

Überblick und Ausblick

Letzte Vorlesung Statistik

Prof. van den Boogaart

Vorlesung Datenanalyse und Statistik

Gliederung

- 1 Rückblick
 - Sortiert nach dem Inhalt der Vorlesung
 - Sortiert nach Daten
- 2 Multivariate Statistik
 - Kovarianzmatrizen
 - Clusteranalyse
 - Hauptkomponentenanalyse
 - Faktorenanalyse
 - Diskriminanzanalyse
- 3 Spezielle Daten
 - Zeitreihenanalyse
 - Geostatistik
 - Kompositionsdaten
 - Zusammenfassung Spezielle Daten

Rückblick

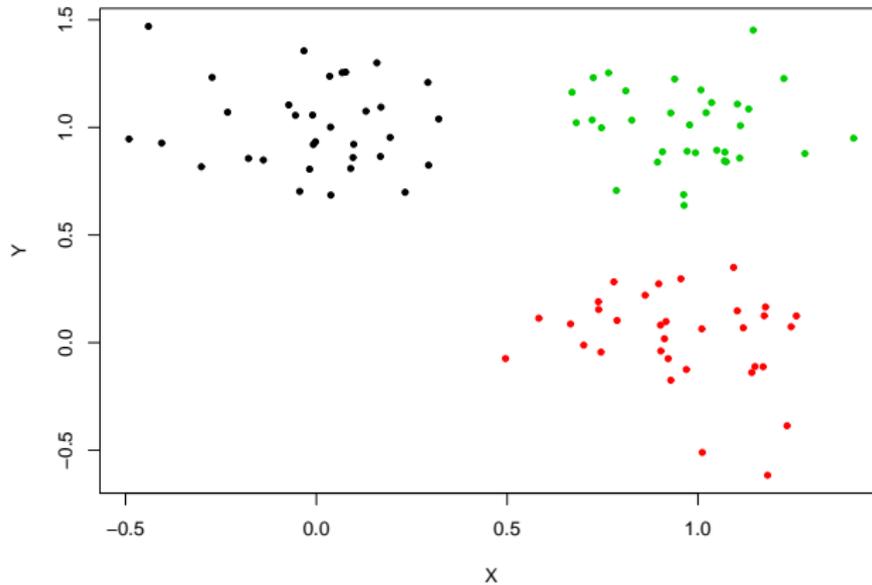
- Daten:
- Statistische Graphik
- Deskriptive Statistik
- Statistische Tests
 - Hypothese, Alternative, Fehler 1. und 2. Art.
 - Nachweis, α -Niveau, p -Werte
 - Bonferroni Korrektur
 - Auswahl der Tests
 - parametrisch, nichtparametrisch, robust
- Lineare Modelle

- **Univariate Statistik**
- **Bivariate Statistik**
- **Multivariate Statistik**
- **Spezielle Daten**
 - *Zeitreihen*
 - *Zufallsfelder*
 - *Kompositionsdaten*
 - ...

