

Raum-zeitliche Analyse von Kindern mit Leukämie in Deutschland, 1987-2007

Arbeitstagung Bayes-Methodik, räumliche
Statistik, Ökologie und Umwelt

Sven Schmiedel, Peter Kaatsch, Maria
Blettner, Joachim Schüz
Dänische Krebsgesellschaft,
Deutsches Kinderkrebsregister

sven@cancer.dk

Lübeck, 3. Dezember 2009

Problem

- Leukämiefälle bei Kindern
- Daten auf Gemeindeebene
- Hypothese: Es gibt keinen allgemeinen Trend zur Clusterbildung (nicht zu verwechseln mit der Suche nach Clustern)
- Annahme: Es gilt die Poissonverteilung für Anzahl Fälle innerhalb Gemeinden
- Wie ist dieses zu überprüfen?



Daten

International Classification of Childhood Cancer, third version	n	%
I Leukemia, myeloproliferative and myelodysplastic diseases	11,946	100.0
I.a Lymphoid leukemia	9,638	80.7
I.b Acute myeloid leukemia	1,642	13.7
I.c Chronic myeloproliferative diseases	145	1.2
I.d Myelodysplastic syndrome and other myeloproliferative diseases	417	3.5
I.e Unspecified and other specified leukaemia	104	0.9

- Alle Kinder < 15 Jahren 1987-2007 in
12.262 Gemeinden, bewohnte Fläche:
353.044km²

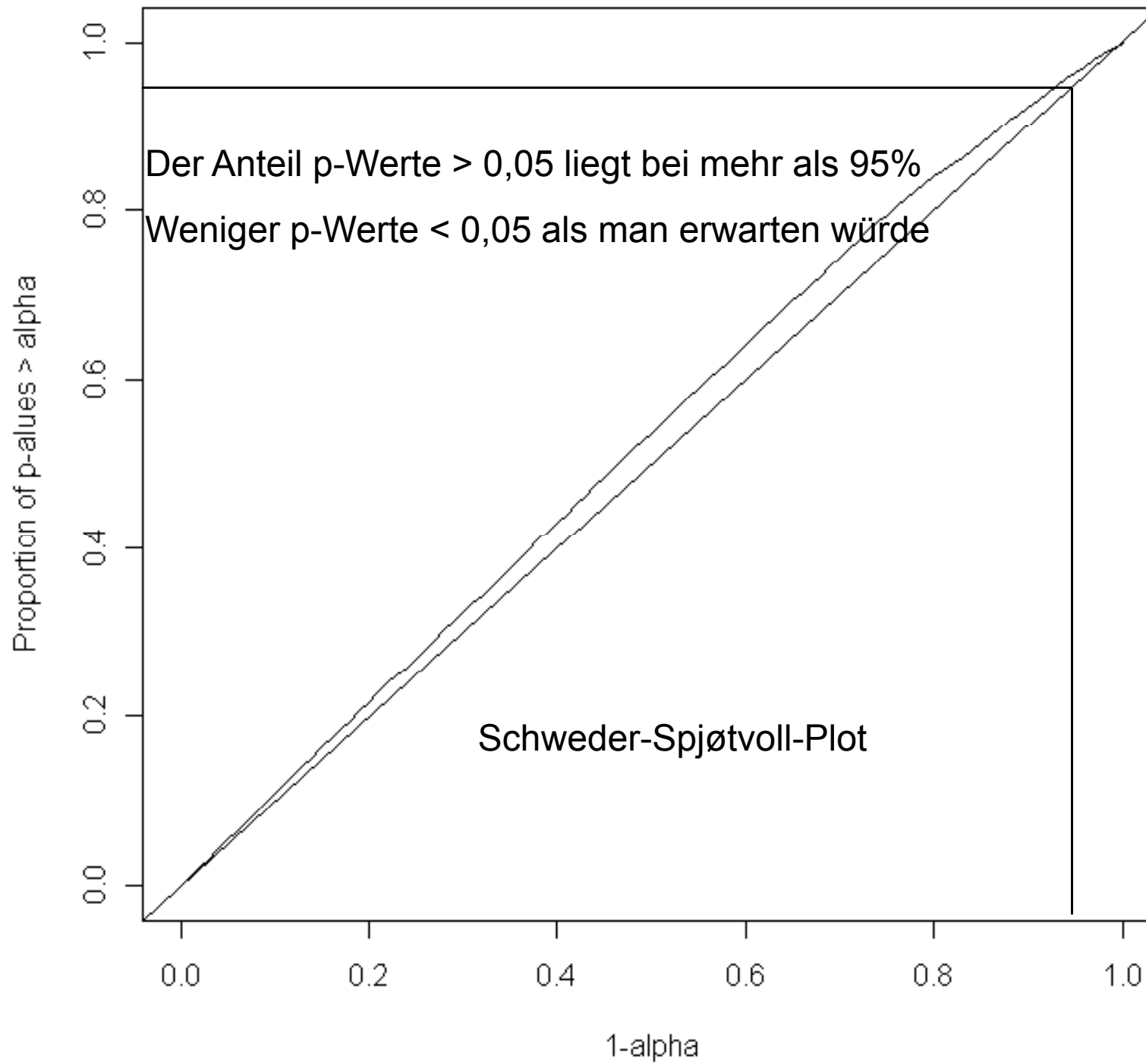
Verteilung (H_0)

