

## Graphische Darstellungen stetiger Variablen

- Histogramme (Parameter: Ankerpunkt, Binbreite)

- Dotplots

- Boxplots

- empirische Verteilungsfunktionen

und

- QQ Plots

- Dichteschätzer (mehrere Methoden und Optionen)

## Mögliche Eigenschaften einer stetigen Variable

- Moden
- Granularität
- Cluster
- Lücken
- Täler
- Ausreißer
- Symmetrie

## Quantil Plots

Entweder für den Vergleich von einer empirischen Verteilung mit einer theoretischen (z.B. sind Modellresiduen annähernd normal?)

oder für den Vergleich empirischer Verteilungen (z.B. wie ähnlich sind zwei Verteilungen?)

## Theoretische Quantile

$$Q(p) = F^{-1}(p) = \inf\{x : F(x) \geq p\}$$

## Empirische Quantile in R (es gibt 9!)

$$Q_i(p) = (1 - \gamma)x_{(j)} + \gamma x_{(j+1)}$$

wo  $x_{(j)}$  ist die  $j$ . Ordnungsstatistik

$$\frac{(j - m)}{n} \leq p < \frac{(j - m + 1)}{n}$$

wo  $m$  hängt vom Typ  $i$  ab und  $n$  ist die Stichprobengröße. Empfohlen wird  $i = 8$ , eine lineare Interpolation zwischen  $x_{(j)}$  und

$$p(j) = \frac{j - \frac{1}{3}}{n + \frac{1}{4}}$$

## Dichteschätzer

Histogramme, Dotplots, und Boxplots betonen die rohen Daten. Dichteschätzer werden eingesetzt, um mögliche Modelle, die die Daten erzeugt haben können, zu schätzen: die Dichte der dahinterstehenden Grundgesamtheit.

Es gibt viele Methoden (z.B. Kerndichteschätzer) mit vielen Optionen (z.B. Kernfunktion, Fensterbreite für Kerndichteschätzer) und verschiedene Implementierungen.

- Wie kann die Fensterbreite gesetzt werden?  
Welcher Standard wird genommen?
- Können Grenzeinschränkungen berücksichtigt werden?
- Wie können Resultate überprüft werden?
- Sind multivariate Dichteschätzungen möglich?

## Dichteschätzer in R

Wie findet man so was in R?

1. ?density estimation
2. R Pakete Webseite  
([cran.r-project.org/src/contrib/PACKAGES.html](http://cran.r-project.org/src/contrib/PACKAGES.html))
3. R Webseite Suche
4. Task Views auf der R Webseite

z.B. density (in Stats), ash (Scott ASH), kde2d(in MASS),  
loffit, ks, sm, KernSmooth, plugdensity, hdrnde, feature,  
GenKern, logspline