Bestätigung von Hypothesen heranzuziehen. Sie lernten vor allen Dingen auch, vorsichtig zu sein beim Überqueren der Straße.

	verkehrs-reiche Stadtumgebung		verkehrsreiche Landumgebung		Mittlere Umge-bung	ruhige Land-Umgebung		sehr ruhige Gegend
Schule	A	B	C	D	E	F	G	H
Ja (würde die Straße überque-ren)	72%	48%	46%	4%	82%	28%	70%	16%
Nein (würde die Straße nicht überqueren)	28%	52%	54%	96%	18%	72%	30%	84%
Entscheidung in weniger als 2 Sekunden	72%	55%	65%	77%	77%	65%	95%	96%
Entscheidung in mehr als 2 Se-kunden	28%	45%	35%	23%	23%	35%	5%	4%
würde hinüber-laufen	19%	14%	0%	4%	0%	0%	0%	8%
zu vorsichtig, hätte mehr als 3 Sekunden Si-cherheit zum Überqueren der Straße, aber tut es nicht	0%	0%	0%	0%	41%	20%	15%	4%
fällt die richtige Entscheidung	81%	86%	100%	96%	59%	80%	85%	88%
Anteil der Nein-Sager, die hin-überlaufen wür-den, wenn je-mand "Ja" sagte	78%	93%	21%	32%	17%	21%	0%	71%
Anteil der Nein-Sager, die dann nicht hinüberlaufen würden	11%	0%	21%	24%	6%	0%	0%	0%

Tabelle 1

Eigenschaften des Korrelationskoeffizienten

Nick Lord, übersetzt von Karl Röttel

Zusammenfassung: Es wird ein auf Ungleichheitsbeziehungen zwischen arithmetischem und geometrischem Mittel beruhender Weg zur Behandlung von Eigenschaften des Korrelationskoeffizienten vorgestellt.

Die folgende Einführung in die Eigenschaften des Produkt-Moment-Korrelationskoeffizienten r ist eine Abänderung der üblichen Berechnungen, die für Schülerniveau passend erscheint. Vor allem wollen wir zeigen, daß r seinen Dienst als Maß für den linearen Zusammenhang zwischen zwei Variablen erfüllt, indem dargelegt wird, daß r zwischen -1 und +1 liegt, und daß r gleich ± 1 nur im Falle vollständiger linearer Abhängigkeit beider Variablen ist. Unser Beweis dafür benötigt nicht mehr als die Ungleichheitsbeziehung von arithmetischem und geometrischem Mittel und scheint weniger abschreckend als die gewöhnlichen Beweiswege in den Lehrbüchern zu sein.

Wie immer motivieren wir die Einführung der Kovarianz $\operatorname{cov}(x,y) = s_{xy}$ durch die Beobachtung, daß das Vorzeichen von $(x - \bar{x}) \cdot (y - \bar{y})$ die Koordinatenebene so in Quadranten einteilt (Abb. 1), daß jeweils $s_{xy} \gg 0$, $s_{xy} \ll 0$, $s_{xy} \approx 0$ für die drei in Abb. 2 dargestellten Verteilungen der Daten ist.

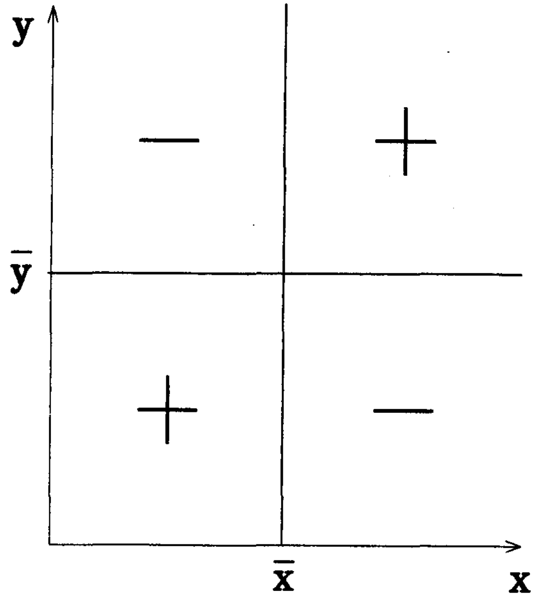
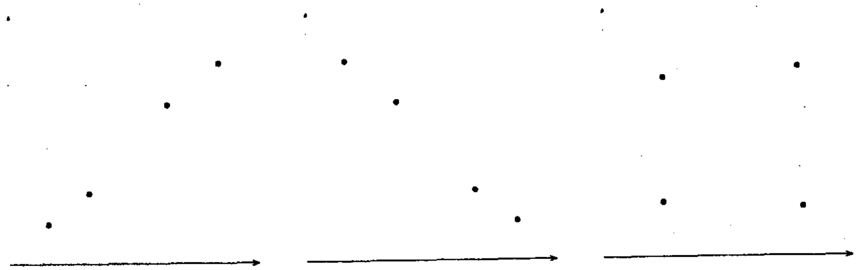


