Emerging Topics in Statistics and Biostatistics

Hon Keung Tony Ng
Daniel F. Heitjan *Editors*

# Recent Advances on Sampling Methods and Educational Statistics

## In Honor of S. Lynne Stokes

Springer

# Emerging Topics in Statistics and Biostatistics

Hon Keung Tony Ng • Daniel F. Heitjan
Editors

# Recent Advances on Sampling Methods and Educational Statistics

## In Honor of S. Lynne Stokes

Springer

*Editors*
Hon Keung Tony Ng
Department of Mathematical Sciences
Bentley University
Waltham, MA, USA

Daniel F. Heitjan
Department of Statistical Science
Southern Methodist University
Dallas, TX, USA

ISSN 2524-7735 ISSN 2524-7743 (electronic)
Emerging Topics in Statistics and Biostatistics
ISBN 978-3-031-14524-7 ISBN 978-3-031-14525-4 (eBook)
https://doi.org/10.1007/978-3-031-14525-4



This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland


S. Lynne Stokes

# Life and Works of S. Lynne Stokes

I was born on December 16, 1950, in Corsicana, the seat of Navarro County, Texas, where six generations of the Stokes family had lived. I was the second-born to a family of teachers. My dad taught mathematics and physics and coached the baseball team at Navarro Junior College, which had been established in 1946 as he and so many others were returning from WWII. My mother taught Spanish and agriculture, neither of which she had ever taken a course in, at the high school in her nearby hometown, Richland. In 1952, my parents decided to pull up roots and head to graduate school at Peabody College, now a part of Vanderbilt University, in Nashville, Tennessee. Their families were horrified that they would move so far away, and especially that a mom of two would take such an unconventional path. But the GI Bill had placed higher education within reach for many families who would now be called "first-generation," including mine. My parents went on to earn doctorates and have careers as college professors, he in mathematics and she in psychology. Their last and longest stint was at Austin Peay State University, where my dad chaired the Math Department for more than 20 years and my mom helped train a generation of school counselors in Clarksville, Tennessee. From this exposure and the joy they had in their careers, I decided at a young age that being a professor was my goal.

I studied mathematics at the University of the South in Suwanee, Tennessee. One of the faculty members, Mac Priestley, agreed to supervise me in an independent study out of Kemeny and Snell's book on Markov chains. From that experience, I decided that enrolling in a statistics PhD program was the right path for me, not realizing that it was actually probability I had been fascinated by. Luckily, I liked statistics even better, which I realized after joining the program at the University of North Carolina.

My years in Chapel Hill are among my fondest memories. My advisor, Norman Johnson, was endlessly encouraging and supportive. He asked me to read Dell and Clutter's 1972 ranked set sampling paper, then recently published, to see if I had any ideas on extensions for my dissertation work. Since that time, I have had the pleasure of discussing and collaborating with many on this topic, including several contributors to this volume.

My first job after school was in the Department of Mathematics at Vanderbilt, which was near my family home. I was one of only two statisticians in a large department. I soon decided I preferred real data and the company of other statisticians, and moved on to the Patuxent Wildlife Research Center in Laurel, Maryland. Patuxent was then a part of the US Fish and Wildlife Service and located in a 16,000-acre refuge of beautiful forest and wetlands in the midst of the Washington DC/Baltimore urban sprawl. There I learned from scratch about birds, and how to model bird-banding and capture-recapture data from the talented biometricians there, including Jim Nichols, a mentor and co-author. This is a skill I transferred from birds to people (at Census) and back to fish and the people who catch them (for NOAA) over the course of my career.

Patuxent changed my life in another way as well. There I met Dan Moulton, a biologist in the bird-banding lab, where he worked between field seasons on Laysan Island in Hawaii, where he was studying and banding Laysan ducks. During his second 6-month field season, we corresponded by letter and audio tape. These could be transported only by military plane or ship as they patrolled the Hawaiian archipelago. Soon after Dan returned from Laysan, we married.

While he was away, I left Patuxent for the US Census Bureau, which was just a short trip around the Beltway. Mary Mulry and I were hired into the Statistical Methods Division by Paul Biemer, whom we had first met at age 20 when all three of us were participants in an undergraduate NSF summer mathematics program at Texas A&M. Mary, Paul, and I have been colleagues, friends, and collaborators for 50 years, and we have NSF to thank for that.

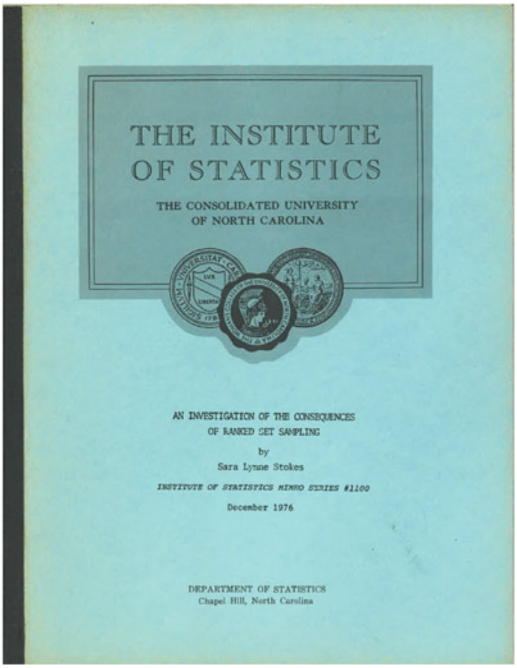
Paul had studied under H. O. Hartley, and he introduced me to sampling theory and measurement error methods. The Census Bureau provided an unlimited supply of real-life problems for non-sampling error research, which has remained a lifelong interest. Fortunately for my career, errors occur whenever data are collected. This allowed me to dabble in many fascinating application areas over the years. Two of these areas, fisheries and education surveys, are well represented in this volume (Brick, Andrews & Foster; Becker & Gozutok).

