

GRTS Survey Designs for a Finite Resource

Thomas Kincaid

May 23, 2011

Contents

1 Preliminaries	1
2 Shapefile attribute data	2
3 Unstratified, equal probability, GRTS survey design	4
4 Stratified, equal probability, GRTS survey design	5
5 Unstratified, unequal probability, GRTS survey design with an oversample	7
6 Unstratified, unequal probability, GRTS survey design with an oversample and a panel structure for survey over time	9

This document presents example GRTS survey designs for a finite resource. The finite resource used in the designs is lakes in the New England region of the U.S. Four survey designs will be presented: (1) an unstratified, equal probability design; (2) a stratified, equal probability design; (3) an unstratified, unequal probability design with an oversample; and (4) an unstratified, unequal probability design with an oversample and a panel structure for survey over time. The sampling frame used for the survey designs is contained in either an ESRI shapefile, a data frame, or an `sp` package object. The frame contains the coordinates for a set of points that define the finite resource in addition to attribute data associated with the points. The coordinate system for the set of points in the sampling frame is an equal area projection rather than latitude and longitude. An equal area projection is used so that calculation of distance between points is valid. Use of the three sources for the sampling frame will be illustrated in the example survey designs.

1 Preliminaries

The initial step is to use the library function to load the `spsurvey` package. After the package is loaded, a message is printed to the R console indicating that the `spsurvey` package was loaded successfully.

Load the `spsurvey` package

```
> library(spsurvey)
```

Version 2.1 of the `spsurvey` package was loaded successfully.

2 Shapefile attribute data

The next step is to read the attribute data from the shapefile. The `read.dbf` function in the `spsurvey` package is used to read the attribute (dbf) file in the shapefile and assign it to a data frame named `att`. The initial six lines in the `att` data frame are printed using the `head` function.

Two attributes, state name and lake area category, that will be used to define, respectively, stratum codes and unequal selection probability (multidensity) categories for the survey designs are examined. State name is contained in a variable named "state", and lake area category is contained in a variable named "area_cat". For lake area category, lakes are classified by surface area measured in hectares. The `table` and `addmargin` functions are used to produce a table displaying number of lakes for each combination of values for the strata and multidensity category variables.

Read the attribute table from the shapefile

```
> att <- read.dbf("reg1_lakes")
```

Display the initial six lines in the attribute data frame

```
> head(att)
```

	lat_dd	lon_dd	xcoord	ycoord	state	area_cat
1	47.41608	-69.23301	2005501	2997939	ME	(1,5]
2	47.38340	-69.04088	2020520	2998539	ME	(1,5]
3	47.31454	-69.05682	2021486	2990894	ME	(500,7e+04]
4	47.37650	-69.13631	2013772	2995782	ME	(10,50]
5	47.34506	-69.09998	2017393	2993215	ME	(5,10]
6	47.34571	-69.11728	2016111	2992918	ME	(1,5]

Display number of lakes cross-classified by the strata and multidensity category variables

```
> addmargins(table(State = att$state, `Lake Area Category` = att$area_cat))
```

	Lake Area Category						
State	(1,5]	(10,50]	(5,10]	(50,500]	(500,7e+04]	[0,1]	Sum
CT	1181	284	270	90	4	483	2312
MA	1658	693	545	209	6	194	3305
ME	1792	1044	693	656	137	202	4524
NH	765	406	331	167	13	43	1725
RI	256	108	85	41	3	11	504
VT	418	138	137	52	12	46	803
Sum	6070	2673	2061	1215	175	979	13173

Lakes in the New England region are displayed in Figure 1. The `plot` function is used to produce the figure.

```
> plot(att$xcoord, att$ycoord, xlab="x-coordinate", ylab="y-coordinate", pch=20,  
+       cex=0.1, col="red")
```

