Projekt LVS-IR-Taubenstein

Projektpartner: Sascha Filimon, Roman Ossner
Gruppenbetreuer: André Klima
Projektgruppe: Alexander Fogus, Lea Vanheyden, Zorana Spasojevic

7. April 2020

Ludwig Maximilians Universität

Inhaltsverzeichnis

1. Hintergrund

- Konfliktsituation zwischen Mensch und Natur/Tierreich im Alpengebiet
- Kooperation des Departments für Geographie an der LMU, Lawinencamp Bayern, Gebietsbetreuer Mangfallgebirge, Alpenregion Tegernsee/Schliersee und DAV Sektion München
- Speziellen Untersuchungen am Spitzingsee (beliebte Gegend für Sportler und Wildtiere)
- Wie verhalten sich die Besucher und wie kann man dieses Verhalten steuern?
- Dazu Untersuchung über die Mitnahme von LVS-Geräten anhand von Checkpoints und manueller Datenerhebung

2. Datengrundlage

- Untersuchungsgegenstand: Wintersportler (vorrangig Skitourengänger & Schneeschuhgeher)
- Untersuchungszeitraum: Wintersaison 18/19
  ↳ Genauer Zeitraum: 21.12.2018 - 12.04.2019
- Checkpoints an zwei Routen (Nord- und Südseite) erfassen:
  ↳ Messungen insgesamt: 31574
  ↳ Personen mit LVS-Gerät: 8468
  ↳ Personen ohne LVS-Gerät: 23106

2. Datengrundlage

2. Datengrundlage

- Weitere Variablen:
  ↳ Datum
  ↳ Uhrzeit
  ↳ Tageslänge
  ↳ Temperatur
  ↳ Schneehöhe
  ↳ Sonnenstrahlung
  ↳ Wochentag
  ↳ Feiertag
  ↳ Lawinenwarnstufe
- Durch manuelle Stichproben wurden die Messungen der Checkpoints als fehlerhaft erkannt

3. Aufgabenstellung

1. **Modell:** Anteil der Skitourengänger mit LVS-Gerät in Abhängigkeit von anderen Faktoren (wie z.B. Uhrzeit, Temperatur, Schneehöhe)
2. Einflussfaktoren von denen die Messfehler abhängen, welcher Art und Struktur
3. Hypothese: Unter Berücksichtigung der Erkenntnisse über die Messfehler
   ↳ Wie beeinflussen die Messfehler die geschätzten Abhängigkeiten?

Daten

Die binäre Zielvariablen y_i sind 0/1-kodiert und bei gegebenen Kovariablen $x_{i1}, ..., x_{ik}$ (bedingt) unabhängig.

Modelle

Die Wahrscheinlichkeit $\pi_i = P(y_i = 1 | x_{i1}, ..., x_{ik})$ und der lineare Prädiktor:

$$\eta_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} = x_i' \beta$$

sind durch eine Responsefunktion $h(\eta) \in [0, 1]$ miteinander verknüft:

$$\pi_i = h(\eta_i).$$

(Quelle: Fahrmeir, Ludwig, et al. Regression; Prof. Dr. Helmut Küchenhoff Vorlesungsskript)

Logit-Modell:

$$\pi = \frac{exp(\eta)}{1 + exp(\eta)} \iff log\frac{\pi}{1-\pi} = \eta.$$

Interpretation: log odds durch lineares Modell beschreibbar.

Probit-Modell:

$$\pi = \Phi(\eta) \iff \Phi^{-1}(\pi) = \eta.$$

Interpretation: z-transformierte Wahrscheinlichkeiten sind durch lineares Modell beschreibbar.

Interpretation der β durch marginale Effekte

(Quelle: Fahrmeir, Ludwig, et al. Regression; Prof. Dr. Helmut Küchenhoff Vorlesungsskript)

- Gruppiertes Logit- oder Probit-Modell würden sich eignen
- Statistische Analysen mit Logit- und Probit-Modellen führen zu ähnlichen Resultaten

5. Probleme

1. Modellwahl
   ↳ Longitudinal-Data-Analysis: z.B. Schneehöhe an Tag 1 korreliert mit Schneehöhe an Tag 2
   ↳ Problem: Daten über gewissen Zeitraum gemessen, an manchen Tagen keine Messungen
2. Modellvergleich
   ↳ Logit- oder Probit-Modell
   ↳ Variablenselektion
3. Überdispersion

5.Probleme

Überdispersion:

Für gruppierte Daten lässt sich die Varianz innerhalb der Gruppe abschätzen durch $\frac{\bar{y}_i(1-\bar{y}_i)}{n_i}$, da $\bar{y}_i$ der ML-Schätzer für π_i basierend auf den Daten der Gruppe i ist.

Problem:

In Anwendungen ist diese *empirische* Varianz oft deutlich größer als die durch ein binomiales Regressionsmodell vorhergesagte Varianz $\frac{\hat{\pi}_i(1-\hat{\pi}_i)}{n_i}$ mit $\hat{\pi}_i = h(x_i'\beta)$.

Lösung:

Einführung eines multiplikativen Überdispersionsparameters $\phi > 1$ in die Varianzformel, d.h. $Var(y_i|\mathbf{x}_i) = \phi\frac{\pi_i(1-\pi_i)}{n_i}$

(Quelle: Fahrmeir, Ludwig, et al. Regression)