

Datenauswertung durch logistische Regression

Ausgabe: 5. Juni 2019

Abgabe: 19. Juni 2019

Für Studenten, die an diesem Kurs zum ersten Mal teilnehmen:

- Diese Hausaufgabe zählt als 20 von 100 der Kurspunkte.

Für Studenten, die diesem Kurs wiederholen:

- Wenn Sie die Hausaufgabe abgeben, zählt sie als 20 von 100 der Kurspunkte und die Klausur zählt als 80 Kurspunkte.
- Wenn Sie die Hausaufgaben nicht abgeben und **dem Dozenten Bescheid geben**, zählt die Klausur als 100 Klausurpunkte.

Sie dürfen die Hausaufgabe allein oder zu zweit bearbeiten. Gruppenarbeit wird nicht differenziert benotet werden.

Der Abgabe Termin ist **Mittwoch 19. Juni 2019**.

Ihre Arbeit besteht aus einem wissenschaftlichen Bericht und Ihre R-Code. Sie dürfen Ihren Bericht entweder in Moodle hochladen oder eine schriftliche Exemplar abgeben. Sie müssen Ihre R-Skript-Datei (R-Code) in Moodle hochladen. Gruppen müssen die Arbeit nur einmal abgeben.

Aufgabe

Ein Datensatz `Telecom.Rda` ist in Moodle gegeben. Sie analysieren die Daten mit einem logistische Regressionsmodell. Sie entwickeln ein passendes Modell durch die Methoden aus Vorlesungen 6-8. *Betrachten Sie nur die Haupteffekte. Ein Wechselwirkungsmodell ist für diese Aufgabe nicht notwendig.* Eine Erklärung des Datensatzes finden Sie auf der nächsten Seite.

Der wissenschaftliche Bericht umfasst zumindest die folgenden Aspekte.

- Eine Beschreibung der Daten einschließlich beschreibender Statistiken und Grafiken.
- Eine Erläuterung aller relevanter Modelle und Ihre Entscheidungen in der Modellierungsvorgehensweise.
- Ihr Endmodell
- Ein Zusammenfassung der Einflüsse aller Einflussgröße im Endmodell.
- Wählen Sie eine Einflussgröße aus dem Endmodells aus und erläutern Sie im Einzelnen, die Relevanz/Bedeutung dieser Einflussgröße.

Ihre R-Skript-Datei sollte:

- Ihre Analyse vollständig durchführen,
- keine irrelevanten Befehle beinhalten und Fehlerfrei sein.
- angemessen Kommentare beinhalten

Geeignete statistische Praxis und eine sinnvolle Modellierungsvorgehensweise ist wichtiger in der Bewertung als die genaue Form Ihres Endmodells.



Der „Telecom“ Datensatz

Die Daten sind eine Stichprobe von Kundendaten einer Telekommunikationsfirma. Der Datensatz enthält 1000 Zeilen und 12 Variablen. Jede Zeile bezieht sich auf einem Kunden (Datenelement). Alle Variablen sind Gruppert (Faktorvariablen).

Die **Zielgröße** ist Churn. „Yes“= der Kunde ist abgewandert, „No“= der Kunde bleibt.

Mögliche Einflussgröße

Gender	Geschlecht
SeniorCitizen	Rentner
Tenure	Kunde seit zumindest 12 Monaten
MultipleLines	Kunde besitzt mehrfache Telefonnummern
InternetService	
OnlineBackup	
StreamingTV	
StreamingMovies	
Contract	Art von Vertrag
PaperlessBilling	papierlose Rechnung
PaymentMethod	Bezahlungsart

