

GRTS Survey Designs for a Finite Resource

Thomas Kincaid

October 11, 2012

Contents

1 Preliminaries	1
2 Shapefile attribute data	2
3 Unstratified, equal probability, GRTS survey design	3
4 Stratified, equal probability, GRTS survey design	6
5 Unstratified, unequal probability, GRTS survey design with an oversample	9
6 Unstratified, unequal probability, GRTS survey design with an oversample and a panel structure for survey over time	11

1 Preliminaries

This document presents example GRTS survey designs for a finite resource. The finite resource used in the designs is lakes in the southern 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.

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

```
> # Load the spsurvey package
> library(spsurvey)
>
```

Version 2.5 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 lake area categories are coded using values such as "(5,10]", which indicates that lake area is greater than five hectares but less than or equal to ten 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

```
> # Read the attribute table from the shapefile
> att <- read.dbf("NE_lakes")
>
```

Display the initial six lines in the attribute data frame

```
> # Display the initial six lines in the attribute data frame
> head(att)
```

	lat_dd	lon_dd	xcoord	ycoord	state	area_cat
1	42.86416	-70.94392	2012313	2474271	MA	(10,50]
2	42.86107	-70.92477	2013905	2474343	MA	(1,5]

```

3 42.86070 -70.99013 2008789 2472920 MA (10,50]
4 42.85068 -70.98077 2009814 2472036 MA (50,500]
5 42.83724 -70.93224 2014014 2471614 MA (5,10]
6 42.80890 -70.93007 2015009 2468603 MA (10,50]

```

```
>
```

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

```

> # Display number of lakes cross-classified by strata and multidensity
> # category
> addmargins(table("State"=att$state, "Lake Area Category"=att$area_cat))

```

	Lake Area Category						
State	(0,1]	(1,5]	(10,50]	(5,10]	(50,500]	(500,7e+04]	Sum
CT	483	1181	284	270	90	4	2312
MA	194	1658	693	545	209	6	3305
RI	11	256	108	85	41	3	504
Sum	688	3095	1085	900	340	13	6121

```
>
```

Lakes in the southern 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")

```

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.

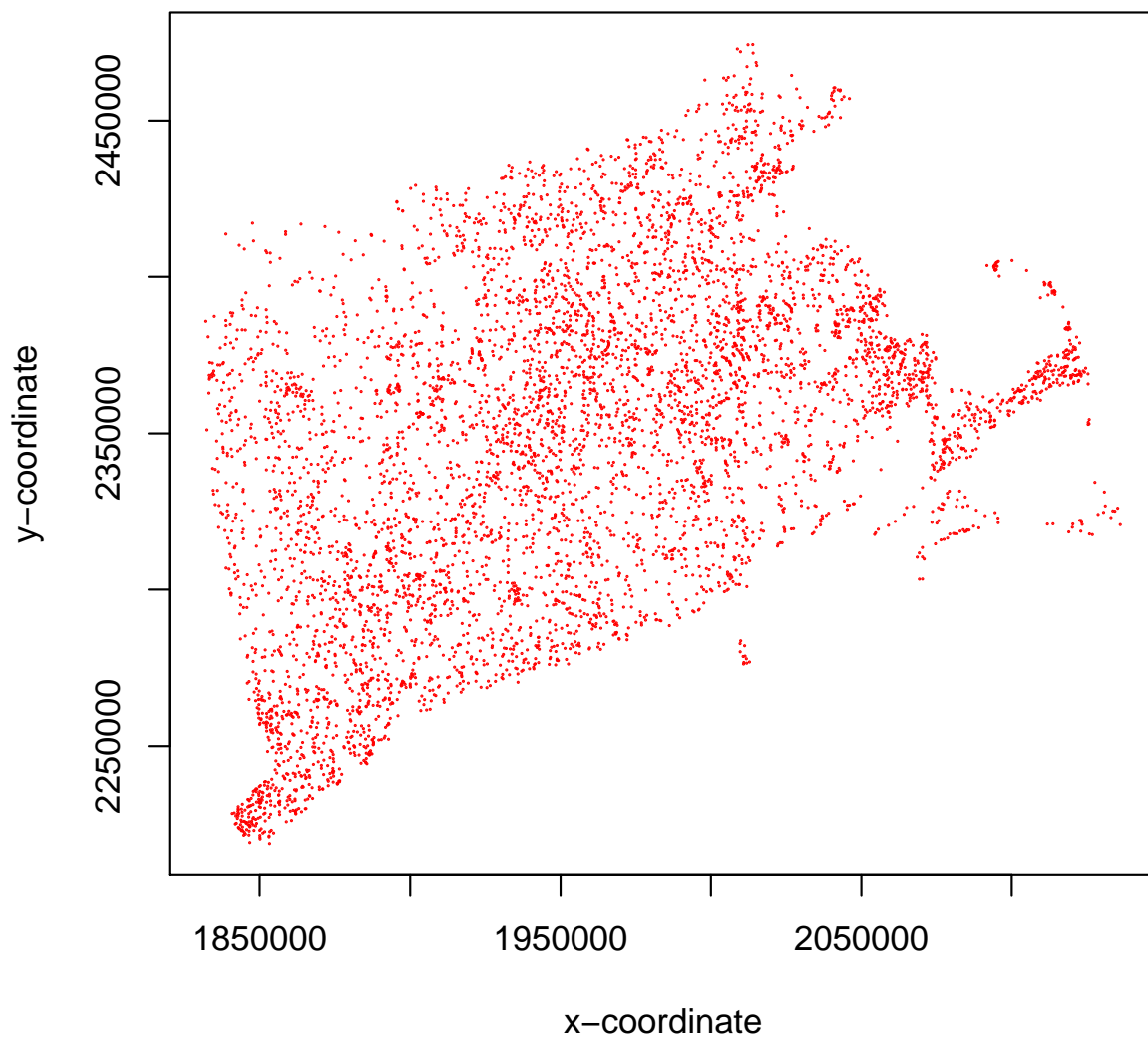


Figure 1: Lakes in the Southern New England Region.

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