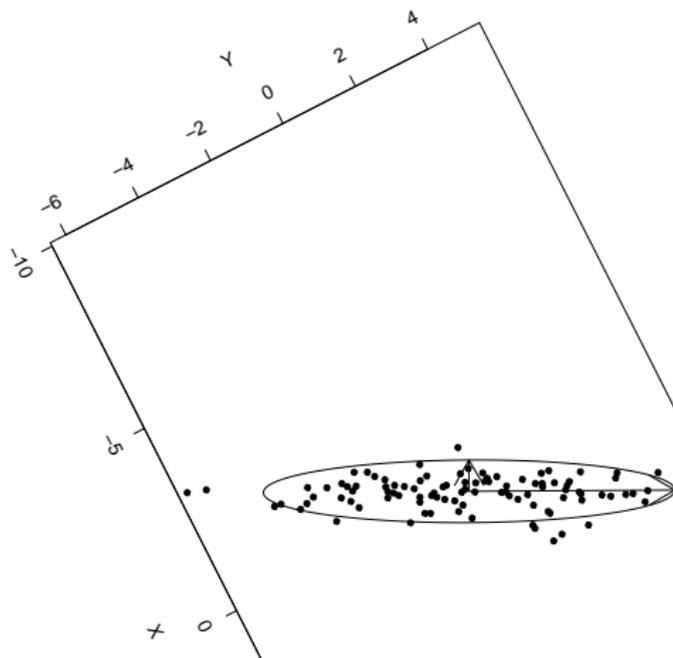
Kovarianzmatrizen

$$\begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \text{cov}(X_1, X_3) & \text{cov}(X_1, X_4) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & \text{cov}(X_2, X_3) & \text{cov}(X_2, X_4) \\ \text{cov}(X_3, X_1) & \text{cov}(X_3, X_2) & \text{var}(X_3) & \text{cov}(X_3, X_4) \\ \text{cov}(X_4, X_1) & \text{cov}(X_4, X_2) & \text{cov}(X_4, X_3) & \text{var}(X_4) \end{pmatrix}$$

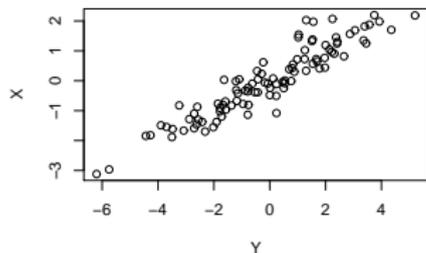
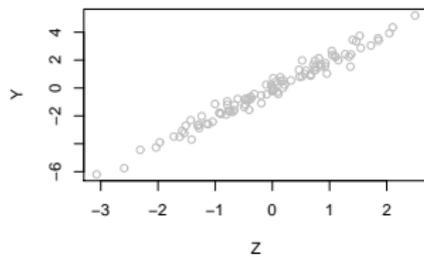
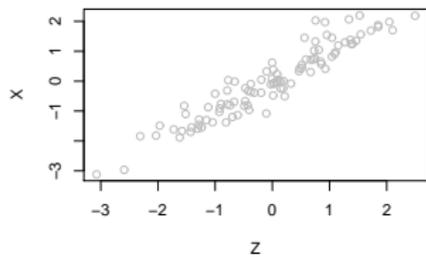
Klusteranalyse II