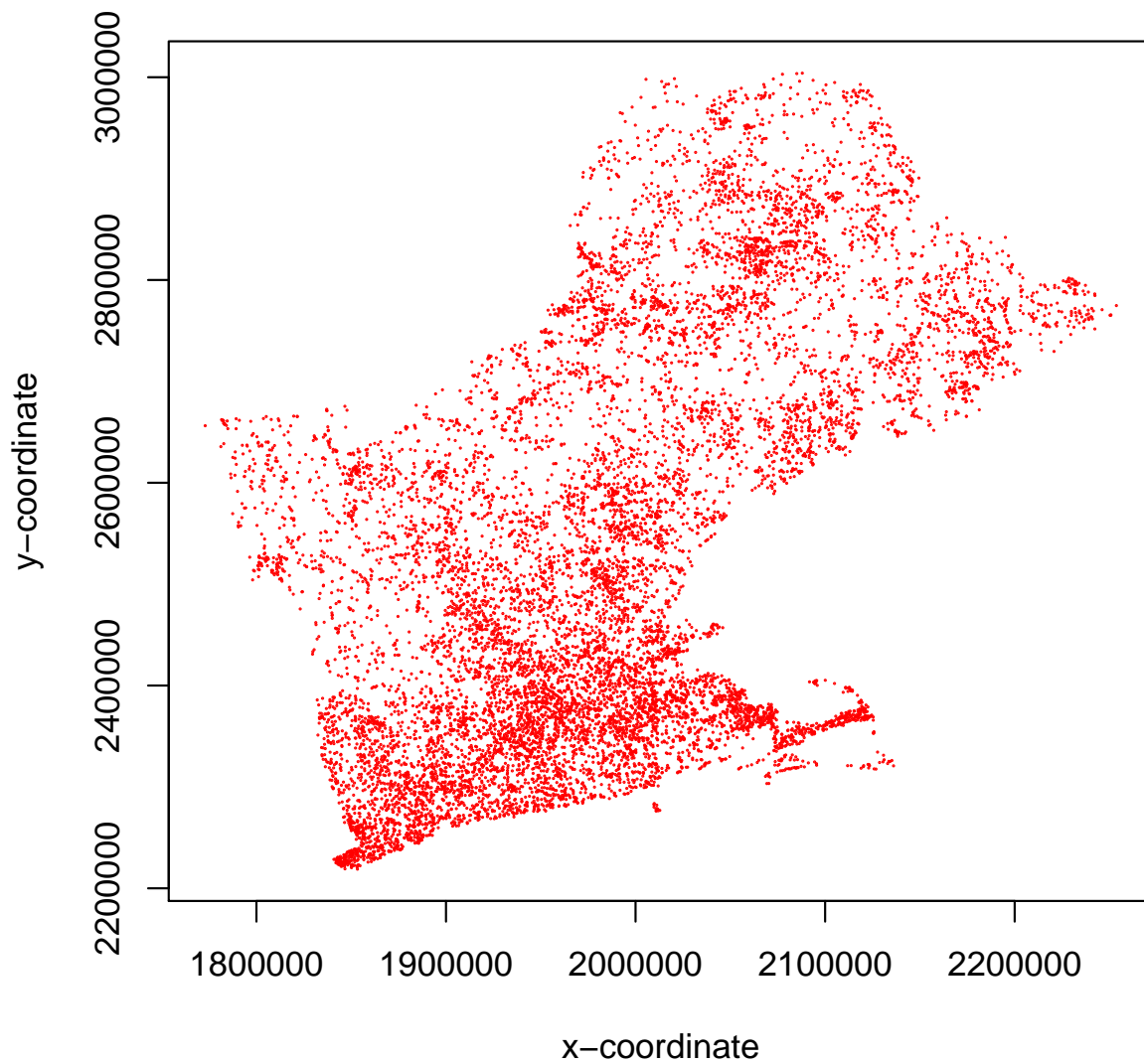


Figure 1: Lakes in the New England Region.

3 Unstratified, equal probability, GRTS survey design

The first survey design is an unstratified, equal probability design. The `set.seed` function is called so that, if necessary, the designs can be replicated.

The initial step is to create a list named `Equaldsgn` that contains information for specifying the survey design. Since the survey design is unstratified, the list contains a single item named "None" that also is a list. The "None" list includes two items: `panel`, which is used to specify the sample size for each panel, and `seltype`, which is used to input the type of random selection for the design. For this example, `panel` is assigned a single value named "PanelOne" that is set equal to 300, and `seltype` is assigned the value "Equal", which indicates equal probability selection.

