

1. Aufgabe Die Datei `800m.csv` enthält die Bestzeiten im 800m-Lauf der Herren für die olympischen Spiele 1920–2008. Berechnen und zeichnen Sie die (kleinste-Quadrate-)Regressionsgerade für Jahr gegen Bestzeit. Wie sieht Ihre Vorhersage für die Bestzeit 2012 aus? Wie sicher sind Sie sich mit Ihrer Schätzung?

2. Aufgabe 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 `qqplots_aufgabe_4.pdf` und raten Sie, welche der neun Datensätze unabhängige Stichproben aus einer Normalverteilung darstellen.

3. Aufgabe Die Datei `qqplots_aufgabe_5.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.

4. Aufgabe Die Datei `daten_aufgabe_6.csv` enthält Wertepaare eines erklärenden Merkmals x und einer Zielgröße y . Versuchen Sie, x und y so umzuskalieren, dass Sie ein möglichst gut passendes Regressionsmodell zur Schätzung von y aus x erhalten. Tipp: Beispiele für gängige Transformationen sind $z \mapsto z^2$, $z \mapsto (a + z)^2$, $z \mapsto z^3$, $z \mapsto e^{z \cdot a}$, $z \mapsto \log(a + z)$, $z \mapsto \sqrt{z + a}$, $z \mapsto \arcsin \sqrt{z}$, $z \mapsto \operatorname{asinh}(z \cdot a)$, wobei a jeweils eine passend gewählte reelle Zahl ist.

5. Aufgabe Folgende Tabelle zeigt Dauer des Studiums (in Semestern) und Einstiegsgehalt (in Tausend €) der Absolventen eines Jahres am Fachbereich Mathematik und Informatik der Yule-Simpson-Paradoxon-Universität:

Semester	12	14	16	12	15	14	13	14	11	13	10	12	14	13	14	15
Gehalt	39.4	38.2	37.4	39.5	32.8	35.3	39.1	35.2	37.9	35.7	41	40.9	34.2	38.4	36.2	38.4
Semester	9	11	9	9	12	13	11	10	10	10	9	10	12	10		
Gehalt	33.7	35.9	36.1	34.2	29.9	31.9	33.3	36.2	33.8	32.9	33.3	35.1	34.2	35.3		

a) Schlägt sich (für diese Absolventen) ein längeres Studium in einem höheren Anfangsgehalt nieder? Bestimmen Sie die Regressiongerade für Studiendauer gegen Anfangsgehalt.

b) Ändert sich Ihr Befund, wenn Sie zusätzlich erfahren, dass die oberen beiden Zeilen der Tabelle sich auf die Absolventen des Fachs Informatik, die unteren beiden sich auf die Absolventen des Fachs Mathematik beziehen, und Sie dieselbe Regression jeweils innerhalb dieser beiden Gruppen durchführen?

6. Aufgabe 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.