- Poisson-Verteilung

$$P_{\lambda}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \text{ hier}$$

$$P_{X_i}(O_i = o) = \frac{X_i^o}{o!} e^{-X_i}, \text{ wobei } O_i \text{ beobachtete}$$

Fälle in Gemeinde $i=1, \dots, n$



Verteilung (H_0)

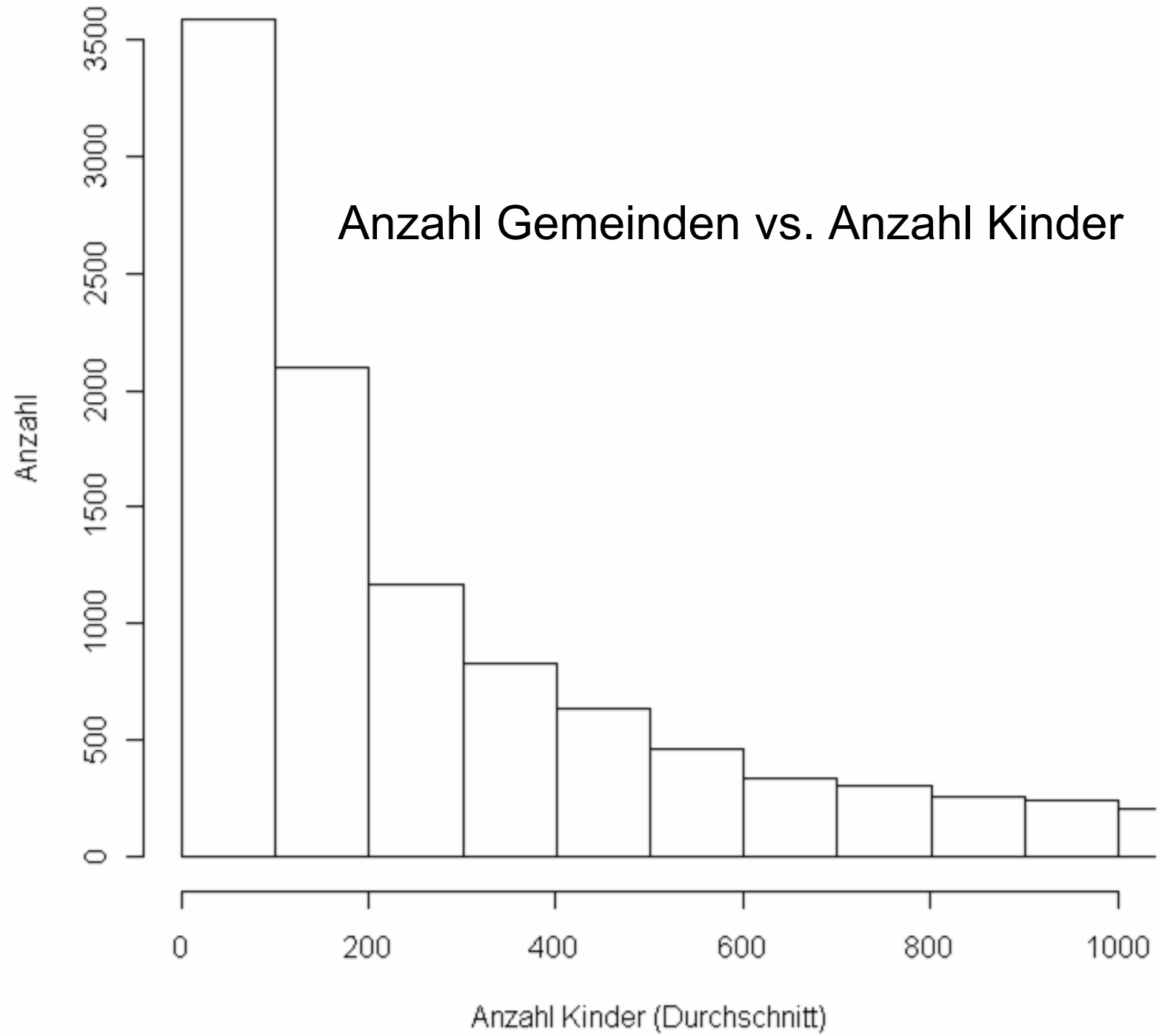
- Voraussetzungen der Poissonverteilung:

