Institut zur Qualitätsentwicklung im Bildungswesen
Forschungsdatenzentrum

WISSENSCHAFTLICHE EINRICHTUNG DER LÄNDER AN DER HUMBOLDT-UNIVERSITÄT ZU BERLIN E.V.

Workshop:	Propensity Score Matching
Lecturers:	Julian Urban (GESIS, Mannheim; Trier University) / Dr. Markus Feuchter (LIfBi, Bamberg)
Data:	Part 1: Tuesday, 03.09.2024, 9:00 – 13:00 Uhr
	Part 2: Wednesday, 04.09.2024, 9:00 – 13:00 Uhr

Abstract *(German version below)*

Educational research encompasses many research questions. Which factors are relevant for educational success, which teaching and learning methods are successful or which groups show particular potential or are disadvantaged? However, answering such questions poses a methodological challenge. Since randomized studies are often not feasible, covariates (e.g. age, social background, intelligence) can bias the effect of interest. If you want to overcome these biases in your studies, propensity score matching (PSM) may be of interest to you.

PSM addresses covariate imbalance between target groups by matching participants based on similarity in covariates (e.g., background variables). Thus, it enhances the quality of your data and analyses by controlling bias caused by cofounds or unequal group sizes. Traditionally, only one grouping variable with two groups can be considered for matching. Recent developments have expanded this method to accommodate more than two groups (e.g., one intervention group and two control groups), as well as more than one grouping variable (e.g., 2x2 designs).

Content.

- An introduction into the statistical and substantive problems of imbalanced group designs for educational research
- A review of the rationale, benefits, and drawbacks of propensity score matching in estimating causal effects in various study designs;
- Utilizing R for the process of computing propensity scores and matching participants;
- A framework for assessing pre- and post-matching balance;
- Guidelines for effectively reporting your matching approach in a manuscript.

Prerequisites.

- Basic knowledge of R (opening and viewing data, manipulating objects, working with functions and packages, ideally worked with R in earlier projects).
- Basic knowledge of (quasi-)experimental/observational study designs and differences between these designs (randomization, possibility to estimate causal effects).
- Basic knowledge of inference statistics (statistical tests such as *t*-test, (M)ANOVA, correlations).
- Basic knowledge of effect sizes (Cohen's *d*, ideally Hedges' *g*).

References.

Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in medicine*, *33*(6), 1057-1069. https://doi.org/10.1002/sim.6004

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance e for the implementation of propensity score matching. *Journal of economic surveys*, *22*(1), 31-72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics, 2(4), 405–420. https://doi.org/10.1080/10618600.1993.10474623

Guo, S., Fraser, M., & Chen, Q. (2020). Propensity score analysis: recent debate and discussion. *Journal of the Society for Social Work and Research*, *11*(3), 463-482. https://doi.org/10.1086/711393

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). The MatchIt package (3.0.2). https://cran.r-hub.io/web/packages/MatchIt/MatchIt.pdf

Jacovidis, J. N. (2017). Evaluating the performance of propensity score matching methods: A simulation study [Doctoral Dissertation]. James Madison University.

Li, M. (2013). Using the Propensity Score Method to Estimate Causal Effects: A Review and Practical Guide. Organizational Research Methods, 16(2), 188–226. https://doi.org/10.1177/1094428112447816

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political analysis*, *27*(4), 435-454. https://doi.org/10.1017/pan.2019.11

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity Score Estimation With Boosted Regression for Evaluating Causal Effects in Observational Studies. 9(4), 403. https://doi.org/10.1037/1082-989X.9.4.403

Powell, M. G., Hull, D. M., & Beaujean, A. A. (2020). Propensity score matching for education data: Worked examples. *The Journal of Experimental Education*, *88*(1), 145-164. https://doi.org/10.1080/00220973.2018.1541850

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55. https://doi.org/10.1093/biomet/70.1.41

Ridgeway, G., McCaffrey, D. F., Morral, A. R., Cefalu, M., Burgette, L. F., Pane, J. D., & Griffin, B. A. (2022). *Toolkit for weighting and analysis of nonequivalent groups: a tutorial for the R TWANG package*. Santa Monica, Calif: Rand. https://cran.r-hub.io/web/packages/twang/vignettes/twang.pdf

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1. https://doi.org/10.1214/09-STS313

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate behavioral research*, *46*(1), 90-118. https://doi.org/10.1080/00273171.2011.540475

Urban, J., Feuchter, M. D., Scherrer, V., Breit, M., & Preckel, F. (2023, April 12). Turn up the Heat! A Tutorial for the MAGMA R-package. https://doi.org/10.31234/osf.io/yfd7a

Zhang, Z., Kim, H. J., Lonjon, G., & Zhu, Y. (2019). Balance diagnostics after propensity score matching. *Annals of translational medicine*, *7*(1). http://dx.doi.org/10.21037/atm.2018.12.10

Zhao, Q. Y., Luo, J. C., Su, Y., Zhang, Y. J., Tu, G. W., & Luo, Z. (2021). Propensity score matching with R: conventional methods and new features. *Annals of translational medicine*, *9*(9). http://dx.doi.org/10.21037/atm-20-3998

Software.

