

Multivariate analysis: multiple logistic regression

Núria Jolis Orriols

2 de juny, 2022

Contents

Step 1 - Obtaining datasets	1
Step 2 - Generate multiple logistic regression models	1
Data A	1
Data B	3
Data C	7
Data D	8
Data E	12

Step 1 - Obtaining datasets

```
dataA <- read.csv("data/datasetA.csv",stringsAsFactors=T)
dataB <- read.csv("data/datasetB.csv",stringsAsFactors=T)
dataC <- read.csv("data/datasetC.csv",stringsAsFactors=T)
dataD <- read.csv("data/datasetD.csv",stringsAsFactors=T)
dataE <- read.csv("data/datasetE.csv",stringsAsFactors=T)
```

Step 2 - Generate multiple logistic regression models

Data A

Full logistic regression model

```
xnames <- names(dataA[2:48])
foriA<- as.formula(paste("outcome ~ ", paste(xnames, collapse= "+")))

full_mlrA<- glm(forA, data = dataA, family = 'binomial')
summary(full_mlrA)
```

Call:

```
glm(formula = foriA, family = "binomial", data = dataA)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.2173	0.0413	0.1478	0.3498	1.8370

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	6.976e+01	7.393e+01	0.944	0.34536
age	-5.110e-02	2.427e-02	-2.105	0.03527 *
genderM	3.963e-01	4.920e-01	0.806	0.42050
BMI	4.046e-02	3.033e-02	1.334	0.18220
hypertensiveYes	-2.283e-02	5.132e-01	-0.044	0.96452
atrialfibrillationYes	7.514e-01	5.296e-01	1.419	0.15593
CHD.with.no.MIYes	1.569e-01	7.662e-01	0.205	0.83771
diabetesYes	1.080e-02	5.576e-01	0.019	0.98454
deficiencyanemiasYes	1.680e+00	5.926e-01	2.836	0.00457 **
depressionYes	4.865e-01	8.208e-01	0.593	0.55337
hyperlipemiaYes	-6.647e-01	5.020e-01	-1.324	0.18548
renal.failureYes	2.971e+00	6.730e-01	4.415	1.01e-05 ***
COPDYes	1.101e+00	8.485e-01	1.297	0.19447
heart.rate	-2.416e-02	1.560e-02	-1.549	0.12134
respiratory.rate	-6.038e-02	5.745e-02	-1.051	0.29327
temperature	7.975e-01	3.783e-01	2.108	0.03500 *
SP.O2	1.491e-02	1.070e-01	0.139	0.88922
urine.output	-1.151e-04	2.042e-04	-0.564	0.57277
hematocrit	4.277e-01	3.384e-01	1.264	0.20635
RBC	-3.311e+00	2.969e+00	-1.115	0.26470
MCH	1.413e+00	1.949e+00	0.725	0.46848
MCHC	-9.962e-01	1.739e+00	-0.573	0.56670
MCV	-5.809e-01	6.714e-01	-0.865	0.38691
RDW	1.153e-01	1.279e-01	0.902	0.36718
leucocyte	-6.903e-02	5.378e-02	-1.284	0.19931
platelets	8.266e-03	2.645e-03	3.125	0.00178 **
neutrophils	9.937e-02	5.960e-02	1.667	0.09549 .
basophils	-3.046e-03	5.659e-03	-0.538	0.59037
lymphocyte	1.586e-01	8.245e-02	1.923	0.05443 .
PT	-9.875e-02	3.058e-01	-0.323	0.74676
INR	4.182e-01	2.651e+00	0.158	0.87468
NT.proBNP	-4.931e-06	1.314e-05	-0.375	0.70745
creatine.kinase	-1.834e-04	2.118e-04	-0.866	0.38655
creatinine	4.241e-03	1.161e-02	0.365	0.71480
urea.nitrogen	-1.923e-02	1.407e-02	-1.367	0.17166
glucose	-3.763e-03	4.505e-03	-0.835	0.40359
blood.potassium	5.540e-01	1.116e+00	0.497	0.61947
blood.sodium	8.729e-01	8.957e-01	0.974	0.32981
blood.calcium	7.724e-01	4.551e-01	1.697	0.08970 .
chloride	-9.535e-01	8.972e-01	-1.063	0.28788
anion.gap	-1.222e+00	9.125e-01	-1.339	0.18051
magnesium.ion	-1.220e+00	8.498e-01	-1.435	0.15117
pH	-7.077e+00	5.903e+00	-1.199	0.23059
bicarbonate	-7.314e-01	9.090e-01	-0.805	0.42101
lactic.acid	8.888e-03	4.274e-02	0.208	0.83526

PCO2	-1.250e-01	4.996e-02	-2.503	0.01231 *
EF	1.340e-02	1.779e-02	0.753	0.45136
m.a.p	2.114e-02	2.742e-02	0.771	0.44064

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 364.60 on 427 degrees of freedom
 Residual deviance: 182.02 on 380 degrees of freedom
 (748 observations deleted due to missingness)
 AIC: 278.02

Number of Fisher Scoring iterations: 9

The full model with dataset A found the variables age, presence of deficiency anemia, presence of renal failure, temperature, platelets and PCO2 to be statistically significant.

Perform stepwise variable selection

The stepAIC function of the MASS package will be used to select the most contributive variables:

