Handout OLS 1 Achim Goerres, Master Methoden

Stellen Sie sich vor, dass der Raum, in dem Sie sitzen, Datenpunkte zu Wahlkreisen enthält. Die Höhe vom Boden zur Decke sei die Wahlbeteiligung (Y) zwischen 0 und 100 %. Die eine Seite des Raumes sei die Arbeitslosenquote in % der Einwohner\*innen im Alter von 16-65 Jahren (X1). Die andere Seite des Raumes sei die Bevölkerungsdichte in Einwohner\*innen pro km² (X2).

1. Bilden Sie einzeln einen Punkt mit einer Faust irgendwo im Raum. Bewegen Sie Ihre Faust: erhöhen Sie Y, während Sie X1 und X2 konstant halten. Erhöhen Sie X1, während Sie Y und X1 konstant halten. Erhöhen Sie X2, während Sie Y und X1 konstant halten.
2. Bilden Sie nun in einer Gruppe von mindestens vier Studierenden eine Punktwolke mit insgesamt 8 Fäusten. Halten Sie die Fäuste ruhig. Überlegen Sie, wo die Regressionslinie hinkommt, die Y vorhersagt und die Abstände zwischen vorhergesagten Y-Werten und den tatsächlichen Y-Werten minimiert.
3. Übung im Plenum mit Dachlatte und großer Datenwolke.
4. Was bedeutet „Drittvariablenkontrolle“ bei einer multiplen Regression?
5. Woran kann ich erkennen, dass ich alle Drittvariablen in einer Regression habe?
6. Warum brauchen wir in der experimentellen Forschung keine multiple Regression?
7. Warum eignet sich die Darstellung mit den Pfeilen?
8. Warum gibt es eine Schätzung für die Unsicherheit?
9. Wie viele Schätzwerte gibt es bei einer Regressionsfunktion mit einer Konstanten und vier unabhängigen Variablen?
10. Bitte laden Sie sich den Open-access-Artikel Goerres/Spies/Kumlin 2018 zur AfD-Wahlneigung herunter. <https://onlinelibrary.wiley.com/doi/pdf/10.1111/spsr.12306> Es sind Umfragedaten unter deutschen Staatsbürgern von 2016 (und teilweise von 2015). Behandeln Sie diese als einfache Zufallsstichprobe der erwachsenen Bevölkerung. Die abhängige Variable reicht von 0 bis 10 (wie wahrscheinlich ist es, dass Sie je die AfD wählen).
11. Schauen Sie sich die deskriptive Statistik zur abhängigen Variable an (erste Zeile, Tabelle A, S. 267). Interpretieren Sie alle vier genannten Zahlen.



1. Wie sind die Variablen Risk of unemployment (wahrgenommene Gefahr einer persönlichen Arbeitslosigkeit) und Risk of poverty (wahrgenommene Gefahr persönlicher Armut) kodiert? Wie sind die Variablen Support for redistributive welfare (Unterstützung einer umverteilenden Sozialpolitik von reich zu arm) und support for life-cycle welfare (Unterstützung für Rente und Gesundheitsversicherung als Verteilung über den Lebenslauf) kodiert?





1. Tabelle 1 (S. 257) enthält vier Regressionen. Wir beschäftigen uns mit der zweiten (Model 2). Interpretieren Sie den R²-Wert.
2. Interpretieren Sie die Konstante.
3. Interpretieren Sie die Punktschätzungen der vier Koeffizienten der beiden Risk- und der beiden Support-Variablen.
4. Interpretieren Sie den p-Wert in Klammern bei den vier Koeffizienten.