# **Mô hình burden ~ res + insured**

## Statement of the problem

Bài toán sử dụng cơ sở dữ liệu “Health Care, Medical Insurance, and Economic Destitution: A Dataset of 1042 Stories”

Làm lại đánh giá áp lực kinh kế lên bệnh nhân như thế nào khi có hoặc không có nhà ở và bảo hiểm y tế

## Dữ liệu và đánh giá mô hình (dataset and estimations)

## Dữ liệu và xây dựng mô hình

|  |
| --- |
| data1<-read.csv("/Statistics/1042/1042data/1042data.csv", header = TRUE)  head(data1) |

Các biến sử dụng:

* Res: bệnh nhân có phải là cư dân ở khu vực khám chữa bệnh không
* Insured: bệnh nhân có chế độ bảo hiểm không
* Burden: áp lực tài chính lên bệnh nhân và gia đình

Xác suất có điều kiện giữa áp lực kinh tế burden và điều kiện nhà ở (res) và bảo hiểm (insured) thể hiện như sau

P ( burden | res , insured ) ∝ P ( res | burden , insured ) P ( burden | insured )

Nếu dựng mô hình quan hệ giữa nhà ở, bảo hiểm và áp lực kinh tế ta có thể thể hiện ở dạng lưới sau:

Insured

~ normal(μinsured, σinsured)

Burden

~ normal(μburden, σburden)

Res

~ normal(μres, σres)

Sử dụng packages bayesvl trên R mô tả mô hình quan hệ ở trên:

|  |
| --- |
| # Add nodes to model  model <- bayesvl()  model <- bvl\_addNode(model, “Burden”, “norm”)  model <- bvl\_addNode(model, “Res”, “cat”)  model <- bvl\_addNode(model, “Insured”, “cat”)  # Add arcs to model  model <- bvl\_addArc(model, “Res”, “Burden”, “slope”)  model <- bvl\_addArc(model, “Insured”, “Burden”, “slope”) |

## Đánh giá mô hình

* Đánh giá mô hình bằng bnlearn

Tính xác suất có điều kiện các biến trong mô hình:

|  |
| --- |
| bvl\_bnBayes(model, data1[c("Res","Burden","Insured")]) |

Kết quả:

|  |
| --- |
| Bayesian network parameters  Parameters of node Burden (multinomial distribution)  Conditional probability table:    , , Insured = No  Res  Burden No Yes  A 0.058351178 0.308659218  B 0.323875803 0.420391061  C 0.580835118 0.252793296  D 0.036937901 0.018156425  , , Insured = Yes  Res  Burden No Yes  A 0.147027601 0.746960486  B 0.244692144 0.216058764  C 0.588641189 0.035714286  D 0.019639066 0.001266464  Parameters of node Res (multinomial distribution)  Conditional probability table:  No Yes  0.4458175 0.5541825  Parameters of node Insured (multinomial distribution)  Conditional probability table:  No Yes  0.3070342 0.6929658 |

Hãy xem xét một bệnh nhân có nhà ở gần bệnh viện nhưng không có bảo hiểm, posterior distribution của áp lực tài chính của bệnh nhận này:

p(Burden|Res=”Yes”,Insured="No")=p(Res="Yes"|Burden,Insured="No")p(Burden|Insured="No")

Nếu vẽ các xác suất có điều kiện này:

|  |
| --- |
| bvl\_bnBarchart(model, data1[c("Res","Burden","Insured")]) |



Nếu chỉ đánh giá sơ bộ trên dữ liệu như trên, có thể dễ nhận thấy trên dữ liệu thống kê đối với người có nhà và bảo hiểm, áp lực tài chính khám chữa bệnh thấp (burden tập trung cao mở mức A).

Đánh giá sức mạnh kết nối của mô hình quan hệ:

|  |
| --- |
| > bvl\_bnStrength(model, data1[c("Res","Burden","Insured")])  *from to strength*  *1 Res Burden 1.682735e-81*  *2 Insured Burden 2.303670e-20* |

Chức năng này gọi package bnlearn để đo lường quan hệ giữa 2 biến trong lưới model thông qua xác suất tương ứng từng cung (arc). Giá trị strength được thể hiện chính là giá trị p-value.

Mô hình hiện tại có 2 arc, các giá trị p-values đều nhỏ hơn 0,05 và support rất tốt từ dữ liệu.

## Mô hình hồi quy bayesian

Giả sử chúng ta chọn 1 mô hình hồi quy tuyến tính đơn giản của biến y với 1 biến covariate x, phương trình toán học sẽ có dạng:

yi = α + βxi + ϵi

Trong đó ϵi ∼ N(0, σ).

Viết cách khác ta có thể có phân phối của y:

yi∼N(α+βxi, σ)

Sử dụng tập dữ liệu khảo sát i ∈ I

Với mô hình quan hệ đã tạo ở trên, nếu viết ở dạng hồi quy tuyến tính, ta có phương trình:

yburden[i]= αburden + βresxres[i] + βinsuredxinsured[i] + ϵburden[i]

Trên phương trình ta thấy có 2 hệ số βres, và βinsured là 2 hệ số hồi quy “slope” là hệ số ảnh hưởng đến hệ số góc của đường hồi quy. Khi tạo mô hình quan hệ các biến ở trên ta đã định nghĩa arc với loại quan hệ “slope”:

model <- bvl\_addArc(model, “Res”, “Burden”, “***slope***”)

model <- bvl\_addArc(model, “Insured”, “Burden”, “***slope***”)

Vì vậy, để tạo mô hình hồi quy trên bayesvl chỉ cần thực hiện lệnh

|  |
| --- |
| # Generate the stan code for model  model\_string <- bvl\_model2Stan(model)  cat(model\_string) |

model\_string chứa code stan cho mô hình hồi quy tuyến tính.

Code stan được tạo:

|  |
| --- |
| data{  // Define variables in data  int<lower=1> Nobs; // Number of observations (an integer)  real Burden[Nobs]; // outcome variable  real Res[Nobs];  real Insured[Nobs];  }  parameters{  // Define parameters to estimate  real<lower=0> sigma\_Burden;  real a\_Burden;  real b\_Res\_Burden;  real b\_Insured\_Burden;    }  transformed parameters{  // Transform parameters  real mu\_Burden[Nobs];  for (i in 1:Nobs) {  mu\_Burden[i] = a\_Burden + b\_Res\_Burden \* Res[i] + b\_Insured\_Burden \* Insured[i];  }  }  model{  // Priors  a\_Burden ~ normal(0,100);  b\_Res\_Burden ~ normal(0,100);  b\_Insured\_Burden ~ normal(0,100);  // Likelihoods  Burden ~ normal(mu\_Burden, sigma\_Burden);  } |

Chạy mô phỏng mô hình:

|  |
| --- |
| dat1042 <- with(data1,  list(Nobs = length(Res),  Res = as.numeric(Res),  Insured = as.numeric(Insured),  Burden = as.numeric(Burden)))  # Fit the model  fit <- bvl\_modelFit(model, dat1042, warmup = 2000, iter = 5000, chains = 4, cores = 4) |