$$\text{Var}(O_i) = E(O_i)$$

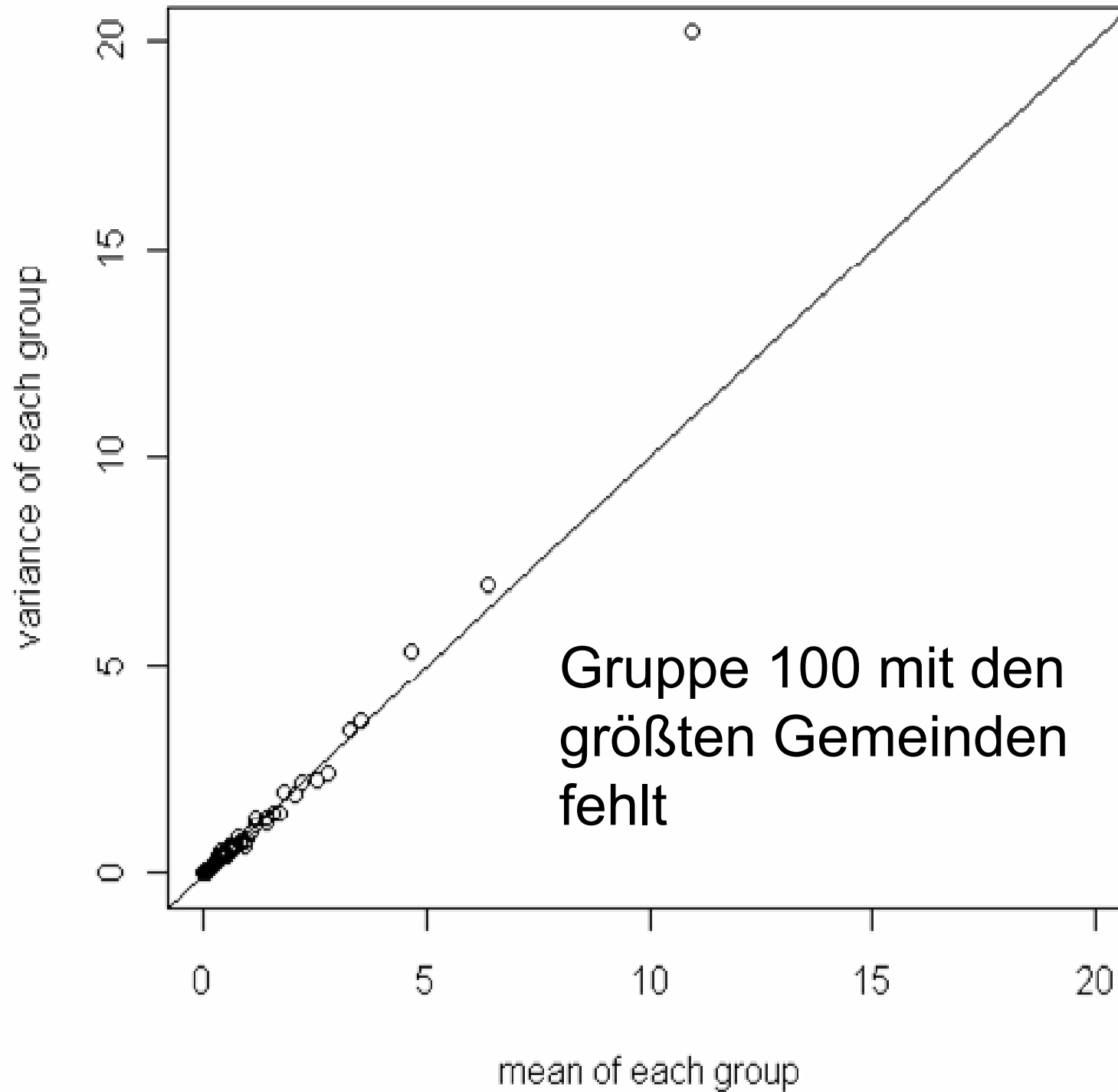
- Wenn $\text{Var}(O_i) > E(O_i)$ spricht man von Überdispersion
- Hinweis auf allgemeinen Trend zur Clusterbildung (oder auch Erhebungsfehler, fehlende Kovariable)

Verteilung (H_0)

- Vergleich empirischer Erwartungswert mit empirischer Varianz innerhalb der Gemeinden
 - Varianz: 32,22
 - Erwartungswert: 0,96
- Problem: Natürlich ist Erwartungswert und Varianz der beobachteten Fälle abhängig von der Anzahl der Einwohner



- Lösung: Einteilung der über 12.000 Gemeinden in 100 Gruppen, um erwartete Anzahl der Fälle innerhalb der Gruppen auf gleichem Niveau zu halten



Teststatistik

- Nutzung der Statistik (Westermeier und Michaelis, 1994)

$$Q = \frac{\frac{1}{n-1} \sum_{i=1}^n (o_i - \bar{o})^2}{\bar{o}}$$

- Nach RA Fisher (1964)

Ergebnis

- Table 1: Distribution of statistic Q in the 100 groups

Min* (1-99)	Max (1-99)	Group 100	Prop* >1
0.74	1.82	2.73	36/97

3 Gruppen, mit 0 beobachteten Fälle

Problem

- Durch Einteilung in 100 Gruppen wird auch 100 Mal getestet
- Besser wäre Statistik, welche auch Einwohnerzahlen mitberücksichtigt

Artikel

Methods for detecting disease clustering, with consideration of childhood leukaemia

Colin R Muirhead Radiation Protection Division, Health Protection Agency, Chilton, Oxon, UK

