

**Aufgabe 1** Die Datei `Lottozahlen.txt` enthält für jede der Zahlen 1 bis 49 die absolute Häufigkeit, mit der diese Zahl in 4644 Ziehungen von 1955 bis 2008 des Lottos 6 aus 49 vorkam (ohne Zusatzzahl, siehe z.B. <http://www.dielottozahlen.de>). Bestimmen Sie für jede der Zahlen ein Konfidenzintervall zum Irrtumsniveau  $\alpha = 0,05$  für die Wahrscheinlichkeit, diese Zahl in einer Ziehung zu sehen. Wieviele dieser Konfidenzintervalle überdecken den theoretischen Wert? Ist das Ergebnis überraschend? – führen Sie beispielsweise mit R eine kleine Simulationsstudie durch.

**Aufgabe 2** Mukoviszidose ist eine menschliche Erbkrankheit, die von einem Gendefekt auf Chromosom 7 hervorgerufen wird. Das Allel für Mukoviszidose ist rezessiv, d.h. nur homozygote Individuen erkranken tatsächlich. Nehmen wir an, eine Population befindet sich (bezüglich dieses Gens) im Hardy-Weinberg-Gleichgewicht, und unter je 3000 Geburten findet sich ein an Mukoviszidose erkranktes Kind.

- Welcher Anteil der Population trägt genau eine Kopie des defekten Allels?
- Nehmen wir an, ein gesundes Paar habe ein an Mukoviszidose erkranktes Kind. Wie wahrscheinlich ist es dann, dass ein weiteres Kind ebenfalls krank sein wird?
- Nehmen wir an, ein gesundes Paar habe bereits ein gesundes Kind. Wie wahrscheinlich ist es dann, dass das zweite Kind krank sein wird?

	$x$	$y$	
	1.2	5.4	Berechnen Sie mit Papier, Bleistift und Taschenrechner die Regressionsgerade für die in der Tabelle gegebenen vier Wertepaare für Merkmal $x$ und Zielgröße $y$ und fertigen Sie eine Skizze an. Hängt $y$ signifikant von $x$ ab?
<b>Aufgabe 3</b>	2.3	2.1	
	3.4	3.1	
	5.0	3.2	

**Aufgabe 4** Die Datei `800m.csv` enthält die Bestzeiten im 800m-Lauf der Herren für die Jahre 1970 bis 2015 in Sekunden<sup>1</sup>. Berechnen und zeichnen Sie die (kleinste-Quadrate-)Regressionsgerade für Jahr gegen Bestzeit. Wie hätten Sie Ende 2015 die Bestzeit für 2016 prognostiziert? Wie sicher sind Sie sich mit Ihrer Schätzung?

**Aufgabe 5** Erzeugen Sie mehrere zufällige Datensätze mit der Normalverteilung und anderen Verteilungen und vergleichen Sie die simulierten Daten mittels Normal-QQ-Plot mit der Normalverteilung (z.B. in R mit `qqnorm(rnorm(15))`). Betrachten Sie dann die Normal-QQ-Plots in der Datei `R2QQPlotsRaten.pdf` und raten Sie, welche der neun Datensätze unabhängige Stichproben aus einer Normalverteilung darstellen.

**Aufgabe 6** Die Datei `2QQPlotsDichtepolygone.pdf` enthält Normal-QQ-Plots für Daten aus 6 verschiedenen Verteilungen. Skizzieren Sie Dichtepolygone, die zeigen, wie die jeweilige Verteilung von einer Normalverteilung (mit entsprechendem Erwartungswert und entsprechender Varianz) abweicht.

<sup>1</sup>Wikipedia, [https://en.wikipedia.org/wiki/800\\_metres](https://en.wikipedia.org/wiki/800_metres), 14.06.2016