

# Kleinräumige extrapolation von Umfragedaten

Kai Husmann, Alexander Lange

Georg-August-Universität Göttingen

11. Juli 2016

# Inhaltsverzeichnis

## Einleitung

## Datengrundlage

- Bürgerumfrage

- Zensus

- Melderegister

## Methoden

- Binäre Modelle

- Kategorielle Modelle

- Geostatistische Modelle

- Splines

# Einleitung

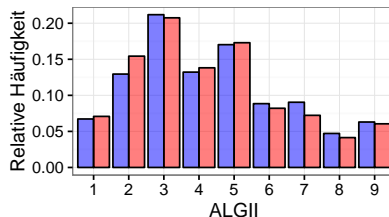
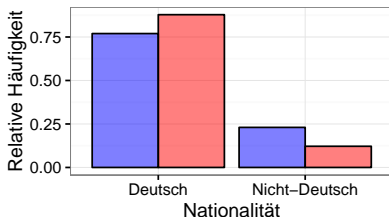
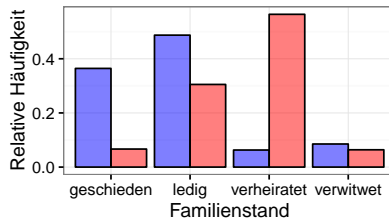
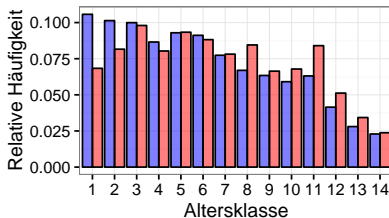
- Daten zur Bevölkerungsstruktur von Stuttgart aus unterschiedlichen Erhebungen
- Datensätze sowohl der Grundgesamtheit, als auch der Stichproben
- Verwendung von sowohl strukturellen- als auch räumlichen Informationen
- Ziel: Durch geeignete Modelle und der Stichproben die Informationen der Grundgesamtheit schätzen

# Bürgerumfrage

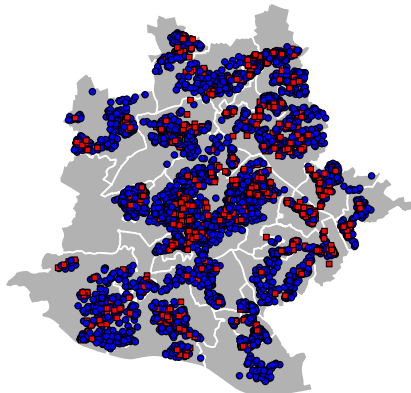
Anzahl Beobachtungen Grundgesamtheit: 470.190

Anzahl Beobachtungen Stichprobe: 4.082

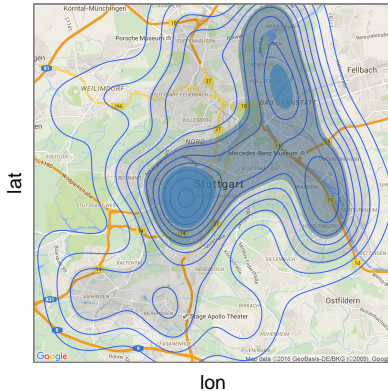
Variable	Skalenniveau	Modus		Verteilung
		G	S	
Altersklasse	Ordinal	18-25	30-35	Logistisch
Geschlecht	Nominal	Frau	Frau	Binom.
Nationalität	Nominal	Deutsch	Deutsch	Binom.
Familienstand	Nominal	Ledig	Verheiratet	Multinom.
Haushaltsgröße	Ordinal	2	2	Logistisch
Wohndauer	Ordinal	>10	>10	Logistisch
ALG II Quote	Ordinal	4% - 6%	4% - 6%	Logistisch
Ein/Zweifamilienhäuser	Ordinal	7	7	Logistisch
Eigentümer	Nominal	(Nicht enthalten)	Mieter	Binom.
Gauß-Krüger X				
Gauß-Krüger Y				