```
step_mlrA <- full_mlrA %>% stepAIC(trace = FALSE)
```

Error in stepAIC(., trace = FALSE): number of rows in use has changed: remove missing values?

The model with the stepAIC function does not converge because of the missing values.

Data B

Full logistic regression model

```
xnamesB <- names(dataB[2:48])
foriB <- as.formula(paste("outcome ~ ", paste(xnamesB, collapse = "+")))
```

```
full_mlrB <- glm(for_iB, data = dataB, family = 'binomial')
summary(full_mlrB)
```

Call:

```
glm(formula = for_iB, family = "binomial", data = dataB)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.2173	0.0413	0.1478	0.3498	1.8370

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	6.976e+01	7.393e+01	0.944	0.34536
age	-5.110e-02	2.427e-02	-2.105	0.03527 *

genderM	3.963e-01	4.920e-01	0.806	0.42050	
BMI	4.046e-02	3.033e-02	1.334	0.18220	
hypertensiveYes	-2.283e-02	5.132e-01	-0.044	0.96452	
atrialfibrillationYes	7.514e-01	5.296e-01	1.419	0.15593	
CHD.with.no.MIYes	1.569e-01	7.662e-01	0.205	0.83771	
diabetesYes	1.080e-02	5.576e-01	0.019	0.98454	
deficiencyanemiasYes	1.680e+00	5.926e-01	2.836	0.00457	**
depressionYes	4.865e-01	8.208e-01	0.593	0.55337	
hyperlipemiaYes	-6.647e-01	5.020e-01	-1.324	0.18548	
renal.failureYes	2.971e+00	6.730e-01	4.415	1.01e-05	***
COPDYes	1.101e+00	8.485e-01	1.297	0.19447	
heart.rate	-2.416e-02	1.560e-02	-1.549	0.12134	
respiratory.rate	-6.038e-02	5.745e-02	-1.051	0.29327	
temperature	7.975e-01	3.783e-01	2.108	0.03500	*
SP.O2	1.491e-02	1.070e-01	0.139	0.88922	
urine.output	-1.151e-04	2.042e-04	-0.564	0.57277	
hematocrit	4.277e-01	3.384e-01	1.264	0.20635	
RBC	-3.311e+00	2.969e+00	-1.115	0.26470	
MCH	1.413e+00	1.949e+00	0.725	0.46848	
MCHC	-9.962e-01	1.739e+00	-0.573	0.56670	
MCV	-5.809e-01	6.714e-01	-0.865	0.38691	
RDW	1.153e-01	1.279e-01	0.902	0.36718	
leucocyte	-6.903e-02	5.378e-02	-1.284	0.19931	
platelets	8.266e-03	2.645e-03	3.125	0.00178	**
neutrophils	9.937e-02	5.960e-02	1.667	0.09549	.
basophils	-3.046e-03	5.659e-03	-0.538	0.59037	
lymphocyte	1.586e-01	8.245e-02	1.923	0.05443	.
PT	-9.875e-02	3.058e-01	-0.323	0.74676	
INR	4.182e-01	2.651e+00	0.158	0.87468	
NT.proBNP	-4.931e-06	1.314e-05	-0.375	0.70745	
creatinine.kinase	-1.834e-04	2.118e-04	-0.866	0.38655	
creatinine	4.241e-03	1.161e-02	0.365	0.71480	
urea.nitrogen	-1.923e-02	1.407e-02	-1.367	0.17166	
glucose	-3.763e-03	4.505e-03	-0.835	0.40359	
blood.potassium	5.540e-01	1.116e+00	0.497	0.61947	
blood.sodium	8.729e-01	8.957e-01	0.974	0.32981	
blood.calcium	7.724e-01	4.551e-01	1.697	0.08970	.
chloride	-9.535e-01	8.972e-01	-1.063	0.28788	
anion.gap	-1.222e+00	9.125e-01	-1.339	0.18051	
magnesium.ion	-1.220e+00	8.498e-01	-1.435	0.15117	
pH	-7.077e+00	5.903e+00	-1.199	0.23059	
bicarbonate	-7.314e-01	9.090e-01	-0.805	0.42101	
lactic.acid	8.888e-03	4.274e-02	0.208	0.83526	
PCO2	-1.250e-01	4.996e-02	-2.503	0.01231	*
EF	1.340e-02	1.779e-02	0.753	0.45136	
m.a.p	2.114e-02	2.742e-02	0.771	0.44064	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 364.60 on 427 degrees of freedom
Residual deviance: 182.02 on 380 degrees of freedom
AIC: 278.02

Number of Fisher Scoring iterations: 9

The full model found the variables age, deficiencyanemia, renal.failure, temperature, platelets and PCO2 to be statistically significant.

