

Understanding Elections through Statistics

Polling, Prediction, and Testing
Second Edition

Ole J. Forsberg



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

Second edition published 2025
by CRC Press
2385 Executive Center Drive, Suite 320, Boca Raton, FL 33431, U.S.A.

and by CRC Press
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

© 2025 Ole J. Forsberg

First edition published by CRC Press 2021

Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged, please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, access www.copyright.com or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. For works that are not available on CCC, please contact mpkbookspermissions@tandf.co.uk

Trademark notice: Product or corporate names may be trademarks or registered trademarks and are used only for identification and explanation without intent to infringe.

ISBN: 978-1-032-68453-6 (hbk)

ISBN: 978-1-032-62186-9 (pbk)

ISBN: 978-1-032-68455-0 (ebk)

DOI: 10.1201/9781032684550

Typeset in Latin Modern
by KnowledgeWorks Global Ltd.

Publisher's note: This book has been prepared from camera-ready copy provided by the authors.

Contents

Preface to the Second Edition	xiii
Acknowledgments	xix
List of Figures	xxi
List of Tables	xxv
About the Author	xxvii
Symbols	xxix
1 Introduction to Polling	1
1.1 The Importance of Political Polling	3
1.2 Polling Terms	5
1.3 Sampling Methods	6
1.4 Biases in Polling: The First Words	10
1.5 Legitimacy in Polling	12
1.6 Experiments in Polling	14
1.7 Conclusion and Looking Ahead	17
1.8 Extensions	18
1.9 Chapter Appendix	19
2 Simple Random Sampling	21
2.1 Simple Random Sampling (SRS)	24
2.2 One Estimator of π : The Sample Proportion	24
2.3 SRS without Replacement	35
2.4 Conclusion	36

2.5	Extensions	37
2.6	Chapter Appendix	38
3	Interval Estimation	43
3.1	Margin of Error	45
3.2	Confidence Intervals	50
3.3	Credible Intervals	55
3.4	Conclusion	61
3.5	Extensions	62
3.6	Chapter Appendix	63
4	Stratified Sampling	68
4.1	Stratified Sampling	71
4.2	The Mathematics of Estimating π	74
4.3	Confidence Intervals	85
4.4	Conclusion	89
4.5	Extensions	90
4.6	Chapter Appendix	91
5	The Bayesian Solution	98
5.1	Overview: Why Bayesian Analysis	100
5.2	Bayes Law	103
5.3	A Particularly Helpful Shortcut	108
5.4	The Bayesian Calculations	110
5.5	Bayesian p-Values	114
5.6	Conclusion: Current and Future Use in Polling	115
5.7	Extensions	116
5.8	Chapter Appendix	117
6	Aggregating Polls	121
6.1	Overview	123
6.2	Simple Averaging of Polls	124
6.3	Weighted Averaging of Polls	126
6.4	Averaging of Polls over Time	128

6.5	Looking Ahead	134
6.6	South Korean 2017 Presidential Election	138
6.7	Conclusion	142
6.8	Extensions	143
6.9	Chapter Appendix	144
7	Going Beyond the Two-Race	151
7.1	Overview	152
7.2	The Multiple Comparisons Problem	153
7.3	The <i>Real</i> Problem	156
7.4	The Bayesian Solution	158
7.5	A Brief Return to Fisher	164
7.6	Conclusion	166
7.7	Extensions	167
7.8	Chapter Appendix	168
8	Testing Elections: Digit Tests	170
8.1	Electoral Forensics	172
8.2	History and Insight	173
8.3	The Benford Test	175
8.4	Extending Newcomb and Benford	180
8.5	Using the Generalized Benford Distribution	183
8.6	Conclusion	190
8.7	Extensions	191
8.8	Chapter Appendix	192
9	Testing Elections: Regression Tests	201
9.1	Differential Invalidation	203
9.2	Regression Modeling	210
9.3	Examining Côte d'Ivoire	220
9.4	Conclusion	222
9.5	Extensions	222
9.6	Chapter Appendix	223

10 Polling Analysis: Brexit Vote, 2016	229
10.1 Know Your Data	230
10.2 Combining the Polls	236
10.3 Discussion: What Went Wrong?	237
10.4 Conclusion	239
11 Election Analysis: Collier County	241
11.1 The CVR Files	242
11.2 Collier County Background	244
11.3 Fundamental Question	248
11.4 Data Preparation and Exploration	250
11.5 Conclusion	256
12 System Analysis: Sri Lanka Since 1994	257
12.1 Differential Invalidation	259
12.2 Methods and Data	259
12.3 Results by Election	262
12.4 Discussion	273
12.5 Conclusion	274
12.6 A Sri Lankan Postscript	275
13 An Afterword: Biases in Polling	276
13.1 The Sources of Bias and Variance	277
13.2 A Concluding Diatribe: The Blame for 2020	281
A A Brief Introduction to R	282
A.1 Installing R on Your Computer	283
A.2 A Quick, Sample Session	284
A.3 A Second Example	291
A.4 Conclusion	294

B Four Probability Distributions 295

- B.1 A Brief Introduction to Probability 295
- B.2 The Binomial Distribution 297
- B.3 The Multinomial Distribution 308
- B.4 The Beta Distribution 309
- B.5 The Dirichlet Distribution 311
- B.6 End of Appendix Materials 315

Bibliography 317

Index 331