Statistical Methods in Medical Research 2006; **15**: 363–383

Modell Potthoff-Whittinghill

- $O_i | X_i \sim \text{Poi}(X_i)$
- $X_i \sim \text{Gamma}(\mu_i = P_i \alpha, \sigma_i^2 = \beta \mu_i)$
- P_i Personenjahre oder Anzahl Personen
- $O_i \sim \text{NBin}(\mu_i, (1 + \beta) \mu_i)$
- Gemeinsame Verteilung von
 $O_1, \dots, O_n | O_T \sim$ gemischte Dirichlet-
multinomial Verteilung ($\sum O_i = O_T$)
- Für $\beta \rightarrow 0$ Multinomialverteilung

Statistik nach Potthoff-Whittinghill

- Unter der Nullhypothese ($\beta = 0$) sind O_i poissonverteilt und die gemeinsame Verteilung ist multinomialverteilt
- Potthoff-Whittinghill (1966) zeigten, dass der Score Test $\beta = 0$ vs. $\beta > 0$ auf folgender Statistik basiert:

$$S = \frac{1}{2} \left[\left(\sum_{i=1}^n \frac{O_i(O_i - 1)}{\boxed{E_i}} \right) - (O_T - 1) \right]$$

Schätzung von β

- Über die log likelihood der gemischten Dirichlet multinomial Verteilung

$$l(\beta) = \sum_{i=1}^n \sum_{j=1}^{O_i} \ln(E_i + \beta(j-1)) - \sum_{k=1}^{O_T} \ln(O_T + \beta(k-1))$$

Vergleich mit anderen Tests

- Wenn Personen(jahre) gleichverteilt entspricht der Test dem Pearson χ^2 -Test
- S ist im Falle von wenigen Personen(jahren) in Gebieten effizienter
- Vergleich mit anderen Tests mittels Simulationen unter verschiedenen Voraussetzungen der Verteilung der Bevölkerung hat gezeigt, dass S robust ist

