MLE01 - Vertiefende statistische Verfahren

2. Übungsblatt SS 2024

Allgemeine Information

Alle Aufgaben sind mit R zu lössen. Die Berechnungen sollen nachvollziehbar und dokumentiert sein. Um die vollständige Punktezahl zu erreichen, müssen alle Ergebnisse und Fragen entsprechend interpretiert bzw. beantwortet werden. Code alleine ist nicht ausreichend! Die Abgabe erfolgt über Moodle entsprechend der Abgaberichtlinien als pdf und Rmd File. Bitte inkludieren Sie namentlich alle beteiligten Gruppenmitglieder sowohl im Bericht als auch im Source Code. Die jeweiligen Datensätze die für diese Übung relevant sind finden Sie ebenfalls in Moodle.

1 Logistische Regressionsanalyse [2P]

Verwenden Sie den Datensatz `birthwt.xlsx`. Dieser Datensatz bezieht sich auf Risikofaktoren im Zusammenhang mit niedrigem Geburtsgewicht von Säuglingen. Eine Beschreibung der einzelnen Variablen entnehmen Sie bitte dem Excel-File.

[1P] a: Erstellen Sie ein Modell, welches das Risiko für niedriges Geburtsgewicht (`low`; Gewicht <2500g ja/nein) in Abhängigkeit verschiedener Faktoren beschreibt. Wie lautet die Modellgleichung?

[1P] b: Überprüfen Sie die Modellvoraussetzungen und bewerten Sie die Güte des Modells. Wie hoch ist die Wahrscheinlichkeit für eine Geburt mit einem Geburtsgewicht <2500g, bei folgenden Daten der Mutter: 38 Jahre alt, 68 kg, weiß, Nichtraucher, 2 Vorgeburten (ptl), keinen Bluthochdruck, keine Reizung der Gebärmutter, ein Arztbesuch im 1. Trimester.

2 Logistische Regressionsanalyse [3P]

Verwenden Sie erneut die Framingham-Herz-Studiendaten in `Framingham.sav`. Die Variable `mi_fchd` beschreibt Patienten die einen hospitalisierten Myokardinfarkt oder eine tödliche koronare Herzkrankheit erlitten haben.

[1.5P] a: Erstellen Sie ein Modell, welches das Risiko für `mi_fchd` in Abhängigkeit von verschiedenen Faktoren beschreibt. Vermeiden Sie nicht relevante bzw. redundante Variablen. Achten Sie auf Ausreißer und fehlende Daten (`NaN, NA's`).

[1P] b: Überprüfen Sie die Modellvorraussetzungen und evaluieren Sie die Performance des finalen Modells hinsichtlich Genauigkeit und AUROC. Vergleichen Sie die Genauigkeit des Modells mit einem Klassifizierungsmodell, dass immer "no" vorhersagt.

[0.5P] b: Interpretieren Sie die OR. Welche Aussagen können Sie bezüglich Risikofaktoren für eine Hospitalisierung durch Myokardinfarkt bzw. tödliche koronare Herzkrankheit treffen? Was sind die Top3 Variablen mit dem stärksten Einfluss auf die Zielvariable?

3 Logistische Regressionsanalyse [2P]

Analysieren Sie Tabelle 4 in der Publikation "**Analysis of factors associated with disease outcomes in hospitalized patients with 2019 novel coronavirus disease**". Den Artikel finden Sie als pdf-Version in Moodle, oder direkt unter https://doi.org/10.1097/CM9.0000000000000775.

Beantworten Sie bitte in eigenen Worten folgende Fragen:

- Was ist die Outcome Variable?
- Welche Variablen wurden in die multivariate Analyse aufgenommen und warum?
- Was sind die Top3 Variablen in der univariaten Analyse mit dem stärksten Einfluss auf die Zielvariable?
- Was sind die Top3 Variablen in der multivariaten Analyse mit dem stärksten Einfluss auf die Zielvariable?
- Welche (medizinische) Bedeutung haben die Variablen im multivariaten Modell?
- Welche Schlussfolgerung ziehen sie für die Praxis?