Prof. Dr. Rupert Lasser
Paul Bergold

Technische Universität München
Zentrum Mathematik

Stochastik für Lehramt Gymnasium – Blatt 6

Wintersemester 2016/17

Lösungshinweise

Hausaufgabe 17

Die Auswertung von Prüfungsergebnissen im Studiengang A und B wurden in folgenden Kontingenztafeln zusammengefasst:

Studiengang A

Examensnote (Y), Semesterzahl (X)

X\Y	1	2	3	4	5	$\sum$
8	3	1	0	0	0	4
9	2	15	39	12	0	68
10	0	1	8	6	5	20
11	0	0	1	1	2	4
12	0	0	2	1	1	4
$\sum$	5	17	50	20	8	100

Studiengang B

Examensnote (Y), Semesterzahl (X)

X\Y	1	2	3	4	5	$\sum$
8	2	1	4	3	0	10
9	5	2	9	6	1	23
10	4	2	8	6	0	20
11	0	1	2	2	1	6
12	1	2	4	3	1	11
$\sum$	12	8	27	20	3	70

1. Berechnen Sie jeweils die mittleren quadratischen Abweichungen σ_X^2 und σ_Y^2.
2. Wie lautet jeweils r_{XY}. Interpretieren Sie die Ergebnisse.

Lösung zu Hausaufgabe 17

Teilaufgabe 1: **Studiengang A**:

$$\bar{x} = \frac{936}{100} = 9.36, \qquad \overline{x^2} = \frac{8824}{100} = 88.24 \qquad \Rightarrow \qquad \sigma_X^2 = \overline{x^2} - \bar{x}^2 \approx 0.6304,$$
$$\bar{y} = \frac{309}{100} = 3.09, \qquad \overline{y^2} = \frac{1043}{100} = 10.43 \qquad \Rightarrow \qquad \sigma_Y^2 = \overline{y^2} - \bar{y}^2 \approx 0.8819.$$

Studiengang B:

$$\bar{x} = \frac{685}{70} \approx 9.7857, \qquad \overline{x^2} = \frac{6813}{70} \approx 97.3286 \qquad \Rightarrow \qquad \sigma_X^2 = \overline{x^2} - \bar{x}^2 \approx 1.5687,$$
$$\bar{y} = \frac{204}{70} \approx 2.9143, \qquad \overline{y^2} = \frac{682}{70} \approx 9.7429 \qquad \Rightarrow \qquad \sigma_Y^2 = \overline{y^2} - \bar{y}^2 \approx 1.2498.$$

Teilaufgabe 2: **Studiengang A**:

$$\|\delta_X\|^2 = n\sigma_X^2 = 63.04, \qquad \|\delta_Y\|^2 = n\sigma_Y^2 = 88.19,$$
$$\langle X, Y\rangle = \sum_{i=1}^{n} x_i y_i = \sum_{j,k} h(X = a_j, Y = b_k) a_j b_k = 2930,$$
$$\langle \delta_X, \delta_Y\rangle = \langle X, Y\rangle - n\bar{x}\bar{y} = 37.76,$$
$$r_{XY} = \frac{\langle \delta_X, \delta_Y\rangle}{\|\delta_X\|\|\delta_Y\|} \approx 0.5064.$$

Studiengang B:

$$\begin{aligned}
\|\delta_X\|^2 &= n\sigma_X^2 \approx 109.809, \quad \|\delta_Y\|^2 = n\sigma_Y^2 \approx 87.486,\\
\langle X, Y\rangle &= \sum_{i=1}^{n} x_i y_i = \sum_{j,k} h(X = a_j, Y = b_k) a_j b_k = 2008,\\
\langle \delta_X, \delta_Y\rangle &= \langle X, Y\rangle - n\bar{x}\bar{y} \approx 11.7143,\\
r_{XY} &= \frac{\langle \delta_X, \delta_Y\rangle}{\|\delta_X\|\|\delta_Y\|} \approx 0.1195.
\end{aligned}$$

Interpretation:
Für Studiengang A gilt $r_{XY} \approx 0.5064$ (mittlere Korrelation). Somit hat Y den Trend, mit wachsendem X anzusteigen. Hingegen ist $r_{XY} \approx 0.1195$ (schwache Korrelation) für Studiengang B. Folglich ist kein Trend für eine lineare Abhängigkeit zu erkennen.