Vorteile

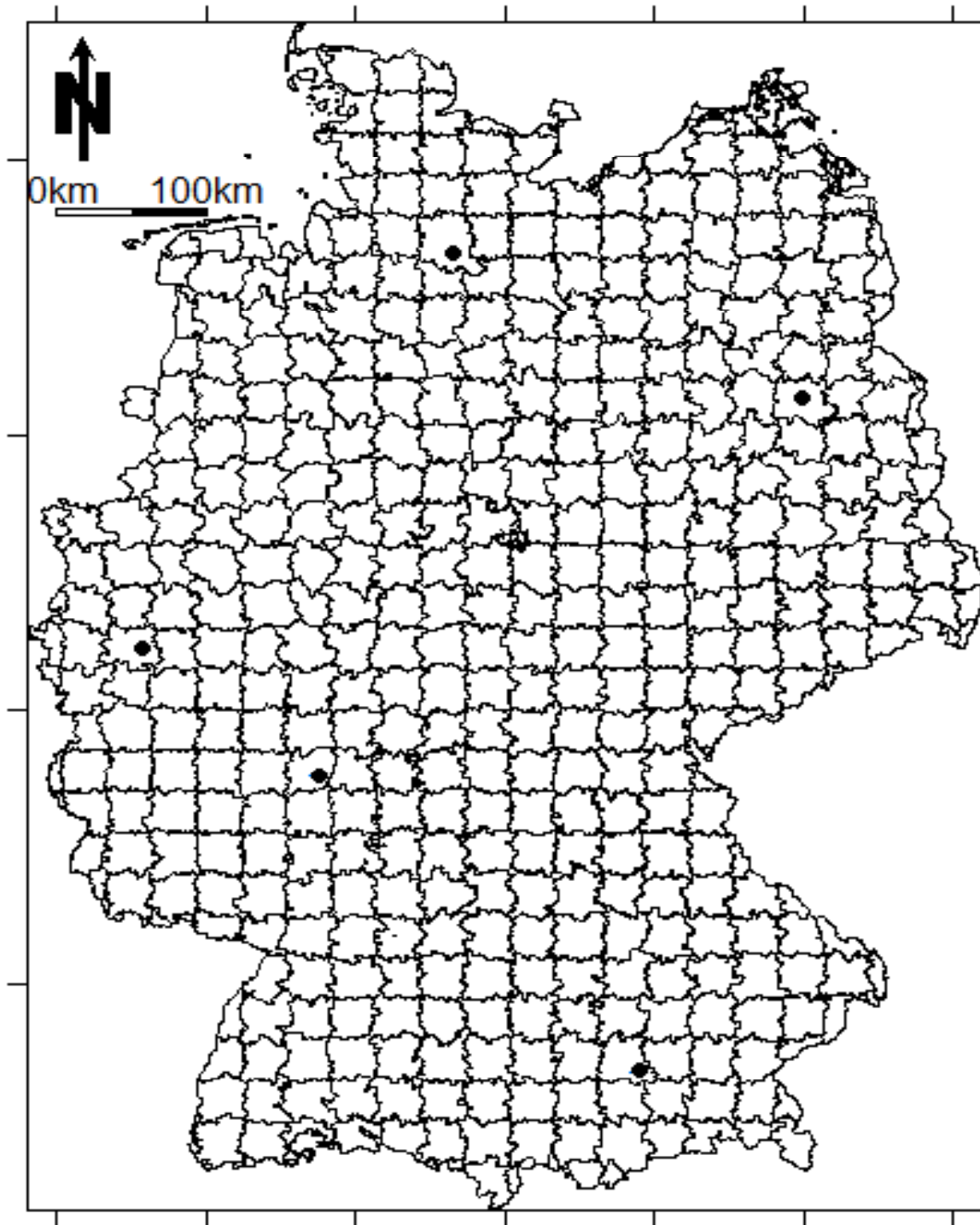
- Hat sich als robust gegenüber verschiedenen Annahmen zur Bevölkerungsverteilung gezeigt
- Berücksichtigung der Bevölkerung der Gemeinden
- Ein Parameter welcher die Überdispersion in interpretierbarer Weise angibt
- Vergleichbarkeit mit Literatur
- Varianz in Komponenten zerlegbar (nicht auf räumliche Struktur beschränkt)

Nachteile

- Orte mit 0 oder 1 Fall werden nicht in Statistik mitberücksichtigt
- Räumlicher Bezug der Daten bleibt unberücksichtigt



- Mit 20km² Raster überdecken
- zusammenlegen der Gemeinden, die ihr Zentrum innerhalb eines Rasters haben



Spatial Scan Statistic

- Kreise mit Zentrum jeder Gemeinde werden gebildet
- Unter H_0 ist das Risiko innerhalb der Kreise so groß wie außerhalb
- Berechnung der Likelihood Funktion für jeden Kreis
- Verwendung des Kreises mit größter Likelihood als „wahrscheinlichstes“ Cluster
- Monte-Carlo Simulation der Verteilung unter H_0