Perform stepwise variable selection

The stepAIC function of the MASS package will be used to select the most contributive variables:

```
step_mlrB <- full_mlrB %>% stepAIC(trace = FALSE)
summary(step_mlrB)
```

Call:

```
glm(formula = outcome ~ age + BMI + atrialfibrillation + deficiencyanemias +
    renal.failure + COPD + heart.rate + temperature + leucocyte +
    platelets + neutrophils + lymphocyte + PT + urea.nitrogen +
    glucose + blood.calcium + chloride + anion.gap + magnesium.ion +
    PCO2, family = "binomial", data = dataB)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.84598	0.07039	0.18636	0.41001	2.10448

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.830771	15.437292	-0.378	0.70565
age	-0.052996	0.019869	-2.667	0.00765 **
BMI	0.039166	0.026619	1.471	0.14120
atrialfibrillationYes	0.699159	0.461376	1.515	0.12968
deficiencyanemiasYes	1.404918	0.517877	2.713	0.00667 **
renal.failureYes	2.419292	0.545191	4.438	9.1e-06 ***
COPDYes	1.385250	0.761775	1.818	0.06900 .
heart.rate	-0.024428	0.012730	-1.919	0.05500 .
temperature	0.483137	0.313881	1.539	0.12375
leucocyte	-0.071654	0.045835	-1.563	0.11798
platelets	0.006761	0.002251	3.004	0.00267 **
neutrophils	0.069262	0.048496	1.428	0.15323
lymphocyte	0.122872	0.067966	1.808	0.07063 .
PT	-0.066624	0.028587	-2.331	0.01978 *
urea.nitrogen	-0.019124	0.011450	-1.670	0.09486 .
glucose	-0.005408	0.003480	-1.554	0.12019
blood.calcium	0.840850	0.366020	2.297	0.02160 *
chloride	-0.077651	0.043106	-1.801	0.07164 .
anion.gap	-0.346357	0.106971	-3.238	0.00120 **
magnesium.ion	-1.204767	0.758323	-1.589	0.11212
PCO2	-0.063569	0.021980	-2.892	0.00383 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 364.6 on 427 degrees of freedom

Residual deviance: 197.6 on 407 degrees of freedom
AIC: 239.6

Number of Fisher Scoring iterations: 7

The function chose a final model with the following formula:

```
outcome ~ age + BMI + atrialfibrillation + deficiencyanemias + renal.failure + COPD + heart.rate +  
temperature + leucocyte + platelets + neutrophils + lymphocyte + PT + urea.nitrogen + glucose +  
blood.calcium + chloride + anion.gap + magnesium.ion + PCO2
```

It found the variables age, deficiencyanemia, renal.failure, temperature, SP.O2, platelets, leucocyte, neutrophils, lymphocyte, blood.calcium, blood.potassium, bicarbonate, pH and PCO2 to be statistically significant.

```
formula(step_mlrB)
```

```
outcome ~ age + BMI + atrialfibrillation + deficiencyanemias +  
renal.failure + COPD + heart.rate + temperature + leucocyte +  
platelets + neutrophils + lymphocyte + PT + urea.nitrogen +  
glucose + blood.calcium + chloride + anion.gap + magnesium.ion +  
PCO2
```

```
results_mlrB<-cbind( "OR" = exp(coef(step_mlrB)), "p-value" = coef(summary(step_mlrB))[,4])  
results_mlrB
```

	OR	p-value
(Intercept)	0.002935811	7.056484e-01
age	0.948383792	7.646358e-03
BMI	1.039942648	1.412010e-01
atrialfibrillationYes	2.012059741	1.296764e-01
deficiencyanemiasYes	4.075191324	6.670910e-03
renal.failureYes	11.237896860	9.100616e-06
COPDYes	3.995823639	6.899541e-02
heart.rate	0.975868046	5.499569e-02
temperature	1.621152215	1.237467e-01
leucocyte	0.930852705	1.179795e-01
platelets	1.006784116	2.668055e-03
neutrophils	1.071717056	1.532311e-01
lymphocyte	1.130740226	7.062882e-02
PT	0.935546889	1.977744e-02
urea.nitrogen	0.981057342	9.486163e-02
glucose	0.994606599	1.201931e-01
blood.calcium	2.318336049	2.160283e-02
chloride	0.925286966	7.164070e-02
anion.gap	0.707260284	1.204327e-03
magnesium.ion	0.299761956	1.121224e-01
PCO2	0.938409785	3.826361e-03

Data C

Full logistic regression model