When Dan took a position at Texas Parks and Wildlife in Austin in 1983, I moved with him and worked remotely for Census, before that was common. This arrangement was facilitated with the help of Kent Marquis, my division chief at Census, and Carl Morris, then in the Mathematics Department at the University of Texas. Kent and Carl had known each other at Rand, proving once again that it helps to get lucky. Soon a faculty position opened for a statistician in the Management Science Department at UT's Business School, and I was again in the right place at the right time. In my 15 years at UT, I expanded the range of problems I worked on with colleagues in fields from finance to demography to operations research.

In 2001, I left UT for the Statistics Department at Southern Methodist University, after a convincing chat with my long-time acquaintance Bill Schucany. I had first met Bill at a Conference of Texas Statisticians meeting shortly after moving to Texas, and had received useful advice from him over the years. SMU was a perfect place for the last 20 years of my career, providing a helpful administration, supportive colleagues, and excellent graduate students. I chaired the department for one term, and then became the inaugural director of SMU's Data Science Institute in my last 2 years there. Several of the contributors to this volume are cherished colleagues and former students from SMU.

My path likely would not have been so straight and well-marked if it had not been for the opportunities that began to open up for women at just the right time for me. I also benefited from introductions provided by supportive male mentors, colleagues, and classmates. I entered the University of the South the first year they accepted women (1969). My entering cohort in the Statistics Department at UNC in 1972 were half women and half men, marking the first year that women who were not wives of students were admitted in significant numbers. I was the fourth woman to receive a PhD in statistics at UNC, three of whom were supervised by Norman Johnson, who may have been influenced by his wife Regina from the UNC Biostatistics Department. At Vanderbilt, I was the first woman to fill a tenure-track position in the Mathematics Department, and at SMU, the first woman chair of the Statistics Department. My network-building began in the NSF program I attended as a 20 year old, which I believe illustrates the value of promoting diversity in such programs for young scholars.

Dallas, TX, USA S. Lynne Stokes
May 2022




S. Lynne Stokes's PhD thesis

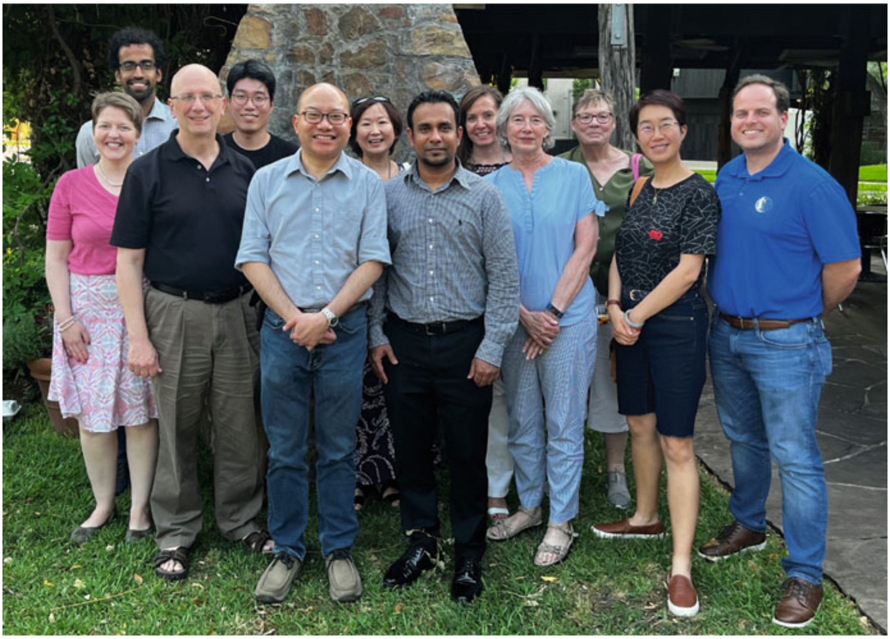

Lynne enjoying the snow at Patuxent Wildlife Research Center, circa 1980


Mary Mulry, Paul Biemer, and Lynne at an NSF program reunion circa 1980

Lynne enjoying Friday morning teatime at SMU in 2007

Celebrating Betsy Becker's election to Fellow at 2008 JSM with an Educational Statistics mentor for both of us, Ingram Olkin

Helena Jia, Lynne, and Bingchen Liu in downtown Princeton during a meeting at ETS in 2017



From left to right: Jessica Wickersham, Raanju R. Sundararajan, Daniel F. Heitjan, Chul Moon, Hon Keung Tony Ng, Xinlei (Sherry) Wang, Mahesh Fernando, Monnie McGee, S. Lynne Stokes, Sheila Crain, Jing Cao, and Charles South in Dallas, Texas, during a department faculty gathering in May 2022

# Awards, Honors, and Publications of S. Lynne Stokes

## Awards and Honors

- Caren Prothro Faculty Service Award, Southern Methodist University (2019)
- Founders Award, American Statistical Association (2013)
- Dedman Family Distinguished Professor, Southern Methodist University (2013)
- United Methodist Church University Scholar/Teacher of the Year Award (2011)
- Don Owen Award, American Statistical Association, San Antonio Chapter (2005)
- Fellow of the American Statistical Association (1998)
- Phi Beta Kappa
- Sigma Pi Sigma

## Publications

### *Refereed Journals and Proceedings*


1. “Investigating Record Linkage for Combining Voluntary Catch Reports with a Probability Sample,” (B. Williams, L. Stokes, and J. Foster), *Fisheries Research*, 251, 106301 (2022).
2. “Predictive modeling of maximum injury severity and potential economic cost in a car accident based on the General estimates system data,” (G. Alkan, R. Farrow, H. Liu, C. Moore, H.K.T. Ng, S. L. Stokes, Y. Xu, Z. Xu, Y. Yan, and Y. Zhang), *Computational Statistics*, 36, 1561–1575 (2021).
3. “The Impact of non-sampling errors on estimators of catch from electronic reporting Systems,” (L. Stokes, B. Williams, R. McShane, and S. Zalsha), *Journal of Survey Statistics and Methodology*, 9, 159–184 (2021).

