Zusammenhang zwischen Meßgrößen

- Kovarianz
- Korrelationskoeffizient
- <u>Lineare Regression</u>
  - <u>Regression Y auf X</u>
    - Lagebestimmung: Abstandsminimierung
    - Berechnung der Koeffizienten
  - Regression X auf Y

Kovarianz

Seien $X = \{x_1, \ldots, x_n\}$ und $Y = \{y_1, \ldots, y_n\}$ zwei gleichverteilte diskrete Zufallsvariablen. Dann stellt die Kovarianz **kov (X, Y)** ein Maß für die gegenseitige (statistische) Abhängigkeit der beiden Zufallsgrößen X und Y dar. Es gilt:

$$\text{kov}(X, Y) = \frac{1}{n} \sum_{j=1}^{n} (x_j - E(X))\,(y_j - E(Y))$$

Korrelationskoeffizient

Der Korrelationskoeffizient **r** wird durch die Division der Kovarianz von **X** und **Y** durch die Standardabweichungen **s(X)** und **s(Y)** berechnet:

$$r = \frac{\mathrm{kov}\,(X, Y)}{s\,(X) * s\,(Y)}$$

Normierung

Die Kovarianz **kov (X, Y)** wird mittels Division durch die Standardabweichungen **s(X)** und **s(Y)** normiert und nimmt Werte zwischen -1 und + 1 an. Es gilt:

$$-1 <= r <= +1$$

Interpretation des Korrelationskoeffizienten (1)

$r = 1$	vollständige positive Korrelation
$0{,}9 < r < 1$	sehr hohe positive Korrelation
$0{,}7 < r <= 0{,}9$	hohe positive Korrelation
$0{,}5 < r <= 0{,}7$	mittlere positive Korrelation
$0{,}2 < r <= 0{,}5$	geringe positive Korrelation
$0 < r <= 0{,}2$	sehr geringe positive Korrelation
$r = 0$	keine Korrelation

Interpretation des Korrelationskoeffizienten (2)

$r = -1$	vollständige negative Korrelation
$-1 < r < -0{,}9$	sehr hohe negative Korrelation
$-0{,}9 <= r < -0{,}7$	hohe negative Korrelation
$-0{,}7 <= r < -0{,}5$	mittlere negative Korrelation
$-0{,}5 <= r < -0{,}2$	geringe negative Korrelation
$-0{,}2 <= r < 0$	sehr geringe negative Korrelation

Regressionsgerade

Wenn die Korrelation (d.h. *der Betrag des Korrelationskoeffizienten*) hinreichend hoch ist, dann darf ein **linearer** Zusammenhang zwischen den Zufallsvariablen **X** und **Y** angenommen werden, d.h. es darf eine lineare Funktion unterstellt werden, die den Zusammenhang der unabhängigen Zufallsvariablen von der abhängigen Zufallsvariablen ausdrückt.

Gleichung der Regressionsgeraden (1)

Wenn Y die abhängige und X die unabhängige Variable ist, dann liegt Regression von Y auf X vor. Dann besitzt die Regressionsgerade die Gleichung:

$$Y = a_1 * X + a_0$$

Optimierung

Die Koeffizienten der Regressionsgeraden $\mathbf{Y = a_1 X + a_0}$ sind so zu bestimmen, daß die Summe der Abstände der Regressionsgeraden zu den Punkten $(\mathbf{x_j, y_j})$ minimal wird (→Extremwertberechnung: Minimum). Die optimale Lage der Regressionsgeraden zu den vorgegebenen Meßpunkten $(\mathbf{x_j, y_j})$ ist gegeben, wenn gilt:

$$a_1 = \frac{\mathrm{kov}\,(X, Y)}{s(X)^2}$$

$$a_0 = E(Y) - a_1 * E(X)$$

Aufgabe 1

In einem Kaufhauskonzern mit n= 10 Filialen sollen die Auswirkungen von Werbeausgaben auf die Umsatzsteigerung untersucht werden. Wir betrachten die Merkmale X „Ausgaben für Werbung“ mit 1000 € als Einheit und Y „Umsatzsteigerung“ mit 1000 € als Einheit.

i	x_i	y_i
1	1,5	20,0
2	2,0	30,0
3	3,5	60,0
4	2,5	50,0
5	0,5	10,0
6	4,5	60,0
7	4,0	50,0
8	5,5	110,0
9	7,5	140,0
10	8,5	170,0

1. Berechnen Sie den Korrelationskoeffizienten und begründen Sie, daß eine hohe Abhängigkeit zwischen den Merkmalen X und Y besteht !
2. Unterstellen Sie einen linearen funktionalen Zusammenhang der Umsatzsteigerung in Abhängigkeit von den Ausgaben für Werbung und berechnen Sie die Gleichung der Regressionsgeraden !

Aufgabe 2

Bei einem Testverkauf werden folgende Indexwerte ermittelt. Der Mengenindex **X** und die zugehörigen Verkaufspreise **Y** sind in der folgenden Tabelle dargestellt:

i	y_i [€]	x_i
1	2.70	73
2	2,50	81
3	2,30	81
4	2,30	85
5	2,10	89
6	2,00	95
7	2.00	99
8	1,95	109
9	1,80	113
10	1.80	120

Ermitteln Sie die (lineare) Nachfragefunktion !

Gleichung der Regressionsgeraden (2)

Wenn X die abhängige und Y die unabhängige Variable ist, dann liegt Regression von X auf Y vor. Dann besitzt die Regressionsgerade die Gleichung:

$$X = b_1 * Y + b_0$$

Optimierung

Die Koeffizienten der Regressionsgeraden $\mathbf{X = b_1\,Y + b_0}$ sind so zu bestimmen, daß die Summe der Abstände der Regressionsgeraden zu den Punkten $(\mathbf{x_j}, \mathbf{y_j})$ minimal wird (→Extremwertberechnung: Minimum). Die optimale Lage der Regressionsgeraden zu den vorgegebenen Meßpunkten $(\mathbf{x_j}, \mathbf{y_j})$ ist gegeben, wenn gilt:

$$b_1 = \frac{\mathrm{kov}\,(X, Y)}{s(Y)^2}$$

$$b_0 = E(X) - b_1 * E(Y)$$