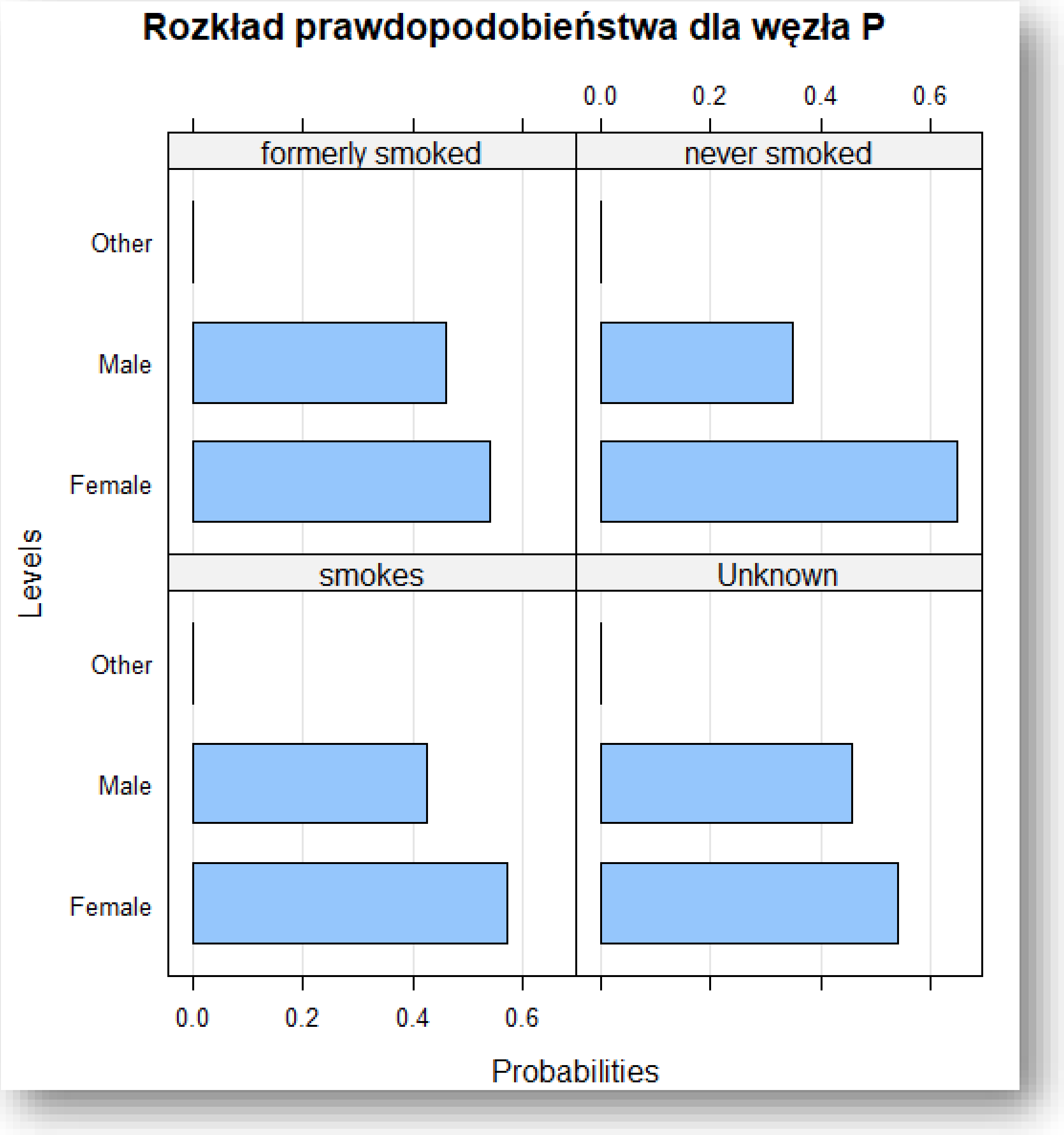
```
  Parameters of node N (multinomial distribution)

Conditional probability table:
 
, , RP = children

   C
N            No Yes
  0 1.00000000    
  1 0.00000000    

, , RP = Govt_job

   C
N            No        Yes
  0 0.92307692 0.88148148
  1 0.07692308 0.11851852

, , RP = Never_worked

   C
N            No Yes
  0 1.00000000    
  1 0.00000000    

, , RP = Private

   C
N            No        Yes
  0 0.95736906 0.88307985
  1 0.04263094 0.11692015

, , RP = Self-employed

   C
N            No        Yes
  0 0.91818182 0.80959097
  1 0.08181818 0.19040903
```