```
xnamesC <- names(dataC[2:40])
foriC<- as.formula(paste("outcome ~ ", paste(xnamesC, collapse= "+")))

full_mlrC<- glm(for iC, data = dataC, family = 'binomial')
summary(full_mlrC)
```

Call:

```
glm(formula = for iC, family = "binomial", data = dataC)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.9320	0.1563	0.2968	0.4817	2.0673

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.330e+01	3.096e+01	-0.753	0.451731
age	-1.633e-02	9.875e-03	-1.654	0.098114 .
genderM	-3.861e-02	2.269e-01	-0.170	0.864853
hypertensiveYes	1.047e-01	2.442e-01	0.429	0.668171
atrialfibrillationYes	-7.878e-02	2.399e-01	-0.328	0.742665
CHD.with.no.MIYes	-1.738e-01	3.719e-01	-0.467	0.640191
diabetesYes	7.934e-02	2.546e-01	0.312	0.755290
deficiencyanemiasYes	6.747e-01	2.583e-01	2.612	0.009011 **
depressionYes	5.987e-01	4.116e-01	1.455	0.145795
hyperlipemiaYes	2.507e-01	2.405e-01	1.043	0.297064
renal.failureYes	1.203e+00	2.746e-01	4.382	1.18e-05 ***
COPDYes	9.717e-01	4.794e-01	2.027	0.042666 *
heart.rate	-2.207e-02	7.491e-03	-2.946	0.003221 **
respiratory.rate	-3.659e-02	2.804e-02	-1.305	0.191959
temperature	2.832e-01	1.792e-01	1.580	0.114048
SP.O2	8.808e-02	5.048e-02	1.745	0.080988 .
urine.output	2.109e-04	1.104e-04	1.911	0.056053 .
hematocrit	1.118e-01	1.627e-01	0.687	0.491947
RBC	-8.877e-01	1.415e+00	-0.627	0.530357
MCH	-4.041e-01	9.497e-01	-0.426	0.670432
MCHC	4.591e-01	8.439e-01	0.544	0.586380
MCV	6.574e-02	3.286e-01	0.200	0.841403
RDW	-6.777e-02	5.823e-02	-1.164	0.244538
leucocyte	-7.032e-02	1.961e-02	-3.587	0.000335 ***
platelets	3.047e-03	1.042e-03	2.925	0.003444 **
PT	-1.939e-02	1.390e-02	-1.394	0.163192
INR	5.777e-03	2.317e-02	0.249	0.803081
NT.proBNP	2.207e-06	8.452e-06	0.261	0.794029
creatinine	1.894e-03	2.028e-03	0.934	0.350255
urea.nitrogen	-1.694e-02	6.800e-03	-2.491	0.012726 *
glucose	-4.851e-04	2.197e-03	-0.221	0.825217
blood.potassium	-5.999e-01	4.093e-01	-1.466	0.142726
blood.sodium	-1.343e-01	2.907e-01	-0.462	0.644084

```

blood.calcium      6.646e-01  2.212e-01  3.005 0.002654 **
chloride           1.275e-01  2.952e-01  0.432 0.665692
anion.gap          2.353e-02  3.002e-01  0.078 0.937537
magnesium.ion      -5.450e-01  4.453e-01 -1.224 0.221065
bicarbonate         1.527e-01  2.975e-01  0.513 0.607679
EF                 -3.906e-03  8.624e-03 -0.453 0.650604
m.a.p              1.813e-02  1.180e-02  1.536 0.124456

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 892.42 on 1104 degrees of freedom
Residual deviance: 637.24 on 1065 degrees of freedom
(71 observations deleted due to missingness)
AIC: 717.24

```

Number of Fisher Scoring iterations: 8

The full model found the variables age, deficiencyanemia, renal.failure, platelets, leucocyte, blood.potassium, blood.calcium and magnesium.ion to be statistically significant.

Perform stepwise variable selection

The stepAIC function of the MASS package will be used to select the most contributive variables:

```
step_mlrC <- full_mlrC %>% stepAIC(trace = FALSE)
```

Error in stepAIC(., trace = FALSE): number of rows in use has changed: remove missing values?

The model with the stepAIC function does not converge because of the missing values.

