Previous Item

Next Item

The questions of the third E-Test are based on the tasks of R-Laboratory 8 and 9. **Please provide numbers in the requested precision within each question. The use of different precision is evaluated as wrong.**

Task 25

At the beginning of your code, set the seed to 2021. Load the dataset *credits.csv* from RWTHmoodle into your workspace. **Make sure that your working directory contains an unmodified version of *credits.csv*. In particular, we recommend downloading a fresh version of *credits.csv* from the RWTHmoodle space.** **4.5 of 5.5 points**

1 of 1 point

In the *credits* dataset, the variable `guaran` stands for wheather the possible credits recipient is "No further guarantor of debtor" (`guaran` $= 1$), "co applicant" (`guaran` $= 2$) or "guarantor" (`guaran` $= 3$). Introduce a new variable `guaran2` with two categories where `guaran2` $= 1$ for `guaran` $= 1$ and `guaran` $= 2$ and `guaran2` $= 2$ otherwise. For how many cases in the data set *credits* it holds that `guaran2=1`? **(requested precision: whole number)**

948 ✓

0.5 of 0.5 points

Split randomly the data in training (65 % of the dat) and testing data (the remaining 35 % of the data). Transform `account, behavior, rate, finance, furthered, home, job, pers` into factor variables. Fit the following logistic regression model on the training data:

`repayment` $\sim$
`time+age+account+behavior+savings+rate+guaran2+finance+furthcred+home+job+pers`

In the following, we will refer to this model as *model b*.

What are the AIC and BIC values for this model? **(requested precision: 1 digit)**

AIC value

663.1 ✓

0.5 points

BIC value

779.5 ✗ 774.4

1 of 1 point

What is the AUC (area under the curve) of *model b*? **(requested precision: 3 digits)**

0.814 ✓

1 of 1 point

Fit the null model corresponding to *model b*. What is the null deviance for this model? **(requested precision: 4 digits)**

797.4835 ✓

0.5 points

Select the best model nested in *model b* in terms of AIC using a backward stepwise selection algorithm. In the following, we will refer to this selected model as *model e*. What is the median of the estimated probabilities for on time credit repayment for the training sample, based on this model? **(requested precision: 4 digits)**

0.776 ✗ 0.7573

0.5 of 0.5 points

Select the best model nested in *model b* in terms of BIC using a backward stepwise selection algorithm. In the following, we will refer to this selected model als *model f*. Test *model f* and *model e* on the test data. What are the values for predicted residual sum of squares (PRESS) for both models ? **(requested precision: 4 digits)**

PRESS for *model e*

62.6048 ✓

0.5 of 0.5 points

PRESS for *model f*

61.4974 ✓

Previous Item

Next Item

Create an inspection request?

Please note that this document will become a part of your exam file!
- Be polite and friendly.
- Describe your rationale as precisely as possible.

Dynexite, 19.07.2021

Request for correction of this item

Dear sir or madam, ...

(This document is valid without signature)

Changes were saved