The `grts` function in the `spsurvey` package is called to select the survey design. The following arguments are included in the call to `grts`: (1) `design`: the named list of stratum design specifications, which is assigned the `Equaldsgn` list; (2) `DesignID`: name for the design, which is used to create a site ID for each site and is assigned the value "EQUAL"; (3) `type.frame`: the type of frame, which is assigned the value "finite" to indicate a finite resource; (4) `src.frame`: source of the frame, which is assigned the value "shapefile" to indicate a shapefile frame; (5) `in.shape`: name of the input shapefile, which is assigned the value "reg1_lakes"; (6) `att.frame`: the data frame of attributes associated with elements in the frame, which is assigned the `att` data frame; and (7) `shapefile`: option to create a shapefile containing the survey design information, which is assigned `FALSE`.

During execution of the `grts` function, messages are printed that indicate the initial number of hierarchical levels used for the GRTS grid, the current number of levels, and the final number of levels. The set of messages is printed for each stratum, and is labeled with the stratum name. For this example, the set of messages is labeled "None", i.e., the name used in the `Equaldsgn` list. Upon completion of the call to `grts`, the initial six sites for the survey design and a design summary are printed.

Call the `set.seed` function so that the design can be replicated

```
> set.seed(4447864)
```

Create the design list

```
> Equaldsgn <- list(None = list(panel = c(PanelOne = 300), seltype = "Equal"))
```

Select the sample

```
> Equalsites <- grts(design=Equaldsgn,
+                   DesignID="EQUAL",
+                   type.frame="finite",
+                   src.frame="shapefile",
+                   in.shape="reg1_lakes",
+                   att.frame=att,
+                   shapefile=FALSE)
```

Stratum: None

Initial number of levels: 5

Current number of levels: 5

Current number of levels: 7

Final number of levels: 7

Print the initial six lines of the survey design

```
> head(Equalsites@data)
```

	siteID	xcoord	ycoord	mdcaty	wgt	stratum	panel	EvalStatus	EvalReason
1	EQUAL-001	2002873	2980126	Equal	43.91	None	PanelOne	NotEval	
2	EQUAL-002	2051755	2787941	Equal	43.91	None	PanelOne	NotEval	
3	EQUAL-003	2009584	2370445	Equal	43.91	None	PanelOne	NotEval	
4	EQUAL-004	1993945	2420687	Equal	43.91	None	PanelOne	NotEval	
5	EQUAL-005	1901755	2536905	Equal	43.91	None	PanelOne	NotEval	
6	EQUAL-006	1992222	2936987	Equal	43.91	None	PanelOne	NotEval	

	lat_dd	lon_dd	xcoord.1	ycoord.1	state	area_cat
1	47.26785	-69.33123	2002873	2980126	ME	(1,5]
2	45.48006	-69.41125	2051755	2787941	ME	(10,50]
3	41.97345	-71.30470	2009584	2370445	MA	(1,5]
4	42.44355	-71.33168	1993945	2420687	MA	(1,5]
5	43.65893	-72.06968	1901755	2536905	NH	(5,10]
6	46.91950	-69.62118	1992222	2936987	ME	(1,5]

Print the survey design summary

```
> dsgnsum(Equalsites)
```

Design Summary: Number of Sites

stratum	
None	Sum
300	300

4 Stratified, equal probability, GRTS survey design

The second survey design is a stratified, equal probability design. The state attribute is used to identify strata. List Stratdsgn is assigned design specifications. Stratdsgn includes six lists, one for each stratum. The names for the lists match the levels of the stratum variable, i.e., the unique values of the state attribute. Each list in Stratdsgn contains two items: panel and seltype. The value for panel is the same as the equal probability design (50), and seltype is assigned "Equal".

