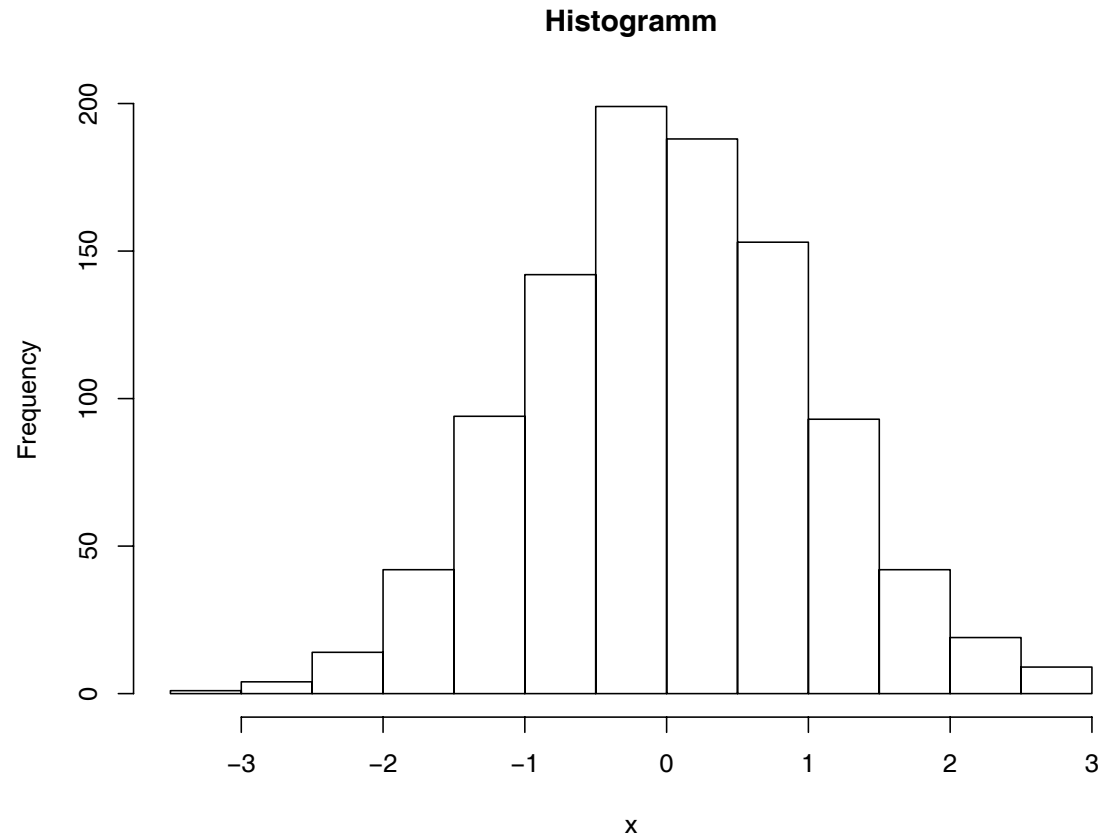


Graphische Auswertung von Daten

- Histogramme
- DichtepLOTS
- Streudiagramme
- Boxplots (Kastengrafik)

