Hinweise

Seminar zu aktuellen Themen der ökologischen Modellbildung (M 409):

„Der Klimawandel: wissenschaftliche Erkenntnis, gesellschaftliche Diskussion, politische Bedeutung"

Anmeldungen bis 15.11.05

Seminar: Forschungswerkstatt „Modellbildung und Simulation"

Wallenfels 20.1. – 22.1.05, Vorbespr. 24.11.05, 14:00 Uhr

Diplomarbeit: „Entwicklung multivariater Indikatoren der Grundwasserbeschaffenheit in Nordbayern"

in Kooperation mit dem *Bayerischen Landesamt für Umwelt* (vorher: *Bayerisches Geologisches Landesamt*)

Tests und Trenderkennung

(Oft) gewünschte oder geforderte Eigenschaften von Zeitreihen:

Eigenschaft	**Anmerkungen**	**Test**
Normalität	gegeb. Transformation der Daten	Kolmogorov-Smirnov, Shapiro-Wilks, χ^2-Test
Ergodizität	pragmatisch: Sättigung der Statistik	*für empirische Daten nicht testbar*
Linearität		*lineare Modelle*
Stationarität = Trendfreiheit	• Stationarität des Mittelwertes („Trend" i.e.S.) • Stationarität der Varianz (Homoskedastizität) • allgemein: Konstanz aller Momente	Mann-Kendall-Test Bartlett-Test, Levene-Test Witt-Test

Mann-Kendall-Test

Ziel: Überprüfung auf Trend des Mittelwerts

Motivation: viele statistische Verfahren setzen Trendfreiheit voraus

Prinzip: Vergleich der Anzahl der Konkordanzen ($x(t_1) > x(t_2)$ für $t_1 > t_2$) und der Diskordanzen ($x(t_1) < x(t_2)$ für $t_1 > t_2$)

Erweiterung: Seasonal Kendall Test (Trendanalyse getrennt z.B. für einzelne Jahreszeiten)

Beziehungen zu anderen Verfahren: Trendfreiheit ist Voraussetzung für viele andere Verfahren

Kendall-Test

= nicht-parametrischer (verteilungsfreier) Test für bivariate Korrelationen

Prinzip: Vergleich der Anzahl der Konkordanzen $\boldsymbol{n_c}$ ($i < j \Rightarrow x_i < y_j$) bzw. der Diskordanzen $\boldsymbol{n_d}$, bzw. der Differenz $S = n_c - n_d$:

$$S = n_c - n_d = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \operatorname{sgn}(x_k - y_j) \qquad \operatorname{sgn}(x) = \begin{cases} +1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

Kendalls $\boldsymbol{\tau}$:
$$\tau = \frac{n_c - n_d}{n(n-1)/2} = \frac{S}{n(n-1)/2}$$

mit $-1 <= \tau <= 1$

$n(n-1)/2$ = Anzahl der Vergleiche

Korrektur für verbundene Ränge

verbundene Ränge (***ties***): identische Werte, die dem gleichen Rang zugeordnet werden

=> erfordern Korrektur für Kendalls $\boldsymbol{\tau}$ - (statt $\tau = \frac{n_c - n_d}{n(n-1)/2}$):

$$\tau = \frac{n_c - n_d}{\sqrt{\left[n(n-1)/2 - \sum_{i=1}^{n_{vRx}} n_{vRx\,i}(n_{vRx\,i}-1)/2 \right] \cdot \left[n(n-1)/2 - \sum_{i=1}^{n_{vRy}} n_{vRy\,i}(n_{vRy\,i}-1)/2 \right]}}$$

mit n_{vRx} , n_{vRy}: Gesamtzahl aller verbundenen Ränge für x bzw. y

und $n_{vRx\,i}$, $n_{vRy\,i}$: Zahl der i-ten verbundenen Ränge

Der Mann-Kendall Test

Anwendung des Kendall-Tests auf Zeitreihen (d.h., sortiert nach Zeit, ohne doppelte Einträge) => Trendtest:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \mathrm{sgn}(x(t_k) - x(t_j)) \qquad \mathrm{sgn}(y) = \begin{cases} +1 & y > 0 \\ 0 & y = 0 \\ -1 & y < 0 \end{cases}$$