Data D

Full logistic regression model

```

xnamesD <- names(dataD[2:48])
foriD<- as.formula(paste("outcome ~ ", paste(xnamesD, collapse= "+")))

```

```

full_mlrD<- glm(forID, data = dataD, family = 'binomial')
summary(full_mlrD)

```

Call:

```
glm(formula = foriD, family = "binomial", data = dataD)
```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-3.0882   0.1426   0.2727   0.4552   2.0554

```


Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.141e+01	3.577e+01	-0.319	0.74967
age	-1.440e-02	1.057e-02	-1.363	0.17301
genderM	-9.468e-02	2.318e-01	-0.408	0.68296
BMI	2.395e-02	1.701e-02	1.408	0.15918
hypertensiveYes	3.078e-02	2.426e-01	0.127	0.89906
atrialfibrillationYes	-1.744e-01	2.421e-01	-0.720	0.47134
CHD.with.no.MIYes	4.345e-02	3.844e-01	0.113	0.90999
diabetesYes	4.878e-02	2.518e-01	0.194	0.84640
deficiencyanemiasYes	6.336e-01	2.605e-01	2.432	0.01500 *
depressionYes	2.866e-01	3.868e-01	0.741	0.45878
hyperlipemiaYes	1.791e-01	2.419e-01	0.740	0.45922
renal.failureYes	1.303e+00	2.799e-01	4.654	3.25e-06 ***
COPDYes	1.239e+00	4.865e-01	2.546	0.01089 *
heart.rate	-2.070e-02	7.386e-03	-2.802	0.00508 **
respiratory.rate	-3.347e-02	2.855e-02	-1.172	0.24109
temperature	2.547e-01	1.835e-01	1.388	0.16507
SP.O2	1.173e-01	4.989e-02	2.352	0.01867 *
urine.output	1.455e-04	1.110e-04	1.310	0.19005
hematocrit	1.350e-02	1.666e-01	0.081	0.93543
RBC	-9.281e-02	1.452e+00	-0.064	0.94902
MCH	-4.682e-01	9.494e-01	-0.493	0.62188
MCHC	5.011e-01	8.443e-01	0.593	0.55287
MCV	1.275e-01	3.307e-01	0.386	0.69978
RDW	-9.984e-02	5.765e-02	-1.732	0.08328 .
leucocyte	-6.005e-02	2.167e-02	-2.771	0.00559 **
platelets	3.487e-03	1.072e-03	3.252	0.00114 **
neutrophils	-4.732e-03	1.695e-02	-0.279	0.78016
basophils	1.006e-03	2.769e-03	0.363	0.71631
lymphocyte	2.260e-02	2.691e-02	0.840	0.40111
PT	-1.328e-02	1.427e-02	-0.931	0.35195
INR	5.189e-03	2.385e-02	0.218	0.82773
NT.proBNP	3.568e-07	8.199e-06	0.044	0.96528
creatinine.kinase	-4.386e-05	1.196e-04	-0.367	0.71375
creatinine	1.905e-03	1.965e-03	0.970	0.33228
urea.nitrogen	-1.215e-02	6.709e-03	-1.811	0.07014 .
glucose	-1.861e-03	2.210e-03	-0.842	0.39954
blood.potassium	-5.556e-01	4.206e-01	-1.321	0.18654
blood.sodium	-1.190e-01	2.861e-01	-0.416	0.67737
blood.calcium	6.698e-01	2.141e-01	3.128	0.00176 **
chloride	1.175e-01	2.904e-01	0.404	0.68588
anion.gap	-5.499e-03	2.964e-01	-0.019	0.98520
magnesium.ion	-5.661e-01	4.540e-01	-1.247	0.21243
pH	-2.611e+00	2.583e+00	-1.011	0.31209
bicarbonate	2.316e-01	2.931e-01	0.790	0.42939
lactic.acid	-2.670e-03	9.607e-04	-2.779	0.00545 **
PCO2	-5.816e-02	1.862e-02	-3.123	0.00179 **
EF	8.000e-05	8.745e-03	0.009	0.99270
m.a.p	1.831e-02	1.187e-02	1.542	0.12310

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 931.77 on 1175 degrees of freedom
Residual deviance: 640.69 on 1128 degrees of freedom
AIC: 736.69

Number of Fisher Scoring iterations: 9

The full model found the variables age, deficiencyanemia, renal.failure, temperature, platelets and PCO2 to be statistically significant.

Perform stepwise variable selection

The stepAIC function of the MASS package will be used to select the most contributive variables:

```
step_mlrD <- full_mlrD %>% stepAIC(trace = FALSE)
summary(step_mlrD)
```

Call:

```
glm(formula = outcome ~ age + BMI + deficiencyanemias + renal.failure +
    COPD + heart.rate + temperature + SP.O2 + RDW + leucocyte +
    platelets + lymphocyte + urea.nitrogen + blood.potassium +
    blood.sodium + blood.calcium + chloride + magnesium.ion +
    bicarbonate + lactic.acid + PCO2 + m.a.p, family = "binomial",
    data = dataD)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.1509	0.1555	0.2865	0.4604	2.1283

