Philipps-Universität Marburg
Fb Mathematik u. Informatik

Prof. Dr. Hajo Holzmann
Daniel Hohmann
Jochen Wilmsmeyer

WS 2013/14

Elementare Stochastik

- Blatt 8 -

Abgabe: Fr, 13.12.13

Aufgabe 38 (4 Punkte)**.** Gegeben sind zwei Urnen mit je N Kugeln, wobei die Kugeln in Urne 1 in L sowie die Kugeln in Urne 2 in R unterscheidbare Gruppen unterteilt werden können. Die Anzahl der Kugeln in den jeweiligen Gruppen sei $l_1, \ldots, l_L$ bzw. $r_1, \ldots, r_R$. Sie ziehen n Mal ohne Rücklegen je eine Kugeln aus jeder Urne.

a. Bestimmen Sie die gemeinsame Verteilung der Zufallsvariablen

$$X_{i,j} : \Omega_{II}^{n,N} \times \Omega_{II}^{n,N} \to \mathbb{N}_0, \qquad i = 1, \ldots L,\ j = 1, \ldots, R,$$

die zählen, wie oft bei den n Durchgängen eine Kugel aus Gruppe i zusammen mit einer Kugel aus Gruppe j gezogen wurde.

b. Zeigen Sie insbesondere, dass im Fall $n = N$ für $x_{ij} \in \mathbb{N}_0$ mit $\sum_{i,j} x_{ij} = n$ gilt

$$\mathrm{P}\Big(\bigcap_{i=1}^{L} \bigcap_{j=1}^{R} \{X_{i,j} = x_{i,j}\}\Big) = \binom{N}{r_1 \ldots r_R}^{-1} \prod_{i=1}^{L} \binom{l_i}{x_{i1} \ldots x_{iR}}.$$

Aufgabe 39 (4 Punkte, L)**.** Seien X und Y Zufallsvariablen mit $\mathrm{E}X^2, \mathrm{E}Y^2 < \infty$.

a. Lösen Sie das Optimierungsproblem

$$\arg\min_{a \in \mathbb{R}} \ \mathrm{E}\big((Y - aX)^2\big).$$

b. Folgern Sie

$$\big(\mathrm{E}(XY)\big)^2 \leq \mathrm{E}(X^2)\mathrm{E}(Y^2),$$

wobei Gleichheit genau dann gilt, wenn ein $a \in \mathbb{R}$ existiert mit $\mathrm{P}(Y = aX) = 1$.

Aufgabe 40 (4 Punkte)**.** Seien $\mathcal{M} = \{1, \ldots, m\}$ ein endliches Alphabet sowie X und Y Zufallsvariablen auf $\mathcal{M}$ mit gemeinsamer Verteilung p und Randverteilungen p_1 und p_2, d.h.

$$\mathrm{P}(X = x, Y = y) = \mathrm{p}(x, y), \quad \mathrm{P}(X = x) = \mathrm{p}_1(x), \quad \mathrm{P}(Y = y) = \mathrm{p}_2(y), \qquad x, y \in \mathcal{M}.$$

Die *gegenseitige Information* von X und Y ist definiert als

$$I(X, Y) = \sum_{x,y \in \mathcal{M}} \mathrm{p}(x, y) \log \frac{\mathrm{p}(x, y)}{\mathrm{p}_1(x)\mathrm{p}_2(y)}.$$

Zeigen Sie: $I(X, Y)$ ist nichtnegativ und genau dann Null, wenn X und Y unabhängig sind.

Aufgabe 41 (4 Punkte)**.** Wir betrachten die Alphabete $\mathcal{M}_k = \{1, \ldots, k\}$.

a. Für welche $k \in \mathbb{N}$ existiert ein Präfix-Code C auf $\mathcal{M}_k$ mit $l_C(x) = x$?

b. Konstruieren Sie einen Präfix-Code C auf

- $\mathcal{M}_5$ mit $l_C(1) = 2$, $l_C(2) = 3$, $l_C(3) = 4$ und $l_C(4) = l_C(5) = 2$.
- $\mathcal{M}_6$ mit $l_C(x) = 3$ für alle $x \in \mathcal{M}_6$.

c. Beschreiben Sie ein allgemeines Verfahren zur Konstruktion eines Präfix-Codes bei vorgegebenen Codelängen l_C, die die Kraft-Ungleichung erfüllen.

Aufgabe 42 (3 Punkte, L)**.** Wie oft muss man im Mittel mit einem fairen Würfel würfeln, bis man alle sechs möglichen Augenzahlen erhalten hat?

Aufgabe 43 (5 Punkte, L)**.** In dem Schulbuch 'Neue Wege - Arbeitsbuch für Gymnasien, Stochastik' (Schroedel-Verlag, 2012) findet man als sogenanntes Basiswissen eine Anleitung zur Anpassung von Geraden an gegebene Daten (siehe unten). Als Übung erscheint zwei Seiten später folgendes Beispiel:

Die Grille als Thermometer

Reiben Grillen ihre vorderen Flügel aneinander, so entsteht ein Ton, sie „zirpen".
Man vermutet eine Beziehung zwischen der Häufigkeit des Zirpens (Zirprate) und der Temperatur in der Umgebung. In der Tabelle sind einige Messdaten festgehalten.