For this survey design, a data frame will be used as the sampling frame. Since it includes spatial coordinates, the att data frame will be used as the frame. The following arguments are included in the call to grts: (1) design: assigned the Stratdsgn list; (2) DesignID: assigned the value "STRATIFIED"; (3) type.frame: assigned the value "finite"; (4) src.frame: assigned the value "att.frame" to indicate that the sampling frame is provided by argument att.frame; (5) att.frame: assigned the att data frame; (6) xcoord: name of the column in the attributes data frame that identifies x-coordinates, which is assigned the value "xcoord"; (7) ycoord: name of the column in the attributes data frame that identifies y-coordinates, which is assigned the value "ycoord"; (8) stratum: name of the column in the attributes data frame that identifies the stratum code for each element in the frame, which is assigned the value "state"; and (9) shapefile: assigned the value FALSE. Upon completion of the call to grts, the initial six sites for the survey design and a design summary are printed.

Create the design list

```
> Stratdsign <- list(CT=list(panel=c(PanelOne=50), seltype="Equal"),
+                     MA=list(panel=c(PanelOne=50), seltype="Equal"),
+                     ME=list(panel=c(PanelOne=50), seltype="Equal"),
+                     NH=list(panel=c(PanelOne=50), seltype="Equal"),
+                     RI=list(panel=c(PanelOne=50), seltype="Equal"),
+                     VT=list(panel=c(PanelOne=50), seltype="Equal"))
```

Select the sample

```
> Stratsites <- grts(design=Stratdsign,
+                    DesignID="STRATIFIED",
+                    type.frame="finite",
+                    src.frame="att.frame",
+                    att.frame=att,
+                    xcoord="xcoord",
+                    ycoord="ycoord",
+                    stratum="state",
+                    shapefile=FALSE)
```

Stratum: CT

```
Current number of levels: 3
Current number of levels: 4
Current number of levels: 5
Final number of levels: 5
```

Stratum: MA

```
Current number of levels: 3
Current number of levels: 5
Final number of levels: 5
```

Stratum: ME

```
Current number of levels: 3
Current number of levels: 5
Final number of levels: 5
```

Stratum: NH

```
Current number of levels: 3
Current number of levels: 5
Final number of levels: 5
```

Stratum: RI

```
Current number of levels: 3
Current number of levels: 5
Final number of levels: 5
```

Stratum: VT

```
Current number of levels: 3
Current number of levels: 5
Final number of levels: 5
```

Print the initial six lines of the survey design

```
> head(Stratsites@data)
```

	siteID	xcoord	ycoord	mdcaty	wgt	stratum	panel	EvalStatus
1	STRATIFIED-001	1869859	2236177	Equal	46.24	CT	PanelOne	NotEval
2	STRATIFIED-002	1911013	2313970	Equal	46.24	CT	PanelOne	NotEval
3	STRATIFIED-003	1906694	2282263	Equal	46.24	CT	PanelOne	NotEval
4	STRATIFIED-004	1930344	2294624	Equal	46.24	CT	PanelOne	NotEval
5	STRATIFIED-005	1851656	2238601	Equal	46.24	CT	PanelOne	NotEval
6	STRATIFIED-006	1872474	2255853	Equal	46.24	CT	PanelOne	NotEval

	EvalReason	lat_dd	lon_dd	xcoord.1	ycoord.1	area_cat
1		41.11857	-73.33996	1869859	2236177	(10,50]
2		41.70514	-72.63349	1911013	2313970	(1,5]
3		41.43965	-72.77708	1906694	2282263	(10,50]
4		41.49531	-72.46404	1930344	2294624	(1,5]
5		41.17784	-73.54560	1851656	2238601	(1,5]
6		41.28379	-73.25348	1872474	2255853	(1,5]

Print the survey design summary

```
> dsgnsum(Stratsites)
```

Design Summary: Number of Sites

stratum						
CT	MA	ME	NH	RI	VT	Sum
50	50	50	50	50	50	300

5 Unstratified, unequal probability, GRTS survey design with an oversample

The third survey design is an unstratified, unequal probability design with an oversample. Lake area classes are used to identify multidensity categories. List Unequaldsgn is assigned design specifications. Since the survey design is unstratified, Unequaldsgn includes a single list named "None" that contains four items: panel, seltype, caty.n, and over. The value for panel is the same as the equal probability design, and seltype is assigned "Unequal" to indicate unequal selection probabilities. The third item, caty.n, assigns sample sizes for each of the six multidensity categories. Note that the sum of sample sizes provided in caty.n must equal the value in panel. The fourth item, over, is assigned the value 120, which specifies an oversample of 120 sites. An oversample is replacement sites for the survey design. The grts function attempts to distribute the oversample proportionately among sample sizes for the multidensity categories. If the oversample proportion for one or more categories is not a whole number, a warning message is printed and the proportion is rounded to the next higher integer. For this example, the oversample is proportionate to the category sample sizes, and the warning message is not printed.

