

Projekt LVS-IR-Taubenstein

Projektpartner: Sascha Filimon, Roman Ossner

Gruppenbetreuer: André Klima

Projektgruppe: Alexander Fogus, Lea Vanheyden, Zorana Spasojevic

22. März 2020

Ludwig Maximilians Universität



1. Hintergrund
2. Datengrundlage
3. Aufgabenstellung
4. Binäre Regressionsmodelle
5. Probleme



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

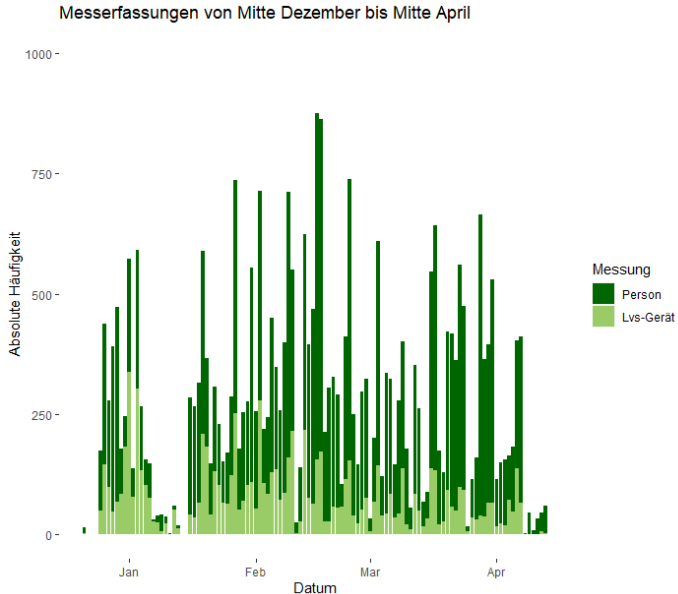


2. Datengrundlage

- Untersuchungsgegenstand: Wintersportler (vorrangig Skitourenzügler & Schneeschuhgeher)
- Untersuchungszeitraum: Wintersaison 18/19
- Checkpoints an zwei Routen (Nord- und Südseite) erfassen:
 - ↳ durch Infrarotsignale vorbeigehende Personen und
 - ↳ eventuell beigeführte LVS-Geräte automatisch



2. Datengrundlage



2. Datengrundlage

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



3. Aufgabenstellung

1. **Modell:** Anteil der Skitourengehnger mit LVS-Gert in Abhngigkeit von anderen Faktoren (wie z.B. Uhrzeit, Temperatur, Schneehhe)
2. Einflussfaktoren von denen die Messfehler abhngen, welcher Art und Struktur
3. Hypothese: Unter Bercksichtigung der Erkenntnisse ber die Messfehler
↳ Wie beeinflussen die Messfehler die geschtzten Abhngigkeiten?



4. Binäre Regressionsmodelle

Daten

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

Modelle

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

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} = \mathbf{x}_i' \boldsymbol{\beta}$$

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

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

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



4. Binäre Regressionsmodelle

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)



4. Binäre Regressionsmodelle

- Gruppiertes Logit- oder Probit-Modell würde sich eignen
- Vergleich:
 - ↳ Statistische Analysen mit Logit- und Probit-Modellen führen zu ähnlichen Resultaten
 - ↳ Log odds Interpretation bei Logit (medizinische Anwendung)
 - ↳ Herleitung über normalverteilte Nutzenfunktion bei Probit (ökonomische Anwendung)
 - ↳ Probit eher ausreisserempfindlich



5. Probleme

1. Korrelation von unabhängigen Variablen
 - ↳ z.B. Schneehöhe an Tag 1 und an Tag 2
2. Variablenselektion und Modellwahl
 - ↳ Hypothesentests wie z.B. Wald- oder Score-Test
 - ↳ Information criteria wie z.B. AIC (Akaike information criterion) und BIC (Bayesian information criterion)
 - ↳ Penalized likelihood model wie z.B. LASSO
3. Longitudinal-Data-Analysis in das Modell miteinbeziehen
 - ↳ Problem: Daten über gewissen Zeitraum gemessen, an manchen Tagen keine Messungen
4. 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 $\hat{\pi}_i = h(\mathbf{x}_i'\beta)$.

Lösung:

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

