Monday, Oct 14

Stratified Cluster Sampling Designs

Another complex sampling design is to *combine* stratified random sampling with cluster sampling to exploit the benefits of both designs.

1. Partition M elements into N clusters.
2. Assign the N clusters into L strata.
3. Use a sampling design (e.g., SRS, PPS) to select some clusters from each strata.

Examples of stratified cluster sampling:

Element	Cluster	Stratum
student	classroom	grade level
tree	plot	elevation
household	block	city/suburb
store	city	region
minute	hour	day/night

Systematic Sampling

Systematic sampling is cluster sampling where the elements are assigned to clusters so that they are "systematically spread out" as opposed to close/adjacent (usually in space or time).

Ordered Systematic Sampling

Suppose we have a population of $M = 12$ elements that can be *ordered* in, say, space or time. How might we put these elements into $N = 4$ clusters of $m_i = 3$ elements each?

It might be convenient to cluster *adjacent* elements.

	Elements											
Sampling Units	$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$	$\mathcal{E}_4$	$\mathcal{E}_5$	$\mathcal{E}_6$	$\mathcal{E}_7$	$\mathcal{E}_8$	$\mathcal{E}_9$	$\mathcal{E}_{10}$	$\mathcal{E}_{11}$	$\mathcal{E}_{12}$
$\mathcal{U}_1$	✓	✓	✓									
$\mathcal{U}_2$				✓	✓	✓						
$\mathcal{U}_3$							✓	✓	✓			
$\mathcal{U}_4$										✓	✓	✓

But alternatively we could "spread out" the elements within each cluster.

This is an example of a 1-in-4 (ordered) systematic sampling design.

Example: A 1-in-5 systematic sample with $M = 15$ elements. The population is

$$\mathcal{P} = \{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4, \mathcal{E}_5 \mid \mathcal{E}_6, \mathcal{E}_7, \mathcal{E}_8, \mathcal{E}_9, \mathcal{E}_{10} \mid \mathcal{E}_{11}, \mathcal{E}_{12}, \mathcal{E}_{13}, \mathcal{E}_{14}, \mathcal{E}_{15}\},$$

	Elements											
Sampling Units	$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$	$\mathcal{E}_4$	$\mathcal{E}_5$	$\mathcal{E}_6$	$\mathcal{E}_7$	$\mathcal{E}_8$	$\mathcal{E}_9$	$\mathcal{E}_{10}$	$\mathcal{E}_{11}$	$\mathcal{E}_{12}$
$\mathcal{U}_1$	✓				✓				✓			
$\mathcal{U}_2$		✓				✓				✓		
$\mathcal{U}_3$			✓				✓				✓	
$\mathcal{U}_4$				✓				✓				✓

and the sampling units are

$$
\begin{aligned}
\mathcal{U}_1 &= \{\mathcal{E}_1, \mathcal{E}_6, \mathcal{E}_{11}\}, \\
\mathcal{U}_2 &= \{\mathcal{E}_2, \mathcal{E}_7, \mathcal{E}_{12}\}, \\
\mathcal{U}_3 &= \{\mathcal{E}_3, \mathcal{E}_8, \mathcal{E}_{13}\}, \\
\mathcal{U}_4 &= \{\mathcal{E}_4, \mathcal{E}_9, \mathcal{E}_{14}\}, \\
\mathcal{U}_5 &= \{\mathcal{E}_5, \mathcal{E}_{10}, \mathcal{E}_{15}\}.
\end{aligned}
$$

Example: A 1-in-3 systematic sample with $M = 15$ elements. The population is

$$\mathcal{P} = \{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3 \mid \mathcal{E}_4, \mathcal{E}_5, \mathcal{E}_6 \mid \mathcal{E}_7, \mathcal{E}_8, \mathcal{E}_9 \mid \mathcal{E}_{10}, \mathcal{E}_{11}, \mathcal{E}_{12} \mid \mathcal{E}_{13}, \mathcal{E}_{14}, \mathcal{E}_{15}\},$$

and the sampling units are

$$
\begin{aligned}
\mathcal{U}_1 &= \{\mathcal{E}_1, \mathcal{E}_4, \mathcal{E}_7, \mathcal{E}_{10}, \mathcal{E}_{13}\}, \\
\mathcal{U}_2 &= \{\mathcal{E}_2, \mathcal{E}_5, \mathcal{E}_8, \mathcal{E}_{11}, \mathcal{E}_{14}\}, \\
\mathcal{U}_3 &= \{\mathcal{E}_3, \mathcal{E}_6, \mathcal{E}_9, \mathcal{E}_{12}, \mathcal{E}_{15}\}.
\end{aligned}
$$

