

# **Proc Surveyselect or the easy way to select samples**

Gitte Churlish

Churlish Consulting

---

# Sample selection (SAS 6)

\* to select 15 in a random fashion;

```
Data temp; Set trees;  
  select = ranuni (5432); run;
```

```
Proc sort data = temp; by select;
```

```
Data sample;  
  set temp (obs = 15);  
  run;
```

# Proc Surveyselect

```
Proc surveyselect data = trees
```

```
Method = SRS n = 15 out = sample1;
```

```
run;
```

- Selects 15 samples from the population.
- Sample1 – contains 15 records with sampling weights

# Methods

- Simple random sample
- Unrestricted random sample
- Systematic random sample
- Sequential random sample
- PPS Sampling (7 methods)

## Use of strata

```
Proc surveyselect data = trees
```

```
Method = SRS n = 15 out = sample1;
```

```
strata lv_d;
```

```
run;
```

- - select 15 trees from live, and 15 trees from the dead strata

# Sample size

```
Proc surveyselect data = trees  
Method = SRS out = sample1;  
strata lv_d;  
Sampsize = 5,15;  
run;
```

- select 5 trees from the dead strata and 15 trees from live
- File must be sorted by strata variables

# Use of file for sample size

```
proc surveysselect data=samp1  
  (where = (lv_d = 'L'))  
  out=select_live  
  method=sys  
  seed = 5763  
  sampsize = no_to_select_live;  
  strata age_grp spp_grp;
```

Note use of SEED – provides repeatability

# Creation of sampsize file

```
data no_to_select_live (keep = spp_grp age_grp _nsize_);
set spp_grp_count_live;
if age_grp = 'Mat' then do ;
    if spp_grp = 'BCHD-dec' then _nsize_ = 1;
    if spp_grp = 'BCHD-BL' then _nsize_ = 10;
    if spp_grp = 'BCHD-C' then _nsize_ = 1;
    if spp_grp = 'BCHD-HW' then _nsize_ = 3;
    if spp_grp = 'FPSL-FD-P' then _nsize_ = 11;
if spp_grp = 'FPSL-LW' then _nsize_ = 3;
if spp_grp = 'FPSL-S' then _nsize_ = 5;
end ;
    if _nsize_ = 0 or _nsize_ = .
        then _nsize_ = 1; /* can't have a zero */
if _nsize_ > count then _nsize_ = count ; run;
```

# PPSWR sampling

- Great for sampling in GIS files
- Method = pps\_wr
- probability proportional to size and with replacement.

# PPS\_WR example

```
proc surveystest data = GIS_polygon_file  
  out = &file_out method = pps_wr  
  outhits seed = &seed  
  sampsize = samp_rate (rename = (max =_nsize_));  
  size polyarea;  
  strata spec_cl vol_cls;          run;
```

NOTE: Size has been added, This indicates the variable to use for proportionality.

## Related Procs

- Proc Surveyfreq
  - Proc surveymeans
  - Proc surveyreg
  - Proc surveylogistic
- 
- Take into account sampling weights