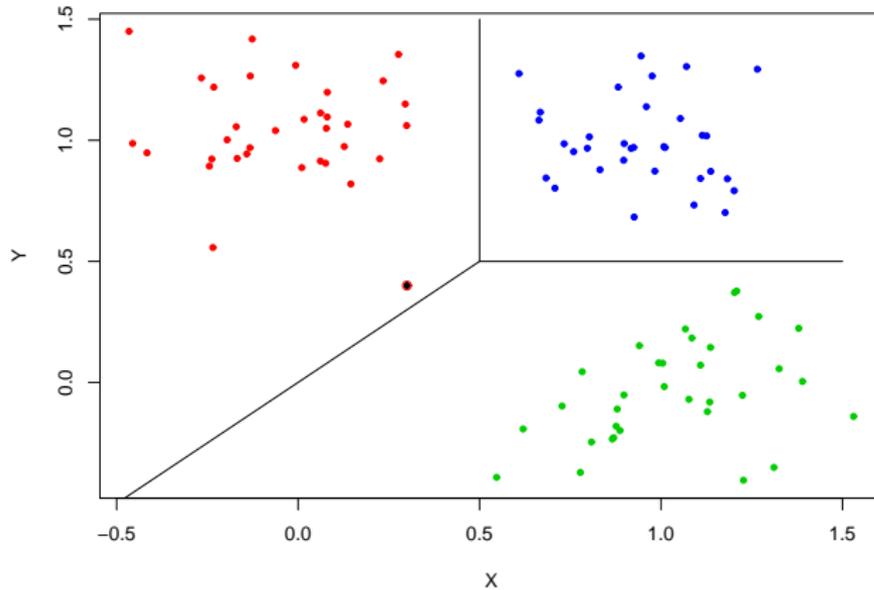
Hauptkomponentenanalyse IV



Faktorenanalyse



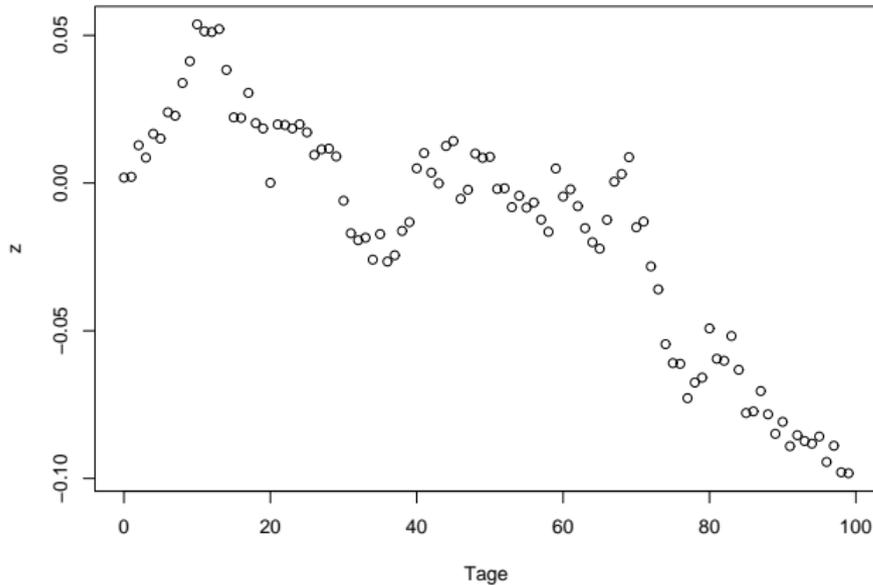
Diskriminanzanalyse



Zusammenfassung MV-Statistik

- Die **Klusteranalyse** versucht Gruppen in den Daten zu finden.
- Die **Hauptkomponentenanalyse** versucht die Hauptrichtung der Streuung im Datensatz aufzufinden.
- Die **Faktorenanalyse** versucht unbeobachtbare gemeinsame Ursachen in den Daten zu entdecken.
- Die **Diskriminanzanalyse** versucht Individuen einer Gruppe zuzuordnen.

Zeitreihen



Eigenschaften von Zeitreihen

- Daten werden in regelmäßigen Zeitabständen erhoben
- In kurzer Zeit ändert sich nicht so viel
- Daten sind also stochastisch abhängig
- Daten verhalten sich oft zyklisch (z.B. immer große Werte im Sommer)

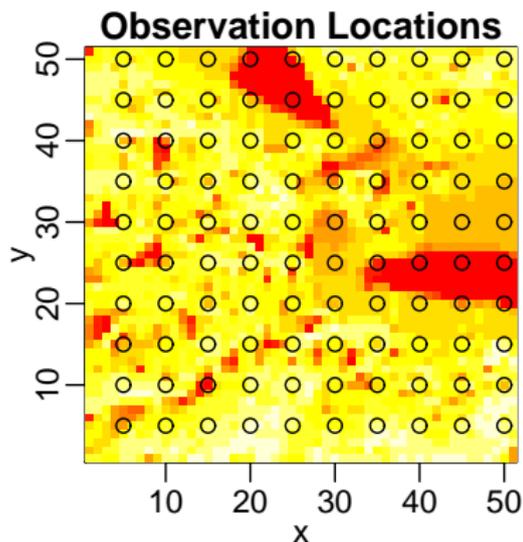
Methoden für Zeitreihen

- Test ob Abhängigkeit tatsächlich vorliegt (z.B. Ansari-Friedmann)
- Abhängigkeit quantifizieren durch Autokovarianzfunktion:

$$c(h) = \text{cov}(Z(t+h), Z(t))$$

- Vorhersage: Wie geht es weiter?
- Welche Gesetze verbergen sich hinter der Zeitreihe?

