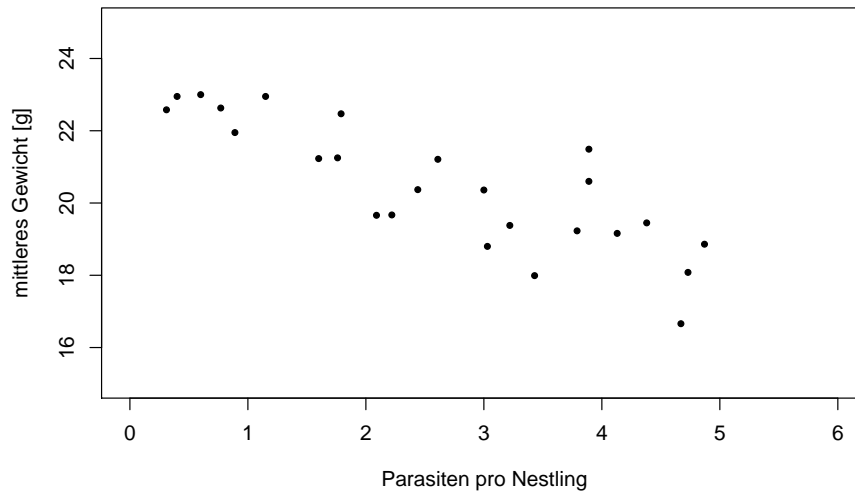


**Aufgabe 1** Nestlinge der Schwalbe *Hirundo pyrrhonota* werden oft von dem Parasiten *Oeciacus vicarius* befallen. In 25 Nestern mit Nestlingen im Alter von 10 Tagen wurde die Anzahl Parasiten pro Nestling und das mittlere Gewicht der Nestlinge bestimmt. Die Beziehung zwischen dem Gewicht  $G$  und der Parasitenzahl  $P$  scheint ungefähr linear zu sein:  $G = a + b \cdot P$ .

**Koerpergewicht und Parasitenbefall (25 Nester)**



- (a) Zeichnen Sie die Regressionsgerade nach Augenmaß ein.
- (b) Schätzen Sie nach Augenmaß den Achsenabschnitt  $a$ .
- (c) Was ist die biologische Bedeutung von  $a$ ? (z.B.: “ $a$  ist das mittlere Gewicht...”)
- (d) Schätzen Sie nach Augenmaß die Steigung  $b$ .
- (e) Was ist die biologische Bedeutung von  $b$ ? (z.B.: “Jeder zusätzliche Parasit...”)
- (f) Schätzen Sie (zur nächsten ganzen Zahl), den Standardfehler  $s$  der Regression.
- (g) In einem Nest mit leichtem Parasitenbefall gibt es 1.5 Parasiten pro Nestling. Schätzen Sie mit Hilfe Ihrer Schätzungen von  $a$  und  $b$  das mittlere Gewicht der Nestlingen. Mit einem wie großen typischen Fehler müssen Sie bei Ihrer Vorhersage rechnen?

**Aufgabe 2**

$x$	$y$
1.2	5.4
2.3	2.1
3.4	3.1
5.0	3.2

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?

**Aufgabe 3** 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?

**Aufgabe 4** 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 5** 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.

**Aufgabe 6** 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 Regressionsgerade 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?