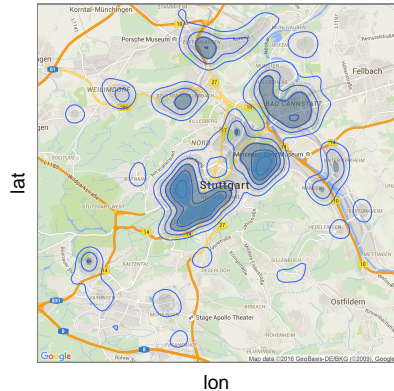
Grundgesamtheit Stichprobe



## Stichprobe



## Grundgesamtheit



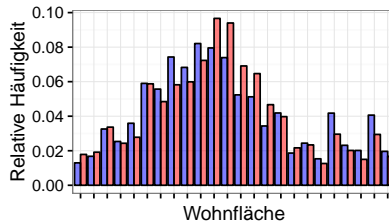
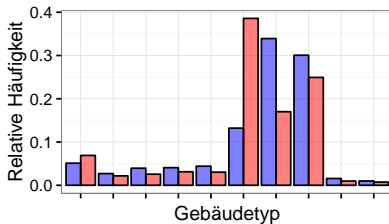
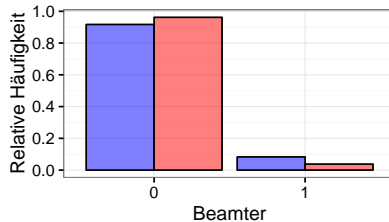
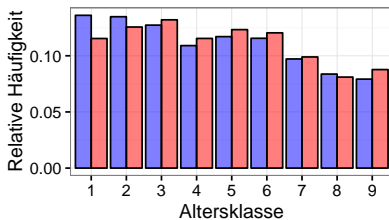
## Zensus

Anzahl Beobachtungen Grundgesamtheit: 380.238

Anzahl Beobachtungen Stichprobe: 14.671

Variable	Skalenniveau	Modus		Verteilung
		G	S	
Altersklasse	Ordinal	18-25	30-35	Logistisch
Geschlecht	Nominal	Mann	Mann	Binom.
Nationalität	Nominal	Deutsch	Deutsch	Binom.
Familienstand	Nominal	Verheiratet	Verheiratet	Multinom.
Haushaltsgröße	Ordinal	2	2	Logistisch
Wohnfläche	Ordinal	70 - <75	75 - <80	Logistisch
Stellung Beruf	Nominal	Selbstst. o. Beschäftigte	Angestellte	Multinom.
Beamter	Nominal	Nein	Nein	Binom./Pois.
Gebäudetyp	Nominal	Mehrfamilienh. 3-6 W.	Mehrfamilienh. >13 W.	Multinom.
Gebäudenutzung	Nominal	Wohnzwecke vermietet	Wohnzwecke vermietet	Multinom.
Gauß-Krüger X				
Gauß-Krüger Y				





Grundgesamtheit Stichprobe

# Melderegister

Anzahl Beobachtungen: 591.887

---

---

**Variable**

---

Altersklasse

---

Geschlecht

---

Familienstand

---

Nationalität

---

---

# Binäre Modelle

- Logit Modell  
 $h(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)}$ , Logistische Verteilung
- Probit Modell  
 $\Phi(\eta) = \int_{-\infty}^{\eta} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2) dx$ , Normalverteilung
- Komplementäres Log-Log Modell  
 $h(\eta) = 1 - \exp(-\exp(-\eta))$ , Extremwertverteilung

# Kategorielle Modelle

- Multinomiales Logit Modell

$$P(y_i = 1) = \frac{\exp(\mathbf{x}'_i \beta)}{1 + \exp(\mathbf{x}'_i \beta)}, \text{ Multinomialverteilung}$$

$$P(y_i = 0) = 1 - P(y_i = 1)$$

- Kumulatives geordnetes Modell

$$\log\left(\frac{P(y_i \leq r)}{P(y_i > r)}\right) = \theta_r + \mathbf{x}'_i \tilde{\beta}, \text{ Logistische Verteilung}$$

- Sequentielles Modell

$$P(y_i = r | y_i \geq r) = F(\theta_r + \mathbf{x}'_i \tilde{\beta}), \text{ Jede kumulative Verteilungsfunktion } F$$

# Geomodelle

- Universelles Kriging  
 $y(\mathbf{s}) = \mu(\mathbf{s}) + W(\mathbf{s}) + \varepsilon(\mathbf{s})$ ,  
mit  $W(\mathbf{s}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

# Splines

- Penalisierter Spline

$$\mathbf{f}_q(\mathbf{z}_q)$$

- Räumliche Effekte als Tensor Produkt Spline

$$f_{spat}(s_x, s_y)$$

- Lineare parametrische Einflüsse

$$\mathbf{x}'_p \boldsymbol{\beta}$$

- Regressionsmodell

$$P(x_i = r) = \beta_0 + \mathbf{x}'_p \boldsymbol{\beta} + \mathbf{f}_q(\mathbf{z}_q) + f_{spat}(s_x, s_y)$$