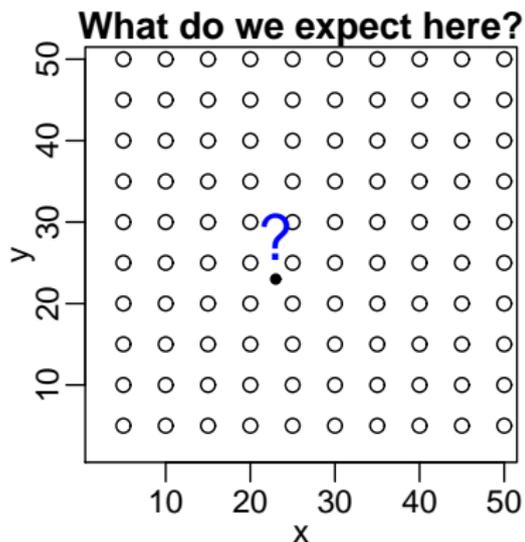
Was ist eine geostatistische Vorhersage?



$$\hat{Z}(s) = \int_A f \left(\sum_{i=1}^n w_i(s) Z(s_i), i = 1, \dots, n \right) ds$$

$Z(s_i)$ = True value at location s_i

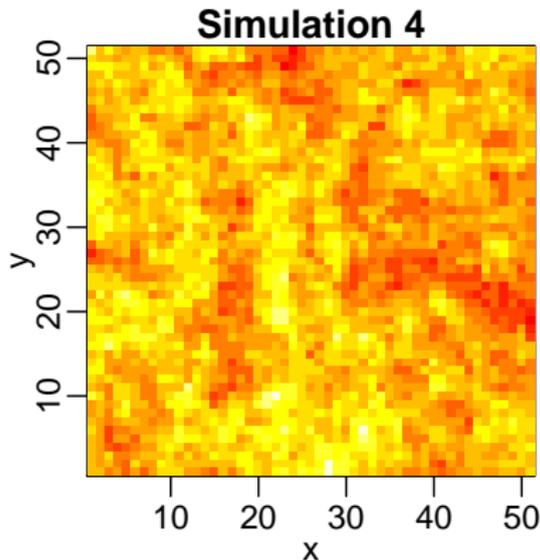
Was ist eine geostatistische Vorhersage?



$$\hat{Z}(s) = \int_A f \left(\sum_{i=1}^n w_i(s) Z(s_i), i = 1, \dots, n \right) ds$$

$w_i(s)$ = Kriging weights

What ist geostatistische Simulation?

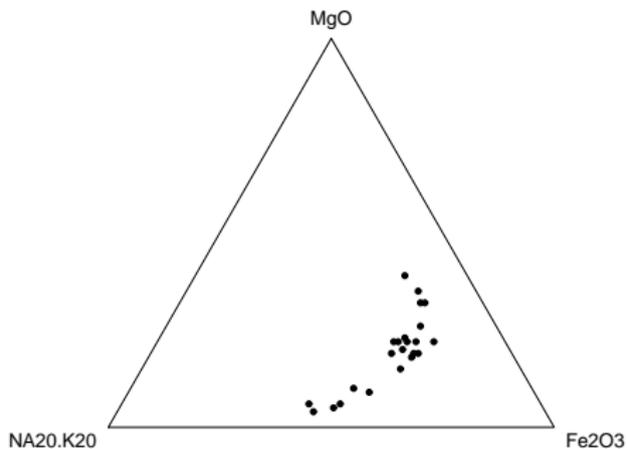


- reality is unknown
- prediction is to smooth
- predict conditional distribution
- simulating possible szenarios
- ... many possible szenaries
- are qualitatively similar
- *but qualitatively different from reality

