Thema	Abstract	Literatur
Updating Intercensal Indicators based on geospatial data	Censuses are fundamental building blocks of most modern-day societies, yet collected every 10 years at best. The thesis assesses different approaches for census updating techniques by incorporating auxiliary information in order to take ongoing subnational population shifts into account. The methods adds satellite imagery as additional data source to derive disaggregated estimates of poverty indicators in Mozambique. The performance of the methods is evaluated using data from two different census periods.	1. T. Koebe, A. Arias-Salazar, N. Rojas-Perilla, and T. Schmid. Intercensal updating using structure-preserving methods and satellite imagery. Journal of the Royal Statistical Society Series A, 185: 170-196, 2022. 2. A.-K. Kreutzmann, S. Pannier, N. Rojas-Perilla, T. Schmid, M. Templ, and N. Tzavidis. The R package emdi for estimating and mapping regionally disaggregated indicators. Journal of Statistical Software 91(7): 1-33, 2019.
Multinomial Mixed Models in Small Area Estimation	The topic deals with small area estimation for the composition of categorical variables. A possible example is, for instance, the estimation of multidimensional poverty in small areas. From a methodological perspective, the thesis deals with small area estimation based on multinomial (mixed) models including the computation of the mean squared error of the estimators. A simulation study could be conducted to investigate the properties of the estimators. Finally, the methods can be applied to poverty data from Mexico.	1. E. Lopez-Vizcaino, M. J. Lombardia, and D. Morales. Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13(2):153-178, 2013. 2. I. Molina, A. Saei, and M. J. Lombardia. Small area estimates of labour force participation under a multinomial logit mixed model. Journal of the Royal Statistical Society: Series A, 170(4):975-1000, 2007.
Nonparametric regression with application to small area problems	Within the field of small area estimation the best linear unbiased predictor (BLUP) or its empirical version, the empirical BLUP (short EBLUP), is a widely used tool to estimate statistics of small areas, even though it is sensitive to model misspecification. More robust estimators like the penalised spline regression, which avoids functional specification, are brought forward as alternatives. Opsomer et al. (2008) suggest an estimator for unit-level data that treats the penalisation coefficient as a random effect, which allows the use of standard EBLUP theory in further analysis. Their estimator is known as nonparametric EBLUP. The nonparametric EBLUP framework could be compared to the standard EBLUP and to Random forests within a simulation study and an application with real data.	1. J. Opsomer, F. Breidt, G. Claeskens, G. Kauermann, and G. Ranalli. Nonparametric small area estimation using penalized spline regression. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 70(1):265-286, 2008. 2. D. Pfeffermann. Small Area Estimation: New Developments and Directions. International Statistical Review/ Revue Internationale de Statistique, 70(1):125–143, 2002.
Poverty Mapping using ELL-Methodology	The World Bank Method, also called ELL (Elbers, Lanjouw and Lanjouw) method, is a small area estimation strategy used to estimate poverty indicators. This estimation technique is implemented in the World Bank software PovMap. In a thesis, a package could be programmed that also makes this estimation technique available to a broader audience via the R programming language. Furthermore, the ELL can be adjusted and analysed based on the comparison to other established methods in the field, such as the empirical best prediction method (EBP).	1. C. Elbers, J. O. Lanjouw, and P. Lanjouw. Micro-level estimation of poverty and inequality. Econometrica 71(1): 355-364, 2003. 2. A.-K. Kreutzmann, S. Pannier, N. Rojas-Perilla, T. Schmid, M. Templ, and N. Tzavidis. The R package emdi for estimating and mapping regionally disaggregated indicators. Journal of Statistical Software 91(7): 1-33, 2019.
Quantile Regression in Complex Design Samples	The Quantile Regression (QR) is an extension of the (multiple) linear regression that specifies the change of conditional quantiles of a dependent variable, instead of the change the conditional mean. The additional information gathered form this method might uncover new insights in datasets such as the one of the Demographic and Health Survey (DHS) of Senegal. QR could be used to estimate the conditional quantiles of wealth indicators such as the Wealth Index Factor Score for women and men, which uses the total value of an individual's assets to determine wealth, instead of commonly used monetary variables.	1. R. Koenker. Quantile Regression. Cambridge University Press: Cambridge, New York, 2005. 2. ANSD. Demographic and Health and Multiple Indicator Cluster Survey (EDS-MICS) 2010-2011.
Robust Small Area Estimation Methods in the Context of Spatial Correlations	The empirical best linear unbiased predictor (EBLUP), which is commonly used in small area statistics, uses the assumptions of normality and uncorrelated random effects. The robust version of EBLUP (REBLUP) can be used when a violation of normality is assumed. The SREBLUP is another extension that additionally considers spatial correlated area effects in the model. A possible thesis could compare the performance of these estimators for different scenarios according to each estimator's underlying assumptions in a model-based simulation study. Real world data that displays spatial dependency and deviation of normality can be used.	1. T. Schmid, and R. Munnich. Spatial robust small area estimation. Statistical Papers, 55(3):653–670, 2014. 2. R. B. Papalia, C. Bruch, T. Enderle, S. Falorsi, A. Fasulo, E. Fernandez Vazquez, M. Ferrante, J.-P. Kolb, R. Münnich, S. Pacei, R. Priam, P. Righi, T. Schmid, N. Shlomo, F. Volk, and T. Zimmermann. Best practice recommendations on variance estimation and small area estimation in business surveys. BLUE-ETS. SP1- Cooperation-Collaborative Project - Small or medium-scale focused research project, 2013.
