Vorlesung 7b

Korrelationskoeffizient und Regressionsgerade

Teil 1

Die Varianz-Kovarianz-Ungleichung

(Buch S. 56)

Wir erinnern an die

Definition der Kovarianz:

Für reellwertige Zufallsvariable X, Y

mit $\mathbf{E}[X^2] < \infty$ und $\mathbf{E}[Y^2] < \infty$ ist

$$\mathbf{Cov}[X, Y] := \mathbf{E}\Big[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])\Big] .$$

Insbesondere ist also

$$\mathbf{Cov}[X, X] = \mathbf{Var}[X].$$

Wichtige Eigenschaften der Kovarianz:

Die Kovarianz ist

- im Fall von zwei gleichen Einträgen *nichtnegativ*:

$$\mathbf{Cov}[X, X] \geq 0$$

- in den beiden Einträgen *symmetrisch*:

$$\mathbf{Cov}[X, Y] = \mathbf{Cov}[Y, X]$$

- *bilinear*, d.h. in jedem einzelnen Eintrag linear:

$$\mathbf{Cov}[c_1 X_1 + c_2 X_2, Y] = c_1 \mathbf{Cov}[X_1, Y] + c_2 \mathbf{Cov}[X_2, Y]$$

Wir beweisen jetzt die

Kovarianz-Varianz-Ungleichung:

$$|\mathbf{Cov}[X, Y]| \le \sqrt{\mathbf{Var}X}\,\sqrt{\mathbf{Var}Y}$$

Mit $G := X - \mu_X, \;\; H := Y - \mu_Y$

folgt die Behauptung aus der

Cauchy-Schwarz Ungleichung:

Für reellwertige Zufallsvariable G, H mit $\mathbf{E}[G^2], \mathbf{E}[H^2] < \infty$

ist

$$\Big|\mathbf{E}[GH]\Big| \le \sqrt{\mathbf{E}[G^2]}\,\sqrt{\mathbf{E}[H^2]}\,.$$

Beweis der Cauchy-Schwarz Ungleichung:

Fall 1: $\mathbf{E}[G^2] > 0$ und $\mathbf{E}[H^2] > 0$.

Zu zeigen ist $\left| E\left[\frac{G}{\sqrt{E[G^2]}} \frac{H}{\sqrt{E[H^2]}} \right] \right| \leq 1.$

Für $U := G/\sqrt{\mathbf{E}[G^2]}$, $V := H/\sqrt{\mathbf{E}[H^2]}$

ist also zu zeigen: $|\mathbf{E}[UV]| \leq 1.$

Aus $\pm 2UV \leq U^2 + V^2$ folgt (mit Monotonie des EW)

$$\pm \mathbf{E}[UV] \leq \frac{1}{2}(\mathbf{E}[U^2] + \mathbf{E}[V^2] = 1.$$

Fall 2: $\mathbf{E}[G^2] = 0$.

Dann folgt

mit dem Satz von der Positivität des Erwartungswertes:

$$\mathbf{P}(G^2 = 0) = 1,$$

also

$$\mathbf{P}(GH = 0) = 1$$

und

$$\mathbf{E}[GH] = 0,$$

und damit

$$\pm\mathbf{E}[GH] = 0 \leq 0 = \sqrt{\mathbf{E}[G^2]}\sqrt{\mathbf{E}[H^2]}. \square$$