- R (Version 4.2 or more recent)
- R Studio

Workshop:	Propensity Score Matching
Dozierende*r:	Julian Urban (GESIS, Mannheim; Universität Trier) / Dr. Markus Feuchter (LIfBi, Bamberg)
Termin:	Teil 1: Dienstag, 03.09.2024, 9:00 – 13:00 Uhr
	Teil 2: Mittwoch, 04.09.2024, 9:00 – 13:00 Uhr

Abstract

Bildungsforschung umfasst viele mögliche Fragstellungen. Welche Faktoren sind relevant für Bildungserfolg, welche Lehr- und Lernmethoden sind erfolgreich oder welche Gruppen zeigen besonderes Potential oder sind benachteiligt? Die Beantwortung solcher Fragestellung bringt jedoch eine methodische Herausforderung mit sich. Da randomisierte Studien oft nicht umsetzbar sind, können Kovariaten (z.B. Alter, soziale Herkunft, Intelligenz) den interessierenden Effekt verzerren. Wenn Sie diese Verzerrungen in Ihren Studien bewältigen wollen, könnte Propensity Score Matching (PSM) für Sie von Interesse sein.

PSM ist eine statistische Methode, die hilft, das Ungleichgewicht in Kovariaten zwischen Zielgruppen zu beheben, indem Teilnehmende verschiedener Gruppen basierend auf der Ähnlichkeit in Kovariaten gematcht werden. Dadurch wird die Qualität Ihrer Daten und Analysen verbessert, indem Bias durch Störvariablen oder ungleiche Gruppengrößen kontrolliert wird. Traditionell konnte nur eine Gruppierungsvariable mit zwei Gruppen für Matching berücksichtigt werden. Neuere Entwicklungen erweitern diese Methode, um mehr als zwei Gruppen (z.B. eine Interventionsgruppe und zwei Kontrollgruppen) sowie mehr als eine Gruppierungsvariable (z.B. 2x2-Designs) zu berücksichtigen.

Inhalte.

- Einführung in die statistischen und inhaltlichen Probleme von nicht randomisierten Gruppendesigns in der Bildungsforschung
- Überblick über die Gründe, Vorteile und Nachteile des Propensity Score Matching bei der Schätzung kausaler Effekte in verschiedenen Studiendesigns
- Verwendung von R für die Berechnung von Propensity Scores und dem Matchen von Studienteilnehmenden

- Bewertung der Balance vor und nach dem Matching;
- Richtlinien für eine effektive Berichterstattung über Ihren Matching-Ansatz in einem Manuskript

Voraussetzungen.

- Grundkenntnisse in R (Öffnen und Anzeigen von Daten, Manipulation von Objekten, Arbeit mit Funktionen und Paketen, idealerweise Arbeit mit R in früheren Projekten).
- Grundkenntnisse über (quasi-)experimentelle/beobachtende Studiendesigns und Verständnis der Unterschiede zwischen diesen Designs (Randomisierung, Möglichkeit zur Schätzung kausaler Effekte).
- Grundkenntnisse der Inferenzstatistik (statistische Tests wie *t*-Test, (M)ANOVA, Korrelationen).
- Grundkenntnisse über Effektgrößen (Cohen's *d*, idealerweise Hedges' *g*).

Literatur.

Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in medicine*, *33*(6), 1057-1069. https://doi.org/10.1002/sim.6004

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance e for the implementation of propensity score matching. *Journal of economic surveys*, *22*(1), 31-72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics, 2(4), 405–420. https://doi.org/10.1080/10618600.1993.10474623

Guo, S., Fraser, M., & Chen, Q. (2020). Propensity score analysis: recent debate and discussion. *Journal of the Society for Social Work and Research*, *11*(3), 463-482. https://doi.org/10.1086/711393

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). The MatchIt package (3.0.2). https://cran.r-hub.io/web/packages/MatchIt/MatchIt.pdf

Jacovidis, J. N. (2017). Evaluating the performance of propensity score matching methods: A simulation study [Doctoral Dissertation]. James Madison University.

Li, M. (2013). Using the Propensity Score Method to Estimate Causal Effects: A Review and Practical Guide. Organizational Research Methods, 16(2), 188–226. https://doi.org/10.1177/1094428112447816

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political analysis*, *27*(4), 435-454. https://doi.org/10.1017/pan.2019.11

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity Score Estimation With Boosted Regression for Evaluating Causal Effects in Observational Studies. 9(4), 403. https://doi.org/10.1037/1082-989X.9.4.403

Powell, M. G., Hull, D. M., & Beaujean, A. A. (2020). Propensity score matching for education data: Worked examples. *The Journal of Experimental Education*, *88*(1), 145-164. https://doi.org/10.1080/00220973.2018.1541850

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55. https://doi.org/10.1093/biomet/70.1.41

Ridgeway, G., McCaffrey, D. F., Morral, A. R., Cefalu, M., Burgette, L. F., Pane, J. D., & Griffin, B. A. (2022). *Toolkit for weighting and analysis of nonequivalent groups: a tutorial for the R TWANG package*. Santa Monica, Calif: Rand. https://cran.r-hub.io/web/packages/twang/vignettes/twang.pdf

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1. https://doi.org/10.1214/09-STS313

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate behavioral research*, *46*(1), 90-118. https://doi.org/10.1080/00273171.2011.540475

Urban, J., Feuchter, M. D., Scherrer, V., Breit, M., & Preckel, F. (2023, April 12). Turn up the Heat! A Tutorial for the MAGMA R-package. https://doi.org/10.31234/osf.io/yfd7a

Zhang, Z., Kim, H. J., Lonjon, G., & Zhu, Y. (2019). Balance diagnostics after propensity score matching. *Annals of translational medicine*, *7*(1). http://dx.doi.org/10.21037/atm.2018.12.10

Zhao, Q. Y., Luo, J. C., Su, Y., Zhang, Y. J., Tu, G. W., & Luo, Z. (2021). Propensity score matching with R: conventional methods and new features. *Annals of translational medicine*, *9*(9). http://dx.doi.org/10.21037/atm-20-3998

Software.

- R (Version 4.2 oder aktueller)
- R Studio