4. “Prevalence of Sexual Victimization among Female and Male College Students: A Methodological Note with Data,” (Jouriles, E. N., Nguyen, J., Krauss, A., Stokes, S. L., and McDonald, R.), *Journal of Interpersonal Violence*, (2020).
5. “A method to correct for frame membership error in dual frame estimators,” (D. Lin, Z. Liu, and L. Stokes), *Survey Methodology*, 45, 543–565 (2019).
6. “Accumulating Evidence of the Impact of Voter ID Laws: Student Engagement in the Political Process,” (K. S. McConville, L. Stokes, and M. Gray), *Statistics and Public Policy*, 5, 1–8 (2018).
7. “Cross-Cultural Issues in Teaching Ethics in a Statistics Curriculum,” (A. Elliott, L. Stokes, and J. Cao) *The American Statistician*, 72, 359–367 (2018).
8. “Comparison of Different Ranking Methods in Wine Tasting,” (J. Cao and S.L. Stokes), *Journal of Wine Economics*, 12, 203–210 (2017).
9. “Estimation of total from a population of unknown size and application to estimating recreational red snapper catch in Texas,” (B. Liu, S.L. Stokes, T. Topping, and G. Stunz), *Journal of Survey Statistics and Methodology*, 5, 350–371 (2017).
10. “Just in time teaching in Statistics Classrooms,” (M. McGee, L. Stokes, and P. Nadolsky), *Journal of Statistics Education*, 24, 16–26 (2016).
11. “A power analysis for fidelity measurement sample size determination,” (L. Stokes and J. Allor) Psychological Methods, 21, 35–46 (2016).
12. Using Ranked Set Sampling with Cluster Randomized Designs for Improved Inference on Treatment Effects.” (X. Wang, J. Lim, and L. Stokes), *Journal of the American Statistical Association*, 111, 1576–1590 (2016).
13. “Analyses of Wine Tasting Data: A Tutorial,” (I. Olkin, Y. Lou, L. Stokes, and J. Cao), *Journal of Wine Economics*, 10, 4–30 (2015).
14. “The National Children’s Study 2014: Commentary on a Recent National Research Council/Institute of Medicine Report Academic Pediatrics,” *Academic Pediatrics*, 14, 545–546 (2014).
15. “Sample Size Calculation for a Hypothesis Test,” (L. Stokes), *Journal of the American Medical Association*, 312, 180–181 (2014).
16. “Kernel Density Estimator from Ranked Set Samples,” (X. Wang, J. Lim, M. Chen, and L. Stokes), *Communications in Statistics – - Theory and Methods*, 43, 2156–2168 (2014).
17. “Methods for Improving Response Rates in Two-Phase Mail Surveys,” (M. Brick, W. Andrews, P. Brick, H. King, N. Mathiowetz, and L. Stokes), *Survey Practice*, 5, 1–6. (2012).
18. “Stranger at the Gate: the Effect of the Plaintiff’s use of an Interpreter on Juror Decision-Making,” (D. Shuman, L. Stokes, and G. Martinez), *Behavioral Sciences and the Law*, 29, 499–512 (2011).
19. “Performance of Weighted Random Effects Model Estimators under Complex Sampling Designs,” (Y. Jia, L. Stokes, I. Harris, and Y. Wang), *Journal of Educational and Behavioral Statistics*, 36, 6–32 (2011).
20. “Evaluation of Wine Judge Performance through Three Characteristics: Bias, Discrimination, and Variation,” (J. Cao and L. Stokes), *Journal of Wine Economics*, 5, 1–11. (2010)

21. "A Bayesian Approach to Ranking and Rater Evaluation: an Application to Grant Reviews," (J. Cao, S. Zhang, and L. Stokes), *Journal of Educational and Behavioral Statistics*, 35, 194–215. (2010).
22. "Data Masking for Disclosure Limitation," (L. Stokes and G. Duncan), *Wiley Interdisciplinary Reviews: Computational Statistics*, 1, 1–10 (2009).
23. "Bayesian IRT guessing models for partial guessing behaviors," (J. Cao and L. Stokes), *Psychometrika*, 73, 209–230 (2008).
24. "A Nonparametric Mean Estimator for Judgment Post-Stratified Data," (X. Wang, J. Lim, and L. Stokes), *Biometrics*, 64, 355–363 (2008).
25. "Judgment Post-Stratification with Multiple Rankers," (L. Stokes, X. Wang, and M. Chen), *Journal of Statistical Theory and Applications*, 6, 344–359 (2007).
26. "Concomitants of multivariate order statistics with application to judgment post-stratification," (X. Wang, L. Stokes, J. Lim, and M. Chen), *Journal of the American Statistical Association*, 101, 1693–1704 (2006).
27. "Forming Post-Strata via Bayesian Treed Capture-Recapture Models," (X. Wang, J. Lim, and L. Stokes), *Biometrika*, 93, 861–876, (2006).
28. "An Estimator of Number of Species from Quadrat Sampling," (P. Haas, Y. Liu, and L. Stokes), *Biometrics*, 62, 135–141 (2006).
29. "Antecedents and consequences of residential choice and school transfer," (T. Falbo, R. Glover, L. Holcombe, and L. Stokes), *Education Policy Analysis Archives*, 13, 29 (2005).
30. "Disclosure Risk vs. Data Utility: The R-U Confidentiality Map as Applied to Topcoding," (G. Duncan and L. Stokes), *Chance*, 17, 16–20 (2004).
31. "Using Spreadsheet Solvers in Sample Design" (L. Stokes and J. Plummer), *Computational Statistics and Data Analysis*, 44, 527–546 (2004).
32. "Using Auxiliary Information for Improving Estimation in the Number of Species Problem," *Statistica Sinica*, 13, 655–671 (2003).
33. "Comment on 'Can a Statistician Deliver?' " *Journal of Official Statistics*, 17, 103–106 (2001).
34. "Acceptance Sampling with Rectification when Classification Errors are Present," (M. Anderson, B. Greenberg, and L. Stokes), *Journal of Quality Technology*, 33, 493–505 (2001).
35. "Editorial: Special issue on Statistical Design and Analysis with Ranked Set Samples," (N. P. Ross and L. Stokes), *Environmental and Ecological Statistics*, 6, 1–6 (1999).
36. "Estimating the Number of Classes in a Finite Population" (P. Haas and L. Stokes), *Journal of the American Statistical Association*, 93, 1475–1487 (1998).
37. "Success rate with repeated cycles of in vitro fertilization-embryo transfer," (D. Meldrum, K. Silverberg, M. Bustillo, and L. Stokes), *Fertility and Sterility*, 69, 1005–1009 (1998).
38. "Do Product Warnings Increase Safe Behavior?: A Meta Analysis" (E. Cox, L. Stokes, E. Murff), *Journal of Public Policy and Marketing*, 25, 195–204 (1997).