For this survey design, an sp package object will be used as the sampling frame. The read.shape function will be used to read the shapefile and assign its output to an sp object named shp. The following arguments are included in the call to grts: (1) design: assigned the Unequaldsgn list;

(2) DesignID: assigned the value "UNEQUAL"; (3) type.frame: assigned the value "finite"; (4) src.frame: assigned the value "sp.object" to indicate that the sampling frame is provided by an sp object; (5) sp.object: name of the sp object, which is assigned the shp object; (6) att.frame: assigned the att data frame; (7) mdcaty: name of the column in the attributes data frame that identifies the unequal probability category for each element in the frame, which is assigned the value "area_cat"; (8) shapefile: assigned the value FALSE. Upon completion of the call to grts, the initial six sites for the survey design and a design summary are printed.

Create the design list

```
> Unequaldsgn <- list(None=list(panel=c(PanelOne=300),
+                                     seltype="Unequal",
+                                     caty.n=c("[0,1]"=50, "(1,5]"=50, "(5,10]"=50,
+                                               "(10,50]"=50, "(50,500]"=50,
+                                               "(500,7e+04]"=50),
+                                     over=120))
```

Read the shapefile

```
> shp <- read.shape("reg1_lakes")
```

Select the sample

```
> Unequalsites <- grts(design=Unequaldsgn,
+                       DesignID="UNEQUAL",
+                       type.frame="finite",
+                       src.frame="sp.object",
+                       sp.object=shp,
+                       att.frame=att,
+                       mdcaty="area_cat",
+                       shapefile=FALSE)
```

Stratum: None

Initial number of levels: 5

Current number of levels: 5

Current number of levels: 7

Current number of levels: 8

Final number of levels: 8

Print the initial six lines of the survey design

```
> head(Unequalsites@data)
```

	siteID	xcoord	ycoord	mdcaty	wgt	stratum	panel	EvalStatus
1	UNEQUAL-001	2118352	2376705	[0,1]	19.58	None	PanelOne	NotEval
2	UNEQUAL-002	1950763	2416446	(10,50]	53.46	None	PanelOne	NotEval
3	UNEQUAL-003	2067486	2816127	(1,5]	121.40	None	PanelOne	NotEval
4	UNEQUAL-004	2043705	2676363	(500,7e+04]	3.50	None	PanelOne	NotEval
5	UNEQUAL-005	1873327	2241300	[0,1]	19.58	None	PanelOne	NotEval
6	UNEQUAL-006	1925486	2336149	[0,1]	19.58	None	PanelOne	NotEval
	EvalReason	lat_dd	lon_dd	xcoord.1	ycoord.1	state		

1	41.77125	-70.01159	2118352	2376705	MA
2	42.50498	-71.85602	1950763	2416446	MA
3	45.68445	-69.11742	2067486	2816127	ME
4	44.53542	-69.89743	2043705	2676363	ME
5	41.15570	-73.28496	1873327	2241300	CT
6	41.86570	-72.39768	1925486	2336149	CT

Print the survey design summary

```
> dsgnsum(Unequalsites)
```

Design Summary: Number of Sites Classified by mdcaty (Multidensity Category) and panel

mdcaty	panel OverSamp	PanelOne	Sum
(1,5]	19	55	74
(10,50]	28	53	81
(5,10]	21	44	65
(50,500]	18	51	69
(500,7e+04]	15	50	65
[0,1]	19	47	66
Sum	120	300	420

6 Unstratified, unequal probability, GRTS survey design with an oversample and a panel structure for survey over time

The fourth survey design is an unstratified, unequal probability design with an oversample and a panel structure for survey over time. List `Paneldsgn` is assigned design specifications. Since the survey design is unstratified, `Paneldsgn` includes a single list named "None" that contains four items: `panel`, `seltype`, `caty.n`, and `over`. A vector identifying sample sizes for five panels is assigned to `panel`. The value "Unequal" is assigned `seltype`, which indicates unequal selection probabilities. The third item, `caty.n`, assigns sample sizes for each of six multidensity categories, where lake area classes are used as the categories. The value 100 is assigned to `over`, which specifies an oversample of 100 sites. For this example, the oversample is not proportionate to the category sample sizes, and the warning message is printed by calling the `warnings` function.

