

Prof. Dr. Antony Unwin, Dr. Ali Ünlü
Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse
Institut für Mathematik
Universität Augsburg
<http://stats.math.uni-augsburg.de/>

Stochastik II — Statistik I (Statistik und Datenanalyse, SS 2009)

Übungsblatt 4

Abgabe: Dienstag 19. Mai 2009, bis spätestens 12.00 Uhr; Briefkasten: Stochastik II

1. Es sind Vorschläge gemacht worden, um zu schätzen, wie groß unter den Erwachsenen der Anteil derer ist, die gegen das Rauchverbot in Bayern sind. Es wird immer dieselbe Frage gestellt: Sind Sie für das bayrische Rauchverbot, ja oder nein?
 - (a) Eine Online-Befragung wird von der Augsburger Allgemeine durchgeführt. Jeder Abonnent darf höchstens einmal teilnehmen.
 - (b) Die Frage wird im nächsten Politbarometer eingebaut.
 - (c) In einhundert zufällig ausgewählten Kneipen werden je zwanzig Leute befragt.

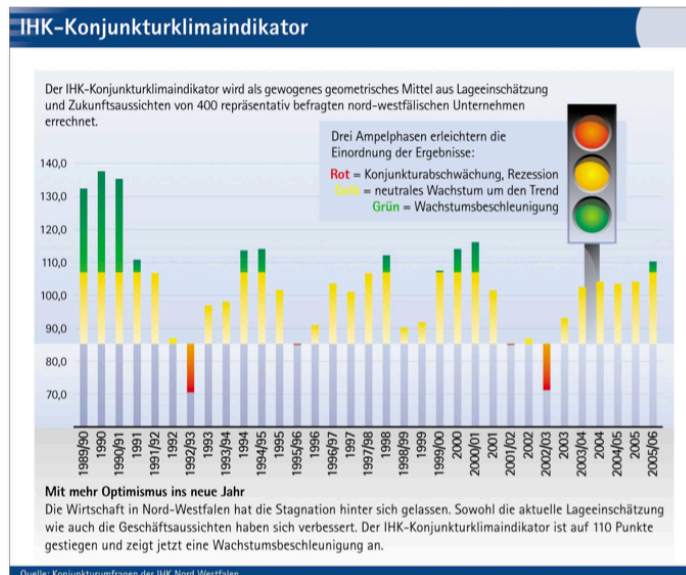
Was halten Sie von den Vorschlägen? Was wird in jedem Fall tatsächlich geschätzt? Welche Annahmen werden implizit gemacht und sind sie vertretbar? Welche statistischen Modelle könnten eingesetzt werden? Würden Sie eine andere Frage empfehlen?

2. Betrachtet wird der Datensatz `Math_Students` auf der Internetseite der Lehrveranstaltung.
 - (a) Laden Sie den Datensatz in R. Erklären Sie die Variablen—was könnte vor allem die Variable `Dnote` bedeuten—und verschaffen Sie sich einen Überblick über den Datensatz.
 - (b) Erstellen Sie für Mathematiker und Wirtschaftsmathematiker geeignete Graphiken, um die Dauer des Grund- und Hauptstudiums darzustellen. Angenommen die Studiendauer sei in jeder der beiden Gruppen normalverteilt, schätzen Sie für jede der Gruppen die nötigen Parameter und berechnen Sie die Wahrscheinlichkeit länger als 6 Semester zu studieren. Erscheinen Ihnen die Ergebnisse sinnvoll, wenn nicht, woran könnte das liegen?
 - (c) Erstellen und beschreiben Sie ein Streudiagramm der Vordiploms- und Hauptdiplomsnote. Finden Sie die Studenten, die sich am meisten verbessert oder verschlechtert haben, und zeichnen Sie diese in der Graphik ein.
3. Betrachtet wird die Online-Umfrage zum Thema Abtreibungsspiele aus Aufgabe 3 von Übungsblatt 3. Bestimmen Sie nach den drei in der Vorlesung besprochenen Konfidenzschätzverfahren
 - (a) 'normal approximativ',
 - (b) 'normal plug-in',
 - (c) 'normal genau'

konkrete Schätzintervalle für den Proportionsparameter p der Binomialverteilung aus dem Umfragebeispiel zum Konfidenzniveau 0.99. Diskutieren Sie Ihre Ergebnisse.

4. Kommentieren Sie die nebenstehende Graphik.

- Was genau wird dargestellt?
- Erläutern Sie die beiden Achsen.
- Wie beurteilen Sie die Farbwahl und ihre Umsetzung?
- Gibt es Probleme hinsichtlich der Interpretation der Graphik bzw. der eingefärbten Bereiche? Was würden Sie ändern?
- Beurteilen Sie die Graphik in Bezug auf Eindeutigkeit, Übersichtlichkeit und Ästhetik.



5. Seien \hat{p} und \hat{p}_B der Maximum-Likelihood-Schätzer bzw. Bayes-Schätzer (apriori Betaverteilung mit Parametern $a > 0$ und $b > 0$) für den Proportionsparameter p einer Binomialverteilung $B(N, p)$.

- Bestimmen Sie die mittleren quadratischen Fehler dieser Schätzer.
- Zeigen Sie, dass für die Wahl $a = b = \sqrt{N/4}$ der apriori Betaverteilung der mittlere quadratische Fehler des Bayes-Schätzers konstant ist, d.h. nicht von p abhängt. Bestimmen Sie hierbei explizit den Bayes-Schätzer und dessen mittleren quadratischen Fehler.

6. Einzeiler-Aufgaben.

- Leiten Sie den Erwartungswert der Poisson-Verteilung her.
- Was macht der R-Befehl `pbinom(3, 10, pnorm(0))` und welcher Wert ergibt sich?