```
> # Call the set.seed function so that the survey designs can be replicate
> set.seed(4447864)
>
```

Create the design list

```
> # 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="NE_lakes",
+                   att.frame=att,
+                   shapefile=FALSE)
```

Stratum: None

Initial number of levels: 5

Current number of levels: 5

Current number of levels: 6

Current number of levels: 7

Final number of levels: 7

Print the initial six lines of the survey design

```
> # Print the initial six lines of the survey design
> head(Equalsites@data)
```

	siteID	xcoord	ycoord	mdcaty	wgt	stratum	panel	EvalStatus
1	EQUAL-001	2114167	2395530	Equal	20.40333	None	PanelOne	NotEval
2	EQUAL-002	1980508	2343135	Equal	20.40333	None	PanelOne	NotEval
3	EQUAL-003	1959920	2408654	Equal	20.40333	None	PanelOne	NotEval
4	EQUAL-004	1907771	2302522	Equal	20.40333	None	PanelOne	NotEval
5	EQUAL-005	2083109	2348787	Equal	20.40333	None	PanelOne	NotEval
6	EQUAL-006	1985761	2418194	Equal	20.40333	None	PanelOne	NotEval

	EvalReason	lat_dd	lon_dd	xcoord.1	ycoord.1	state	area_cat
1		41.94336	-69.99879	2114167	2395530	MA	(1,5]
2		41.80361	-71.73038	1980508	2343135	RI	(1,5]
3		42.41690	-71.77158	1959920	2408654	MA	(10,50]
4		41.61294	-72.70519	1907771	2302522	CT	(1,5]
5		41.61511	-70.51295	2083109	2348787	MA	(1,5]
6		42.44075	-71.43626	1985761	2418194	MA	(5,10]

```
>
```

Print the survey design summary

```
> # Print the survey design summary
> dsngsum(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

```
> Stratdsgn <- list(CT=list(panel=c(PanelOne=125), seltype="Equal"),
+                   MA=list(panel=c(PanelOne=125), seltype="Equal"),
+                   RI=list(panel=c(PanelOne=50), seltype="Equal"))
```

Select the sample

```
> Stratsites <- grts(design=Stratdsgn,
+                   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: 4

Current number of levels: 5

Current number of levels: 6

Final number of levels: 6

Stratum: MA

Current number of levels: 4

Current number of levels: 6

Final number of levels: 6

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

Print the initial six lines of the survey design

```
> # Print the initial six lines of the survey design
> head(Stratsites@data)
```

	siteID	xcoord	ycoord	mdcaty	wgt	stratum	panel	EvalStatus
1	STRATIFIED-001	1956475	2298201	Equal	18.496	CT	PanelOne	NotEval
2	STRATIFIED-002	1941900	2358465	Equal	18.496	CT	PanelOne	NotEval
3	STRATIFIED-003	1922626	2337816	Equal	18.496	CT	PanelOne	NotEval
4	STRATIFIED-004	1880092	2258457	Equal	18.496	CT	PanelOne	NotEval
5	STRATIFIED-005	1935045	2301159	Equal	18.496	CT	PanelOne	NotEval
6	STRATIFIED-006	1955161	2348226	Equal	18.496	CT	PanelOne	NotEval

	EvalReason	lat_dd	lon_dd	xcoord.1	ycoord.1	area_cat
1		41.46867	-72.14784	1956475	2298201	(0,1]
2		42.02273	-72.13751	1941900	2358465	(5,10]
3		41.88642	-72.42635	1922626	2337816	(50,500]
4		41.29023	-73.15698	1880092	2258457	(1,5]
5		41.54160	-72.38965	1935045	2301159	(5,10]
6		41.90456	-72.01248	1955161	2348226	(10,50]