39. “Estimation of the CDF of a Finite Population using a Calibration Sample” (M. Luo, L. Stokes, and T. Sager), *Environmental and Ecological Statistics*, 15, 346–352 (1997).
40. “Considerations of Cost Trade-Offs in Insurance Solvency Surveillance Policy” (J. Lamm-Tennant, L. Starks, L. Stokes), *Journal of Banking and Finance*, 20, 835–852 (1996).
41. “Repetitive Testing in the Presence of Inspection Errors,” (B. Greenberg and L. Stokes), *Technometrics*, 37, 102–111 (1995).
42. “Sampling-Based Estimation of the Number of Distinct Values of an Attribute,” (P. Haas, J. Naughton, S. Sehadri, and L. Stokes), *VLDB 95: Proceedings of the International Conference on Very large Databases* (U. Dayal, P. Gray, S. Nishio, Eds.), 311–322 (1995).
43. “Parametric Ranked Set Sampling,” (L. Stokes), *Annals of the Institute of Statistical Mathematics*, 47, 465–482 (1995).
44. “Reliability of Coherence of Causal, Diagnostic, and Joint Subjective Probabilities,” (K. Wright, L. Stokes and J. Dyer), *Decision Sciences*, 25, 691–709 (1994).
45. “Estimating Nonconformance Rate after Zero-Defect Sampling with Rectification,” (B. Greenberg and L. Stokes), *Technometrics*, 34, 203–215 (1992).
46. “An Empirical Bayes Approach to Estimating Loss Ratios,” (J. Lamm-Tennant, L. Starks, and L. Stokes), *Journal of Risk and Insurance*, 59, 426–442 (1992).
47. “Estimating the Size of a Subdomain: An Application in Auditing,” (L. Stokes), *Journal of Business and Economic Statistics*, 8, 337–346 (1990).
48. “Developing an Optimal Call Scheduling Strategy for a Telephone Survey” (B. Greenberg and L. Stokes), *Journal of Official Statistics*, 6, 421–435 (1990).
49. “The Optimal Design of Quality Control Samples to Detect Interviewer Cheating,” (P. Biemer and L. Stokes), *Journal of Official Statistics*, 5, 23–40 (1989).
50. “Combining Multiple Risk Assessments for Construction Risk Identification,” (D. Ashley, L. Stokes, and Y.H. Peng), *Proceedings of the 7th International Conference on Offshore Mechanics and Arctic Engineering Conference*, American Society of Mechanical Engineers, 183–192 (1988).
51. “Characterization of a Ranked Set Sample with Application to Estimating Distribution Functions,” (L. Stokes and T. Sager), *Journal of the American Statistical Association*, 83, 374–381 (1988).
52. “Estimation of Interviewer Effects for Categorical Items in a Random Digit Dial Telephone Survey,” (S. L. Stokes), *Journal of the American Statistical Association*, 83, 623–630 (1988).
53. “Estimation of the Correlated Component of Response Variance for Categorical Variables,” (L. Stokes and M. Mulry), *Journal of Official Statistics*, 3, 389–401 (1987).
54. “Optimal Design of Interviewer Variance Estimates in Complex Surveys,” (P. Biemer and L. Stokes), *Journal of the American Statistical Association*, 80, 158–166 (1985).

55. “The Jolly-Seber Method Applied to Age-Stratified Populations,” (S. L. Stokes), Journal of Wildlife Management, 48, 1053–1059 (1984).
56. “Additional Comments on the Assumption of Homogeneous Survival Rates in Modern Bird Banding Estimation Models,” (J. Nichols, L. Stokes, J. Hines, and M. Conroy), Journal of Wildlife Management, 46, 953–962 (1982).
57. “Remarks on the Use of Mark-recapture Methodology in Estimating Avian Population Size” (J. Nichols, B. Noon, L. Stokes, and J. Hines), *Studies in Avian Biology*, 6, 121–136 (1981).
58. “Estimation of Variance Using Judgment Ordered Ranked Set Samples,” (S. L. Stokes), *Biometrics*, 36, 35–42 (1980).
59. “Inferences on the Correlation Coefficient in Bivariate Normal Populations from Ranked Set Samples,” (S. L. Stokes), *Journal of the American Statistical Association*, 75, 989–995 (1980).
60. “Ranked Set Sampling with Concomitant Variables,” (S. L. Stokes), *Communications in Statistics – Theory and Methods*, 6, 1207–1211 (1977).

