

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

Stochastik für Lehramt

Übungsblatt 5

Abgabe: Montag 31. Mai 2010, bis spätestens 10.00 Uhr; Briefkasten: Stochastik für Lehramt

Die Aufgaben können auch in Gruppen bearbeitet und abgegeben werden! Für jede Aufgabe gibt es 5 Punkte. Lösungen in R können in Form eines Skriptes (Textdatei) per email an die jeweiligen Übungsgruppenleiter geschickt werden!

Grundlegendes:

- (a) Was wird durch $e^{-4 \cdot 2} - e^{-4 \cdot 6}$ berechnet?
- (b) Was ist eine Faltung?
- (c) Wofür könnte in der Praxis eine Gleichverteilung eingesetzt werden?

-
1. Ein Manager wurde Opfer der Finanzkrise und muss nun auf andere Weise sein Geld verdienen. Da er keine Würstchen verkaufen möchte, kommt er auf eine Idee. Er kauft sich jeden Tag in München ein Bayern-Ticket für 28 Euro. Damit darf er den ganzen Tag innerhalb Bayerns ohne weitere Kosten Bahn fahren. Er fährt einmal stündlich von Augsburg nach München (oder zurück) und nimmt bis zu 4 Fahrgäste für jeweils 5 Euro mit seinem Ticket mit.
 - (a) Nehmen Sie an, er hat zwischen zwei Fahrten nur begrenzt Zeit und kann daher jeweils nur n Leute fragen, von denen jeder mit einer Wahrscheinlichkeit von p zusagt. Wie könnte man die Wahrscheinlichkeit dafür, dass er k Personen auf einer Fahrt mitnimmt, modellieren, wenn
 - i. $n \leq 4$?
 - ii. $n > 4$?(Tipp zu ii): *Fallunterscheidung, negative Binomialverteilung*)
 - (b) Berechnen Sie den durchschnittlichen Gewinn eines Monats (20 Tage 12 Fahrten) für $n = 4$ und $p = 0.6$.
 - (c) Welchen Wert wird der Gewinn mit 90%-iger Wahrscheinlichkeit nicht unterschreiten? Dabei darf eine Normalverteilung mit den zu $n = 4$ und $p = 0.6$ passenden Parametern verwendet werden.
 2. Laden Sie den Datensatz `Erdbeben` in R. Der Datensatz beschreibt die Zeit (in Tagen) zwischen aufeinanderfolgenden Aufzeichnungen von Erdbeben. Dabei wurden Erdbeben weltweit betrachtet, und es wurden diejenigen Erdbeben aufgenommen, die eine Stärke größer als 7.5 auf der Richterskala aufwiesen oder bei denen über 1 000 Menschen getötet wurden. Aufzeichnungen starteten am 16. Dezember 1902 und endeten am 4. März 1977. Es gab insgesamt 63 Erdbeben; in dem Datensatz sind die 62 "Wartezeiten" zwischen aufeinanderfolgenden Erdbeben notiert. [Quelle: The Open University (1981). *S237 The Earth: Structure, Composition and Evolution*. Milton Keynes, The Open University Press.]
 - (a) Erstellen Sie ein Histogramm der Wartezeiten. Was fällt auf? Welche Verteilung(en) wäre(n) hier denkbar?