```
>
```

Print the survey design summary

```
> # Print the survey design summary
> dsgnsum(Stratsites)
```

Design Summary: Number of Sites

```
stratum
  CT  MA  RI Sum
125 125  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]"=120, "(5,10]"=50,
+                                "(10,50]"=50, "(50,500]"=25,
+                                "(500,7e+04]"=5),
+                                over=120))
```

Read the shapefile

```
> # Read the shapefile
> shp <- read.shape("NE_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: 6

Current number of levels: 7

Current number of levels: 8

Current number of levels: 9

Final number of levels: 9

Print the initial six lines of the survey design

```
> # Print the initial six lines of the survey design
> head(Unequalsites@data)
```

	siteID	xcoord	ycoord	mdcaty	wgt	stratum	panel	EvalStatus
1	UNEQUAL-001	1978339	2290209	(1,5]	25.79167	None	PanelOne	NotEval
2	UNEQUAL-002	2004648	2332660	(1,5]	25.79167	None	PanelOne	NotEval
3	UNEQUAL-003	1918520	2396464	(500,7e+04]	2.60000	None	PanelOne	NotEval
4	UNEQUAL-004	1885626	2325033	(0,1]	13.76000	None	PanelOne	NotEval
5	UNEQUAL-005	2000332	2417941	(50,500]	13.60000	None	PanelOne	NotEval
6	UNEQUAL-006	2042863	2346316	(1,5]	25.79167	None	PanelOne	NotEval
	EvalReason	lat_dd	lon_dd	xcoord.1	ycoord.1	state		
1		41.35079	-71.91667	1978339	2290209	CT		
2		41.65830	-71.47985	2004648	2332660	RI		
3		42.40370	-72.29898	1918520	2396464	MA		
4		41.85599	-72.89953	1885626	2325033	CT		
5		42.40514	-71.26482	2000332	2417941	MA		
6		41.68822	-70.99051	2042863	2346316	MA		

>

Print the survey design summary

```
> # Print the survey design summary
> dsgnsum(Unequalsites)
```

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

	panel		
mdcaty	OverSamp	PanelOne	Sum
(0,1]	20	49	69
(1,5]	43	124	167
(10,50]	20	44	64
(5,10]	24	52	76
(50,500]	13	22	35
(500,7e+04]	0	9	9
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 `Paneldsign` is assigned design specifications. Since the survey design is unstratified, `Paneldsign` 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 `Paneldsign` 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]"=120, "(5,10]"=50,
+                                               "(10,50]"=50, "(50,500]"=25,
+                                               "(500,7e+04]"=5),
+                                     over=100))
```

Select the sample

```
> Panelsites <- grts(design=Paneldsgn,
+                    DesignID="UNEQUAL",
+                    type.frame="finite",
+                    src.frame="shapefile",
+                    in.shape="NE_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

```
> # 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

```
> # Print the initial six lines of the survey design
> head(Panelsites@data)
```

	siteID	xcoord	ycoord	mdcaty	wgt	stratum	panel	EvalStatus
1	UNEQUAL-001	1892325	2299346	(0,1]	13.76	None	Annual	NotEval
2	UNEQUAL-002	1850801	2376753	(10,50]	21.70	None	Annual	NotEval
3	UNEQUAL-003	1994373	2330947	(5,10]	18.00	None	Annual	NotEval
4	UNEQUAL-004	1985881	2411784	(5,10]	18.00	None	Annual	NotEval
5	UNEQUAL-005	1967384	2364767	(5,10]	18.00	None	Annual	NotEval
6	UNEQUAL-006	1949323	2389161	(10,50]	21.70	None	Annual	NotEval

	EvalReason	lat_dd	lon_dd	xcoord.1	ycoord.1	state
1		41.61877	-72.89566	1892325	2299346	CT
2		42.37949	-73.16050	1850801	2376753	MA
3		41.66686	-71.60533	1994373	2330947	RI
4		42.38504	-71.45487	1985881	2411784	MA
5		42.02031	-71.81833	1967384	2364767	CT
6		42.27197	-71.95670	1949323	2389161	MA

>

Print the survey design summary

```
> # Print the survey design summary
> dsgnsum(Panelsites)
```

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

	panel							
mdcaty	Annual	OverSamp	Year1	Year2	Year3	Year4	Year5	Sum
(0,1]	10	12	10	8	6	10	8	64
(1,5]	13	41	17	25	27	23	17	163
(10,50]	11	19	5	7	4	6	11	63
(5,10]	11	19	11	7	9	8	8	73
(50,500]	5	9	4	3	4	2	4	31
(500,7e+04]	0	2	3	0	0	1	2	8
Sum	50	102	50	50	50	50	50	402

>