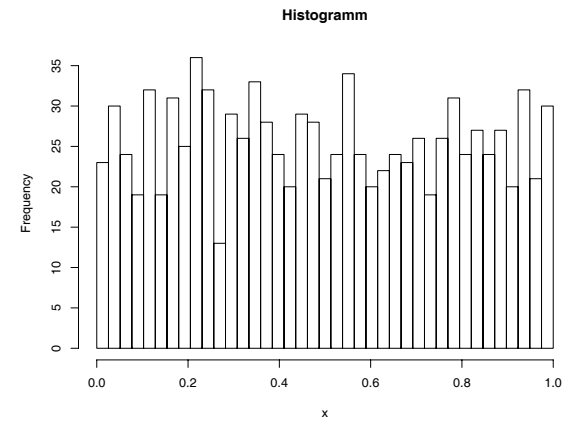
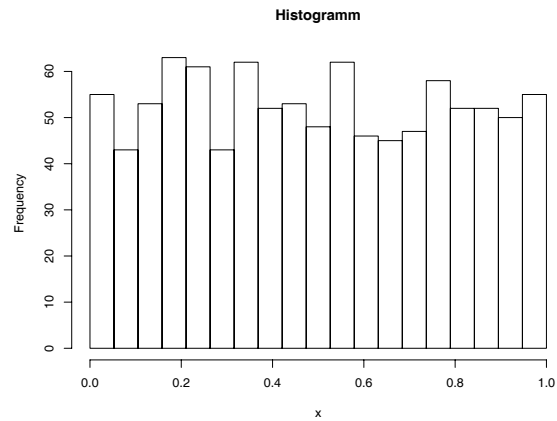
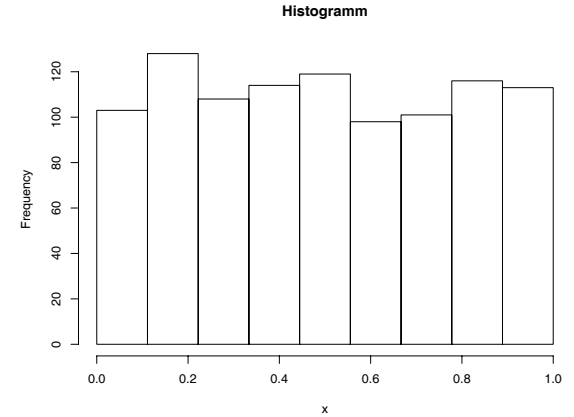
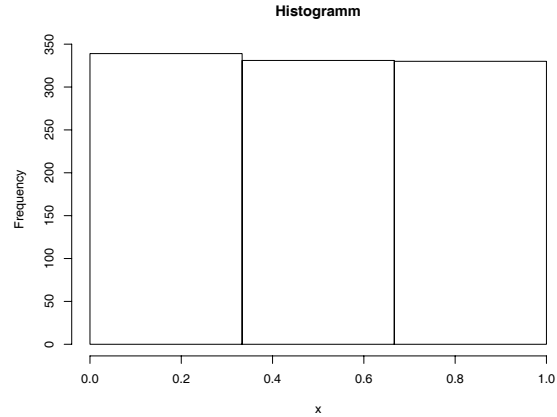
Histogramme

- $X_i \sim N(0,1)$
- $n=1000$
- Beachte:
Wahl der
Balkengrenzen
beeinflusst
Aussehen