- (b) Bestimmen Sie die mittlere Wartezeit zwischen aufeinanderfolgenden Erdbeben. Was ist die Wahrscheinlichkeit dafür, irgendwo auf dieser Welt an irgendeinem Tag ein Erdbeben zu beobachten?
- (c) Gehen Sie von der Exponentialverteilung mit dem in b) berechneten Erwartungswert aus. Was ist die Wahrscheinlichkeit dafür, dass für mehr als 5 Jahre (inklusive einem Schaltjahr) kein Erdbeben irgendwo auf dieser Welt eintritt? Vergleichen Sie diese Wahrscheinlichkeit mit der relativen Häufigkeit der Wartezeiten (in den Daten), die größer als 5 Jahre sind? Erklären Sie warum diese Wahrscheinlichkeit so gering ist?
3. Eine Gruppe von 100 Schülern nimmt an einem Test teil, in dem jeder Schüler Fragen zu Mathematik, Deutsch und Geschichte beantworten muss. Die Anzahl der Punkte, die jeder Schüler erreichen kann sei in den drei Gebieten jeweils unabhängig normalverteilt: $X_M \sim N(25, 5^2)$, $X_D \sim N(36, 6^2)$ und $X_G \sim N(36, 6^2)$
- (a) Wie wahrscheinlich ist es, dass ein Schüler in Mathematik mindestens drei Punkte mehr erreicht, als in Deutsch?
- (b) Wie wahrscheinlich ist es, dass die durchschnittliche Anzahl der Punkte der ganzen Gruppe in jedem Teilgebiet unter dem Erwartungswert liegt? Wie wahrscheinlich ist es, dass sie sogar mindestens 25% darunter liegt?
- (c) Ist es wahrscheinlicher oder unwahrscheinlicher als das Ereignis in b), dass die Gruppe in der Summe aller drei Fächer im Schnitt unter den Erwartungen zurückbleibt?
4. (Fortsetzung von Aufgabe 3) Sie haben die Auswertung des Tests für ein Institut durchgeführt und dabei insbesondere die Durchschnitte angegeben. Ein Kritiker bemerkt, dass die Anzahl der Punkte für jeden Schüler in den drei Bereichen in Wirklichkeit gar nicht normalverteilt sei.
- (a) Wie begründen Sie, dass ihre Berechnungen für die Durchschnitte trotzdem näherungsweise korrekt sind?
- (b) Wovon hängt es ab, wie gut diese Annäherung ist?
- (c) **(Zusatz)** Simulieren Sie in R für die Punkte in Mathematik Stichproben im Umfang von 30 und 1000 Schülern jeweils mit einer Normalverteilung und einer anderen Verteilung ihrer Wahl (z.B. Poisson) und berechnen Sie den Durchschnitt. Wiederholen Sie dies 1000 Mal, zeichnen Sie jeweils ein Histogramm der Ergebnisse und interpretieren Sie diese.
Tipp: `replicate(1000, mean(rpois(n = 30, lambda = 25)))`

5. Die nachfolgende Grafik visualisiert den Benzinverbrauch verschiedener Verkehrsmittel.

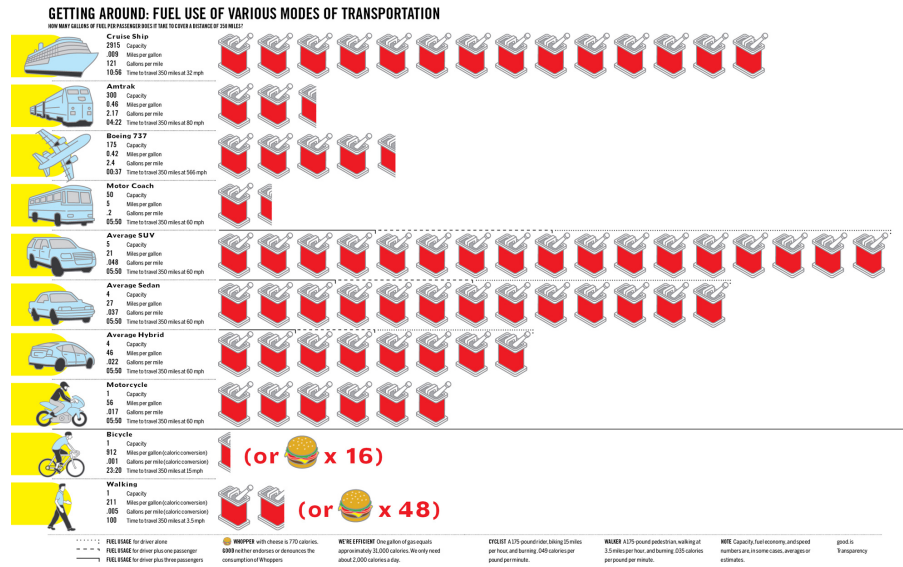


Abbildung 1: <http://awesome.good.is/transparency/web/trans0209gettingaround.html>

- Beurteilen Sie die Grafik in Bezug auf Suggestivität, Eindeutigkeit und Ästhetik.
- Was wird dargestellt? Ist die Darstellung gelungen? Diskutieren Sie die Grafik in Ihrer Übung!
- Wie würden Sie selbst die Daten visualisieren? Würden Sie Informationen hinzufügen oder weglassen?

(R) Bearbeiten Sie folgende kurze Fragestellungen in R!

- Berechnen Sie einen Wert T , unter dem mit einer Wahrscheinlichkeit von 2.5% die Realisation einer Zufallsvariable $X \sim N(100, 10^2)$ liegt (q_{norm}).
- Berechnen Sie für $\mu = 100$ und $\sigma = 10$ die Wahrscheinlichkeiten, dafür, dass $\mu - k\sigma < X < \mu + k\sigma$ für $k = 2$ und $k = 3$.