## *Book Chapters*

1. “Measuring treatment fidelity with reliability and validity across a program of intervention research: Practical and theoretical considerations,” (Allor, J. H. and Stokes, L.), In G. Roberts, S. Vaughn, S. N. Beretvas, and V. Wong (Eds.), *Measuring and Modeling Treatment Fidelity in Studies of Educational Intervention*, New York: Routledge Taylor & Francis Group (2017).
2. “Interviewer Effects,” in *Encyclopedia of Research Methods for the Social Sciences*, M. Lewis-Beck, A. Brayman and T.F. Liao, Editors, Sage Publications (2003).
3. “Identifying and Adjusting for Recall Error with Application to Fertility Surveys,” (T. Pullum and L. Stokes), Chapter 31 (pp. 711–732), *Survey Measurement and Process Quality*, John Wiley and Sons (1997).
4. “A Cost-Effective Approach for Regulating Insurance Company Solvency,” (J. Lamm-Tennant, L. Starks, and L. Stokes), in *The Financial Dynamics of the Insurance Industry*, 153–167, E.I Altman and I.T. Vanderhoof, Editors, Irwin Professional Publishing, New York (1995).
5. “Some Recent Results on the Modeling and Estimation of Measurement Errors in Surveys,” (with P. Biemer and L. Stokes), Chapter 24 (pp. 487–516) in *Measurement Errors in Surveys*, John Wiley & Sons (1991).
6. “A New Approach to Identifying Sources of Interviewer Effects in Telephone Surveys,” (L. Stokes and M. Yeh), Chapter 22 (pp. 357–373) in *Telephone Survey Methodology*, Robert M. Groves, Editor, John Wiley and Sons, Inc. (1988).
7. “Ranked Set Sampling,” in *Encyclopedia of Statistical Sciences*, N. Johnson and S. Kotz, Editors, John Wiley & Sons, 585–588 (1986).

# Preface

When our colleague Lynne Stokes announced her intention to transition to emerita status at the end of the 2022 academic year, our initial reactions were dismay—at losing a valued colleague—and surprise—that she would walk away while still at the top of her game. How can you retire, Lynne; what will you do? And what will our department do without you?

After reconciling ourselves to the coming new reality, we decided that we should do something special to commemorate Lynne's remarkable career and recognize this momentous life change. A symposium, we thought—but Lynne said she did not want a symposium. Well then, a party hosting current and past colleagues and students. No, Lynne said, no party. Perhaps an intimate dinner with the faculty? No again. A Texas barbecue? A Lynne-themed Friday tea time? No and no. Well how about a *festschrift*?

And that is how this book came to be.

So we made the rounds of Lynne's many students, co-authors, and past and current colleagues, who were universally eager to contribute papers in areas where she has worked over the years. We express our sincere gratitude to all of them for writing chapters of such high quality on a tight deadline. Special thanks are also due to the referees, many of them authors as well, for their constructive reviews. And we acknowledge the team from Springer Nature Group—Laura Aileen Briskman, Kirthika Selvaraju, Faith Su, and Amelie von Zumbusch—who have gently guided the project from inception to production.

Most importantly, we are grateful to our colleague and friend Lynne Stokes for blessing this work and for supporting our efforts with her characteristic energy, generosity, and humility. It is our great pleasure to present her with this book on the occasion of her transition to the next phase of a most interesting and well-lived life.

Waltham, MA, USA — Hon Keung Tony Ng
Dallas, TX, USA — Daniel F. Heitjan
June 2022

# Contents

# Contributors

**Soohyun Ahn** Department of Mathematics, Ajou University, Gyeonggi, Korea

**Talha Alvi** Department of Psychology, Southern Methodist University, Dallas, TX, USA

**William R. Andrews** NOAA Fisheries, Silver Spring, MD, USA

**Betsy Jane Becker** College of Education, Florida State University, Tallahassee, FL, USA

**Paul P. Biemer** RTI International, Research Triangle Park, NC, USA

**J. Michael Brick** Westat, Rockville, MD, USA

**G. Gordon Brown** SAS Institute, Cary, NC, USA

**Jennifer Brown** University of Canterbury, School of Mathematics and Statistics, Christchurch, New Zealand

**Jing Cao** Department of Statistical Science, Southern Methodist University, Dallas, TX, USA

**John Foster** NOAA Fisheries, Silver Spring, MD, USA

**Ahmet Serhat Gözütok** Measurement and Evaluation in Education Educational Sciences, Ereğli Faculty of Education, Zonguldak Bülent Ecevit University, Zonguldak, Turkey

**Yue Jia** Educational Testing Service, Princeton, NJ, USA

**Jiae Kim** Department of Statistics, Indiana University, Bloomington, IN, USA

**Olena Kravchuk** University of Adelaide, School of Agriculture, Food and Wine, Adelaide, SA, Australia

**Yi-Hsuan Lee** Educational Testing Service, Princeton, NJ, USA

**Johan Lim** Department of Statistics, Seoul National University, Seoul, Korea

**Bingchen Liu** Educational Testing Service, Princeton, NJ, USA

**Steven N. MacEachern** Department of Statistics, The Ohio State University, Columbus, OH, USA

**John Mazzeo** Educational Testing Service, Princeton, NJ, USA

**Vincent T. Mule Jr.** U.S. Census Bureau, Suitland, MD, USA

**Mary H. Mulry** U.S. Census Bureau, Suitland, MD, USA

**James D. Nichols** Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL, USA

**Linh H. Nghiem** School of Mathematics and Statistics, University of Sydney, Sydney, NSW, Australia
Research School of Finance, Actuarial Studies and Statistics, Australian National University, Canberra, ACT, Australia

**Omer Ozturk** Department of Statistics, The Ohio State University, Columbus, OH, USA

**Bivin Philip Sadler** Master of Science in Data Science, Southern Methodist University, Dallas, TX, USA

**S. Lynne Stokes** Department of Statistical Science, Southern Methodist University, Dallas, TX, USA

**Benjamin A. Tabak** Department of Psychology, Southern Methodist University, Dallas, TX, USA

**Zachary Wallmark** Department of Musicology and Ethnomusicology, University of Oregon, Eugene, OR, USA

**Xinlei Wang** Department of Statistical Science, Southern Methodist University, Dallas, TX, USA

**Christopher Wiesen** The Odum Institute for Research in Social Sciences, University of North Carolina, Chapel Hill, NC, USA

