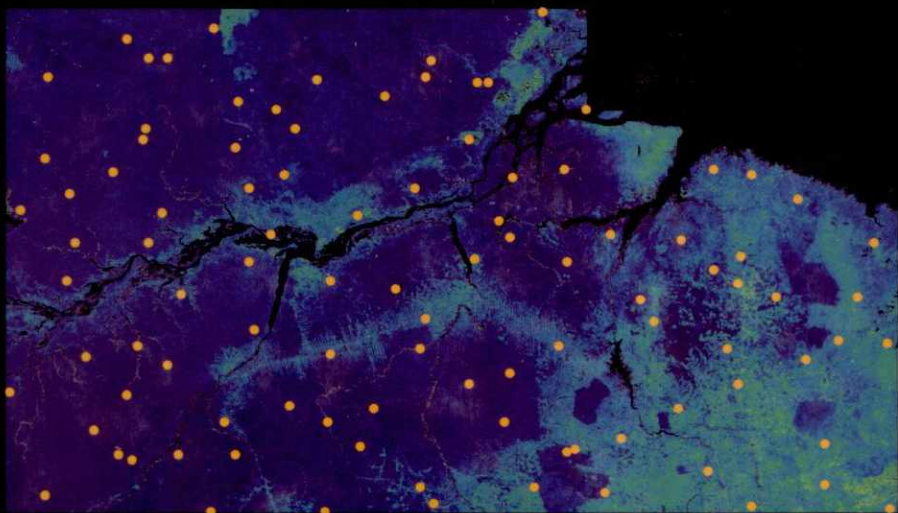


The R Series

Spatial Sampling with R



Dick J. Brus



CRC Press

Taylor & Francis Group

A CHAPMAN & HALL BOOK

Contents

Preface	xi
1 Introduction	1
1.1 Basic sampling concepts	1
1.1.1 Population parameters	4
1.1.2 Descriptive statistics vs. inference about a population	6
1.1.3 Random sampling vs. probability sampling	6
1.2 Design-based vs. model-based approach	7
1.3 Populations used in sampling experiments	8
1.3.1 Soil organic matter in Voorst, the Netherlands	10
1.3.2 Poppy fields in Kandahar, Afghanistan	11
1.3.3 Aboveground biomass in Eastern Amazonia, Brazil	12
1.3.4 Annual mean air temperature in Iberia	13
I Probability sampling for estimating population parameters	17
2 Introduction to probability sampling	19
2.1 Horvitz-Thompson estimator	21
2.2 Hansen-Hurwitz estimator	23
2.3 Using models in design-based approach	24
3 Simple random sampling	27
3.1 Estimation of population parameters	30
3.1.1 Population proportion	34
3.1.2 Cumulative distribution function and quantiles	34
3.2 Sampling variance of estimator of population parameters	36
3.3 Confidence interval estimate	40
3.3.1 Confidence interval for a proportion	43
3.4 Simple random sampling of circular plots	44
3.4.1 Sampling from a finite set of fixed circles	44
3.4.2 Sampling from an infinite set of floating circles	45
4 Stratified simple random sampling	49
4.1 Estimation of population parameters	51

4.1.1	Population proportion, cumulative distribution function, and quantiles	54
4.1.2	Why should we stratify?	56
4.2	Confidence interval estimate	60
4.3	Allocation of sample size to strata	60
4.4	Cum-root- f stratification	64
4.5	Stratification with multiple covariates	67
4.6	Geographical stratification	69
4.7	Multiway stratification	73
4.8	Multivariate stratification	73
5	Systematic random sampling	79
5.1	Estimation of population parameters	83
5.2	Approximating the sampling variance of the estimator of the mean	84
6	Cluster random sampling	93
6.1	Estimation of population parameters	97
6.2	Clusters selected with probabilities proportional to size, without replacement	102
6.3	Simple random sampling of clusters	103
6.4	Stratified cluster random sampling	106
7	Two-stage cluster random sampling	109
7.1	Estimation of population parameters	112
7.2	Primary sampling units selected without replacement	118
7.3	Simple random sampling of primary sampling units	119
7.4	Stratified two-stage cluster random sampling	121
8	Sampling with probabilities proportional to size	123
8.1	Probability-proportional-to-size sampling with replacement	125
8.2	Probability-proportional-to-size sampling without replacement	128
8.2.1	Systematic pps sampling without replacement	128
8.2.2	The pivotal method	130
9	Balanced and well-spread sampling	137
9.1	Balanced sampling	137
9.1.1	Balanced sample vs. balanced sampling design	138
9.1.2	Unequal inclusion probabilities	139
9.1.3	Stratified random sampling	145
9.1.4	Multiway stratification	148
9.2	Well-spread sampling	152
9.2.1	Local pivotal method	152
9.2.2	Generalised random-tessellation stratified sampling	154
9.3	Balanced sampling with spreading	158

