

Statistische Methoden der Datenanalyse

Professor Markus Schumacher

Freiburg / Sommersemester 2009

- Motivation
- Syllabus
- Informationen zur Vorlesung
- Literatur
- Organisation der Übungen und Scheinkriterium

- Deskriptive Statistik (Erinnerung)

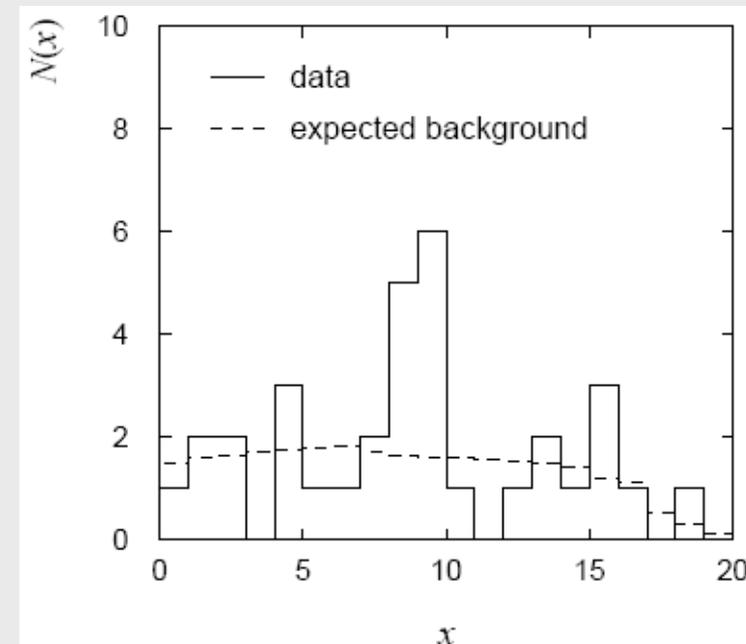
Motivation

- Ziel der Physik:
Erkennen der grundlegenden Gesetze, die das Verhalten und Struktur der materiellen Welt bestimmen
- Wie? durch Wechselspiel zwischen
Theorie (Modellbildung) und
Experiment (Messungen u. Datenbeschreibung)
- Verbindung: durch quantitative Datenanalyse mit statistischen Methoden
- Modellbildung in der Theorie:
 - Objekte die beschrieben werden (bekannte und postulierte)
Gibt es weitere unbekannte Objekte?
 - Gesetzmäßigkeiten/Verteilungen abhängig von Parametern
Stimmen die Gesetze? Welche Werte haben die Parameter?

Aufgaben der stat. Datenanalyse

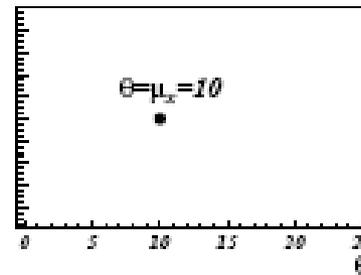
- Darstellung und Beschreibung der Daten
- Bestimmung des „besten“ Wertes für einen unbekannt Parameter
- Bestimmung eines Intervalls oder von Grenzen innerhalb der ein unbekannter Parameter mit einer gewisser Konfidenz liegen sollte
- Quantifizierung der Übereinstimmung zwischen Messdaten und Modell (zwischen verschiedenen Messdatensätzen)
- Vergleich mehrerer Hypothesen bzgl. Übereinstimmung mit den Daten
- Entscheiden auf der Grundlage von Daten

- richtige und wichtige Fragen stellen
- Ergebnisse kritisch hinterfragen
- bei Interpretation Annahmen und Methode klar erklären