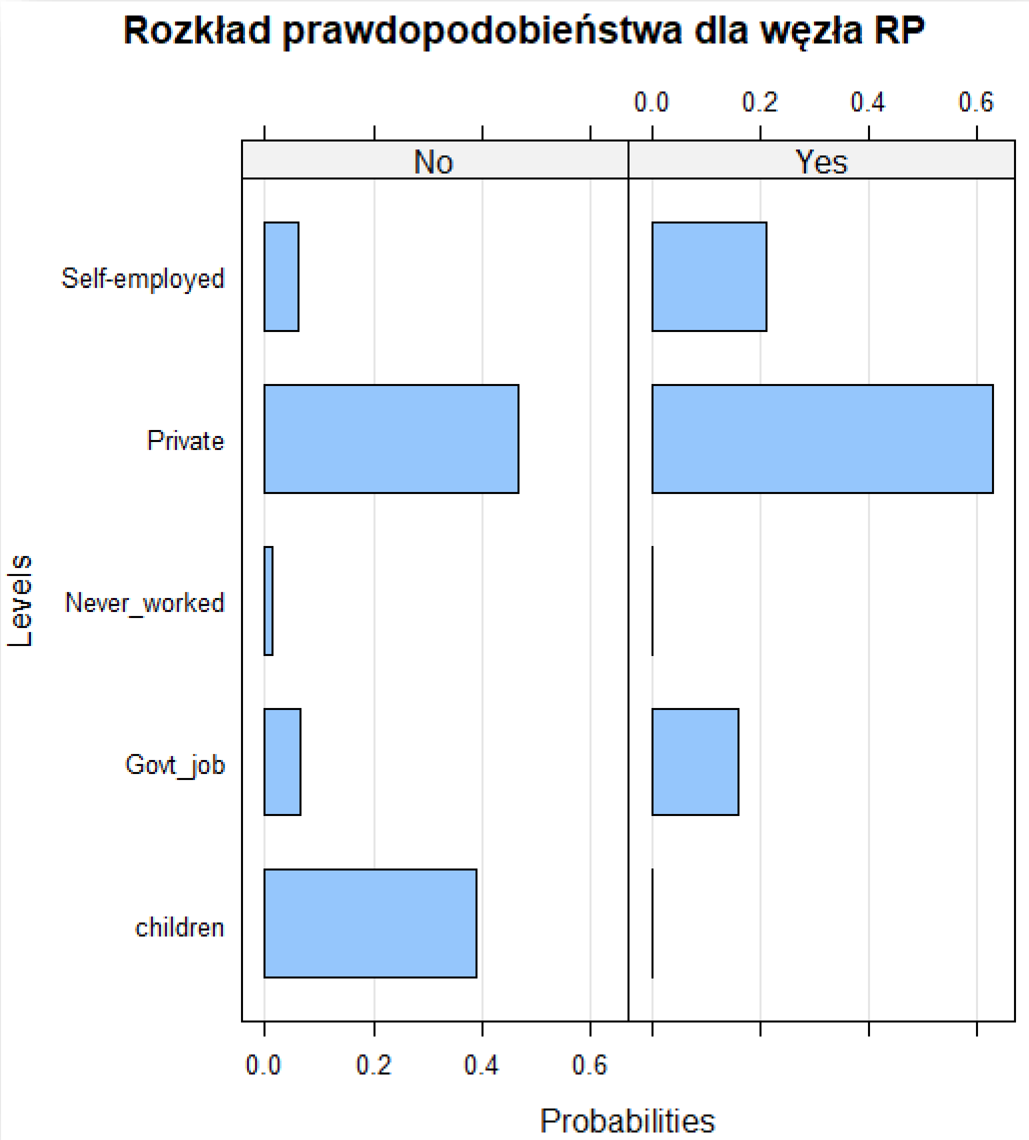
```
  Parameters of node C (multinomial distribution)

Conditional probability table:

     W
C           0.08       0.16       0.24       0.32        0.4       0.48       0.56       0.64
  No  1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000
  Yes 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
     W
C           0.72        0.8       0.88          1       1.08       1.16       1.24       1.32
  No  1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000
  Yes 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
     W
C            1.4       1.48       1.56       1.64       1.72        1.8       1.88          2
  No  1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000
  Yes 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
     W
C              3          4          5          6          7          8          9         10
  No  1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000
  Yes 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
     W
C             11         12         13         14         15         16         17         18
  No  1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 0.98333333
  Yes 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.01666667
     W
C             19         20         21         22         23         24         25         26
  No  0.96000000 0.93220339 0.87234043 0.84444444 0.89062500 0.65454545 0.57894737 0.61290323
  Yes 0.04000000 0.06779661 0.12765957 0.15555556 0.10937500 0.34545455 0.42105263 0.38709677
     W
C             27         28         29         30         31         32         33         34
  No  0.58181818 0.48148148 0.50980392 0.36363636 0.34177215 0.33802817 0.19298246 0.27941176
  Yes 0.41818182 0.51851852 0.49019608 0.63636364 0.65822785 0.66197183 0.80701754 0.72058824
     W
C             35         36         37         38         39         40         41         42
  No  0.16666667 0.23076923 0.18421053 0.12500000 0.12676056 0.21621622 0.16438356 0.14084507
  Yes 0.83333333 0.76923077 0.81578947 0.87500000 0.87323944 0.78378378 0.83561644 0.85915493
     W
C             43         44         45         46         47         48         49         50
