

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 6

**Abgabe:** Montag 7. 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:

Seien  $X$  und  $Y$  unabhängig identisch normalverteilte Zufallsvariablen mit  $\mu = 12$  und  $\sigma = 4$ .

- Wie sind  $2 \cdot X$  und  $X + Y$  verteilt?
  - Wie berechnet man mit der Tabelle einer Standardnormalverteilung (sh. letzte Seite) einen Wert  $k$ , so dass  $P(X \leq k) = 0.88$ ?
  - Wie berechnet man mit der Tabelle einer Standardnormalverteilung (sh. letzte Seite) einen Wert  $k$ , so dass  $P(X \leq k) = 0.3$ ?
  - (Zusatz)** Wie geht man in b) und c) in R vor?
- 

1. In einer klinischen Studie über Leukämie wird die Anzahl weißer Blutkörperchen von 30 Patienten jeweils vor und nach einer medikamentösen Behandlung gemessen. Sie erhalten den Auftrag, die Ergebnisse grafisch darzustellen.

- Ein Freund rät Ihnen, zwei Histogramme der Daten zu erstellen, eines vor der Behandlung und eines nach der Behandlung. Worauf wäre dabei zu achten? Beschreiben Sie genau, wie Sie dabei vorgehen würden.
- Ist diese Art der grafischen Darstellung im gegebenen Kontext sinnvoll? Was hätte man stattdessen visualisieren sollen?
- (Zusatz)** Simulieren Sie in R zwei Stichproben mit dem Umfang 30 und den Mittelwerten 400 und 500, sowie Standardabweichungen von 40 bzw. 60. Erzeugen Sie Ihren eigenen Visualisierungsvorschlag aus b)! Sind Sie zufrieden? :)

2. Untersuchen Sie die folgende interaktive Grafik im Internet:

<http://www.gapminder.org/world>

- Beurteilen Sie die Grafik in Bezug auf Suggestivität, Eindeutigkeit und Ästhetik.
- Was wird dargestellt? Welche Variablen fließen in die Grafik ein? Ist die Darstellung gelungen?
- Welche interaktiven Möglichkeiten bietet die Grafik? Welche Möglichkeiten würden Sie sich wünschen?
- Untersuchen Sie den Verlauf für einzelne Länder (z.B. China, Indien, Deutschland,...) mit Hilfe des Schiebereglers. Kann man von einem allgemeinen Zusammenhang sprechen?
- Würden Sie die Grafik im Unterricht einsetzen?

3. Nachfolgende Grafik ist ein Streudiagramm der Länge von Filmen bis 600 Minuten (y) gegen das Produktionsjahr (x). Die Daten stammen von imdb.com.