Hausaufgabe 18
Betrachten Sie die Merkmale $X = (x_1, x_2)^T, Y = (y_1, y_2)^T \in \mathbb{R}^2$.

1. Bestimmen Sie δ_X und δ_Y.
2. Berechnen Sie $\langle \delta_X, \delta_Y\rangle$, sowie $\|\delta_X\| \cdot \|\delta_Y\|$.
3. Folgern Sie, dass $r_{XY} \in \{-1, 1\}$. Wie lässt sich diese Aussage geometrisch begründen? (Hinweis: Für ein beliebiges Merkmal $X \in \mathbb{R}^n$ gilt $\delta_X \perp \mathbf{1}$, d.h. der Abweichungsvektor steht senkrecht auf der Geraden $\mathbb{R} \cdot \mathbf{1}$.)

Lösung zu Hausaufgabe 18
Teilaufgabe 1: Es gilt

$$\bar{x} = \frac{x_1 + x_2}{2}, \quad \bar{y} = \frac{y_1 + y_2}{2}.$$

Somit folgt

$$\begin{aligned}
\delta_X &= (x_1 - \bar{x}, x_2 - \bar{x})^T = \frac{1}{2}(x_1 - x_2, x_2 - x_1)^T,\\
\delta_Y &= (y_1 - \bar{y}, y_2 - \bar{y})^T = \frac{1}{2}(y_1 - y_2, y_2 - y_1)^T.
\end{aligned}$$

Teilaufgabe 2: Wir berechnen weiter:

$$\begin{aligned}
\langle \delta_X, \delta_Y\rangle &= \langle X, Y\rangle - n\bar{x}\bar{y} = x_1y_1 + x_2y_2 - 2 \cdot \frac{x_1 + x_2}{2}\frac{y_1 + y_2}{2}\\
&= \frac{1}{2}(x_1y_1 + x_2y_2 - x_1y_2 - x_2y_1) = \frac{1}{2}(x_1 - x_2)(y_1 - y_2),\\
\|\delta_X\|^2 &= \frac{1}{4}\left((x_1 - x_2)^2 + (x_2 - x_1)^2\right) = \frac{1}{2}(x_1 - x_2)^2,\\
\|\delta_Y\|^2 &= \frac{1}{2}(y_1 - y_2)^2,\\
\Rightarrow \quad \|\delta_X\|\|\delta_Y\| &= \frac{1}{2}|x_1 - x_2||y_1 - y_2| = \frac{1}{2}|(x_1 - x_2)(y_1 - y_2)|.
\end{aligned}$$

Teilaufgabe 3: Sei nun $x_1 \neq x_2$ sowie $y_1 \neq y_2$. Dann erhalten wir mit den Ergebnissen aus Teilaufgabe 1 und 2:

$$r_{XY} = \frac{\langle \delta_X, \delta_Y \rangle}{\|\delta_X\|\|\delta_Y\|} = \frac{(x_1 - x_2)(y_1 - y_2)}{|(x_1 - x_2)(y_1 - y_2)|} \in \{-1, 1\}.$$

In $\mathbb{R}^2$ sind die Abweichungsvektoren δ_X und δ_Y linear unabhängig, da sowohl δ_X als auch δ_Y senkrecht auf der Geraden $\mathbb{R} \cdot \mathbf{1}$ (hier $y = x$) steht. Somit schließen die Vektoren einen Winkel von $\varphi = 0$ oder $\varphi = \pi$ ein. Da $\cos(0) = 1$ bzw. $\cos(\pi) = -1$ folgt $r_{XY} \in \{-1, 1\}$.

Alternativ kann man begründen:
Da durch die Punkte (x_1, y_1) und (x_2, y_2) die Gerade $R_{Y(X)}$ eindeutig festgelegt wird, besteht stets ein linearer Zusammenhang, also $|r_{XY}| = 1$, zwischen den Merkmalen X und Y. Insbesondere gilt für die Steigung der Trendgeraden

$$a^* = \frac{\langle \delta_X, \delta_Y \rangle}{\|\delta_X\|^2} = \frac{(x_1 - x_2)(y_1 - y_2)}{(x_1 - x_2)^2} = \frac{y_1 - y_2}{x_1 - x_2}.$$

Dies ist die gewohnte Formel zur Berechnung der Steigung einer Geraden durch die Punkte (x_1, y_1) und (x_2, y_2).