Kompositionsdaten

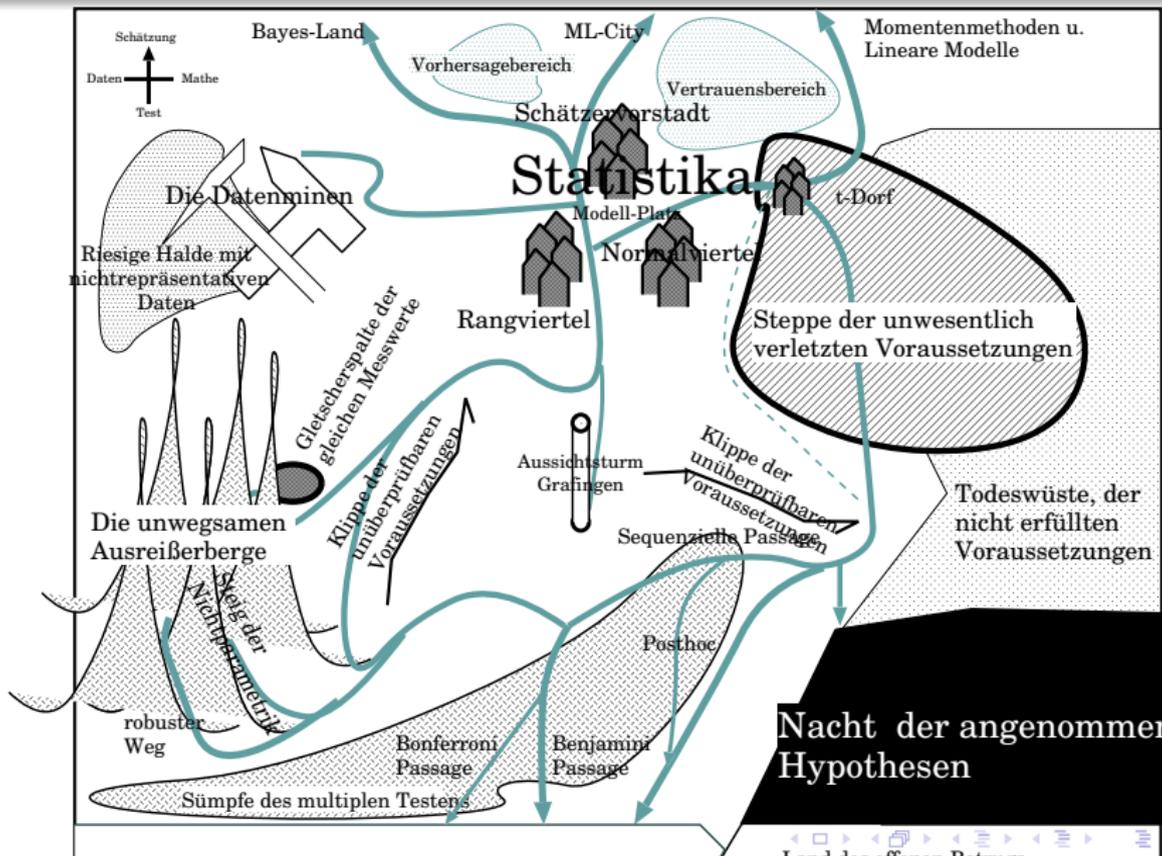
- Komponenten addieren zu 100%
- Komponenten sind also abhängig und automatisch negativ korreliert.
- Komponenten sind einzeln Anteile, aber hängen zusammen.

Ternäre Diagramme



Zusammenfassung SD

- **Zeitreihenanalyse**
Daten mit zeitlichen Abhängigkeiten
- **Geostatistik**
Daten mit räumlichen Abhängigkeiten
- **Kompositionsdaten**
Werte addieren zu 1 bzw. Summe egal.



Ich wünsche Ihnen

eine bestandene Klausur

schöne Ferien!

ein erfolgreiches Leben