Comparing area-level to unit-level models in model-based settings	First, different methods (unit-level models under limited data vs. area-level models (with and without aggregation)) will be compared in 4 model-based settings. If possible, it should then be considered how extreme settings might look like in which one of the methods is most convincing. A little application to real-wold data is also possible.	1. Rao, J. N. K. and I. Molina (2015). Small Area Estimation (Second Edition). Hoboken: John Wiley & Sons. 2. Corral, Paul; Molina, Isabel; Cojocaru, Alexandru; Segovia, Sandra. 2022. Guidelines to Small Area Estimation for Poverty Mapping. Washington, DC : World Bank.
Working with saeTrafo: Implementation of an analytical MSE for log-transfromed unit-level models	Based on the publication of Molina, I. and N. Martín (2018). Empirical best prediction under a nested error model with log transformation. The Annals of Statistics 46(5), 1961–1993, this analytical MSE is to be implemented for the saeTrafo package. An application to EUSILC data is also intended for this method.	1. Würz, N. (2022). saeTrafo: Transformations for Unit-Level Small Area Models. R package version 1.0.0. 2. Molina, I. and N. Martín (2018). Empirical best prediction under a nested error model with log transformation. The Annals of Statistics 46(5), 1961–1993.
Logistic and multinomial models on the example of gender issues	Topic addresses small-area estimates for the composition of categorical variables. Gender issues in male-dominated occupations can be studied. From a methodological perspective, the thesis deals with estimation based on logistic and multinomial models. Data from NEPS, DZHW or data from Alumnae Tracking can be applied.	1)Christof Wolf, Henning Best (2010): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden : VS, Verlag für Sozialwissenschaften. 2) Mood, C. (2010). Logistic Regression: Why We Cannot Do What We Think We Can Do, And What We Can Do About It. European Sociological Review, 26, 67–82. https://doi.org/10.1093/esr/jcp006 3) Förtsch, S.; Gärtig-Daugs, A.; Buchholz, S.; Schmid, U. (2018): Keep it going Girl! An Empirical Analysis of Gender Differences and Inequalities in Computer Sciences, International Journal of Gender, Science and Technology, 10 (2), pp. 265-286. 4) Förtsch, S., & Gärtig-Daugs, A. (2019): Trust yourself: You have the IT-Factor! Career coaching for female computer scientists, International Journal of Gender, Science and Technology, 11 (3), pp. 490-527
Statistical misuse and consequences	Statisticial literacy: Selection from typical problem clusters of statistical literacy, such as confounding, lack of measures of significance and/or goodness of fit, correlation versus causality, relative versus absolute risk, various kinds of biases (e.g., self-selection, response bias, attrition, Hawthorne effect, social desirability bias), immortal time bias, etc.	1)Milo Schield. Statistical literacy: Thinking critically about statistics. Of Significance, 1(1):15–20, 1999. 2)DJordje Kadijevi'c & Max Stephens. Modern statistical literacy, data science, dashboards, and automated analytics and its applications. Teachingof Mathematics, 23(1):71–80, 2020. 3) Eric Sowey & Peter Petocz. A panorama of statistics: Perspectives, puzzles and paradoxes in statistics. John Wiley & Sons, 2017.
Convergence in Distribution: Examining the Gelman-Rubin Diagnostic	Convergence in distribution is factually impossible to prove. The GD diagnostic has become a widely-accepted measure to assess convergence in distribution. Investigate the Accuracy of the GD diagnostic using an extensive MC simulation study.	Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7: 457–472
Diagnostics for mass imputation: R package	To date none of the imputation packges in R (or Stata, SPSS;...) can visualize missing data in meaningful way if the number of observations/ variables is large.	Meinfelder, F. (2013): Datenfusion: Theoretische Implikationen und praktische Umsetzung. In: Weiterentwicklung der amtlichen Haushaltsstatistiken. (eds. Riede, T.; Ott, N.; Bechthold, S.), pp. 83-100, 1st ed., Berlin, GWI Wissenschaftspolitik Infrastrukturentwicklung.
Inference for the Potential Outcome framework	Different matching strategies impact the variance term of the test statistic. With and without replacement matching as well as weighting affect how independent matched samples are.	Bai, Yuehao; Romano, Joseph P.; Shaikh, Azeem M (2019) Inference in experiments with matched pairs, Econstor Working Paper, https://www.econstor.eu/bitstream/10419/211112/1/1663702217.pdf
Multiple Imputation of categorical variables using an extension to Predictive Mean Matching	PMM has become a popular MI method and is the default imputation method in 'mice' for 'numeric'-type variables. However, no extension for non-ordered categorical (nominal-scale) variables is available. One problem is the identification of a nearest neighbour over k categories. This can be overcome by using the KL divergence to match vial the empirical distribution of donor and recipient predictions. Your task is to implement this method in R and compare it with alternative approaches.	Koller-Meinfelder, F. (2009): Analysis of Incomplete Survey Data – Multiple Imputation via Bayesian Bootstrap Predictive Mean Matching, Dissertation, Otto-Friedrich-Universität Bamberg.
Comparing Generalized Linear Mixed Models and Transformations in the SAE Context	Two main approaches to deal with non normal and a non linear relationship between the independent and the dependenent variable are generalized linear mixed models (GLM) and transformations. In a possible thesis the researcher could compare the pros and cons of both approaches in a simulatins study.	1) Morales, Domingo, et al. "A course on small area estimation and mixed models." Methods, theory and applications in R (2021). 2) Chandra, Hukum, and Ray Chambers. "Small area estimation under transformation to linearity." (2008).