Mathematik für BiologInnen
WS 2005

3.Block

A.o. Univ.-Prof. DI Dr. Michael Hintermüller

michael.hintermueller@uni-graz.at

Karl-Franzens Univeristät Graz

Institut für Mathematik und Wissenschaftl. Rechnen

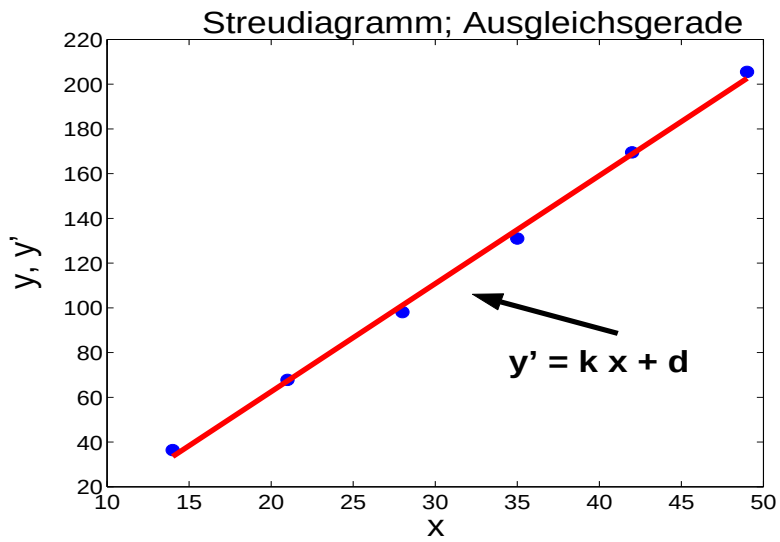
Spezielle Funktionen

Von der Beobachtung zur Funktion.

- Wie hängt y von x ab? Annahme: Datenmaterial aus Beobachtung vorhanden, z.B. in Form einer Wertetabelle.
- Beispiel: Um das Wachstum von Sonnenblumen in Abhängigkeit vom Alter x (in Tagen) zu studieren, wurde die Pflanzenhöhe y (in cm) in regelmäßigen Zeitabständen gemessen:

x/Tage	14	21	28	35	42	49
y/cm	36.4	67.8	98.1	131.0	169.5	205.5

- Veranschaulichung in einem rechtwinkeligen Koordinatensystem (**Streudiagramm**):



Spezielle Funktionen

- Streudiagramme erlauben eine rasche Beurteilung, in welcher Weise die Variablen zusammenhängen.
- Oft läßt die Punkteverteilung im Streudiagramm einen durch eine einfache Kurve beschreibbaren Trend erkennen.
- Die Punkteverteilung wird durch die Trendkurve (Ausgleichskurve) im allgemeinen nicht exakt erfasst.
- Die durch die Ausgleichskurve fixierte Zuordnung nennt man **Ausgleichsfunktion** oder **Regressionsfunktion**.
- In unserem Beispiel ist die Beschreibung des Zusammenhangs zwischen x und y mittels einer linearen Ausgleichsfunktion (**Regressionsgerade**) der Gestalt

$$y' = k \cdot x + d, \quad k, d \in \mathbb{R}$$

  sinnvoll. Im allgemeinen kann der Zusammenhang zwischen x und y auch einem nichtlinearen Trend entsprechen.
- Die freien **Parameter** (k, d in unserem Beispiel) müssen mit Hilfe der gegebenen Daten fixiert werden (**Methode der kleinsten Fehlerquadrate**).

(Affine) Lineare Funktionen

Zwei-Punktform, Interpolation. Sei $f : D \to W$ gegeben. Weiter seien $x_1, x_2 \in D$, $\Delta x = x_2 - x_1$ und $\Delta y = y_2 - y_1 = f(x_2) - f(x_1)$. Charakteristisch für (affine) lineare Funktionen (Geraden) ist, dass der **Differenzenquotient (Anstieg)**

$$k = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

konstant ist. Sei nun $x \neq x_1$, $x \neq x_2$ und $y = f(x)$. Dann gilt:

$$\frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1}.$$

Auflösen nach y liefert die **Zwei-Punktform**

$$y = y_1 + \frac{y_2 - y_1}{x_2 - x_1}(x - x_1).$$

Diese Vorgangsweise entspricht der Lösung einer Interpolationsaufgabe: Von einer Funktion g kennt man an zwei Stellen x_1, x_2 die Funktionswerte $g(x_1), g(x_2)$. Gesucht ist nun eine Näherung des Funktionswerts $g(x)$ an einer Stelle x zwischen x_1 und x_2. Liegt x außerhalb des von x_1 und x_2 definierten Intervalls, so spricht man von **Extrapolation**.

Lineare Funktionen

Beispiel. Die CO_2-Konzentration in der irdischen Atmosphäre belief sich im Jahre 1950 auf 306ppm um im Jahre 1970 auf 321ppm (1ppm $CO_2 = 1\text{cm}^3$ CO_2 pro 1m^3 Luft). Welchen CO_2-Gehalt erhält man für das Jahr 1965 bei linearer Interpolation? Welcher Wert ergibt sich für 1980 bei linearer Extrapolation?