**Benjamin Williams** Department of Business Information and Analytics, Daniels College of Business, University of Denver, Denver, CO, USA

# Part I
# Ranked-Set Sampling, Judgement Post-stratified Sampling, and Capture-Recapture Methods

# Predictive Modelling and Judgement Post-stratification

**Steven N. MacEachern and Jiae Kim**

**Abstract** Predictive modelling has come to the forefront of statistics in recent years as interest in forecasting the results of experiments and interventions has increased. We now routinely see forecasts in the news media that include point predictions, an assessment of variation to accompany the prediction and even a full predictive distribution. In the area of ranked set sampling, Stokes and coauthors' work on the use of measured order statistics, and their concomitants provided a crucial step that allows one to pass from the subjective assessment of ranks of responses within a set to the use of covariates. The transition also allows one to make use of formal models for a response given measured covariates to improve upon the basic ranked set sampling estimators while retaining the robustness properties of the method. This chapter pursues the use of predictive distributions in the context of ranked set sampling. We find that the predictive viewpoint naturally leads us away from imposing a strict ranking on the units in a set to expressing a distribution over ranks for each unit in the set. In turn, this change suggests the use of judgement post-stratification rather than ranked set sampling. It also yields novel estimators which are shown to outperform the standard estimators.

## 1 Ranked Set Sampling and Judgement Post-stratification

Stokes' pioneering work (Stokes, 1977) brought measured covariates to ranked set sampling (RSS). Briefly restating her work and establishing notation, consider a set of $nH$ units that are partitioned at random into $n$ sets, each of size $H$. The units are presumed to form a random sample from some distribution. Within a given set, we

S. N. MacEachern (✉)
Department of Statistics, The Ohio State University, Columbus, OH, USA
e-mail: snm@stat.osu.edu

J. Kim
Department of Statistics, Indiana University, Bloomington, IN, USA
e-mail: jiaekim@iu.edu

 
H. K. T. Ng, D. F. Heitjan (eds.), *Recent Advances on Sampling Methods and Educational Statistics*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-14525-4_1

begin with $(X_h, Y_h), h = 1, \ldots, H$. These units are ranked on the $X_h$, so that $X_{(r:H)}$ is the $r$th order statistic in the set. The measured response, $Y_{[r:H]}$, associated with this unit is its concomitant. To draw a RSS of size $n$ from such a population, sample sizes $n_h$, $h = 1, \ldots, H$, are specified, with $\sum_{h=1}^{H} n_h = n$. One unit is drawn from each of the $n$ sets; in $n_h$ sets, the unit ranked $h$ is selected. The resulting sample is a RSS.

The earliest description of RSS appears in McIntyre (1952) (republished as McIntyre, 2005). In McIntyre's description of the technique, ranking is based on the subjective judgement of an experimenter who examines each set of $H$ units, specifying the ranks of the units in the set. Once the units in each set have been ranked, the sample is drawn as described above and the response of interest, $Y$, is measured on the $n$ sampled units. Extending our notation to capture both set and rank within set, the mean of the $nH$ units is

$$\bar{Y} = (nH)^{-1} \sum_{i=1}^{n} \sum_{h=1}^{H} Y_{ih} \ , \tag{1}$$

where $Y_{ih}$ is the response of the unit with rank $h$ in set $i$. Suppressing the notation for the rank, define $Y_i$ to the be $i$th of the $n$ sampled units. Provided $n_h > 1$ for all $h$,

$$\bar{Y}_{rss} = H^{-1} \sum_{h=1}^{H} \bar{Y}_h \ , \tag{2}$$

where $\bar{Y}_h$ is the sample mean of the $n_h$ sampled units with rank $h$. The RSS estimator is unbiased: $E[\bar{Y}_{rss} \mid \bar{Y}] = \bar{Y}$ for any collection of $nH$ units. Furthermore, when the units are a random sample from a distribution with mean $\mu = E[Y]$, $E[\bar{Y}_{rss}] = E[\bar{Y}] = \mu$. The goal of RSS is to estimate $\mu$. Stokes and Sager (1988) cast estimation of a cumulative distribution function as estimation of a proportion (mean) for all cut points on the real line.

RSS with estimation following (2) is robust to variation in the specifics of how the ranks are created. When created subjectively, better ranking leads to greater separation of the means of the rank classes (or strata), in turn leading to greater reduction in variance relative to estimators based on a random sample from the population. When ranks arise from a measured covariate, the same holds. Sound experimental practice includes blinding the ranker to which units will be fully measured. When implemented, the estimator is unbiased for $\mu$ as long as the ranks can be determined before the responses of the selected units are measured. If the ranking process makes modest use of the measured responses, the bias is typically small.

Judgement post-stratification (JPS) is a common variant of ranked set sampling. To draw a JPS sample, a collection of $n$ units is selected for full measurement (both $X$ and $Y$) from the population as described above. For each fully measured unit, a set is filled out by independently drawing an additional $H - 1$ units. For these supplemental units, only $X$ is measured. The end result is $n$ sets of $H$ units.

Within a set, $X_h$ is measured for all $H$ units, while $Y_h$ is measured for a single unit. Upon ranking the units, conceptually, we have the pairs $(X_{(r:H)}, Y_{[r:H]})$, for $r = 1, \ldots, H$. In practice, most of the responses are missing and we have only one measured response, $Y_{[r:H]}$, for some $r$. When units are ranked on the basis of a measured covariate, the name *judgement* post-stratification is a misnomer. The name stems from the original work on the technique (MacEachern et al., 2004) where ranking was based on subjective judgement about the units.

An equivalent description of JPS exists. As in the RSS, we could form $n$ sets, each consisting of $H$ units. Instead of specifying the $n_h$, $h = 1, \ldots, H$, we select a single unit at random from each set for full measurement. With ranks based on the measured covariate, the $n_h$, $h = 1, \ldots, H$, are random variables. The vector $(n_1, \ldots, n_H)$ follows a multinomial distribution with $n$ trials and parameter vector $(1/H, \ldots, 1/H)$.