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.613e+01	8.984e+00	-1.796	0.072523	.
age	-1.861e-02	9.508e-03	-1.957	0.050317	.
BMI	2.793e-02	1.586e-02	1.761	0.078222	.
deficiencyanemiasYes	6.758e-01	2.455e-01	2.753	0.005906	**
renal.failureYes	1.318e+00	2.662e-01	4.950	7.41e-07	***
COPDYes	1.272e+00	4.739e-01	2.684	0.007268	**
heart.rate	-2.479e-02	6.748e-03	-3.673	0.000240	***
temperature	2.634e-01	1.725e-01	1.527	0.126880	
SP.O2	1.211e-01	4.706e-02	2.574	0.010053	*
RDW	-1.313e-01	4.847e-02	-2.709	0.006748	**
leucocyte	-5.315e-02	2.001e-02	-2.656	0.007901	**
platelets	3.132e-03	9.814e-04	3.192	0.001415	**
lymphocyte	3.580e-02	1.692e-02	2.116	0.034356	*
urea.nitrogen	-1.396e-02	6.213e-03	-2.246	0.024676	*
blood.potassium	-6.424e-01	2.738e-01	-2.347	0.018946	*
blood.sodium	-1.374e-01	6.150e-02	-2.234	0.025463	*
blood.calcium	6.928e-01	2.060e-01	3.364	0.000769	***
chloride	1.280e-01	5.510e-02	2.323	0.020168	*
magnesium.ion	-7.166e-01	4.345e-01	-1.649	0.099097	.
bicarbonate	2.254e-01	5.166e-02	4.363	1.28e-05	***
lactic.acid	-2.641e-03	9.283e-04	-2.845	0.004436	**
PCO2	-4.935e-02	1.270e-02	-3.885	0.000102	***

```
m.a.p          2.157e-02  1.116e-02   1.934 0.053114 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 931.77  on 1175  degrees of freedom
Residual deviance: 654.37  on 1153  degrees of freedom
AIC: 700.37
```

Number of Fisher Scoring iterations: 6

The function chose a final model with the following formula:

```
outcome ~ age + BMI + deficiencyanemias + renal.failure + COPD + heart.rate + temperature + SP.O2
+ RDW + leucocyte + platelets + lymphocyte + urea.nitrogen + blood.potassium + blood.sodium +
blood.calcium + chloride + magnesium.ion + bicarbonate + lactic.acid + PCO2 + m.a.p
```

It found the variables age, deficiencyanemia, renal.failure, temperature, SP.O2, platelets, leucocyte, neutrophils, lymphocyte, blood.calcium, blood.potassium, bicarbonate, pH and PCO2 to be statistically significant.

```
formula(step_mlrD)
```

```
outcome ~ age + BMI + deficiencyanemias + renal.failure + COPD +
  heart.rate + temperature + SP.O2 + RDW + leucocyte + platelets +
  lymphocyte + urea.nitrogen + blood.potassium + blood.sodium +
  blood.calcium + chloride + magnesium.ion + bicarbonate +
  lactic.acid + PCO2 + m.a.p
```

```
results_mlrD<-cbind( "OR" = exp(coef(step_mlrD)), "p-value" = coef(summary(step_mlrD))[,4])
results_mlrD
```

	OR	p-value
(Intercept)	9.840785e-08	7.252309e-02
age	9.815627e-01	5.031657e-02
BMI	1.028326e+00	7.822176e-02
deficiencyanemiasYes	1.965517e+00	5.905536e-03
renal.failureYes	3.734415e+00	7.414917e-07
COPDYes	3.568146e+00	7.268223e-03
heart.rate	9.755196e-01	2.399739e-04
temperature	1.301313e+00	1.268800e-01
SP.O2	1.128785e+00	1.005339e-02
RDW	8.769505e-01	6.747946e-03
leucocyte	9.482408e-01	7.900620e-03
platelets	1.003137e+00	1.414831e-03
lymphocyte	1.036447e+00	3.435571e-02
urea.nitrogen	9.861404e-01	2.467560e-02
blood.potassium	5.260369e-01	1.894595e-02
blood.sodium	8.716113e-01	2.546309e-02
blood.calcium	1.999267e+00	7.692005e-04
chloride	1.136572e+00	2.016832e-02
magnesium.ion	4.884110e-01	9.909677e-02

bicarbonate	1.252800e+00	1.283204e-05
lactic.acid	9.973621e-01	4.435837e-03
PCO2	9.518447e-01	1.022856e-04
m.a.p	1.021809e+00	5.311395e-02

Data E

Full logistic regression model

```
xnamesE <- names(dataE[2:48])
foriE<- as.formula(paste("outcome ~ ", paste(xnamesE, collapse= "+")))
```

```
full_mlrE<- glm(forE, data = dataE, family = 'binomial')
summary(full_mlrE)
```