Antwort. Für 1965 erhält man

$$306 + \frac{321 - 306}{1970 - 1950}(1965 - 1950) = 317.25.$$

Für 1980 erhält man

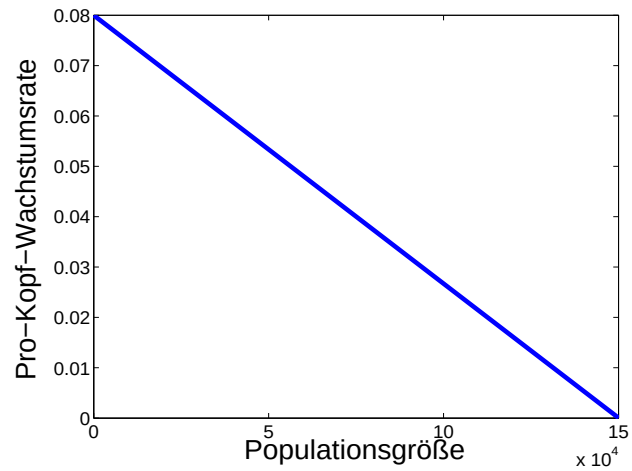
$$306 + \frac{321 - 306}{1970 - 1950}(1980 - 1950) = 328.5.$$

Lineare Funktionen

- **Hauptform.** $y = k \cdot x + d$ mit $d = y_1 - k \cdot x_1$ (y-Achsenabschnitt).
- Die zeitliche Änderung R der Größe N einer Population ist durch $R = rN(1 - N/C)$ gegeben; r ist die natürliche Zuwachsrate und $C > 0$ die Kapazität des Lebensraumes. D.h.: Die sogenannte Pro-Kopf-Wachstumsrate $r' = R/N =$ $= r(1 - N/C)$ bei positivem r ist eine lineare fallende Funktion der Populationsgröße. In der (N, r')-Ebene wird also die Abhängigkeit der Pro-Kopf-Wachstumsrate von der Populationsgröße durch eine fallende Gerade dargestellt. Für eine Blauwalpoluation mit $r = 0.08$ p.a. und $C = 150000$ Tieren ergibt sich

$$r' = (-r/C)N + r \quad \text{(Hauptform)},$$

wobei $k = -r/C = -5.33 \cdot 10^{-7}$ p.a. und $d = r = 0.08$ p.a.



Regressionsgerade

- **Problemstellung.** Wir nehmen an, dass zwischen zwei Größen x und y eine lineare Beziehung besteht. Anstelle von exakten Wertepaaren (x, y) liegen Messungen (x_i, y_i), $i = 1, \ldots, n$, vor. Die Meßpunkte (x_i, y_i) streuen um die wahren Werte mehr oder weniger stark ("Punktwolke"). Das Ziel ist nun, in der (x, y)-Ebene eine Gerade zu finden, welche die gegebenen Punkte (x_i, y_i), $i = 1, \ldots, n$, möglichst gut annähert. Wir wählen also den Ansatz

$$y = k \cdot x + d,$$

wobei k, d noch zu bestimmende Parameter sind.
Jedem x_i-Wert läßt sich so ein y-Wert zuordnen:

$$y_i' = k \cdot x_i + d, \quad i = 1, \ldots n.$$

Die Abweichung vom gegebenen Meßwert y_i wird als Residuum bezeichnet:

$$r_i = y_i - y_i' = y_i - k \cdot x_i - d, \quad i = 1, \ldots, n.$$

Es ist naheliegend, k, d so zu wählen, dass r_i, $i = 1, \ldots, n$, in einem geeigneten Sinn möglichst klein wird.

Regressionsgerade

- **Schätzung der Geradenparameter.** Ein häufig verwendetes Maß für die Güte der Anpassung ist

$$Q(k,d) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - y_i')^2 = \sum_{i=1}^{n} (y_i - kx_i - d)^2.$$

Ziel ist nun, k und d so zu bestimmen, dass Q möglichst klein wird $\Rightarrow$ **Methode der kleinsten Fehlerquadrate**.
Man erhält die Regressionsgerade

$$y' = \frac{s_{xy}}{s_x^2}(x - \bar{x}) + \bar{y},$$

wobei $\bar{x}$ und $\bar{y}$ die Mittelwerte sind und

$$s_x^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right) \quad \text{(Varianz)},$$

$$s_{xy} = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)\right) \quad \text{(Kovarianz)}.$$

Regressionsgerade

- **Beurteilung der Anpassungsgüte.** Die Anpassungsgüte wird durch das Bestimmtheitsmaß

$$B = \left(\frac{s_{xy}}{s_x s_y}\right)^2 = 1 - \frac{S_{rr}}{(n-1)s_y^2}$$

  beurteilt. Dabei ist

$$S_{rr} = (n-1)s_y^2 \left(1 - \left(\frac{s_{xy}}{s_x s_y}\right)^2\right)$$

  der optimale Wert von Q.

- Es gilt: $B \in [0, 1]$.
  - Ist $B = 1$, dann ist $S_{rr} = 0$. Alle Meßpunkte liegen auf der Ausgleichsgeraden.
  - Ist $B = 0$, dann gilt

$$\frac{S_{rr}}{n-1} = s_y^2.$$

    Die Regressionsgerade verläuft parallel zur x-Achse, und die Varation von y kann durch x nicht erklärt werden.

- **Regression durch den Nullpunkt**. In diesem Fall wählen wir den Ansatz $y = k \cdot x$.