Whichever description of JPS is used, the data that are used for estimation consist of $n$ independent and identically distributed (IID) vectors $(Y_i, R_i)$, where $Y_i$ is the measured value and $R_i$ is the rank of the unit within its set. For estimation, we parallel the technique of post-stratification from survey sampling. Conditioning on the observed $n_h$ and using the estimator in Eq. (2), an estimator for $\mu$ can be obtained as

$$\hat{\mu}_{jps1} = H^{-1} \sum_{h=1}^{H} \bar{Y}_h = H^{-1} \frac{\sum_{i=1}^{n} Y_i I_{ih}}{\sum_{i=1}^{n} I_{ih}} , \tag{3}$$

where shorthand notation has $I_{ih} = I(R_i = h)$. The within-rank sample size is $n_h = \sum_{i=1}^{n} I_{ih}$, $h = 1, \ldots, H$. The resulting estimator is unbiased for $\mu$, conditional on all of the $n_h > 0$. Various patches exist to define the estimator when one or more $n_h = 0$. Frey and Feeman (2012) and Frey (2016) developed methods to reduce the mean square error of $\hat{\mu}_{jps1}$ by allowing some conditional bias in the estimator.

To extend the technique to two rankers, the data used for estimation consist of the vectors $(Y_i, R_{1i}, R_{2i})$. The information from both rankers is used to form the estimator

$$\hat{\mu}_{jps2} = H^{-1} \sum_{h=1}^{H} \frac{\sum_{i=1}^{n} Y_i p_{ih}}{\sum_{i=1}^{n} p_{ih}} , \tag{4}$$

where $p_{ih} = [I(R_{1i} = h) + I(R_{2i} = h)]/2$. The notation $p_{ih}$ reflects an empirical estimate of the probability that the $i$th fully measured unit has rank $h$. The method is easily extended to more than two rankers and to rankers of differing quality.

The move from RSS to JPS has several advantages. For one, it allows the experimenter to use a conventional design (based on a random sample from the population), with estimates improved by the use of covariates measured on additional units. A second advantage is that JPS can be used in situations where the units are not actually ranked. This may be due to disagreements between multiple rankers as in MacEachern et al. (2004), or it may be due to the presence of more than one informative covariate, as in Wang et al. (2006). Wolfe provided an insightful review of RSS, JPS and related techniques (Wolfe, 2012).

## 2 Multivariate Order Statistics and JPS

In Wang et al. (2006), Stokes and coauthors posed the intriguing question of how to use multiple covariates to convey information about the ranks of units for use in JPS. Their solution is to rank on each of the distinct covariates. In the case of a continuous bivariate covariate, $(X_1, X_2)$, each of the units in the set would be assigned a pair of ranks—one for $X_1$ and the other for $X_2$. This pair of ranks defines the post-stratum (or rank class) of the unit. For a set of size $H$, there are $H^2$ post-strata. We denote these post-strata with $\mathbf{r} = (r_1, r_2)$, where $r_1, r_2 \in \{1, \ldots, H\}$. We focus on a bivariate covariate but note that the technique extends to covariates of greater dimension. Figure 1 illustrates the situation for a bivariate order statistic for set size $H = 5$.

The increase in the number of post-strata from $H$ to $H^2$ necessitates reconsideration of the basic post-stratification estimator (3). Marginally, each covariate for the measured unit will have rank $r_i = h$ with probability $1/H$ for $i = 1, 2$ and $h = 1, \ldots, H$. The joint distribution of $\mathbf{R}$ leads to the stratum probability $\pi_{\mathbf{r}} = P(\mathbf{R} = \mathbf{r})$. In general, these probabilities can be found via numerical integration if the model for $(X_1, X_2)$ is fully specified. Some of the $\pi_{\mathbf{r}}$ may be much smaller than $H^{-2}$, leading to a large probability that the estimator is undefined.

Wang et al. (2006) handled this issue by appealing to a parametric model as an aid to estimation. The authors defined $\mu_{[\mathbf{r}]} = E[Y \mid \mathbf{R} = \mathbf{r}]$. The value of $\mu_{[\mathbf{r}]}$ can be found by numerical integration over the conditional distribution of $Y \mid \mathbf{R}$. Once the

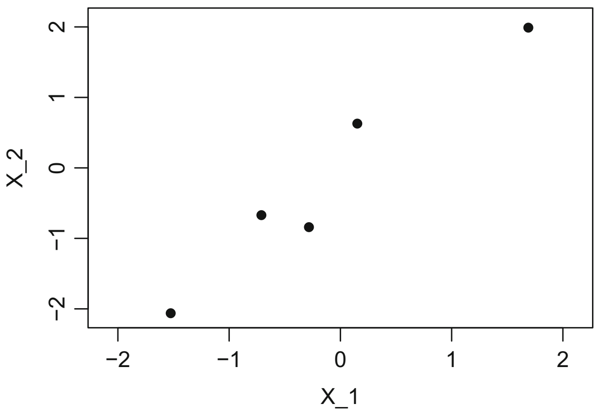



**Fig. 1** Covariate pairs for a set of size $H = 5$. The bivariate rank vectors are (1, 1), (2, 3), (3, 2), (4, 4), and (5, 5). The ranks based on $X_1$ and $X_2$ agree for three of the five items and disagree for two. Extreme differences in the ranks may be very rare

stratum means are in place, they are connected to the mean of $Y$ via the expression $\mu = \sum_{\mathbf{r}} \pi_{\mathbf{r}} \mu_{[\mathbf{r}]}$. It is helpful to introduce the difference between the stratum mean and the overall mean, $\delta_{[\mathbf{r}]} = \mu_{[\mathbf{r}]} - \mu$. The authors suggested estimation by ordinary least squares applied to a model for $\mu$, with observations in stratum $\mathbf{r}$ offset by $\delta_{[\mathbf{r}]}$. The data are $(Y_i, \mathbf{r}_i)$, $i = 1, \ldots, n$, and the estimator is

$$\hat{\mu}_{oLS} = n^{-1} \sum_{i=1}^{n} (Y_i - \delta_{[\mathbf{r}_i]}) \ . \tag{5}$$

The estimator $\hat{\mu}_{oLS}$ can be viewed in two stages: In the first, each observation is bias-corrected by subtracting its $\delta_{[\mathbf{r}]}$; in the second, the sample mean of the bias-corrected observations is computed. Partitioning the sample into strata reduces the within-stratum variances. Removing bias and then using the sample mean ensures that each observation receives equal weight in the estimator. Together, these two stages lead to substantial variance reduction, especially for relatively large set sizes.