Poisson Likelihood

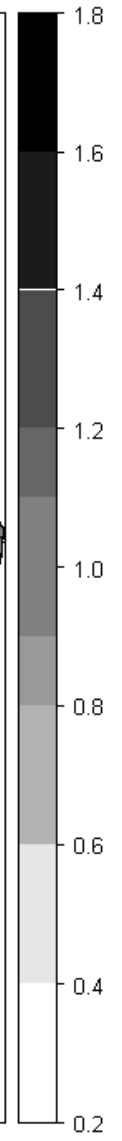
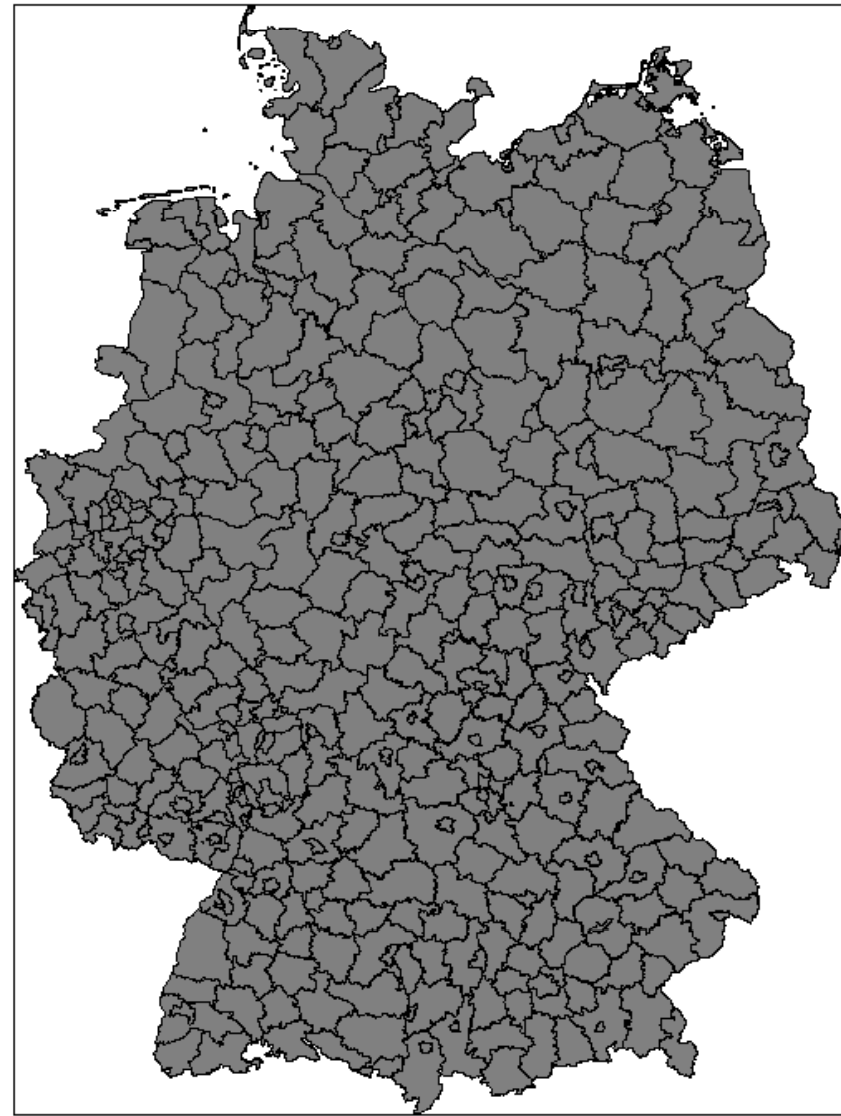
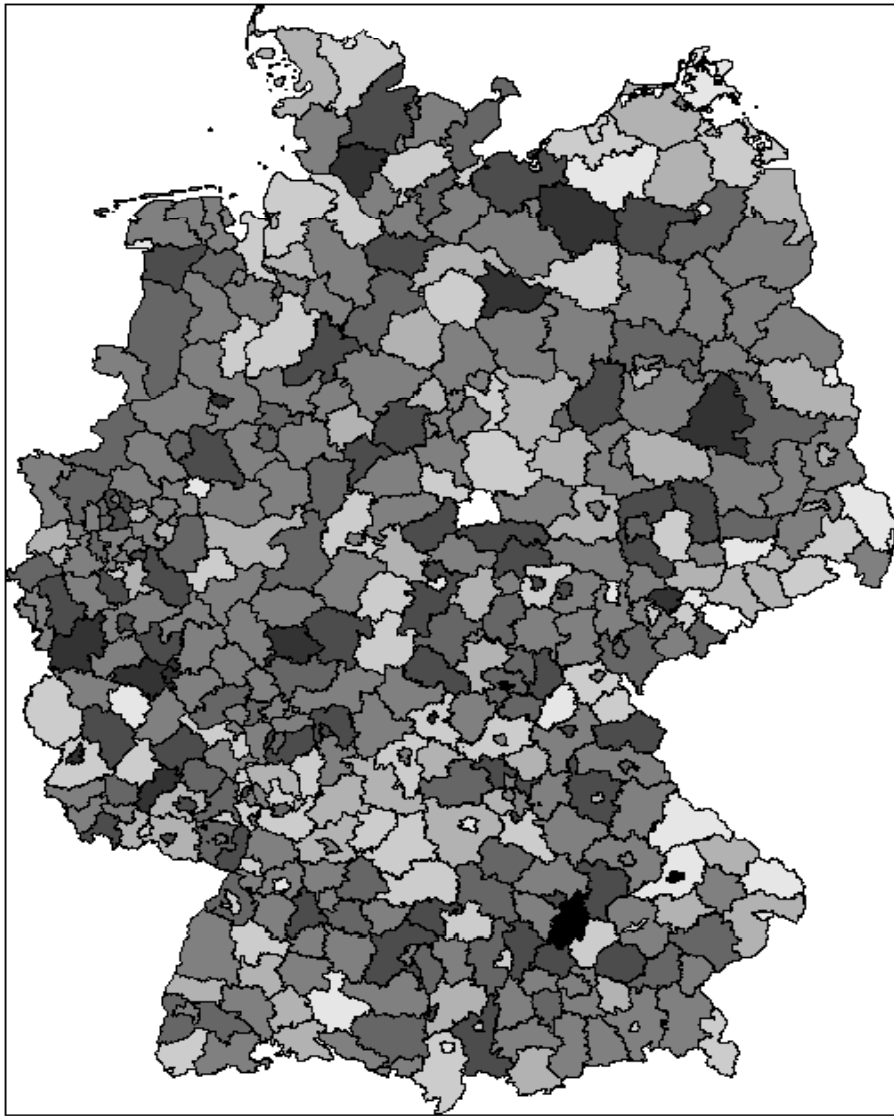
- $[c/\mu]^c \times [(C-c)/(C-\mu)]^{C-c}$
- c = Anzahl Fälle im Kreis
- μ = Erwartete Fälle im Kreis
- C = Gesamtanzahl Fälle

