



Data Analysis by Resampling: Concepts and Applications

Clifford E. Lunneborg

Suranaree University of Technology



31051090600029

D U N B U R Y P R E S S



Table of Contents

Part I: Resampling Concepts

Introduction		1
Concepts 1:	Terms and Notation	5
	Cases, Attributes, Scores, and Treatments	5
	Experimental and Observational Studies	7
	Data Sets, Samples, and Populations	7
	Parameters, Statistics, and Distributions	8
	Distribution Functions	9
Applications 1:	Cases, Attributes, and Distributions	13
	Attributes, Scores, Groups, and Treatments	13
	Distributions of Scores and Statistics	13
	Exercises	14
Concepts 2:	Populations and Random Samples	19
	Varieties of Populations	19
	Random Samples	24
Applications 2:	Random Sampling	27
	Simple Random Samples	27
	Exercises	30
Concepts 3:	Statistics and Sampling Distributions	33
	Statistics and Estimators	33
	Accuracy of Estimation	34
	The Sampling Distribution	34
	Bias of an Estimator	37
	Standard Error of a Statistic	37
	RMS Error of an Estimator	38
	Confidence Interval	38
Applications 3:	Sampling Distribution Computations	43
	Exercises	43
Concepts 4:	Testing Population Hypotheses	45
	Population Statistical Hypotheses	45

	Population Hypothesis Testing	46
Applications 4:	Null Sampling Distribution p-Values	51
	The p-Value of a Directional Test	52
	The p-Value of a Nondirectional Test	53
	Exercises	55
Concepts 5:	Parametrics, Pivotal, and Asymptotics	57
	The Unrealizable Sampling Distribution	57
	Sampling Distribution of a Sample Mean	58
	Parametric Population Distributions	60
	Pivotal Form Statistics	60
	Asymptotic Sampling Distributions	63
	Limitations of the Mathematical Approach	65
Applications 5:	CI for Normal Population μ and σ^2	67
	CI for a Normal Population Mean	67
	CI for a Normal Population Variance	69
	Nonparametric CI Estimation	71
	Exercises	72
Concepts 6:	Limitations of Parametric Inference	75
	Range and Precision of Scores	75
	Size of Population	76
	Size of Sample	76
	Roughness of Population Distributions	77
	Parameters and Statistics of Interest	77
	Scarcity of Random Samples	78
	Resampling Inference	78
Applications 6:	Resampling Approaches to Inference	83
	Exercises	83
Concepts 7:	The Real and Bootstrap Worlds	85
	The Real World of Population Inference	85
	The Bootstrap World of Population Inference	87
	Real-World Population Distribution Estimates	88
	Nonparametric Population Estimates	96
	Sample Size and Distribution Estimates	98
Applications 7:	Bootstrap Population Distributions	99
	Nonparametric Population Estimates	99
	Exercises	103
Concepts 8:	The Bootstrap Sampling Distribution	105
	The Bootstrap Conjecture	105

	Complete Bootstrap Sampling Distributions	106
	Monte Carlo Bootstrap Distributions	108
	The Bootstrap Estimate of Standard Error	110
	The Bootstrap Estimate of Bias	111
	Simple Bootstrap CI Estimates	112
Applications 8:	Bootstrap SE, Bias, and CI Estimates	119
	Example	119
	Exercises	122
Concepts 9:	Better Bootstrap CIs: The Bootstrap-t	125
	Pivotal Form Statistics	125
	The Bootstrap- t Pivotal Transformation	127
	Forming Bootstrap- t CIs	129
	Estimating the Standard Error of an Estimate	132
	Range of Application of the Bootstrap- t	135
	Iterated Bootstrap CIs	136
Applications 9:	SEs and CIs for Trimmed Means	137
	Definition of the Trimmed Mean	137
	Importance of the Trimmed Mean	138
	A Note on Outliers	139
	Determining the Trimming Fraction	140
	Sampling Distribution of the Trimmed Mean	141
	Applications	143
	Exercise	154
Concepts 10:	Better Bootstrap CIs: BCA Intervals	157
	Bias-Corrected and Accelerated CI Estimates	157
	Applications of the BCA CI	166
	Better Confidence Interval Estimates	166
Applications 10:	Using CI Correction Factors	167
	Requirements for a BCA CI	167
	Implementations of the BCA Algorithm	168
	Exercise	179
Concepts 11:	Bootstrap Hypothesis Testing	181
	CIs, Null Hypothesis Tests, and p-Values	181
	Bootstrap- t Hypothesis Testing	186
	Bootstrap Hypothesis Testing Alternatives	187
	CI Hypothesis Testing	190
	Confidence Intervals or p-Values?	193
Applications 11:	Bootstrap p-Values	195
	Computing a Bootstrap- t p-Value	195