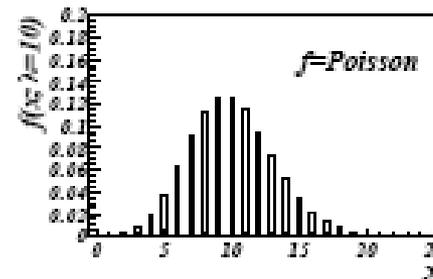


Bsp.: Theorie = Poissonverteilung mit Mittelwert 10

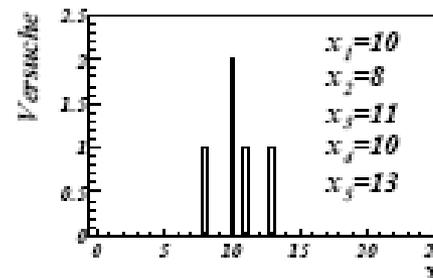
Modell
mit Parameter θ



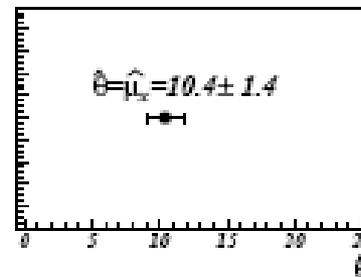
Observable x
mit Verteilung $f(x; \theta)$



Stichprobe
 X_1, \dots, X_N



Schätzwert $\hat{\theta}$
Fehler von $\hat{\theta}$



Syllabus (vorläufig)

- Grundlegende Konzepte
- Ausgewählte Wahrscheinlichkeitsverteilungen
- Die Monte-Carlo-Methode
- Schätzung von Parametern
 - grundlegende Prinzipien und Eigenschaften
 - Maximum-Likelihood -Methode
 - Methode der kleinsten Quadrate
- Hypothesentest
 - grundlegendes Konzept
 - verschiedene Testmethoden
 - Klassifizierung von Ereignissen
- Vertrauensintervalle, Ausschlussgrenzen

Informationen zur Vorlesung

- Wahlpflichtfach II → 3 SWS Vorlesung und 2 SWS Übung
- Vorlesungszeiten:
 - jede Woche Montags 11 bis 13 Uhr (13 Vorlesungen)
April 20.,27. Mai 4.,11.,18,25.
Juni 8.,15.,22.,29. Juli 6.,13.,20.
 - jede 2. Woche Freitags 11 bis 13 Uhr (7 Vorlesungen)
April 24., Mai 8. ,22. , Juni 12. 26., Juli 20. 24.
- Medien: hauptsächlich Tafel unterstützt durch Folien
- Webseite: <http://terascale.physik.uni-freiburg.de/?Lehre/>

- **Statistical Data Analysis**

Autor: Glen D. Cowan

Taschenbuch, Englisch, 216 pages

Oxford University Press, 1998, ISBN-10: 0198501552, ISBN-13: 978-0198501558

Preis: 30,99 Euro

see also www.pp.rhul.ac.uk/~cowan/sda

- **Statistics: A Guide to the Use of Statistical Methods
in the Physical Sciences**

Autor: Roger J. Barlow

Taschenbuch, Englisch, 222 pages

Wiley VCH, 1999, ISBN-10: 0471922951, ISBN-13: 978-0471922957

Preis: 34,99Euro

see also hepwww.ph.man.ac.uk/~roger/book.html

- **Statistische und numerische Methoden der Datenanalyse**

Autoren: Volker Blobel, Erich Lohrmann

Taschenbuch, Deutsch, 358 Seiten

Teubner Verlag, 1998, ISBN-10: 3519032430, ISBN-13: 978-3519032434

Preis: 31,50 Euro

- **Datenanalyse. Mit statistischen Methoden und Computerprogrammen.**
Autor: Siegmund Brandt
Gebundene Ausgabe, Deutsch, 646 Seiten
Spektrum Akad. Verlag, 1999, ISBN-10: 3827401585, ISBN-13: 978-3827401588
Preis: EUR 26,00
- **Probability and Statistics in Particle Physics**
Autoren: G. Frodeson, O. Skjeggstad, H. Tofte,
Gebundene Ausgabe, Englisch, 501 Seiten
Universitetsforlaget, 1979, ISBN-10: 8200019063, ISBN-13: 978-8200019060
leider nicht mehr verfügbar

und viele andere.