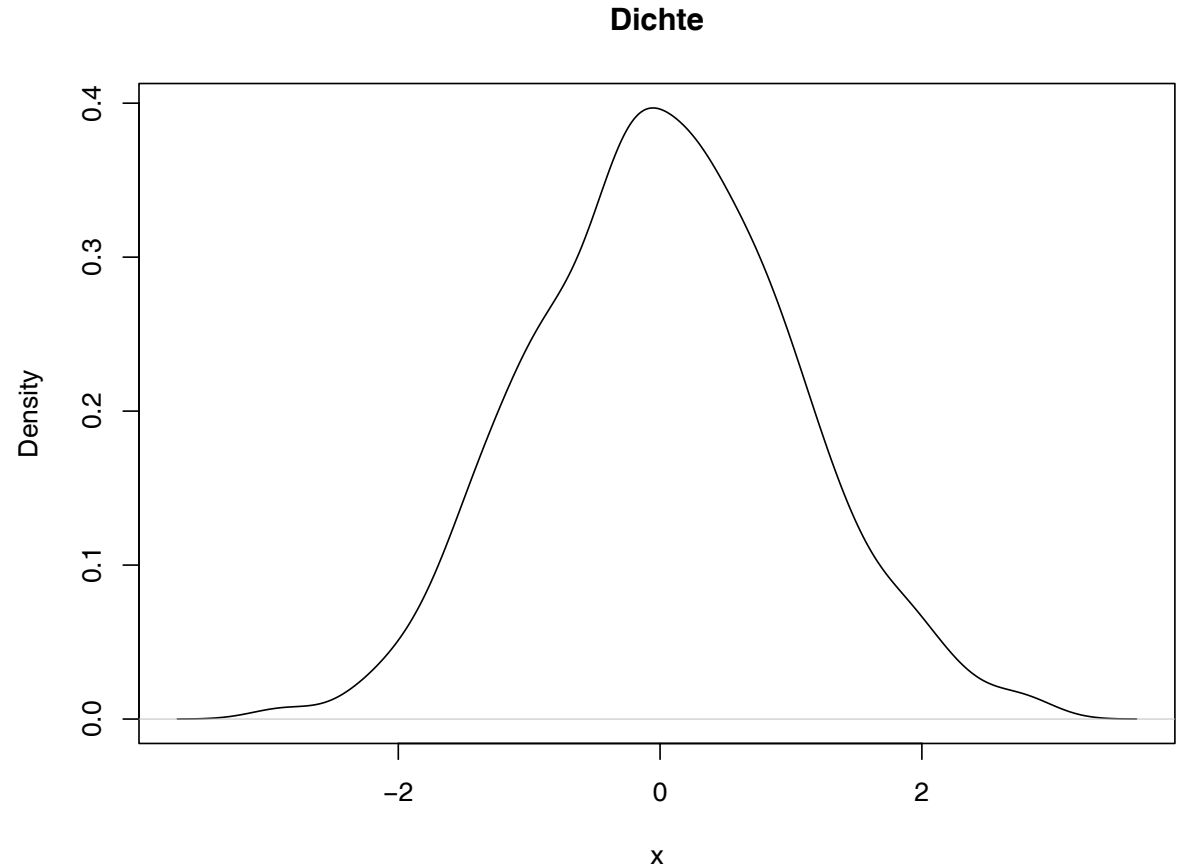
Wahl der Grenzen

- $X_i \sim U(0,1)$
- $n=1000$



Dichteplots

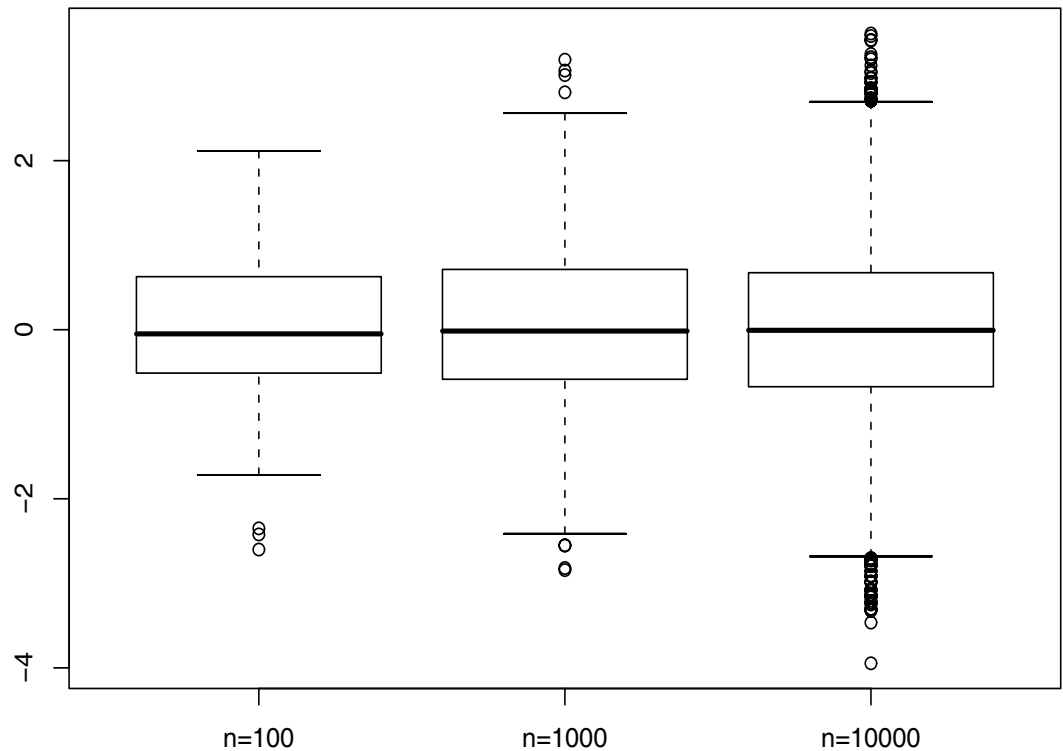
- $X_i \sim N(0,1)$
- $n=1000$



Boxplots

- $X_i \sim N(0,1)$
- Strich: Median
- Box: Quartile -> enthält 50% der Daten
- Whisker: min und max der Daten
- Kreise: Ausreißer

- Ablesbar:
