Beratungspraxis

Eine Schranke für die Korrelation zweier Zufallsvariablen bei gegebener Korrelation mit einer dritten

von **Ralf Runde**, Dortmund

FACHÜBERGREIFENDE BERATUNG war mein erster Gedanke, als neulich ein befreundeter Wirtschaftswissenschaftler bei mir anrief. Er hätte da ein statistisches Problem (was sonst?). Er hätte trotz wochenlangem Bücherwälzen ***absolut*** nichts in der einschlägigen Literatur gefunden und ich wäre seine letzte Hoffnung.

Seine Frage: bei drei reellwertigen Zufallsvariablen X, Y und Z mit endlicher Varianz seien zwei Korrelationskoeffizienten bekannt, sagen wir ρ_{XY} und ρ_{YZ}. Kann man aufgrund dieser Kenntnis irgendwelche Aussagen über die (unbekannte) Korrelation zwischen X und Z, also ρ_{XZ} machen?

Um die Antwort vorwegzunehmen: es lassen sich leicht nichttriviale untere und obere Schranken für ρ_{XZ} bestimmen und zwar so:

Die Korrelationsmatrix

$$K_{XYZ} := \mathrm{Korr}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 1 & \rho_{XY} & \rho_{XZ} \\ \rho_{YX} & 1 & \rho_{YZ} \\ \rho_{XZ} & \rho_{ZY} & 1 \end{pmatrix} ,$$

ist als Kovarianzmatrix der standardisierten Zufallsvariablen positiv definit. Damit ist die Determinante von K_{XYZ} größer als Null. Nutzen wir diesen Sachverhalt aus, so folgt nach einigen einfachen Umformungen:

$$\rho_{XZ}^2 - 2\rho_{XY}\rho_{YZ}\rho_{XZ} + \rho_{XY}^2 + \rho_{YZ}^2 - 1 < 0$$

$$\Rightarrow \begin{cases} \rho_{XZ} > \rho_{XY}\rho_{YZ} - \sqrt{(1-\rho_{XY}^2)(1-\rho_{YZ}^2)} \\ \rho_{XZ} < \rho_{XY}\rho_{YZ} + \sqrt{(1-\rho_{XY}^2)(1-\rho_{YZ}^2)} \end{cases} .$$

Berücksichtigen wir nun noch die „natürlichen" Grenzen -1 und 1 für den Korrelationskoeffizienten, so lautet das Ergebnis

$$\max\{ -1 \, , \, \rho_{XY}\rho_{YZ} - \sqrt{(1-\rho_{XY}^2)(1-\rho_{YZ}^2)} \} \; < \; \rho_{XZ}$$

$$< \; \min\{ 1 \, , \, \rho_{XY}\rho_{YZ} + \sqrt{(1-\rho_{XY}^2)(1-\rho_{YZ}^2)} \} \; .$$

Leider scheinen diese einfachen Schranken der bisherigen Fachliteratur keine Erwähnung wert zu sein. Ich kann mir diese Lücke nur so erklären, daß das Problem entweder zu einfach ist, oder daß bisher noch niemand damit konfrontiert worden ist.

P.S.: Vor Kurzem bekam ich erneut einen Anruf dieses Kollegen. Begeistert teilte er mir mit, welche große Hilfe diese Schranken für ihn wären, und daß fast 30% der von ihm seither berechneten Korrelationen doch tatsächlich innerhalb dieser Grenzen lägen.

Ein altes chinesisches Sprichwort sagt: „Frage den Fremden lieber einmal zuviel, wohin genau er möchte, bevor Du ihm den falschen Weg weist."

Ich glaube, in diesem Fall wäre für die empirische Korrelation ein Konfidenzintervall weitaus sinnvoller gewesen als - wie hier - Schranken basierend auf den theoretischen Korrelationen zwischen X und Y bzw. Y und Z.