



Statistik II

Übungsblatt 10

Abgabe: Do. 29.1.2004, 10.00 Uhr, Briefkasten: Statistik II.

Bei jeder Aufgabe können maximal 5 Punkte erreicht werden.

1. Betrachtet wird der Datensatz "Math Students". Er enthält Informationen über Studenten der Mathematik an der LMU München.
 - (a) Stellen Sie die Daten graphisch dar
 - i. Den Zusammenhang zwischen den Einflussvariablen.
 - ii. Die Diplomnote in Abhängigkeit der Einflussvariablen.
 - (b) Erstellen Sie ein geeignetes Modell für die Diplom Note, und interpretieren Sie es.
2. Betrachtet wird der Datensatz "Hits", der die Anzahl der Zugriffe auf den Webserver <http://www.rosuda.org> zählt.
 - (a) Plotten Sie die Daten und zeichnen Sie die KQ-Gerade ein.
 - (b) Berechnen Sie lowess-Glättungen mit einem Span $f = 5, 10, 20$ und 50% .
 - (c) Diskutieren Sie die Ergebnisse für verschiedene f . Welchen Wert für f bevorzugen Sie, warum?
 - (d) Erklären Sie warum die periodische Struktur nicht mit einem lowess Glätter erfasst werden kann!
3. Stellen Sie die "Tipping Rate" aus dem "Tipping" Datensatz in einem Histogramm dar. Entfernen Sie vorher alle Werte größer als 0.4.
 - (a) Wählen Sie einen geeigneten Ankerpunkt und Bin-Breite, so dass die Daten im Histogramm möglichst gut wiedergegeben werden.
 - (b) Berechnen Sie einen Kerndichteschätzer (z.B. mit dem R-Befehl `density`) mit der Bandbreite 0.025 und 0.05 und tragen sie diese beiden Schätzungen in 3 Histogramme ein, jeweils mit einem
 - i. Rechteckkern
 - ii. Dreieckskern
 - iii. Normalkern
 - (c) Sind Sie mit Ihrer Wahl aus (a) immer noch zufrieden, oder würden Sie nun andere Werte wählen?
4. In einem Experiment werden 9 Studenten zufällig ausgewählt und ihre Körpergröße bestimmt. Die Körpergröße sei normalverteilt mit Erwartungswert $\mu = 180$ cm und Standardabweichung $\sigma = 8.4$ cm. Simulieren Sie die Ergebnisse dieses Experiments und berechnen Sie für μ ein 95 % Konfidenzintervall. Wiederholen Sie diese Simulation insgesamt 100 mal.
 - (a) Wie unterscheidet sich diese Simulation von einem Bootstrap?
 - (b) Zeichnen Sie die erhaltenen Konfidenzintervalle in ein gemeinsames Koordinatensystem.
 - (c) Wieviele der Intervalle enthalten μ ?
 - (d) Würden Sie erwarten, daß alle Intervalle μ enthalten?
 - (e) Haben alle Intervalle die gleiche Länge?

5. Wir betrachten erneut den Datensatz "Tipping":
- (a) Berechnen Sie den Median $x_{0.5}$ der "Tipping Rate" (hier werden wieder alle Werte betrachtet).
 - (b) Erstellen Sie 100 Bootstrap Samples der Tipping Rate (z.B. mit dem R-Befehl `sample`) und berechnen Sie jeweils $x_{0.5(i)}$, $i = 1, \dots, 50$.
 - (c) Wie groß ist die durchschnittliche Abweichung der Bootstrap Stichproben vom Median der Ursprungsstichprobe, und wie groß ihr Standardfehler?
 - (d) Wie interpretieren Sie die Ergebnisse aus (c)?