Universität Köln
Angewandte Mathematik
Johannes Müller
Michael Kurth

Mathematik für Biologen I, WS 01/02

`http://www.mi.uni-koeln.de/mi/Forschung/Kuepper/biologie.html`

Mathematisches Institut
Raum 116/117
Tel: 0221/4702891

Lösungshinweise zu Übungsblatt Nr. 5

Aufgabe 1 (4 Punkte):

In dieser Aufgabe wollen wir die Formel für die lineare Regression herleiten. Sei dazu wie in der Vorlesung

$$f(a,b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

(a) Berechnen Sie zunächst $\frac{\partial}{\partial b} f(a,b)$ und setzen dies gleich Null.

(b) Zusammen mit der Formel aus der Vorlesung für $\frac{\partial}{\partial a} f(a,b)$ ergeben sich nun zwei Gleichungen für die beiden Unbekannten a und b, aus denen man die bekannte Lösung herleite.

Lösung:

(a) Zunächst ist also $\frac{\partial}{\partial b} f(a,b)$ zu berechnen. Dazu halten wir in $f(a,b)$ die Variable a fest und differenzieren nach b. Es folgt

$$\begin{aligned}
\frac{\partial}{\partial b} f(a,b) &= \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - (ax_i + b))^2 \quad \text{(nach Definition von } f\text{)} \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial b} (y_i - (ax_i + b))^2 \quad \text{(nach Summenregel)} \\
&= \sum_{i=1} 2\,(y_i - (ax_i + b)) \frac{\partial}{\partial b} (y_i - (ax_i + b)) \quad \text{(nach Kettenregel)} \\
&= -2 \sum_{i=1}^{n} (y_i - (ax_i + b)) \\
&= -2 \left(\sum_{i=1}^{n} y_i - a \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} b \right) \\
&= 2 \left(nb + a \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \right)
\end{aligned}$$

(b) Wir dividieren nun die Gleichung $\frac{\partial}{\partial b} f(a,b) = 0$ durch $2n$ und erhalten

$$b + \frac{a}{n} \sum_{i=1}^{n} x_i - \frac{1}{n} \sum_{i=1}^{n} y_i = 0.$$

Aus der Vorlesung ist die Gleichung

$$\frac{a}{n} \sum_{i=1}^{n} x_i^2 + \frac{b}{n} \sum_{i=1}^{n} x_i - \frac{1}{n} \sum_{i=1}^{n} x_i y_i = 0$$

bekannt. Wir haben also zwei Gleichungen für die beiden Unbekannten a und b. Auflösen der ersten Gleichung nach b und Einsetzen des Ausdrucks in die zweite Gleichung liefert

$$\begin{aligned}
& \frac{a}{n} \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\frac{a}{n} \sum_{i=1}^{n} x_i - \frac{1}{n} \sum_{i=1}^{n} y_i \right) \sum_{i=1}^{n} x_i - \frac{1}{n} \sum_{i=1}^{n} x_i y_i = 0 \\
\Leftrightarrow \quad & \frac{a}{n} \sum_{i=1}^{n} x_i^2 - a \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)^2 + \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right) \left(\frac{1}{n} \sum_{i=1}^{n} y_i \right) - \frac{1}{n} \sum_{i=1}^{n} x_i y_i = 0 \\
\Leftrightarrow \quad & a \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)^2 \right) = \frac{1}{n} \sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right) \left(\frac{1}{n} \sum_{i=1}^{n} y_i \right).
\end{aligned}$$

Demnach lautet die Lösung für a

$$a = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2}.$$

Daraus lässt sich nun b berechnen als

$$b = \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) - a\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right),$$

womit die Formeln aus der Vorlesung bewiesen wären.

Aufgabe 2 (4 Punkte):

Käfer werden 5 Stunden Karbondisulfid ausgesetzt. Wir bezeichnen mit c_i die Konzentration, mit n_i die Zahl der Käfer und mit r_i die Zahl der getöteten Käfer.

