

statistische Verfahren WS 2017/2018

Projekt 7 - Kriminalität

Reda Ihtassine (Matrikelnummer) Ingo Schäfer (165 220)

Jena, am 23. März 2018

Inhaltsverzeichnis

1	Einleitung	1
2	Material und Methoden	2
3	Resultate	3
4	Diskussion	4

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

latex $\mathbf{C}=(n-1)^{-1} \times \mathbf{D}^{\mathbf{T}} \times \mathbf{D} s=2$ *[p. 12]baltagli:1

Dem Projekt liegen Kriminalitätsdaten des US-amerikanischen Bundesstaats North Carolina zugrunde, welche in dem Zeitraum von 1981 bis 1987 erhoben wurden. Diese Daten wurden schon mehrfach mit Hilfe von verschiedenen Methoden (Hausmann, 2SLS, ...) von Anderen untersucht ¹.

Diese Arbeit versucht ein geeignetes allgemeines lineares Modell zu erarbeiten, mit dem gute Abschätzungen erzielt werden können.

¹1

2 Material und Methoden

2.1 Material

Der Datensatz besteht aus einer .csv-Datei. In ihr sind die unterschiedlichen 90 Counties von North Carolina zeilenweise aufgelistet. Die Spalten sind (mögliche) Eigenschaftsvektoren. In der Arbeit von Baltagli ² werden noch einige Eigenschaften mehr aufgelistet, als in dieser Arbeit betrachtet wurden. Daher hier eine kleine Übersicht über alle möglichen Einflussgrößen:

Alle Eigenschaftsvektoren sind logarithmisch mit Ausnahme der Region und der Zeit. Die erste Spalte beinhaltet die Zielgröße *crimes*, also die Anzahl aller Straftaten in dem jeweiligen County über den Zeitraum von 1981-1987.

Weiterhin wurde die Arrestwahrscheinlichkeit P_A hinzugefügt. Sie berechnet sich aus $P_A = \frac{\text{Arrestierungen}}{\text{textDelikte}}$. Sie wird abgekürzt *prbarr* geschrieben. Daneben gibt es auch die Überzeugungswahrscheinlichkeit P_C . Sie gibt das Verhältnis zwischen tatsächlichen Arrestierungen und den gestandenen Straftaten an und wird daher berechnet mit $P_P = \frac{\text{Anzahl tatsächlicher Arrestierungen}}{\text{Anzahl gestandener Straftaten}}$. Sie wird bezeichnet als *prbpris*.

Eine weitere Eigenschaft ist die Fähigkeit des Countys ein Verbrechen auch zu ermitteln. In dem Datensatz spiegelt sich dies in der Variable *polpc* wieder. Sie gibt das Polizei-pro-Kopf-Verhältnis an.

Ein weiteres wichtiges Merkmal ist die Bevölkerungsdichte (*density*). Sie stellt das Verhältnis *franzahlbevölkerungFlchedesCountysinsquaremiles* dar.

Darüber hinaus wird das Verhältnis von Minderheiten zu der Gesamtanzahl Einwohner in der Variable *pctmin* ausgedrückt.

pctymale ist eine Eigenschaft, die den Anteil der jungen männlichen Bevölkerung zur Gesamtbevölkerung anzeigt.

Die letzten fünf Variablen geben den durchschnittlichen Bruttolohn in den Bereichen Baugewerbe (*wcon*), Staatsangestellte (*wsta*), Dienstleistungssektor (*wser*), Handel (*wtrd*) und Bankgeschäften (*textitwfr*) wieder.

2.2 Methoden

²1

3 Resultate

1:

Vorgehensweise:

- negative binomialverteilung statt gauß-verteilung, begründen - siehe quelle! - 5 unterschiedliche Herangehensweisen um ein geeignetes Modell zu finden, alle kurz erklären
- besondere Rolle von *region*
- vergleich der modelle funktionsweise knapp erläutern(aic, cross_validation, cor())
- vorstellen 5 gewinnermodelle, den gewinner

2 :

- beschreibungstest() - funktion(ggf funktionen eingriffigererennamengeben)
- welche einstellungen erzielt gute ergebnisse?
- einmal mit einfachem modell zeigen : mDensity...(< -warum mDensity?) - einmal mit gewinnermodell - - > wie gut ist ergebnis? 1/4 berleitung zu diskussion...

4 Diskussion