Vorlesung am 10.11.2008

Dichteschätzung:

Eine Funktion

$$f : \mathbb{R} \to \mathbb{R}$$

(d.h., eine Zuweisung, die jeder reellen Zahl $x \in \mathbb{R}$ eine eindeutig bestimmte reelle Zahl $f(x) \in \mathbb{R}$ zuordnet) heißt **Dichte**, falls gilt:

1. $f(x) \geq 0$ für alle $x \in \mathbb{R}$,

2. $\int_{-\infty}^{\infty} f(x)\, dx = 1.$

Anschaulich: Die Funktionswerte sind nichtnegativ und der Flächeninhalt zwischen Funktion und x-Achse ist Eins.

Eine Dichte beschreibt approximativ eine Menge von Datenpunkten $x_1, \ldots, x_n$, falls in jedem Intervall I für die Anzahl n_I der in I enthaltenen Datenpunkte gilt:

$$\frac{n_I}{n} \approx \int_I f(x)\,dx.$$

Anschaulich: Die relative Anzahl n_I/n der Datenpunkte im Intervall I stimmt ungefähr mit dem Flächeninhalt zwischen Funktion und x-Achse im Intervall I überein.

Eine solche Dichte kann mit Hilfe des sogenannten **Kerndichteschätzers** konstruiert werden . . .

Statistische Maßzahlen:

Fassen Datenpunkte $x_1, \ldots, x_n \in \mathbb{R}$ in wenige Zahlen zusammen.

Diese beschreiben zum einen die *"Mitte" der Datenpunkte*, z.B. durch das (empirische) arithmetische Mittel

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n},$$

sowie die *"Schwankung" der Datenpunkte um den Mittelwert*, z.B. durch die (empirische) Varianz

$$\frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1}.$$

Im Boxplot werden einige dieser Maßzahlen graphisch übersichtlich dargestellt (u.a. Median und Interquartilabstand).