Konzentration	Käfer insgesamt	getötet
$c_1 = 1.6907$	$n_1 = 59$	$r_1 = 6$
$c_2 = 1.7242$	$n_2 = 60$	$r_2 = 13$
$c_3 = 1.7552$	$n_3 = 62$	$r_3 = 18$
$c_4 = 1.7842$	$n_4 = 56$	$r_4 = 28$
$c_5 = 1.8113$	$n_5 = 63$	$r_5 = 52$
$c_6 = 1.8369$	$n_6 = 59$	$r_6 = 53$
$c_7 = 1.8610$	$n_7 = 62$	$r_7 = 61$
$c_8 = 1.8839$	$n_8 = 60$	$r_8 = 60$

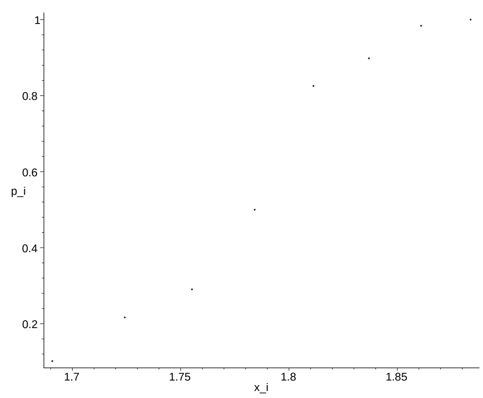
Beschreiben Sie den Zusammenhang zwischen der relativen Anzahl getöteter Käfer und der Konzentration des Karbondisulfids, indem Sie ein logit-Modell zu Grunde legen. Bei welcher Konzentration stirbt die Hälfte der Käfer (LD50-Wert)? Zeichnen Sie weiter die Daten mitsamt der errechneten Funktion in ein Diagramm.

Lösung:

Sei x die Konzentration des Karbondisulfids und $p(x)$ der relative Anteil getöteter Käfer. Zur Logit-Analyse benötigen wir nun die folgenden Messdaten

Konzentration $x_i \hat{=} c_i$	rel. Anzahl getöteter Käfer $p_i = p(x_i) = r_i/n_i$
1.6907	$6/59 = 0.1016949$
1.7242	$13/60 = 0.2166667$
1.7552	$18/62 = 0.2903226$
1.7842	$28/56 = 0.5000000$
1.8113	$52/63 = 0.8253968$
1.8369	$53/59 = 0.8983051$
1.8610	$61/62 = 0.9838710$
1.8839	$60/60 = 1.0000000$

Die Daten sehen in einem Diagramm wie folgt aus:



An der Verteilung der Daten erkennt man, wieso hier ein logit-Modell verwendet wird. Wir nehmen also an, daß die Funktion p folgende Form hat

$$p(x) = \operatorname{logit}(x) = \frac{\exp(ax+b)}{1+\exp(ax+b)}.$$

Um die Konstanten a und b mittels linearer Regresion bestimmen zu können, benötigen wir die Umkehrfunktion von logit (vgl. 4.Übung bzw. Vorlesung). Es ist

$$ax + b = \ln\left(\frac{p(x)}{1-p(x)}\right) =: y(x)$$

Um nun die Regressionsgerade bestimmen zu können, benötigen wir die Werte $y_i = y(x_i)$. Dabei müsen wir auf den letzten Wert in der Tabelle verzichten, da dort $p(x) = 1$ ist und somit $y(x)$ nicht definiert ist. Die entsprechenden Werte lauten dann

x_i	y_i
1.6907	-2.1785324
1.7242	-1.2851982
1.7552	-0.8938179
1.7842	0.0000000
1.8113	1.5533484
1.8369	2.1785324
1.8610	4.1108739
1.8839	Nicht definiert

Zur Bestimmung von a und b werden der Tabelle folgende Werte entnommen:

$$\frac{1}{7}\sum_{i=1}^{7} x_i = 12.4635/7 = 1.7805$$

$$\frac{1}{7}\sum_{i=1}^{7} y_i = 3.4852062/7 = 0.4978866$$

$$\frac{1}{7}\sum_{i=1}^{7} x_i^2 = 22.21375911/7 = 3.173394159$$

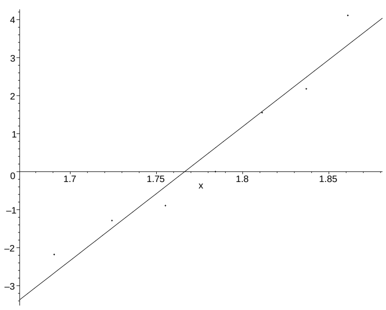
$$\frac{1}{7}\sum_{i=1}^{7} x_i y_i = 6.997649808/7 = 0.9996642583$$