  No  0.10000000 0.14666667 0.09411765 0.09677419 0.05333333 0.10606061 0.06329114 0.10843373
  Yes 0.90000000 0.85333333 0.90588235 0.90322581 0.94666667 0.89393939 0.93670886 0.89156627
     W
C             51         52         53         54         55         56         57         58
  No  0.10465116 0.07777778 0.04705882 0.04597701 0.08433735 0.07792208 0.08421053 0.11764706
  Yes 0.89534884 0.92222222 0.95294118 0.95402299 0.91566265 0.92207792 0.91578947 0.88235294
     W
C             59         60         61         62         63         64         65         66
  No  0.03750000 0.09722222 0.03947368 0.05405405 0.02702703 0.05660377 0.06451613 0.06666667
  Yes 0.96250000 0.90277778 0.96052632 0.94594595 0.97297297 0.94339623 0.93548387 0.93333333
     W
C             67         68         69         70         71         72         73         74
  No  0.00000000 0.08510638 0.16666667 0.02222222 0.04918033 0.02222222 0.10869565 0.05000000
  Yes 1.00000000 0.91489362 0.83333333 0.97777778 0.95081967 0.97777778 0.89130435 0.95000000
     W
```

```
C            75         76         77         78         79         80         81         82
  No  0.05660377 0.04000000 0.02380952 0.12745098 0.15294118 0.10000000 0.13333333 0.14285714
  Yes 0.94339623 0.96000000 0.97619048 0.87254902 0.84705882 0.90000000 0.86666667 0.85714286

  Parameters of node RP (multinomial distribution)

Conditional probability table:

               C
RP                      No        Yes
  children      0.39100740 0.00000000
  Govt_job      0.06659078 0.16104981
  Never_worked  0.01252134 0.00000000
  Private       0.46727376 0.62749776
  Self-employed 0.06260672 0.21145243

  Parameters of node S (multinomial distribution)

Conditional probability table:

                 RP
S                    children    Govt_job Never_worked     Private Self-employed
  formerly smoked 0.018922853 0.202435312  0.000000000 0.181538462   0.253968254
  never smoked    0.078602620 0.429223744  0.636363636 0.414017094   0.404151404
  smokes          0.002911208 0.182648402  0.000000000 0.185641026   0.151404151
  Unknown         0.899563319 0.185692542  0.363636364 0.218803419   0.190476190

>
```