Für die H_0-Hypothese (= "es gibt keinen Trend") gilt dann:

$$E(S) = \quad \bar{S} = 0$$

$$\mathrm{var}(S) = \quad \sigma^2(S) = \frac{n(n-1)\cdot(2n+5)}{18}$$

=> normalverteilt

=> Ableitung der Testgröße: Abweichung der beobachteten (normierten) S von den laut $\mathbf{H_0}$ erwarteten

Der Mann-Kendall Test

beachte: Korrektur für verbundene Ränge (Ranggleichheit) notwendig

=> statt

$$\operatorname{var}(S)=\sigma^2(S)=\frac{n(n-1)\cdot(2n+5)}{18}$$

$$\operatorname{var}(S)=\sigma^2(S)=\frac{n(n-1)\cdot(2n+5)-\boxed{\sum_{j=1}^{p} n_{vR\,j}(n_{vR\,j}-1)\cdot(2n_{vR\,j}+5)}}{18}$$

$n_{vR\,j}$ = Anzahl der Werte jeweils gleicher Ränge,

p = Anzahl der Gruppen verbundener Ränge

=> Teststatistik: $\tau=\frac{S}{D}$ (*für große n annähernd normalverteilt*),

wobei D = maximal mögliche Anzahl der Konkordanzen:

$$D=\sqrt{\frac{1}{2}(n(n-1)-\frac{1}{2}\sum_{j=1}^{p} n_{vRj}(n_{vRj}-1)} \cdot\sqrt{\frac{1}{2}n(n-1)}$$

Saisonaler Mann-Kendall Test

n Beobachtungen pro Saisonteil (z.B. fester Tag im Jahr),
m Saisonteile pro Saison (z.B. 365 Tage/Jahr)

x_{ig} *i*-te Beobachtung im *g*-ten Saisonteil

$$S_g=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\operatorname{sgn}(x_{jg}-x_{ig}),\ \sigma_g=\sqrt{n(n-1)(2n+5)/18}$$

$$S=\sum_{g=1}^{m} S_g,\quad \sigma_S^2=\sum_{g=1}^{m}\sigma_g^2+\sum_{\substack{g,h\\ g\neq h}}\operatorname{cov}(S_g S_h)$$

- Entmaskierung von Gesamttrends
- Trends in einzelnen Saisonteilen (z.B. Monaten)

Trendbeseitigung

Linearer Trend des Mittelwertes:

- $\tilde{x}(t) = x(t) - x'(t)$ mit der Regressionsgeraden $x'(t) = a + b \cdot x(t)$

Saisonaler Trend des Mittelwerts:

- stufenweise: $\tilde{x}(t) = x(t) - \bar{x}_j$ mit $j = 1, \ldots, p$ Saisons

- $\tilde{x}(t) = x(t) - x'(t)$ mit der Regression $x'(t) = \sum_{j=1}^{n_f} a_j \cdot \sin\left(b_j \cdot \left[x(t) - c_j\right]\right)$
  für n_f ausgewählte Frequenzen

Tests auf Homoskedastizität

Eine Zeitreihe heißt homo-(hetero-)skedastisch, wenn ihre Varianz fensterweise konstant (variabel) ist.

Übliche Tests:

- Bartlett-Test
- Levene-Test
- ...

Bartlett-Test

- Bartlett (1937)
- Voraussetzung: x ist (annähernd) normalverteilt
- Die Zeitreihe wird in p Fenster aufgeteilt mit jeweils n_i Werten
- Bartlett-Testgröße $$T = \frac{(n-p)\cdot \ln \sigma_p^2 - \sum_{i=1}^{p}(n_i - 1)\cdot \ln \sigma_i^2}{1 + \frac{1}{3(p-1)}\cdot\left(\sum_{i=1}^{p}\frac{1}{n_i - 1} - \frac{1}{n-p}\right)}$$