Call:

```
glm(formula = foriE, family = "binomial", data = dataE)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.0993	0.1513	0.2814	0.4645	2.0271

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.095e+01	3.407e+01	-0.908	0.363729
age	-1.504e-02	1.034e-02	-1.455	0.145787
genderM	-6.444e-02	2.280e-01	-0.283	0.777410
hypertensiveYes	3.657e-02	2.419e-01	0.151	0.879826
atrialfibrillationYes	-1.453e-01	2.411e-01	-0.603	0.546566
CHD.with.no.MIYes	8.627e-02	3.789e-01	0.228	0.819889
diabetesYes	5.759e-02	2.467e-01	0.233	0.815440
deficiencyanemiasYes	7.090e-01	2.575e-01	2.753	0.005899 **
depressionYes	3.109e-01	3.827e-01	0.812	0.416625
hyperlipemiaYes	2.197e-01	2.392e-01	0.919	0.358321
renal.failureYes	1.285e+00	2.768e-01	4.643	3.44e-06 ***
COPDYes	1.140e+00	4.839e-01	2.356	0.018462 *
hematocrit	1.501e-02	1.649e-01	0.091	0.927482
RBC	-8.504e-02	1.434e+00	-0.059	0.952709
MCH	-3.886e-01	9.460e-01	-0.411	0.681248
MCHC	4.594e-01	8.421e-01	0.546	0.585376
MCV	9.757e-02	3.291e-01	0.296	0.766878
RDW	-9.985e-02	5.803e-02	-1.721	0.085311 .
leucocyte	-6.734e-02	2.144e-02	-3.142	0.001680 **
platelets	3.404e-03	1.056e-03	3.223	0.001269 **
NT.proBNP	-2.542e-07	8.091e-06	-0.031	0.974935
creatinine	2.046e-03	2.011e-03	1.018	0.308853
urea.nitrogen	-1.400e-02	6.571e-03	-2.131	0.033100 *
blood.potassium	-5.661e-01	4.151e-01	-1.364	0.172607
blood.sodium	-1.537e-01	2.887e-01	-0.532	0.594569
chloride	1.511e-01	2.930e-01	0.516	0.605970
anion.gap	3.595e-02	2.983e-01	0.121	0.904074

```

magnesium.ion      -6.077e-01  4.484e-01  -1.355  0.175338
bicarbonate         2.037e-01  2.951e-01   0.690  0.489919
EF                  -1.082e-03  8.582e-03  -0.126  0.899631
BMI                 2.174e-02  1.541e-02   1.411  0.158164
heart.rate          -1.925e-02  7.286e-03  -2.643  0.008227 **
respiratory.rate    -3.530e-02  2.836e-02  -1.245  0.213140
temperature         2.695e-01  1.839e-01   1.465  0.142844
SP.O2               1.142e-01  4.967e-02   2.299  0.021491 *
urine.output        1.515e-04  1.111e-04   1.364  0.172704
neutrophils         -7.221e-03  1.540e-02  -0.469  0.639034
basophils           -6.952e-04  1.924e-03  -0.361  0.717853
lymphocyte          6.653e-03  2.362e-02   0.282  0.778228
PT                  -1.416e-02  1.390e-02  -1.019  0.308315
INR                  4.715e-03  1.634e-02   0.288  0.772981
creatine.kinase     -9.007e-05  1.341e-04  -0.672  0.501668
glucose             -1.487e-03  2.214e-03  -0.672  0.501856
blood.calcium       7.278e-01  2.155e-01   3.377  0.000732 ***
pH                  2.145e-01  2.037e+00   0.105  0.916119
lactic.acid         -2.581e-03  9.461e-04  -2.728  0.006368 **
PCO2                -2.104e-02  1.260e-02  -1.670  0.094839 .
m.a.p               1.974e-02  1.195e-02   1.651  0.098711 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 931.77  on 1175  degrees of freedom
Residual deviance: 651.04  on 1128  degrees of freedom
AIC: 747.04

```

Number of Fisher Scoring iterations: 8

The full model found the variables age, deficiencyanemia, renal.failure, temperature, platelets and PCO2 to be statistically significant.

Perform stepwise variable selection

The stepAIC function of the MASS package will be used to select the most contributive variables:

```

step_mlrE <- full_mlrE %>% stepAIC(trace = FALSE)
summary(step_mlrE)

```

Call:

```

glm(formula = outcome ~ age + deficiencyanemias + renal.failure +
    COPD + MCH + MCHC + RDW + leucocyte + platelets + urea.nitrogen +
    blood.potassium + blood.sodium + chloride + magnesium.ion +
    bicarbonate + BMI + heart.rate + respiratory.rate + temperature +
    SP.O2 + urine.output + blood.calcium + lactic.acid + PCO2 +
    m.a.p, family = "binomial", data = dataE)

```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-3.1281   0.1565   0.2900   0.4710   2.1122