Die obigen kenne ich am besten und mag ich am meisten.

Organisation der Übungen und Scheinkriterium

- 2 SWS Übungen jede Woche, 12 Termine
- Termin: Mittwoch 14 s.t. bis 16 s.t. Uhr
- ungerade Woche : Übungsblatt mit Anwesenheits- und Hausaufgaben
 - Anwesenheit: vertiefende einfache Rechnungen
 - Hausaufgaben: Abgabe in der folgenden Woche
- gerade Woche: Übungsblatt nur mit Anwesenheitsaufgaben
 - praktische Übungen am Computer im CIP-Pool (ROOT, C++)
- Übungsleiter: Matthew Beckingham und Markus Warsinsky
- Scheinkriterium:
 - regelmässige und aktive Teilnahme an Übungen
 - erfolgreiche Hausaufgaben (50% der Punkte)
 - erfolgreiche Klausur am Ende des Semesters

Deskriptive Statistik (Erinnerung)

- Messung liefert Stichproben von Messdaten: $x_i \quad i = 1, n$
- Ziel:
 - übersichtliche Darstellung → Graphik
 - Bestimmung weniger charakteristischer Zahlen
- Typen von Daten:
 - numerische:
 - diskrete: abzählbar viele z.B. ganze Zahlen
 - kontinuierliche: meist reelle Zahl
 - qualitative: Farben, Marken, ...

Darstellung

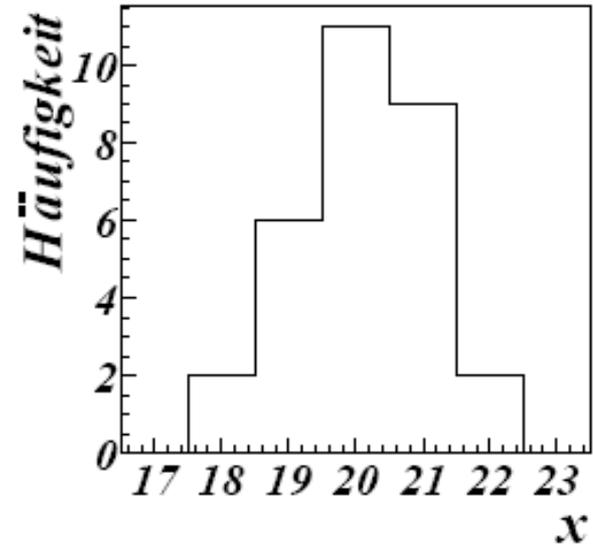
● Tabelle

x[00] = 18.21	x[15] = 20.21
x[01] = 20.63	x[16] = 19.6
x[02] = 19.24	x[17] = 19.69
x[03] = 20.24	x[18] = 19.3
x[04] = 19.21	x[19] = 20.49
x[05] = 21.25	x[20] = 19.62
x[06] = 19.69	x[21] = 20.67
x[07] = 20.73	x[22] = 20.56
x[08] = 20.43	x[23] = 18.62
x[09] = 18.05	x[24] = 20.46
x[10] = 21.52	x[25] = 18.64
x[11] = 21.41	x[26] = 22.22
x[12] = 18.95	x[27] = 21.16
x[13] = 19.83	x[28] = 19.54
x[14] = 20.81	x[29] = 21.27

Balkendiagramm



Histogramm



- in Physik: meist Histogramme auch mehrdimensionale
Wahl des „Binning“: - statistische Fluktuation in Bins klein
- keine Struktur „verstecken“
- andere: Säulendiagramm, Kuchendiagramm

Charakteristische Größen

- „Schwerpunkt“:

meist: arithmetischer Mittelwert

selten: harmonisches, geometrisches Mittel, ...

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

- Streuung:

meist: Varianz = mittlere quadratische Abweichung von arithmetischem Mittel

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)^2$$

teilweise: FWHM volle Breite
auf halber Höhe

selten weitere Größen:

Schiefe, Wölbung (später)

mittlere absolute Abweichung vom Mittelwert)

$$\sigma_S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$