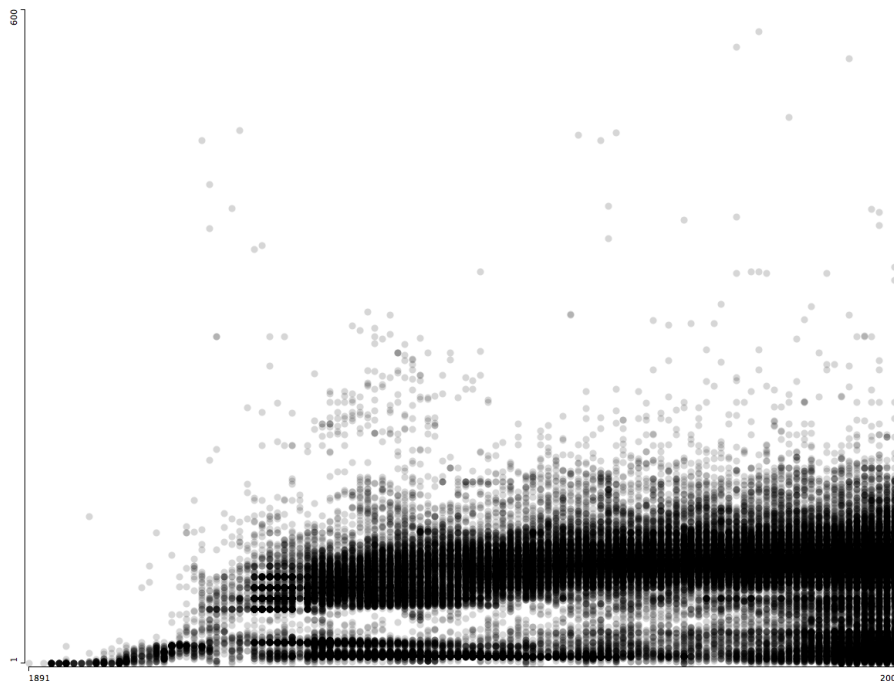


Abbildung 1: <http://www.imdb.com>

- (a) Beschreiben Sie die Grafik. Welche Auffälligkeiten gibt es und wie ließen sie sich erklären?
- (b) Sind neue Filme länger als alte?
- (c) Wie wird sich die Zusammensetzung der Filme über die Jahre hinweg geändert haben? Wie sieht es insbesondere in den letzten Jahren aus?
- (d) Welcher zusätzliche Effekt kommt im Streudiagramm zum Einsatz? Wozu könnte er nützlich sein? Ist er hier ausreichend?
4. Bei einem IQ-Test wurden die Werte 84, 89, 92, 94, 99, 101, 102, 102, 109, 111, 113, 115, 120, 122, 139 erzielt. Bearbeiten Sie die nachfolgenden Aufgabenstellungen per Hand oder mir R.
- (a) Berechnen Sie für die Daten *Mittelwert*, *Median*, *Standardabweichung* und *Varianz*, *Interquartilsabstand* sowie *Modus* und *Spannbreite*.
- (b) Welche einzelne Statistik würden Sie wählen, um die Daten zu beschreiben? Welches Paar würden Sie vorschlagen?
- (c) Wie 'reagieren' die Statistiken darauf, wenn man den höchsten Wert entfernt? Wie, wenn man eine '102' entfernt? Was schließen Sie daraus in Bezug auf die Eigenschaften der Statistiken?

5. Das Streudiagramm in der unteren Abbildung zeigt die tatsächlichen Abflugs- und Ankunftszeiten aller Flüge zwischen ATL (Atlanta) und DFW (Dallas) im Jahr 2007.

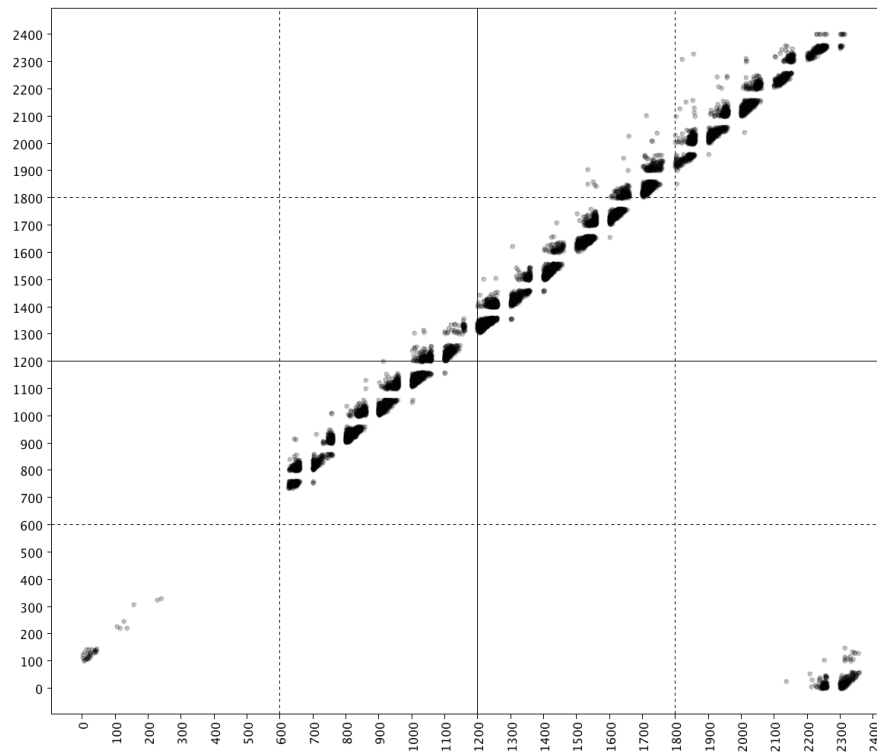


Abbildung 2: Streudiagramm Ankunftszeit (y) gegen Abflugzeit (x)