Daraus lässt sich nun die Regressionsgerade bestimmen:

$$a = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} = \frac{0.9996642583 - 1.7805 \cdot 0.4978866}{3.173394159 - 1.7805^2} = 35.21480135$$

$$b = \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) - a\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = 0.4978866 - 35.21480135 \cdot 1.7805 = -62.2020672$$

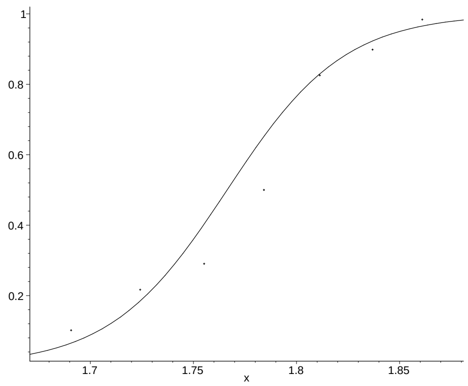
Also lautet die Regressionsgerade $35.21480135x - 62.2020672$, die folgende Abbildung zeigt sie zusammen mit den Daten (x_i, y_i)



Insgesamt haben wir nun mit Hilfe der Logit-Analyse den Zusammenhang

$$p(x) = \frac{\exp(35.21480135x - 62.2020672)}{1 + \exp(35.21480135x - 62.2020672)}$$

gefunden. Zur Güte dieser Berechnung vergleiche folgendes Diagramm:



Es bleibt die Frage nach dem $LD50$-Wert. Gesucht ist also das x mit $p(x) = c = 0.5$. Aus der 4.Übung ist die Umkehrfunktion von logit bekannt. Es gilt

$$x = -\frac{1}{a}\left(\ln\left(\frac{1-c}{c}\right) + b\right) = -\frac{1}{35.21480135}(\ln(1) - 62.2020672) = \frac{62.2020672}{35.21480135} = 1.766361$$

Bei der Konzentration von circa 1.766 sterben also die Hälfte der Käfer. Man beachte, daß aufgrund der Versuchsreihe der $LD50$-Wert bei 1.7842 liegt.

Aufgabe 3 (vorrechnen):

Die Daten der Maus-Elefantenkurve betreffen die Abhängigkeit von Sauerstoffverbrauch und Körpergewicht.

Tierart	Körpergewicht $[g]$	O_2-Verbrauch $[ml/(g\,h)]$
Maus	25	1.65
Erdhörnchen	96	1.03
Ratte	290	0.87
Hund	11700	0.33
Schaf	42700	0.22
Mensch	70000	0.21
Pferd	650000	0.11
Elefant	3833000	0.07

Daten aus: R. Eckert, Tierphysiologie. Thieme, 1986. pp.622

Welcher funktionale Zusammenhang der Form Verbrauch=α Gewicht$^\beta$ beschreibt die obigen Daten ? Welchen Sauerstoffverbrauch hat eine Katze mit einem Gewicht von 2500 g nach diesem Modell ? Der tatsächliche Wert liegt übrigens bei 0.68.

Veranschaulichen Sie sich den Zusammenhang zwischen Verbrauch und Gewicht in einem Koordinatensystem. War obiger Zusammenhang sinnvoll gewählt ?