  mit $\sigma_p^2 = \frac{1}{n-p}\cdot \sum_{i=1}^{p}(n_i - 1)\cdot \sigma_i^2$

- T ist χ^2-verteilt mit p-1 Freiheitsgraden

Levene-Test

- Levene (1960)
- weniger empfindlich gegen Verstoß gegen die Annahme der Normalverteilung
- Die Zeitreihe wird in p Fenster aufgeteilt mit jeweils n_i Werten
- Levene-Testgröße $$W = \frac{(n-p)\cdot \sum_{i=1}^{p} n_i(\bar{z}_{i.} - \bar{z}_{..})^2}{(p-1)\cdot \sum_{i=1}^{p}\sum_{j=1}^{n_i}(z_{ij} - \bar{z}_{i.})^2}$$

  mit $z_{ij} = \left|x_{ij} - \bar{x}_{i.}\right|$

  (robuster: Median statt arithm. Mittel)

- ist F-verteilt mit (p-1, n-p) Freiheitsgraden

Stationaritätstest nach Witt et al. (1998)

Ziel: Überprüfung der Konstanz der Momente (Verteilung) innerhalb der Zeitreihe

Motivation: Voraussetzung vieler statistischer Verfahren

Prinzip: Vergleich der Histogramme für Fenster bestimmter Länge

Voraussetzungen: keine

Beziehungen zu anderen Verfahren: Stationarität ist Voraussetzung für viele andere Verfahren

Stationaritätstest nach Witt et al. (1998)

- Aufteilung der Zeitreihe in p Fenster der jeweils gleichen Länge l

  $x_\tau^j = x_{(j-1)l+\tau},\ j = 1,\dots,n;\quad \tau = 1,\dots,l$

- Histogramm der Werte in jedem Fenster mit r Klassen (*bins*). Werte des i-ten Fensters in der k-ten Klasse:

  $\left\{ x^i_{\rho_1}, x^i_{\rho_2}, \dots, x^i_{\rho_{R_k^i}} \right\}$ R_k^i : Anzahl der Elemente der Klasse (Bedingung: $R_k^i > 20$ in allen Bins)

- Teststatistik für p Fenster

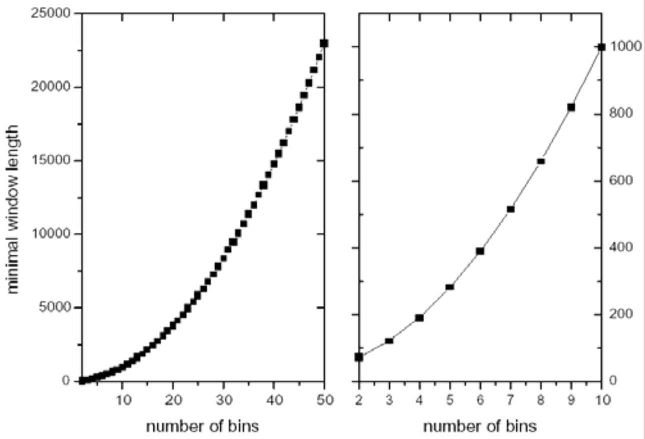
$$t_{A,p} = p^2 l^2 \sum_{k=1}^{r} \frac{\sum_{i=1}^{p} (R_k^i - R_k / p)^2}{\sum_{j=1}^{p} \sigma^2 (R_k^j)} \qquad \left(R_k = \sum_{i=1}^{p} R_k^i\right)$$

ist asymptotisch χ^2-verteilt mit $(r-1)(p-1)$ Freiheitsgraden

Erforderliche Datenmenge

Für eine verlässliche Abschätzung der χ^2-Verteilung

bei *p* Klassen gilt für die minimale Fensterlänge: $l > \frac{9 \cdot p^3}{p-1}$



Aufgabe

1. Bestimmen Sie im gegebenen Datensatz die Trends der Mittelwerte jeweils durch Regression der Messwerte mit der Zeitachse.

2. Führen Sie für jeden Parameter eine Trendanalyse mittels des Kendall-Tests durch.

3. Teilen Sie den Datensatz in 5 Fenster gleicher Länge, erstellen für einen Parameter in jedem Fenster ein Histogramm oder die kumulative Häufigkeitsverteilung, und stellen die Ergebnisse in einer Grafik dar.