10	Model-assisted estimation	161
10.1	Generalised regression estimator	162
10.1.1	Simple and multiple regression estimators	164
10.1.2	Penalised least squares estimation	172
10.1.3	Regression estimator with stratified simple random sampling	174
10.1.3.1	Combined regression estimator	178
10.2	Ratio estimator	181
10.2.1	Ratio estimators with stratified simple random sampling	185
10.2.2	Poststratified estimator	187
10.3	Model-assisted estimation using machine learning techniques	190
10.3.1	Predicting with a regression tree	192
10.3.2	Predicting with a random forest	195
10.4	Big data and volunteer data	198
11	Two-phase random sampling	201
11.1	Two-phase random sampling for stratification	203
11.2	Two-phase random sampling for regression	206
12	Computing the required sample size	211
12.1	Standard error	212
12.2	Length of confidence interval	213
12.2.1	Length of confidence interval for a proportion	214
12.3	Statistical testing of hypothesis	215
12.3.1	Sample size for testing a proportion	216
12.4	Accounting for design effect	218
12.5	Bayesian sample size determination	219
12.5.1	Bayesian criteria for sample size computation	220
12.5.1.1	Average length criterion	221
12.5.1.2	Average coverage criterion	222
12.5.1.3	Worst outcome criterion	222
12.5.2	Mixed Bayesian-likelihood approach	223
12.5.3	Estimation of population mean	223
12.5.4	Estimation of a population proportion	226
13	Model-based optimisation of probability sampling designs	231
13.1	Model-based optimisation of sampling design type and sample size	232
13.1.1	Analytical approach	233
13.1.1.1	Bulking soil aliquots into a composite sample	238
13.1.2	Geostatistical simulation approach	240
13.1.3	Bayesian approach	245
13.2	Model-based optimisation of spatial strata	252

14 Sampling for estimating parameters of domains	263
14.1 Direct estimator for large domains	264
14.2 Model-assisted estimators for small domains	266
14.2.1 Regression estimator	268
14.2.2 Synthetic estimator	271
14.3 Model-based prediction	273
14.3.1 Random intercept model	274
14.3.2 Geostatistical model	279
14.4 Supplemental probability sampling of small domains	284
15 Repeated sample surveys for monitoring population parameters	289
15.1 Space-time designs	289
15.2 Space-time population parameters	290
15.3 Design-based generalised least squares estimation of spatial means	293
15.3.1 Current mean	296
15.3.2 Change of the spatial mean	297
15.3.3 Temporal trend of the spatial mean	297
15.3.4 Space-time mean	298
15.4 Case study: annual mean daily temperature in Iberia	299
15.4.1 Static-synchronous design	300
15.4.2 Independent synchronous design	302
15.4.3 Serially alternating design	304
15.4.4 Supplemented panel design	306
15.4.5 Rotating panel design	308
15.4.6 Sampling experiment	310
15.5 Space-time sampling with stratified random sampling in space	312
II Sampling for mapping	315
16 Introduction to sampling for mapping	317
16.1 When is probability sampling not required?	317
16.2 Sampling for simultaneously mapping and estimating means	319
16.3 Broad overview of sampling designs for mapping	321
17 Regular grid and spatial coverage sampling	325
17.1 Regular grid sampling	325
17.2 Spatial coverage sampling	327
17.3 Spatial infill sampling	331
18 Covariate space coverage sampling	335
18.1 Covariate space infill sampling	338
18.2 Performance of covariate space coverage sampling in random forest prediction	340

19 Conditioned Latin hypercube sampling	345
19.1 Conditioned Latin hypercube infill sampling	351
19.2 Performance of conditioned Latin hypercube sampling in random forest prediction	351
20 Spatial response surface sampling	357
20.1 Increasing the sample size	365
20.2 Stratified spatial response surface sampling	367
20.3 Mapping	371
21 Introduction to kriging	375
21.1 Ordinary kriging	375
21.2 Block-kriging	381
21.3 Kriging with an external drift	383
21.4 Estimating the semivariogram	388
21.4.1 Method-of-moments	388
21.4.2 Maximum likelihood	389
21.5 Estimating the residual semivariogram	391
21.5.1 Iterative method-of-moments	392
21.5.2 Restricted maximum likelihood	393
22 Model-based optimisation of the grid spacing	397
22.1 Optimal grid spacing for ordinary kriging	397
22.2 Controlling the mean or a quantile of the ordinary kriging variance	399
22.3 Optimal grid spacing for block-kriging	402
22.4 Optimal grid spacing for kriging with an external drift	403
22.5 Bayesian approach	407
23 Model-based optimisation of the sampling pattern	413
23.1 Spatial simulated annealing	413
23.2 Optimising the sampling pattern for ordinary kriging	414
23.3 Optimising the sampling pattern for kriging with an external drift	420
23.4 Model-based infill sampling for ordinary kriging	423
23.5 Model-based infill sampling for kriging with an external drift	424
24 Sampling for estimating the semivariogram	429
24.1 Nested sampling	429
24.2 Independent sampling of pairs of points	436
24.3 Optimisation of sampling pattern for semivariogram estimation	441
24.3.1 Uncertainty about semivariogram parameters	442
24.3.2 Uncertainty about the kriging variance	445

24.4	Optimisation of sampling pattern for semivariogram estimation and mapping	450
24.5	A practical solution	456
25	Sampling for validation of maps	459
25.1	Map quality indices	461
25.1.1	Estimation of map quality indices	463
25.2	Real-world case study	464
25.2.1	Estimation of the population mean error and mean squared error	465
25.2.2	Estimation of the standard error of the estimator of the population mean error and mean squared error	467
25.2.3	Estimation of model efficiency coefficient	469
25.2.4	Statistical testing of hypothesis about population ME and MSE	470
26	Design-based, model-based, and model-assisted approach for sampling and inference	473
26.1	Two sources of randomness	473
26.2	Identically and independently distributed	476
26.3	Bias and variance	480
26.4	Effective sample size	482
26.5	Exploiting spatial structure in design-based approach	487
26.6	Model-assisted vs. model-dependent	488
A	Answers to exercises	495
	Bibliography	515
	Index	529