Lösung:

Wir nehmen also einen funktionalen Zusammenhang der Form Verbrauch=a Gewicht^b an und erkenne, daß wir es mit einer doppeltlogarithmischen Transformation zu tun haben (vgl. das Bachforellenbeispiel aus der Vorlesung). Logarithmieren der Modellgleichung führt auf

$$\ln(\text{Verbrauch}) = \ln(\alpha \text{Gewicht}^\beta) = \ln(\alpha) + \beta \ln(\text{Gewicht}).$$

Setzen wir nun $a = \beta$ und $b = \ln(\alpha)$, so ergibt sich

$$\ln(\text{Verbrauch}) = a \ln(\text{Gewicht}) + b.$$

Zur Berechnung von a und b benötigen wir folgende Tabelle

Tierart	Gewicht $[g]$	O_2-Verb.	$x = \ln(\text{Gewicht})$	$y = \ln(O_2 - \text{Verbrauch})$
Maus	25	1.65	3.218876	0.5007753
Ratte	290	0.87	5.669881	-0.1392621
Hund	11700	0.33	9.367344	-1.108663
Schaf	42700	0.22	10.66195	-1.514128
Mensch	70000	0.21	11.15625	-1.560648
Pferd	650000	0.11	13.38473	-2.207275
Elefant	3833000	0.07	15.15916	-2.65926

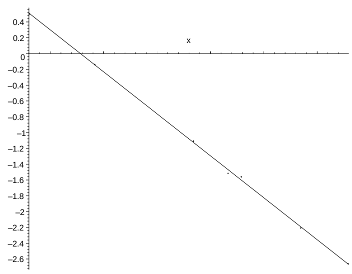
Daraus erhält man:

$$\frac{1}{7}\sum_{i=1}^{7} x_i = 9.802598714, \qquad \frac{1}{7}\sum_{i=1}^{7} y_i = -1.241208686$$
$$\frac{1}{7}\sum_{i=1}^{7} x_i^2 = 111.0494383, \quad \frac{1}{7}\sum_{i=1}^{7} x_i y_i = -16.1390511$$

Nun berechnen sich a und b zu

$$a = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} = -0.2655333953$$
$$b = \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) - a\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = 1.361708633$$

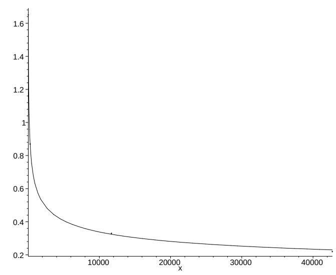
Die Regressionsgerade ist also $y(x) = -0.2655333953x + 1.361708633$ (vgl. folgendes Diagramm):



Daraus ergibt sich mit $\alpha = \exp(b) = 3.902856157$ und $\beta = a = -0.2655333953$ für den Zusammenhang zwischen Verbrauch und Gewicht:

$$\text{Verbrauch} = -0.2655333953\ \text{Gewicht}^{3.902856157}.$$

Man betrachte dazu folgendes Diagramm, wobei hier nur die ersten 4 Daten eingezeichnet sind. Man erkennt die Schwierigkeiten aufgrund der grossen Unterschiede der einzelnen Gewichte. Dies verdeutlicht auch, warum in diesem Fall die doppeltlogarithmische Darstellung so vorteilhaft ist.



Wenn man dieses Modell zu Grunde legt, müsste die Katze einen Verbrauch von $\alpha 2500^{\beta} = 0.4887828934$ im Gegensatz zu dem realen wert von 0.68 haben. Dieser relativ hohe Fehler ist darauf zurüchzufúhren, daß im Gewichtsbereich der Katze wenig Daten vorlagen.

Aufgabe 4 (vorrechnen):

Bei Brillenschötchen (*Biscutella laevigata*) wurden Anzahl der Stengelblätter am Hauptsproß und Sproßhöhe gemessen. Dabei ergab sich folgende Tabelle:

Anzahl Stengelblätter	Sproßhöhe [cm]
3	7
5	13
6	14
8	17
11	20
14	25

Berechnen Sie die lineare Regressionsgerade und tragen diese zusammen mit den Messdaten in ein Koordatensystem ein.