Anwendung

- Keine auffälligen p-Werte

Besag Modell

- Einbeziehung der räumlichen Struktur in das Modell:
Sei $o_i \sim \text{Poisson}(e_i \cdot \theta_i)$ mit $\log(\theta_i) = \mu + u_i + v_i$
- μ : Gesamtlevel
- v_i : Komponente ohne Struktur
- u_i : Komponente unter Einbeziehung räumlicher Struktur
- Gaussian Markov random field Modell für u_i :
 - $u_i | u_{j \neq i} \sim N(\bar{u}_i, \sigma_u^2 / m_i)$
 - \bar{u}_i : Mittelwert der Nachbarn
 - m_i : Anzahl Nachbarn
- Unabhängige zufällige Effekte $v_i \sim N(0, \sigma_v^2)$
- Unbekannte Hyperparameter σ_u^2 and $\sigma_v^2 \rightarrow$ hyperpriors



Literatur

- Muirhead CR. 2006. Methods for detecting disease clustering, with consideration of childhood leukaemia, *Stat Methods Med Res*, 15: 363-383.
- Westermeier T, Michaelis J. 1995. Applicability of the Poisson distribution to model the data of the German Children's Cancer Registry, *Radiat Environ Biophys*, 34: 7-11.
- Fisher RA. 1964. The significance of deviations from expectation in a Poisson series, *Biometrics*, 20: 265-272.
- Potthoff RF, Whittinghill M. 1966. Testing for homogeneity, *Biometrika*, 53: 167-190.