For this survey design, a shapefile will be used as the sampling frame. The following arguments are included in the call to `grts`: (1) `design`: assigned the `Paneldsgn` list; (2) `DesignID`: assigned the value "UNEQUAL"; (3) `type.frame`: assigned the value "finite"; (4) `src.frame`: assigned the value "shapefile"; (5) `in.shape`: assigned the value "reg1_lakes"; (6) `att.frame`: assigned the `att` data frame; (7) `mdcaty`: assigned the value "area_cat"; and (8) `shapefile`: assigned the value FALSE. Upon completion of the call to `grts`, the initial six sites for the survey design and a design summary are printed.

Create the design list

```
> Paneldsgn <- list(None=list(panel=c(Annual=50, Year1=50, Year2=50, Year3=50,
+                                     Year4=50, Year5=50),
+                               seltype="Unequal",
```

```
+
+             caty.n=c("[0,1]"=50, "(1,5]"=50, "(5,10]"=50,
+                   "(10,50]"=50, "(50,500]"=50,
+                   "(500,7e+04]"=50),
+             over=100))
```

Select the sample

```
> Panelsites <- grts(design=Paneldsgn,
+                   DesignID="UNEQUAL",
+                   type.frame="finite",
+                   src.frame="shapefile",
+                   in.shape="reg1_lakes",
+                   att.frame=att,
+                   mdcaty="area_cat",
+                   shapefile=FALSE)
```

Stratum: None

Initial number of levels: 5

Current number of levels: 5

Current number of levels: 7

Current number of levels: 8

Final number of levels: 8

Print the warning message

```
> warnings()
```

Warning message:

```
In grts(design = Paneldsgn, DesignID = "UNEQUAL", type.frame = "finite", :
```

Oversample size is not proportional to category sample sizes for stratum "None".

Print the initial six lines of the survey design

```
> head(Panelsites@data)
```

	siteID	xcoord	ycoord	mdcaty	wgt	stratum	panel	EvalStatus
1	UNEQUAL-001	2085642	2633214	[0,1]	19.58	None	Annual	NotEval
2	UNEQUAL-002	2043490	2785734	(500,7e+04]	3.50	None	Annual	NotEval
3	UNEQUAL-003	2021486	2990894	(500,7e+04]	3.50	None	Annual	NotEval
4	UNEQUAL-004	1987175	2419772	(10,50]	53.46	None	Annual	NotEval
5	UNEQUAL-005	2090281	2730607	[0,1]	19.58	None	Annual	NotEval
6	UNEQUAL-006	2093807	2874103	(5,10]	41.22	None	Annual	NotEval
	EvalReason	lat_dd	lon_dd	xcoord.1	ycoord.1	state		
1		44.06025	-69.53846	2085642	2633214	ME		
2		45.48147	-69.52065	2043490	2785734	ME		
3		47.31454	-69.05682	2021486	2990894	ME		
4		42.45116	-71.41461	1987175	2419772	MA		
5		44.88875	-69.14252	2090281	2730607	ME		
6		46.11868	-68.58204	2093807	2874103	ME		

Print the survey design summary

```
> dsgnsum(Panelsites)
```

Design Summary: Number of Sites Classified by mdcaty (Multidensity Category)
and panel

mdcaty	panel		Annual	OverSamp	Year1	Year2	Year3	Year4	Year5	Sum
(1,5]			10	15	7	12	11	5	6	66
(10,50]			8	19	6	9	7	14	12	75
(5,10]			6	17	10	8	6	5	4	56
(50,500]			9	19	6	4	6	10	7	61
(500,7e+04]			7	13	10	8	11	6	10	65
[0,1]			10	19	11	9	9	10	11	79
Sum			50	102	50	50	50	50	50	402