

Kolloquium „Statistische Methoden in der empirischen Forschung“

Wann: 25. November 2014, 17:00 – 18:30 Uhr

Wo: Lebenswissenschaftliche Fakultät, Humboldt-Universität zu Berlin,
Hörsaal 2, 2. Etage, Invalidenstr. 42, 10115 Berlin

Arne Bathke (Universität Salzburg)

Analyse multivariater Daten, aber ohne Normalverteilung - geht das überhaupt?

Wenn Hypothesentests oder andere Methoden der schließenden Statistik auf multivariate Daten (multiple Endpunkte) angewandt werden, dann muss man gewöhnlich zwei recht schwerwiegende Annahmen treffen: gleiche Varianz-Kovarianz-Matrizen in den verschiedenen Gruppen, und multivariate Normalverteilung der Daten. Beide Annahmen sind leider in der Praxis schwer bzw. gar nicht zu überprüfen. Aber die Verletzung nur einer der Annahmen hat zur Folge, dass z.B. die klassische MANOVA nicht verwendet werden sollte. Als Alternative zu dieser parametrischen Inferenzmethode stellen wir einen rein nicht-parametrischen Ansatz zum Vergleich multivariater Stichproben vor.

Es ist hier nicht nur keine multivariate Normalverteilung der Daten notwendig, sondern die unterschiedlichen Zielvariablen dürfen sogar völlig unterschiedlich skaliert sein (binär/dichotom, ordinal, quantitativ). Die ermittelten P-Werte basieren auf F-Approximationen und Permutationstests. Zur Erleichterung der tatsächlichen Anwendung präsentieren wir das R-Paket `npmv`, welches zusätzlich zu den globalen Tests auch identifiziert, welche Konstellationen von Zielvariablen und Faktorstufen signifikant sind. Dies geschieht unter Verwendung eines multiplen Testalgorithmus, der das multiple Niveau einhält (control of the familywise error rate).