```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.128e+01	9.919e+00	-2.146	0.031910	*
age	-1.712e-02	9.255e-03	-1.850	0.064273	.
deficiencyanemiasYes	7.529e-01	2.447e-01	3.076	0.002095	**
renal.failureYes	1.325e+00	2.653e-01	4.994	5.9e-07	***
COPDYes	1.144e+00	4.752e-01	2.407	0.016101	*
MCH	-8.059e-02	4.887e-02	-1.649	0.099145	.
MCHC	1.981e-01	1.055e-01	1.878	0.060405	.
RDW	-1.078e-01	5.151e-02	-2.093	0.036360	*
leucocyte	-7.181e-02	1.906e-02	-3.767	0.000165	***
platelets	3.326e-03	1.002e-03	3.320	0.000900	***
urea.nitrogen	-1.595e-02	6.096e-03	-2.617	0.008874	**
blood.potassium	-5.118e-01	2.784e-01	-1.838	0.066052	.
blood.sodium	-1.182e-01	6.159e-02	-1.920	0.054917	.
chloride	1.141e-01	5.511e-02	2.071	0.038333	*
magnesium.ion	-7.449e-01	4.301e-01	-1.732	0.083265	.
bicarbonate	1.734e-01	4.991e-02	3.475	0.000510	***
BMI	2.070e-02	1.460e-02	1.418	0.156183	.
heart.rate	-1.911e-02	7.035e-03	-2.717	0.006592	**
respiratory.rate	-3.977e-02	2.742e-02	-1.450	0.147029	.
temperature	2.646e-01	1.744e-01	1.517	0.129220	.
SP.O2	1.121e-01	4.770e-02	2.350	0.018793	*
urine.output	1.536e-04	1.074e-04	1.430	0.152652	.
blood.calcium	7.819e-01	2.057e-01	3.800	0.000145	***
lactic.acid	-2.574e-03	9.260e-04	-2.780	0.005438	**
PCO2	-2.248e-02	9.826e-03	-2.288	0.022146	*
m.a.p	2.127e-02	1.133e-02	1.878	0.060416	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 931.77 on 1175 degrees of freedom
Residual deviance: 660.58 on 1150 degrees of freedom
AIC: 712.58

Number of Fisher Scoring iterations: 6

The function chose a final model with the following formula:

outcome ~ age + BMI + atrialfibrillation + deficiencyanemias + renal.failure + COPD + heart.rate +
temperature + leucocyte + platelets + neutrophils + lymphocyte + PT + urea.nitrogen + glucose +
blood.calcium + chloride + anion.gap + magnesium.ion + PCO2

It found the variables age, deficiencyanemia, renal.failure, temperature, SP.O2, platelets, leucocyte, neutrophils, lymphocyte, blood.calcium, blood.potassium, bicarbonate, pH and PCO2 to be statistically significant.

```
formula(step_mlrE)
```

```
outcome ~ age + deficiencyanemias + renal.failure + COPD + MCH +  
MCHC + RDW + leucocyte + platelets + urea.nitrogen + blood.potassium +
```

```
blood.sodium + chloride + magnesium.ion + bicarbonate + BMI +
heart.rate + respiratory.rate + temperature + SP.O2 + urine.output +
blood.calcium + lactic.acid + PCO2 + m.a.p
```

```
results_mlrE<-cbind( "OR" = exp(coef(step_mlrE)), "p-value" = coef(summary(step_mlrE))[,4])
results_mlrE
```

	OR	p-value
(Intercept)	5.722608e-10	3.191005e-02
age	9.830221e-01	6.427285e-02
deficiencyanemiasYes	2.123071e+00	2.094956e-03
renal.failureYes	3.762257e+00	5.900092e-07
COPDYes	3.138312e+00	1.610051e-02
MCH	9.225741e-01	9.914542e-02
MCHC	1.219133e+00	6.040478e-02
RDW	8.978055e-01	3.635957e-02
leucocyte	9.307094e-01	1.650687e-04
platelets	1.003332e+00	8.996779e-04
urea.nitrogen	9.841739e-01	8.873747e-03
blood.potassium	5.994315e-01	6.605163e-02
blood.sodium	8.885004e-01	5.491705e-02
chloride	1.120916e+00	3.833346e-02
magnesium.ion	4.747646e-01	8.326452e-02
bicarbonate	1.189382e+00	5.103316e-04
BMI	1.020912e+00	1.561829e-01
heart.rate	9.810685e-01	6.592315e-03
respiratory.rate	9.610117e-01	1.470290e-01
temperature	1.302950e+00	1.292200e-01
SP.O2	1.118587e+00	1.879283e-02
urine.output	1.000154e+00	1.526522e-01
blood.calcium	2.185554e+00	1.446029e-04
lactic.acid	9.974293e-01	5.438301e-03
PCO2	9.777694e-01	2.214570e-02
m.a.p	1.021500e+00	6.041613e-02