Stratified Sampling

By David R. Larsen

Stratification for sampled areas is on of the most useful techniques in resource sampling. Stratification should be applied when a difference in the statistic of interest is expected to vary by the sampled strata. If this is true stratification can:

- Increase the precision of the sample estimates
- Or decrease the number of samples needed to equal unstratified estimates.

For example consider the following area:

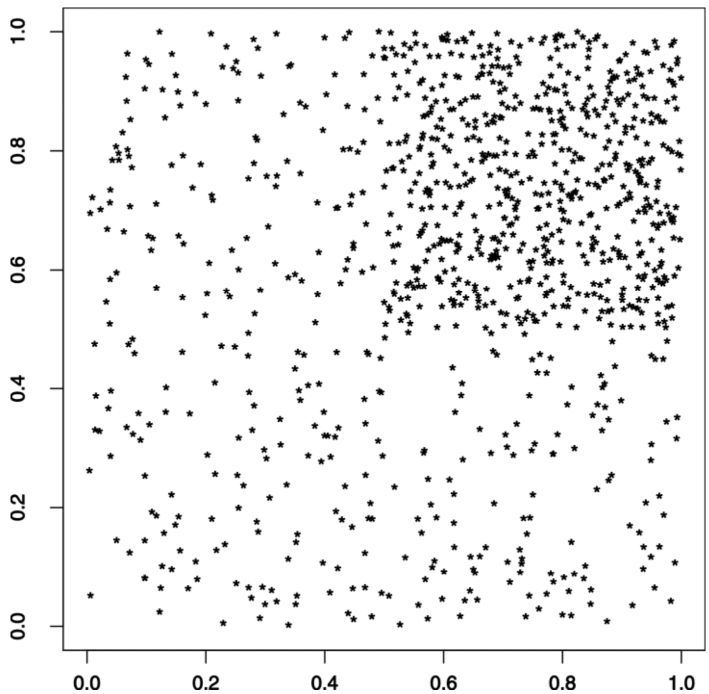

Figure 1. Population with potential strata

If this population were sampled for density without considering the differences in the regions of the population the estimate of variance will be higher. If this area were stratified and then sampled by strata the overall estimate will have lower variance or few samples may be required.

Consider the following histogram of a population that was generated from two normal populations (Population 1, $\mu=4$, $\sigma=2$; and Population 2 $\mu=10$, $\sigma=1$). The sample means are plotted as solid lines, with the overall mean thicker than the two strata means. The sample standard deviations (s) are plotted as dashed line with the overall line thicker that the strata lines. This graph illustrates the increase in sample standard deviation because of pooling the two strata versus calculating the strata individually.

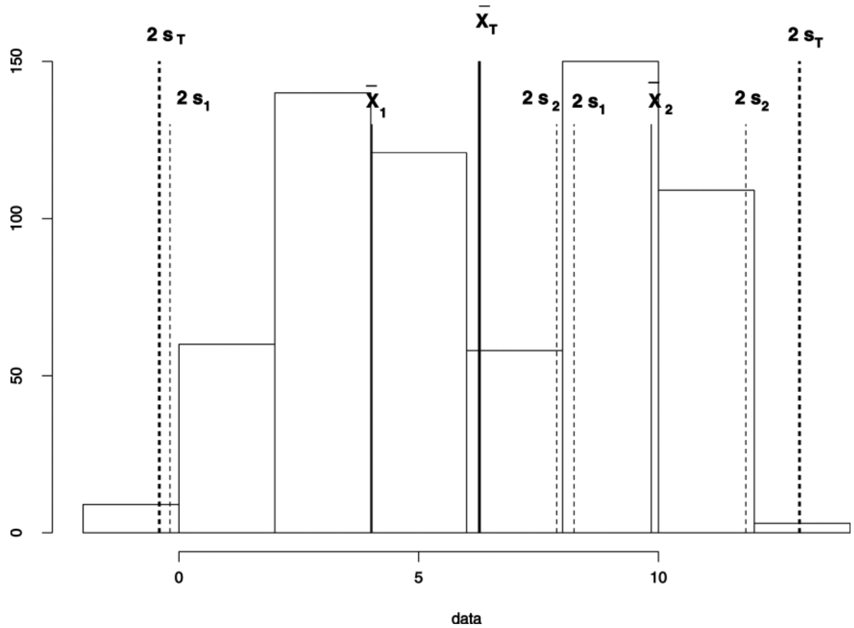

Figure 2. Example of pooling two strata

Table 1. Simple statistics for the two strata

Statistic	n	Mean	Standard Deviation
Overall ignoring strata	650	4.10	3.397
Strata 1	400	3.99	2.023
Strata 2	250	10.03	1.009
Overall by strata		6.31	1.700

Estimate of the mean using stratum means:

$$\bar{x} = \sum_{j=1}^{M} \frac{N_j}{N} \bar{x}_j$$

$$\bar{x} = \left(\frac{400}{650}\right)3.99 + \left(\frac{250}{650}\right)10.03$$

$$\bar{x} = 6.31$$

Estimate of the total for ***X*** for the entire population:

$$\hat{X} = \left(\sum_{j=1}^{M} N_j \bar{x}_j \right)$$

$$\hat{X} = (400 \cdot 3.99) + (250 \cdot 10.03)$$

$$\hat{X} = 4{,}103.5$$

Standard Error of the mean:

$$s_{t\bar{x}}^2 = \sum_{j=1}^{M} \left(\frac{N_j}{N} \right)^2 \frac{s_j^2}{n_j}$$

$$s_{t\bar{x}}^2 = \left(\left(\frac{400}{650} \right)^2 \frac{2.023^2}{400} \right) + \left(\left(\frac{250}{650} \right)^2 \frac{1.009^2}{250} \right)$$

$$s_{\bar{x}}^2 = 0.004477$$

The standard error of the total estimate:

$$s_X^2 = N^2 s_X^2$$
$$s_X^2 = 650^2 \cdot 0.004477$$
$$s_X^2 = 1{,}891.53$$

Given these definitions:

M = number of strata in the population
n = total number of sampling units measured for all strata
n_j = total number of sampling units measured in the *j*th stratum
N = total number of sampling units in the population
N_j = total number of sampling units in the *j*th stratum
X_{ij} = quantity ***X*** measured on the *i*th sampling unit of the *j*th stratum
x = mean of ***X*** for the *j*th stratum
x_j= estimated mean of ***X*** for the population
P_j = proportion of the total area in the *j*th stratum
X= estimated total of ***X*** for the population
s^2_x= variance of ***X*** for the *j*th stratum
s^2_j= estimated variance for the mean for the population
s^2_x= estimated variance of

Also See:

Chapter 13 - Sampling Designs in Forest Inventories in:

Husch, B., T. W. Beers, and J. A. Kershaw, Jr. 2003. Forest Mensuration. John Wiley and Sons, Inc. Hoboken, New Jersey. 443 pp. ISBN 0-471-01850-3

Chapter 5 – Stratified Random Sampling in:

Shiver, B. D. and B. E. Borders. 1996. Sampling Techniques for Forest Resources Inventory. John Wiley and Sons, New York. 356 pp. ISBN 0-471-10940-1

Chapter 6 - Sample Designs - Random Sampling pages 200-236, in:

Krebs, C. J. 1989. Ecological Methodology. Harper and Row, Publishers. New York. 654 pp.