5. Rozkłady prawdopodobieństw dla poszczególnych węzłów




Rozkład prawdopodobieństwa dla węzła P
formerly smoked
never smoked
smokes
Unknown
Other
Male
Female
Other
Male
Female
Levels
0.0
0.2
0.4
0.6
0.0
0.2
0.4
0.6
Probabilities



6. Prawdopodobieństwa warunkowe

- Prawdopodobieństwo udaru dla osoby mającej nadciśnienie będąc w przedziale wiekowym 75-84.

```
junction <-compile(as.grain(siec))

warunek <- setEvidence(junction,
                       nodes=c("N"),
                       states=c("1"))

querygrain(warunek, nodes = c("W"))$W
```

```
W
         0-18         19-24         25-34         35-44
0.0001265663 0.0005062650 0.0018352107 0.0072142767
        45-54         55-64         65-74         75-84
0.0213896975 0.0740412606 0.2776230857 0.3259081129
        85-94           95+
0.2684470320 0.0229084926
```

Prawdopodobieństwo obliczone za pomocą wzoru Bayesa:

$$P(N = 1|W = 75 - 84) = \frac{P(W = 75 - 84|N = 1) * P(N = 1)}{P(W = 75 - 84)}$$

- Prawdopodobieństwo udaru dla kobiet , które są samozatrudnione, palące i będące kiedykolwiek zamężne

```
querygrain(junction, nodes = c("P","RP", "C", "S"), type="joint")
```

```
, , RP = Self-employed, C = Yes

        S
P         formerly smoked never smoked     smokes    Unknown
  Female     1.899240e-02   0.03642505 0.012034403 0.01430952
  Male       1.620526e-02   0.01964997 0.008972553 0.01211859
  Other      3.981635e-05   0.00000000 0.000000000 0.00000000
```

Prawdopodobieństwo obliczone za pomocą wzoru Bayesa:

$$\begin{aligned} &P(S = Smokes \wedge \mathrm{RP} = \mathrm{Self-employed} \wedge P = Female \wedge \mathrm{C} = \mathrm{Yes}) \\ &\quad = \mathrm{P(P = Female) * P(C = Yes|P = Female)} \\ &\quad * \mathrm{P(C = Yes|P = Female) * P(RP = Self-employed|P = Female)} \\ &\quad * \mathrm{P(S = Smokes|RP = Self-employed)} \end{aligned}$$

- Prawdopodobieństwo udaru dla mężczyzn z przedziału wiekowego 45-54, którzy byli kiedykolwiek żonaci

```
warunek <- setEvidence(junction,
                       nodes=c("P","W"),
                       states=c("Male","45-54"))

querygrain(warunek, nodes = c("C"))$C
```

```
C
       No       Yes
0.3525138 0.6474862
```

- Prawdopodobieństwo udaru dla mężczyzn, którzy palą, pracują w sektorze prywatnym, nie mają nadciśnienia, byli/są żonaci i mają 71 lat

```
#p4
options(max.print = 22300)
querygrain(junction, nodes = c("W","C","N", "RP", "S","P"), type="joint")
```

```
, , RP = Private, N = 0, C = Yes, W = 71

        S
P         formerly smoked never smoked       smokes
  Female     6.154078e-04 0.0016914855 0.0006688899
  Male       5.250964e-04 0.0009124938 0.0004987077
  Other      1.290163e-06 0.0000000000 0.0000000000
```

$$
\begin{aligned}
P(P = Male \wedge \mathrm{S} = \mathrm{Smokes} \wedge \mathrm{RP} = \mathrm{Private} \wedge \mathrm{N} = 0 \wedge \mathrm{C} = \mathrm{Yes} \wedge \mathrm{W} = 71) \\
= \mathrm{P}(\mathrm{W} = 71) * \mathrm{P}(\mathrm{C} = \mathrm{Yes}|\mathrm{W} = 71) * \mathrm{P}(\mathrm{N} = 0|\mathrm{C} = \mathrm{Yes}) \\
* \mathrm{P}(\mathrm{RP} = \mathrm{Private}|\mathrm{C} = \mathrm{Yes}) * \mathrm{P}(\mathrm{S} = \mathrm{Smokes}|\mathrm{RP} = \mathrm{Private}) \\
* \quad \mathrm{P}(\mathrm{P} = \mathrm{Male}|\mathrm{S} = \mathrm{Smokes})
\end{aligned}
$$

7. Podsumowanie

Projekt miał na celu wykorzystanie sieci Bayesa do analizy danych dotyczących udaru mózgu. Poprzez konstrukcję struktury sieci, dopasowanie jej do danych i obliczanie prawdopodobieństw warunkowych, mogliśmy uzyskać wgląd w zależności między różnymi zmiennymi dotyczącymi udarów mózgu. Wykorzystując tą sieć mogliśmy np. obliczyć prawdopodobieństwo udaru dla osób, które mają nadciśnienie i są osobami starszymi.