Wertebereich,
Symmetrie



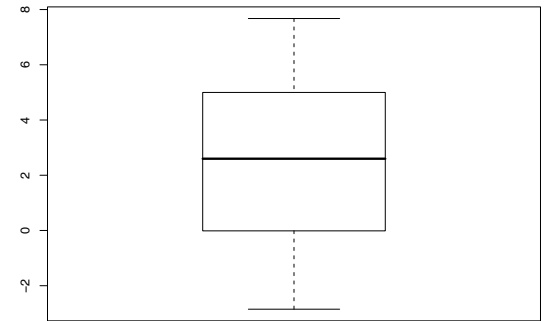
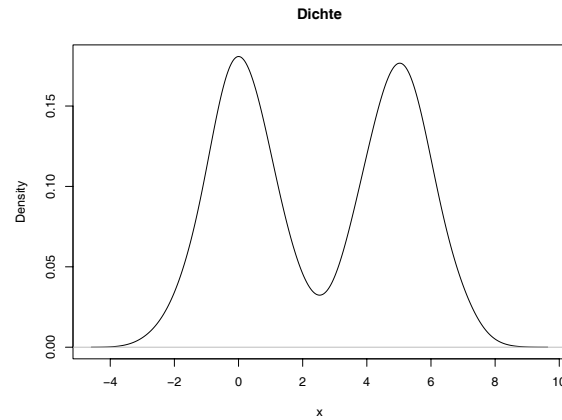
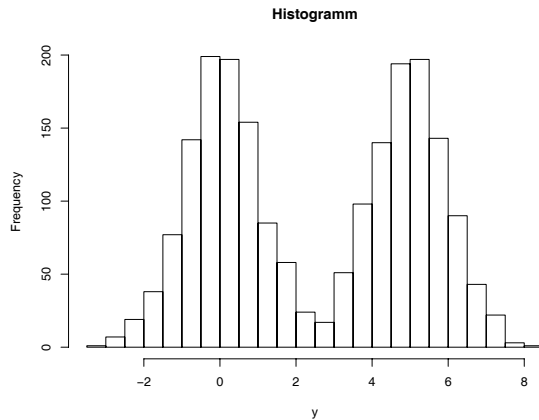
Berechnung der Whisker in R:

Extremster Datenpunkt, der nicht weiter als der 1.5-fache Abstand zwischen den Quartilen von der Box entfernt ist.

Alternativ: 2,5% - und 97,5%-Quantile der Daten.

