AssignmentIV

Silpa Soni Nallacheruvu (19980824-5287) Hernan Aldana (20000526-4999)

2024-12-09

Summary

Exercise 4:1

Question 1

Report the model selection process briefly. Based on your chosen model, which factors affect the probability of not surviving? Report odds ratios with confidence intervals for the most important variables/factors, and interpret them. Use the variable names from the table (not V3, V4, etc.).

Approach:

- Model Selection Process:
  - Use a stepwise selection with AIC to identify a parsimonious model.
- Analysis of the Final Model:
- Extract coefficients, odds ratios, and their 95% confidence intervals for significant variables.

-Ensure variable names are replaced with their descriptions (e.g., Age, Sex, etc.) instead of column names.

- Interpretation:
  - Interpret the results form the AIC, odds ratios, and confidence intervals, to determine the best model.

Code and results :

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

##
## Call:
## glm(formula = Survival ~ ConsciousnessLevel + TypeOfAdmission +
##     Age + Cancer + Patient + BloodCarbonDioxide + BloodPH + BloodPressure,
##     family = binomial, data = data_ca4)
##
## Coefficients:
```

```
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -4.7353420  1.6104573  -2.940 0.003278 **
## ConsciousnessLevel  2.6208042  0.6859650   3.821 0.000133 ***
## TypeOfAdmission     3.0547147  0.9339217   3.271 0.001072 **
## Age                 0.0385864  0.0133655   2.887 0.003889 **
## Cancer              2.3388380  0.8671971   2.697 0.006997 **
## Patient            -0.0020714  0.0008783  -2.359 0.018345 *
## BloodCarbonDioxide -2.4646334  1.0619854  -2.321 0.020299 *
## BloodPH             2.0884994  0.9031831   2.312 0.020757 *
## BloodPressure      -0.0099893  0.0070360  -1.420 0.155682
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 130.19  on 191  degrees of freedom
## AIC: 148.19
##
## Number of Fisher Scoring iterations: 6
```

```
## Waiting for profiling to be done...
```

```
##                              Variable    OddsRatio     CI.Lower    CI.Upper
## (Intercept)               (Intercept)  0.008779445 0.0002996746   0.1795376
## ConsciousnessLevel ConsciousnessLevel 13.746774142 4.3144886742  65.2810381
## TypeOfAdmission       TypeOfAdmission 21.215131928 4.3561881520 189.1540175
## Age                               Age  1.039340550 1.0141827727   1.0692793
## Cancer                         Cancer 10.369181076 1.9513659807  66.5483329
## Patient                       Patient  0.997930702 0.9961324780   0.9995944
## BloodCarbonDioxide BloodCarbonDioxide  0.085040009 0.0080634712   0.5539141
## BloodPH                       BloodPH  8.072791946 1.4001269889  53.6032965
```

Conclusion :

Model selection

- From the output the final model includes the following variables: ConsciousnessLevel, TypeOfAdmission, Age, Cancer, Patient, BloodCarbonDioxide, BloodPH, and BloodPressure.
- These variables where selected using a stepwise AIC, which ensures a balance between model complexity and goodness of fit.

** Significant Variables **

- Variables with a p-value < 0.05 are considered significant predictors of survival
  - ConsciousnessLevel
  - TypeOfAdmission
  - Age
  - Cancer
  - BloodCarbonDioxide,
  - BloodPH

Odds Ratios and Confidence Intervals:

The odds ratio table

Output :

Observation :