Lösung:

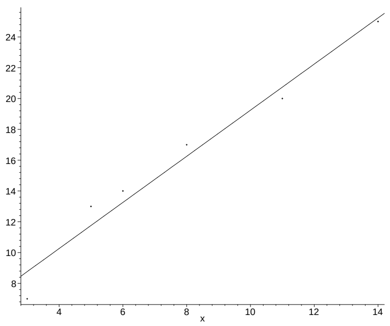
Sei x_i die Anzahl der Stengelblätter und y_i die Sproßhöhe in cm.Wir berechnen

$$\frac{1}{6}\sum_{i=1}^{6} x_i = 47/6 = 7.833333333, \qquad \frac{1}{6}\sum_{i=1}^{6} y_i = 96/6 = 16$$
$$\frac{1}{6}\sum_{i=1}^{6} x_i^2 = 451/6 = 75.16666667, \quad \frac{1}{6}\sum_{i=1}^{6} x_i y_i = 876/6 = 146$$

Darus erhalten wir a und b wie folgt:

$$a = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} = 744/497 = 1.496981891$$
$$b = \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) - a\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = 2124/497 = 4.273641851$$

Damit ergibts ich die Regressionsgerade zu $y(x) = 1.496981891\,x + 4.273641851$. Zusammen mit den Messdaten ergibt sich folgendes Bild:



Aufgabe 5 (Wiederholung):

Eine Bakterienkultur der Masse M_0 werde in einer Nährlösung gezüchtet. Sie vermehre sich täglich um 15%. Zu wissenschaftlichen Testzwecken werde aber auch in täglichen Abständen die Menge M entnommen.

(a) Ermitteln Sie eine allgemeine Berechnungsformel für die Menge M_n der Bakterien nach n Tagen.

(b) Wie groß ist im Falle $M_0 = 21g$, $M = 3g$ die Bakterienmasse nach 20 und nach 30 Tagen? Was passiert für $n \to \infty$? Ist dies realistisch ?

(c) Wie hoch muss die tägliche Entnahme sein, um die Masse der Bakterien konstant zu halten ?

Lösung: Dieser Aufgabe liegt das selbe Prinzip zu Grunde wie in Aufgabe 3 der 3.Übung ("Caesium-Isotop").

(a) Sei $q = 0.15 = 15\%$. Nach einem Tag hat die Bakterienkultur die Masse

$$M_1 = M_0 + qM_0 - M = (1+q)M_0 - M.$$

Nach zwei Tagen gilt

$$M_2 = M_1 + qM_1 - M = (1+q)M_1 - M = (1+q)\left((1+q)M_0 - M\right) - M = (1+q)^2 M_0 - M - (1+q)M.$$

Induktiv folgt

$$M_{n+1} = (1+q)^{n+1} M_0 - M \sum_{k=0}^{n} (1+q)^k$$

Mit Hilfe der geometrischen Reihe folgt

$$M_{n+1} = (1+q)^{n+1} M_0 - \frac{1-(1+q)^{n+1}}{1-(1+q)} M = (1+q)^n M_0 + \frac{1-(1+q)^{n+1}}{q} M$$

(b) Mit obiger Formel berechnen wir

$$M_{20} = 36.3665374, \qquad M_{30} = 86.211772$$

Für $n \to \infty$ wächst die Masse über alle Schranken; denn

$$M_{n+1} = (1+q)^{n+1} M_0 + \frac{1-(1+q)^{n+1}}{q} M = (1+q)^n (M_0 - \frac{M}{q}) + \frac{M}{q}.$$

In unserem Fall ist nun $M_0 - \frac{M}{q} = 20 > 0$ und da $(1+q)^n \to \infty$ für $n \to \infty$ gilt dies auch für M_n. Normalerweise gibt es aufgrund begrenzter Nährstoffressourcen eine obere Schranke für die Masse einer Bakterienkultur. Daher sit dieses Modell unrealistisch. Um den Wachstum zu begrenzen, müsste ein quadratischer Term in Analogie zu der logistischen Differentialgleichung in das Modell eingeführt werden.

(c) Um die Masse der Bakterien konstant zu halten, müsste $M_1 = M_0$ gelten. Also

$$M_1 = (1+q)M_0 - M = M_0 \quad \Leftrightarrow \quad M = qM_0$$

Wählt man also $M = qM_0$, so bleibt die Masse der Bakterienkultur über die Zeit konstant.