In a refinement, Wang et al. (2006) suggested consideration of a weighted least squares estimator that takes within-stratum variances into account. The within-stratum variances are computed on the basis of numerical integration. This estimator takes the form

$$\hat{\mu}_{wLS} = \frac{\sum_{i=1}^{n} \sigma_{\mathbf{r}_i}^{-2} (Y_i - \delta_{[\mathbf{r}_i]})}{\sum_{i=1}^{n} \sigma_{\mathbf{r}_i}^{-2}} \ . \tag{6}$$

In the event that the $\delta_{[\mathbf{r}]}$ and $\sigma_{\mathbf{r}_i}^2$ are estimated, we place hats over them to denote this. In the framework of bias-corrected estimators, $\hat{\mu}_{oLS}$ and $\hat{\mu}_{wLS}$ are excellent performers—the mean and an optimally weighted mean. Wang et al. (2006) demonstrated superior performance of these new estimators when the class of models (multivariate normal distributions) is correct and the parameters in the model are known or are estimated.

The theory developed in Wang et al. (2006) implies that the weighted average of the offsets is zero for every model for which $\mu$ exists. That is,

$$\sum_{\mathbf{r}} \pi_{\mathbf{r}} \delta_{[\mathbf{r}]} = 0 \ . \tag{7}$$

This is a delicate expression, as it is naturally satisfied when both the $\pi_{\mathbf{r}}$ and the $\delta_{[\mathbf{r}]}$ are correctly specified. Asymptotically, we expect the expression to hold if we replace these two quantities with consistent estimators of them. If not, one would expect the expression (7) to evaluate to something other than 0, leaving us with a Fisher-consistent estimator of a quantity near, but not exactly equal to, $\mu$.[1]

[1]Huber (1981), in his study of robustness, found a need to redefine consistency when the distribution that generates the data might not lie in a tidy parametric family. His definition of Fisher consistency focuses on functionals of the empirical distribution converging to a well-defined population quantity. This quantity often differs from the nominal target of inference.

An open question is whether one can develop estimators that are nearly as stable as $\hat{\mu}_{oLS}$ and $\hat{\mu}_{wLS}$ and yet show more robustness to violations of the model that is implicit in their construction. In the sequel, we develop estimators that show greater robustness to departures from the joint model for $\mathbf{X}$ and from the model for $Y|\mathbf{X}$. In certain circumstances, our estimators show greater stability than do those of Wang et al. (2006).

## 3 Consistency of JPS Estimators

The literature on RSS and JPS demonstrates the consistency of the estimators $\bar{Y}_{rss}$ and $\hat{\mu}_{jps1}$ in (2) and (3), respectively, under minimal conditions. These traditional estimators borrow heavily from the design-based perspective of survey sampling, where (approximate) unbiasedness is prized. Small variance is the secondary consideration. Modern work with surveys adjusts the balance, relying more heavily on models, especially where missing data is a concern (Lohr, 2010). With this perspective, a bit more bias is allowed, provided it is accompanied by a substantial reduction in variance. Simulations are used to evaluate the estimators' performance when the model does not hold. Wang et al. (2006) pursued this path.

We work in the infinite population setting where we collect IID sets, observing a single member of each set. As such, we envision that the data come from some distribution which we refer to as the "true model". In addition, there is a model used to construct the estimator. We assume that $\mu$ exists under both models. Consistency concerns arise when the true model and that used for analysis differ.

To set the framework for our consideration of robustness, we split the models into two parts. The first is the conditional distribution of $Y \mid \mathbf{R}$. The second is the distribution of $\mathbf{R}$ for the unit that is to be fully measured. The true and analysis models may differ in one or both of these aspects. A given estimator may be robust to differences in one portion of the model but not to differences in the other portion of the model. We consider each of the estimators in turn, presenting a heuristic argument for or against consistency. Our statements are to be taken loosely; simulations appearing in a later section support our claims. Formal statements and proofs of these results await another venue.

We briefly note that the estimators $\hat{\mu}_{jps1}$ and $\hat{\mu}_{jps2}$ are consistent for $\mu$. These estimators do not rely on a model, and so we need not consider the gap between the true and analysis models. Consistency was established in MacEachern et al. (2004).

The estimators based on parametric models, $\hat{\mu}_{oLS}$ and $\hat{\mu}_{wLS}$, may or may not be consistent. We begin with $\hat{\mu}_{oLS}$. For a given stratum $\mathbf{r}$, an offset observation, $Y - \delta_{[\mathbf{r}]} = Y - \mu_{[\mathbf{r}]} + \mu$, has mean $\mu$—provided the true and analysis models agree for the distribution of $Y|(\mathbf{R} = \mathbf{r})$ so that $\mu_{[\mathbf{r}]}$ has the same value under the two models and the offset has been correctly specified (or will be estimated consistently). Averaging across the strata, we see that the estimator targets the quantity $\mu - \sum_{\mathbf{r}} \pi_{\mathbf{r}} \delta_{[\mathbf{r}]}$. The estimator will be consistent for $\mu$ if (7) holds so that the average offset is zero. It is clear that this will be the case when the distribution on $\mathbf{R}$ and the conditional mean