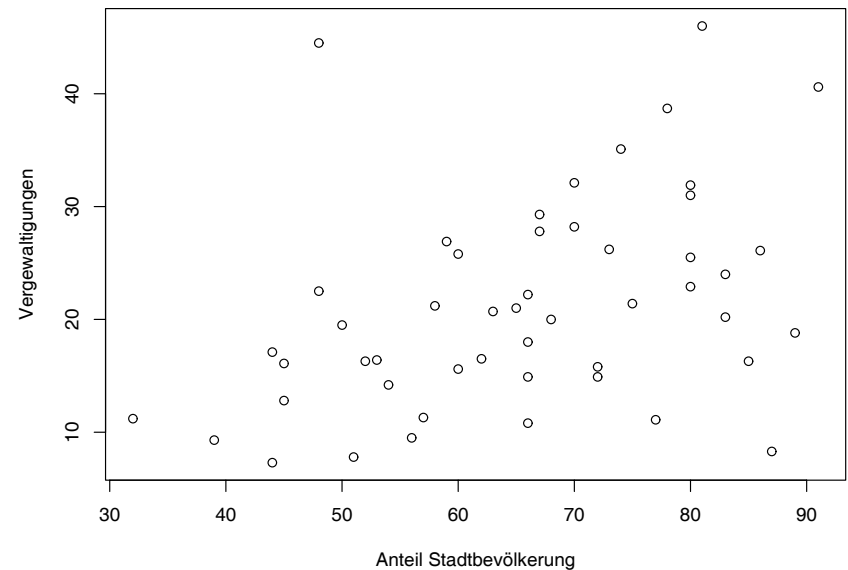
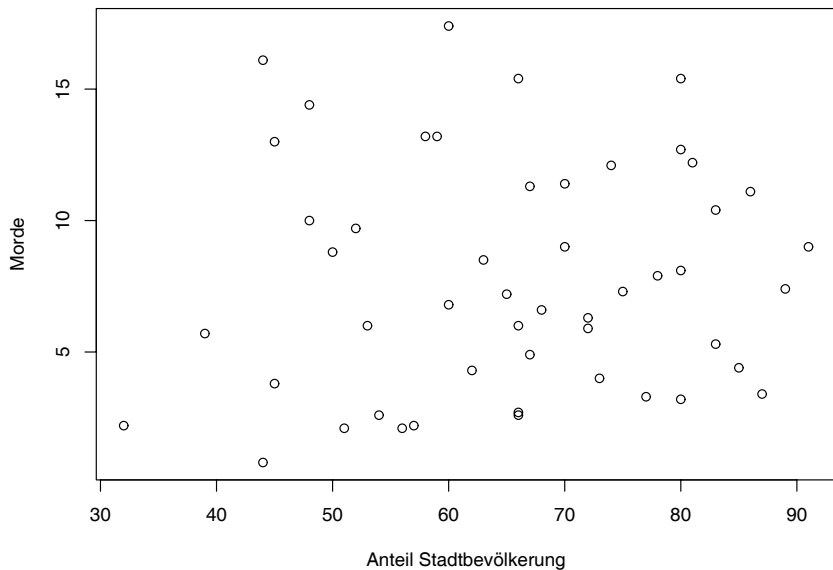
Multimodale Verteilungen

- $X_i \sim N(0, 1)$ für $i = 1, \dots, 1000$ und
 $X_i \sim N(5, 1)$ für $i = 1001, \dots, 2000$



