

Kolloquium „Statistische Methoden in der empirischen Forschung“

Wann: 11. Dezember 2018, 17:00 – 18:30 Uhr

Wo: Robert Koch-Institut | Nordufer 20 | 13353 Berlin (Wedding),
S41, S42, U9 Westhafen | U9, Bus 142 Amrumer Str

Janine Witte (Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, Bremen)

Kovariablen-Selektion für kausale Inferenz: verschiedene Ansätze im Vergleich

Kovariablen-Selektion kann verschiedene Ziele verfolgen. Soll ein kausaler Effekt geschätzt werden, sollten die Kovariablen in erster Linie danach ausgewählt werden, ob sie für Confounding adjustieren – hierin unterscheidet sich Kovariablen-Selektion für kausale Inferenz von Kovariablen-Selektion zur optimalen Vorhersage. Zusätzlich beeinflussen die gewählten Kovariablen aber auch die Effizienz des Schätzers, seine Robustheit gegenüber Modellmisspezifikation und die Interpretation des geschätzten Effekts – hier wiederum bestehen Ähnlichkeiten zur Selektion für Vorhersagen. Die Wahl der Selektionsstrategie hängt zudem davon ab, wie viel Vorwissen vorhanden ist und welche (parametrischen) Annahmen vertretbar sind. So ist es nicht verwunderlich, dass mittlerweile eine Vielzahl an teilweise sehr unterschiedlichen Ansätzen vorgeschlagen wurde, um Kovariablen zu selektieren. Darunter sind neben klassischen Modellwahlverfahren wie der schrittweisen Regression beispielsweise Kriterien, die allein auf Expertenwissen beruhen, und Machine-Learning-Algorithmen.

In diesem Vortrag werde ich sechs Klassen von Kovariablen-Selektionsmethoden vorstellen, die sich im Wesentlichen darin unterscheiden, welche Kovariablen sie unter idealen Bedingungen auswählen würden. Allen ist gemeinsam, dass sie auf kausale Annahmen angewiesen sind, die sich nicht empirisch überprüfen lassen. Ich werde auf Vor- und Nachteile eingehen und das unterschiedliche Verhalten der Methoden mit Hilfe von kausalen Diagrammen und Simulationen illustrieren. Meine Ergebnisse verdeutlichen, dass dieselbe Selektionsstrategie das Verhalten verschiedener Adjustierungsmethoden (z. B. lineare Regression, Matching, Propensity-Score-Weighting) unterschiedlich beeinflussen kann. Selektionsmethode und Adjustierungsmethode sollten daher immer in Kombination betrachtet werden.