Temperatur in °C	16	18	19	20	22	23	25
Anzahl des Zirpens pro 15 s	20	21	23	27	30	34	39

a. Führen Sie für dieses Beispiel folgende Schritte aus der o.g. Anleitung durch:

- Streudiagramm zeichnen
  Erproben Sie dabei die Wirkung von mindestens zwei verschiedenen Skalierungen der Achsen.
- Methode nach Augenmaß
  Bestimmen Sie anschließend die Gleichung der gewählten Gerade durch Ablesen.
- Methode der Kleinsten Quadrate
  Nutzen Sie die Ergebnisse von Aufgabe 36.
- Residuendiagramm

b. Quasi als Fortsetzung des letzten Schrittes aus Teil a soll die Korrelation qualitativ beurteilt werden. Berechnen Sie dazu den folgenden sog. *empirischen Korrelationskoeffizienten*

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \,,$$

wobei $\bar{x}$ und $\bar{y}$ die empirischen Mittel der Daten x_i und y_i seien.

Faustregel: Je näher $|r|$ an 1 liegt, umso stärker ist der lineare Zusammenhang.

Basiswissen **Anpassung von Geraden an Daten – Regressionsgerade**

Ziel

Beziehungen zwischen zwei quantitativen Merkmalen X und Y entdecken, untersuchen und beschreiben.

Besteht ein Zusammenhang zwischen Gewicht und Körpergröße von Fußballern?

Vermutung

Oft liegt aus dem Sachzusammenhang eine Vermutung vor.

Je größer, desto schwerer. Linearer Zusammenhang?

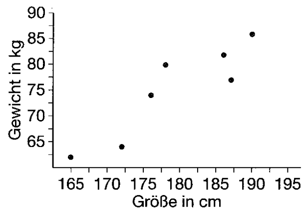
Datentabelle erstellen

Zur Stützung der Vermutung werden Beobachtungen/Messungen vorgenommen und in einer Tabelle festgehalten.

Größe in cm	190	187	186	165	172	176	178
Gewicht in kg	86	77	82	62	64	74	80

Streudiagramm zeichnen

Zur Tabelle wird ein Streudiagramm erstellt. Ein Paar (x|y) von Messwerten erscheint als Punkt im Diagramm, alle Paare bilden eine „Punktwolke".

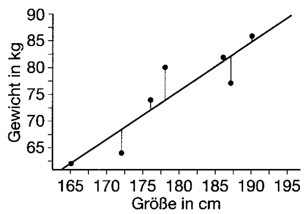


Modellierung des vermuteten linearen Zusammenhangs mit einer Ausgleichsgeraden

Zur Anpassung der Geraden gibt es mehrere Methoden. Dabei liefern die vertikalen Abstände der Datenpunkte von der Ausgleichsgeraden („Residuen") wertvolle Hilfe.

Methode nach Augenmaß

Nach Augenmaß wird eine Gerade eingefügt, die von den Datenpunkten „möglichst wenig" abweicht.



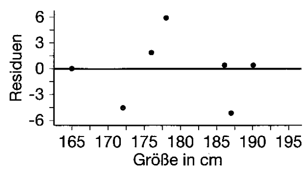
Methode der kleinsten Quadrate

Man berechnet die Steigung a und den Achsenabschnitt b der *Regressionsgeraden* $y = ax + b$ so, dass die Summe der Quadrate der Residuen minimiert wird.

Diese Methode wird in der Statistik favorisiert. Die danach bestimmte Ausgleichsgerade wird als **Regressionsgerade** bezeichnet. GTR und Statistik-Software können **Regressionsgeraden** berechnen.

Residuendiagramm als Hilfe zur Anpassung

Hier werden die vertikalen Abstände der Datenpunkte von der Ausgleichsgeraden eigens dargestellt. Bei einer guten Anpassung sollten die Residuen möglichst nahe an der Nulllinie liegen, sich nach oben und unten ausgleichen und kein systematisches Muster zeigen.



Interpretation und Auswertung

Im Modell kann zu einer Merkmalsausprägung x die zugehörige Merkmalsausprägung y berechnet werden. Abweichungen können festgestellt und untersucht werden.

Ein 1,81 m großer Spieler hat ein Gewicht von durchschnittlich 76,9 kg. Der 1,78 m große Spieler ist mit 80 kg vergleichsweise schwer.

Prognosen können erstellt werden, insbesondere bei Zeitreihen. Diese sind im Allgemeinen nur in der Nähe des erhobenen Datenbereichs möglich.

Wichtig:
Ein linearer Zusammenhang bei realen Daten kann rein statistisch sein. Ob dies auch ein kausaler Zusammenhang ist, bedarf zusätzlicher Untersuchungen von Sachexperten.

Hinweis

Bachelor- und Masterstudenten bearbeiten die Aufgaben 38-41. Lehramtsstudenten bearbeiten die drei mit 'L' gekennzeichneten Aufgaben sowie eine weitere Aufgaben nach Wahl.