Abb. 1 Einteilung der Koordinatenebene



$s_{xy} >> 0$ $s_{xy} << 0$ $s_{xy} \approx 0$

Abb. 2 Prinzipskizze Verteilungen

(Vorbereitendes Erarbeiten der Verschlüsselungen oder Standardisierungen von Normalverteilungen kann über diesen etwas schwierigen Schritt hinweghelfen.)

Wir definieren den Korrelationskoeffizienten r als Kovarianz der standardisierten Werte:

$$\operatorname{cov}\left(\frac{(x-\bar{x})}{s_x}, \frac{(y-\bar{y})}{s_y}\right) = \frac{1}{n}\left(\sum \frac{(x-\bar{x})}{s_x} \cdot \frac{(y-\bar{y})}{s_y}\right) = \frac{s_{xy}}{s_x \cdot s_y}$$

Um die Eigenschaften von r herzuleiten, benötigen wir als einziges algebraisches Werkzeug die Ungleichheitsbeziehung zwischen den Mittelwerten in der einfachsten Gestalt, d. h.:

Für zwei beliebige Zahlen a und b ist

$$-\frac{1}{2}(a^2 + b^2) \leqq ab \leqq \frac{1}{2}(a^2 + b^2),$$

wobei eines der Gleichheitszeichen genau dann gilt, wenn a = b oder a = -b. Der Beweis erfolgt durch Umordnen von $(a \pm b)^2 \geqq 0$.

Wenden wir dies auf die standardisierten Wertepaare an, erhalten wir

$$-\frac{1}{2}\left(\frac{(x-\bar{x})^2}{s_x^2} + \frac{(y-\bar{y})^2}{s_y^2}\right) \leqq \frac{(x-\bar{x})(y-\bar{y})}{s_{xy}} \leqq \frac{1}{2}\left(\frac{(x-\bar{x})^2}{s_x^2} + \frac{(y-\bar{y})^2}{s_y^2}\right)$$

und nach Summierung über alle Zahlenpaare

$$-\frac{1}{2}(n+n) \leqq \frac{n \cdot s_{xy}}{s_x \cdot s_y} \leqq \frac{1}{2}(n+n),$$

womit $-1 \leqq r \leqq 1$ gezeigt ist.

Zudem gilt Gleichheit rechts und links in der letzten Ungleichung genau dann, wenn sich rechts und links dieselben Terme gegenüberstehen, was bedeutet, daß entweder

$$\frac{(x-\bar{x})}{s_x} = \frac{(y-\bar{y})}{s_y}$$

oder

$$\frac{(x-\bar{x})}{s_x} = -\frac{(y-\bar{y})}{s_y}$$

für alle Wertepaare ist, d. h. ein vollständiger linearer Zusammenhang von x und y besteht.

So ist r eine derart standardisierte Zahl, daß

$$-1 \leqq r \leqq 1$$

gilt, womit wir voll berechtigt sind, sie als Maß für das Ausmaß des linearen Zusammenhanges zwischen zwei Variablen zu betrachten, da sie die "extremen" Werte +1 und -1 genau dann annimmt, wenn die Variablen in linearer Beziehung stehen.

Bemerkung: Ein Beweis der Eigenschaften von r ist im Wesen gleichbedeutend mit einem Beweis der Cauchy-Schwarzschen Ungleichung; so kann das oben Vorgeführte als Beweis für die Ungleichheitsbeziehung der Mittelwerte gelten. Tatsächlich wurde unser Beweis durch einen der Beweise der Hölderschen Ungleichung, einer Verallgemeinerung der Cauchy-Schwarzschen Ungleichung, inspiriert.

Statistik für deutsche Hochschulen

Information der Zentralvergabe für Studienplätze - Sommersemester 91

"...

3.16 **Durchschnittsnote**

Tragen Sie hier die allgemeine Durchschnittsnote Ihrer Hochschulzugangsberechtigung bzw. Qualifikation ein. Die Durchschnittsnote muß auf eine Stelle nach dem Komma bestimmt sein, andernfalls ist eine gesonderte Bescheinigung erforderlich.

..."