- Warum gibt es Blöcke von Daten?
- Warum gibt es Daten unten rechts in der Graphik? Was muss unternommen werden, um einen sinnvollen Plot zu bekommen?
- In welcher Zeitspanne bewegt sich die Dauer eines Fluges?
- Im Datensatz wird auch zwischen der tatsächlichen Ankunftszeit und der geplanten Ankunftszeit unterschieden. Wie könnte ein Streudiagramm dieser Daten aussehen?

---

6. [R]

Es soll der Datensatz *Cars*, der auf der Lehrstuhlwebseite verfügbar ist, betrachtet werden.

- Erstellen mit der Funktion `summary` eine Zusammenfassung des Datensatzes.
- Erstellen Sie einen `table` der Variablen *Type* und *Drive*.
- Was macht der Befehl `tapply(1/HighwayMilesPerGallon, Type, mean)`?

## Verteilungstabelle der Standardnormalverteilung

$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$
0.000	0.5000	0.750	0.7734	1.500	0.9332	2.250	0.9878
0.025	0.5100	0.775	0.7808	1.525	0.9364	2.275	0.9885
0.050	0.5199	0.800	0.7881	1.550	0.9394	2.300	0.9893
0.075	0.5299	0.825	0.7953	1.575	0.9424	2.325	0.9900
0.100	0.5398	0.850	0.8023	1.600	0.9452	2.350	0.9906
0.125	0.5497	0.875	0.8092	1.625	0.9479	2.375	0.9912
0.150	0.5596	0.900	0.8159	1.650	0.9505	2.400	0.9918
0.175	0.5695	0.925	0.8225	1.675	0.9530	2.425	0.9923
0.200	0.5793	0.950	0.8289	1.700	0.9554	2.450	0.9929
0.225	0.5890	0.975	0.8352	1.725	0.9577	2.475	0.9933
0.250	0.5987	1.000	0.8413	1.750	0.9599	2.500	0.9938
0.275	0.6083	1.025	0.8473	1.775	0.9621	2.525	0.9942
0.300	0.6179	1.050	0.8531	1.800	0.9641	2.550	0.9946
0.325	0.6274	1.075	0.8588	1.825	0.9660	2.575	0.9950
0.350	0.6368	1.100	0.8643	1.850	0.9678	2.600	0.9953
0.375	0.6462	1.125	0.8697	1.875	0.9696	2.625	0.9957
0.400	0.6554	1.150	0.8749	1.900	0.9713	2.650	0.9960
0.425	0.6646	1.175	0.8800	1.925	0.9729	2.675	0.9963
0.450	0.6736	1.200	0.8849	1.950	0.9744	2.700	0.9965
0.475	0.6826	1.225	0.8897	1.975	0.9759	2.725	0.9968
0.500	0.6915	1.250	0.8944	2.000	0.9772	2.750	0.9970
0.525	0.7002	1.275	0.8988	2.025	0.9786	2.775	0.9972
0.550	0.7088	1.300	0.9032	2.050	0.9798	2.800	0.9974
0.575	0.7174	1.325	0.9074	2.075	0.9810	2.825	0.9976
0.600	0.7257	1.350	0.9115	2.100	0.9821	2.850	0.9978
0.625	0.7340	1.375	0.9154	2.125	0.9832	2.875	0.9980
0.650	0.7422	1.400	0.9192	2.150	0.9842	2.900	0.9981
0.675	0.7502	1.425	0.9229	2.175	0.9852	2.925	0.9983
0.700	0.7580	1.450	0.9265	2.200	0.9861	2.950	0.9984
0.725	0.7658	1.475	0.9299	2.225	0.9870	2.975	0.9985