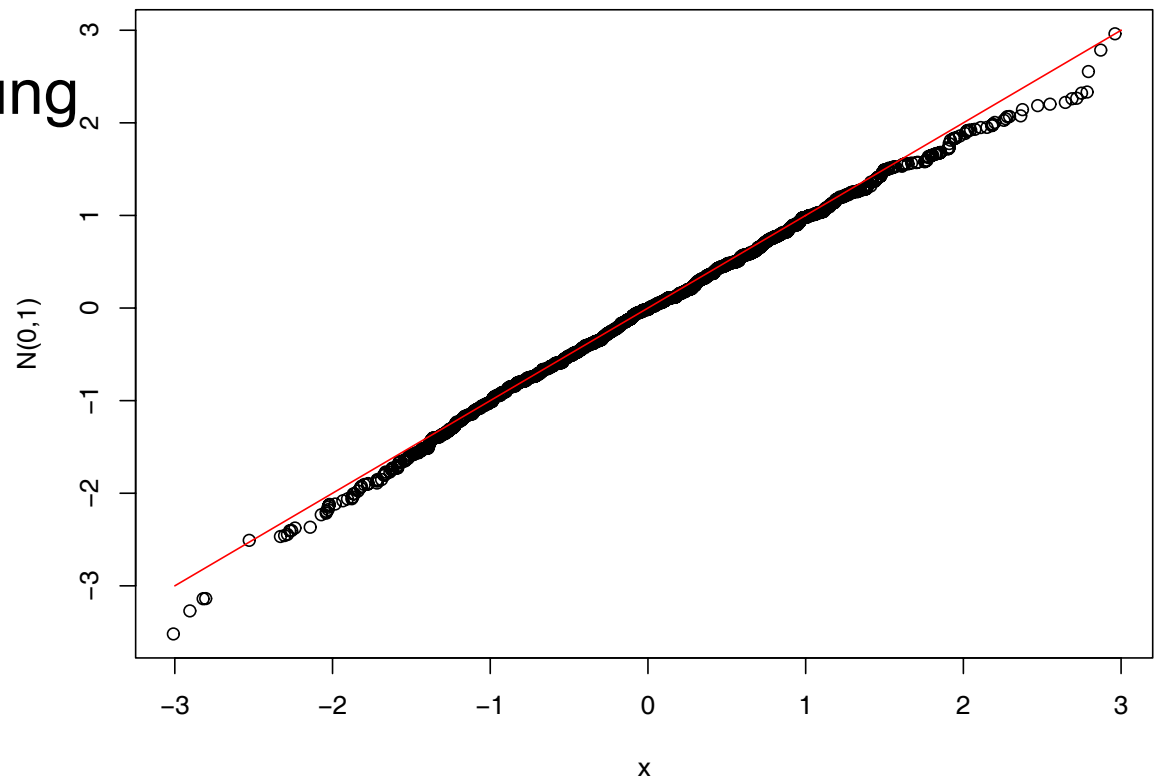
Streudiagramme

- Zur Untersuchung der Abhängigkeit verschiedener Merkmale
- Datensatz USArrests in R (Punkte=Staaten)



Q-Q-Plots

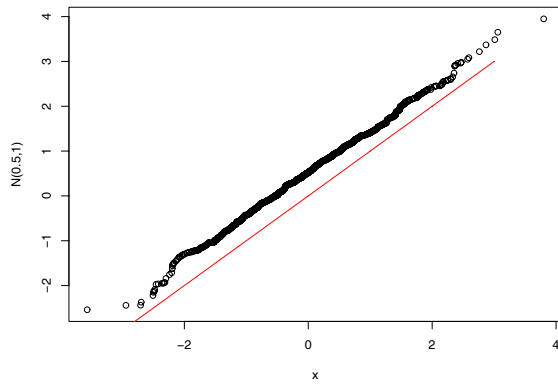
- $X_i \sim N(0,1)$, $n=1000$
- Plotte Stichprobenquantile gegen Quantile des Modells
- Bei guter Anpassung
-> Gerade $y=x$



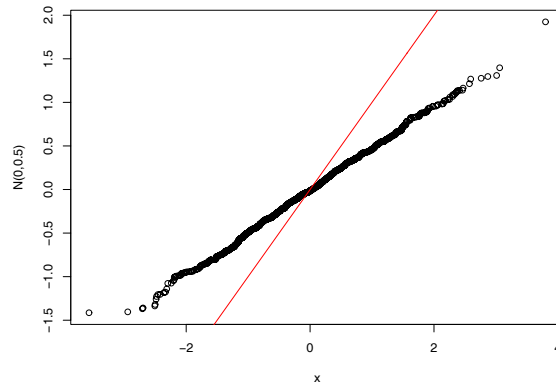
Q-Q-Plots

- $X_i \sim N(0,1)$, $n=1000$

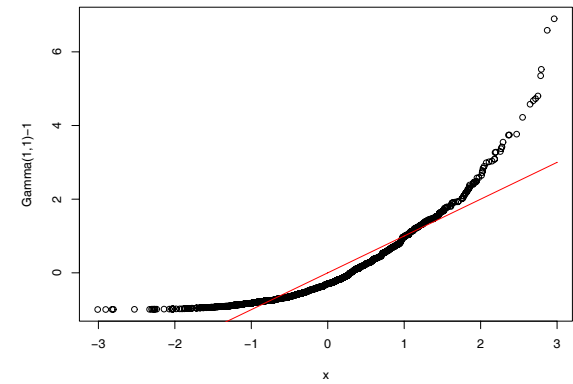
• $N(0.5,1)$



$N(0,0.5)$



Gammavert.



Software

- R: frei verfügbar unter www.r-project.org