	Fixed- α CIs and Hypothesis Testing	198
	Computing a BCA CI p-Value	200
	Exercise	202
Concepts 12:	Randomized Treatment Assignment	205
	Two Functions of Randomization	205
	Randomization of Sampled Cases	206
	Randomization of Available Cases	206
	Statistical Basis for Local Causal Inference	207
	Population Hypotheses Revisited	213
Applications 12:	Monte Carlo Reference Distributions	217
	Serum Albumen in Diabetic Mice	217
	Resampling Stats Analysis	219
	SC Analysis	222
	S-Plus Analysis	223
	Exercises	224
Concepts 13:	Strategies for Randomizing Cases	227
	Independent Randomization of Cases	227
	Completely Randomized Designs	231
	Randomized Blocks Designs	233
	Restricted Randomizations	235
	Constraints on Rerandomization	236
Applications 13:	Implementing Case Rerandomization	237
	Completely Randomized Designs	237
	Randomized Blocks Designs	243
	Independent Randomization of Cases	248
	Restricted Randomization	251
	Exercises	254
Concepts 14:	Random Treatment Sequences	257
	Between- and Within-Cases Designs	257
	Randomizing the Sequence of Treatments	258
	Causal Inference for Within-Cases Designs	260
	Sequence Randomization Strategies	265
Applications 14:	Rerandomizing Treatment Sequences	269
	Analysis of the AB-BA Design	269
	Sequences of $K > 2$ Treatments	276
	Exercises	285
Concepts 15:	Between- and Within-Cases Designs	289
	Between/Within Designs	289
	Between/Within Resampling Strategies	293

	Doubly Randomized Available Cases	294
Applications 15:	Interactions and Simple Effects	299
	Simple and Main Effects	299
	Exercise	308
Concepts 16:	Subsamples: Stability of Description	309
	Nonrandom Studies and Data Sets	309
	Local Descriptive Inference	310
	Description Stability and Case Homogeneity	311
	Subsample Descriptions	312
	Employing Subsample Descriptions	316
	Subsamples and Randomized Studies	318
Applications 16:	Structured and Unstructured Data	319
	Half-Samples of Unstructured Data Sets	319
	Subsamples of Source-Structured Cases	323
	Exercises	325

Part II: Resampling Applications

Introduction		331
Applications 17:	A Single Group of Cases	335
	Random Sample or Set of Available Cases	335
	Typical Size of Score Distribution	334
	Variability of Attribute Scores	342
	Association Between Two Attributes	349
	Exercises	356
Applications 18:	Two Independent Groups of Cases	359
	Constitution of Independent Groups	359
	Location Comparisons for Samples	361
	Magnitude Differences, CR and RB Designs	368
	Magnitude Differences, Nonrandom Designs	372
	Study Size	375
	Exercises	381
Applications 19:	Multiple Independent Groups	385
	Multiple Group Parametric Comparisons	385
	Nonparametric K -Group Comparisons	389
	Comparisons among Randomized Groups	393
	Comparisons among Nonrandom Groups	401
	Adjustment for Multiple Comparisons	404
	Exercises	407

Applications 20:	Multiple Factors and Covariates	411
	Two Treatment Factors	411
	Treatment and Blocking Factors	423
	Covariate Adjustment of Treatment Scores	424
	Exercises	440
Applications 21:	Within-Cases Treatment Comparisons	447
	Normal Models, Univariate and Multivariate	447
	Bootstrap Treatment Comparisons	449
	Randomized Sequence of Treatments	452
	Nonrandom Repeated Measures	455
	Exercises	458
Applications 22:	Linear Models: Measured Response	465
	The Parametric Linear Model	465
	Nonparametric Linear Models	474
	Prediction Accuracy	498
	Linear Models for Randomized Cases	501
	Linear Models for Nonrandom Studies	503
	Exercises	506
Applications 23:	Categorical Response Attributes	513
	Cross-Classification of Cases	513
	The 2×2 Table	529
	Logistic Regression	531
	Exercises	546
Postscript:	Generality, Causality, and Stability	551
	Study Design and Resampling	552
	Resampling Tools	556
References		557
Index		561