Example: A 1-in-3 systematic sample with $M = 14$ elements. The population is

$$\mathcal{P} = \{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3 \mid \mathcal{E}_4, \mathcal{E}_5, \mathcal{E}_6 \mid \mathcal{E}_7, \mathcal{E}_8, \mathcal{E}_9 \mid \mathcal{E}_{10}, \mathcal{E}_{11}, \mathcal{E}_{12} \mid \mathcal{E}_{13}, \mathcal{E}_{14}\},$$

and the sampling units are

$$
\begin{aligned}
\mathcal{U}_1 &= \{\mathcal{E}_1, \mathcal{E}_4, \mathcal{E}_7, \mathcal{E}_{10}, \mathcal{E}_{13}\}, \\
\mathcal{U}_2 &= \{\mathcal{E}_2, \mathcal{E}_5, \mathcal{E}_8, \mathcal{E}_{11}, \mathcal{E}_{14}\}, \\
\mathcal{U}_3 &= \{\mathcal{E}_3, \mathcal{E}_6, \mathcal{E}_9, \mathcal{E}_{12}\}.
\end{aligned}
$$

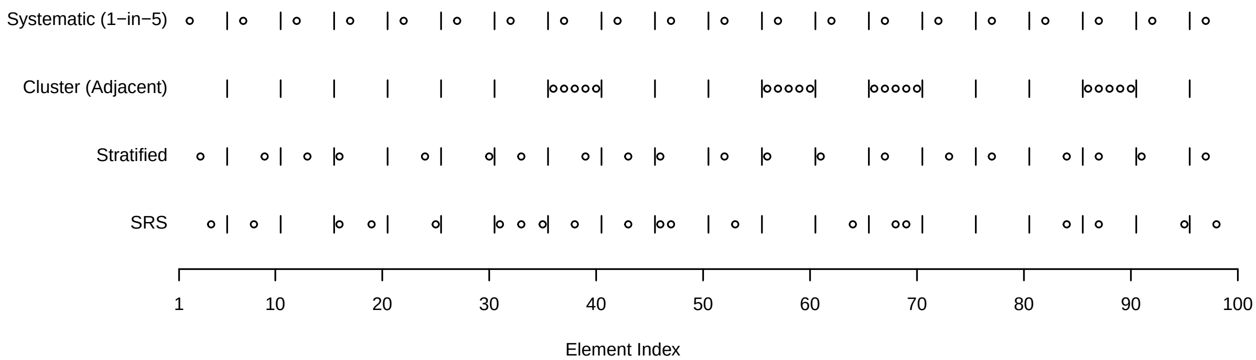
How do we a sampling unit with ordered systematic sampling?

1. Randomly select a number between 1 and k. Call this number j.
2. Select elements $\mathcal{E}_j, \mathcal{E}_{j+k}, \mathcal{E}_{j+2k}, \mathcal{E}_{j+3k}, \ldots$.

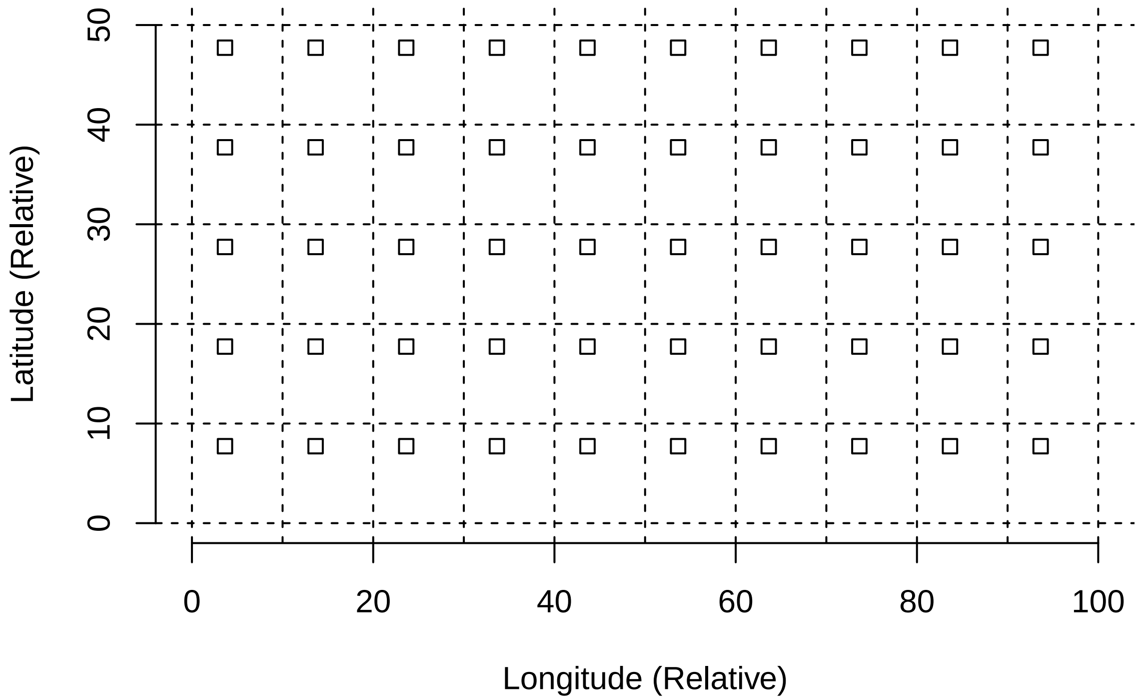
Why use systematic sampling?

1. Simple protocol (e.g., “observe every 10th customer/fish/minute”).
2. As in other cluster sampling designs, useful when there is no sampling frame.
3. Samples can be more representative, leading to lower variance than cluster sampling *with adjacent elements* and simple random sampling.

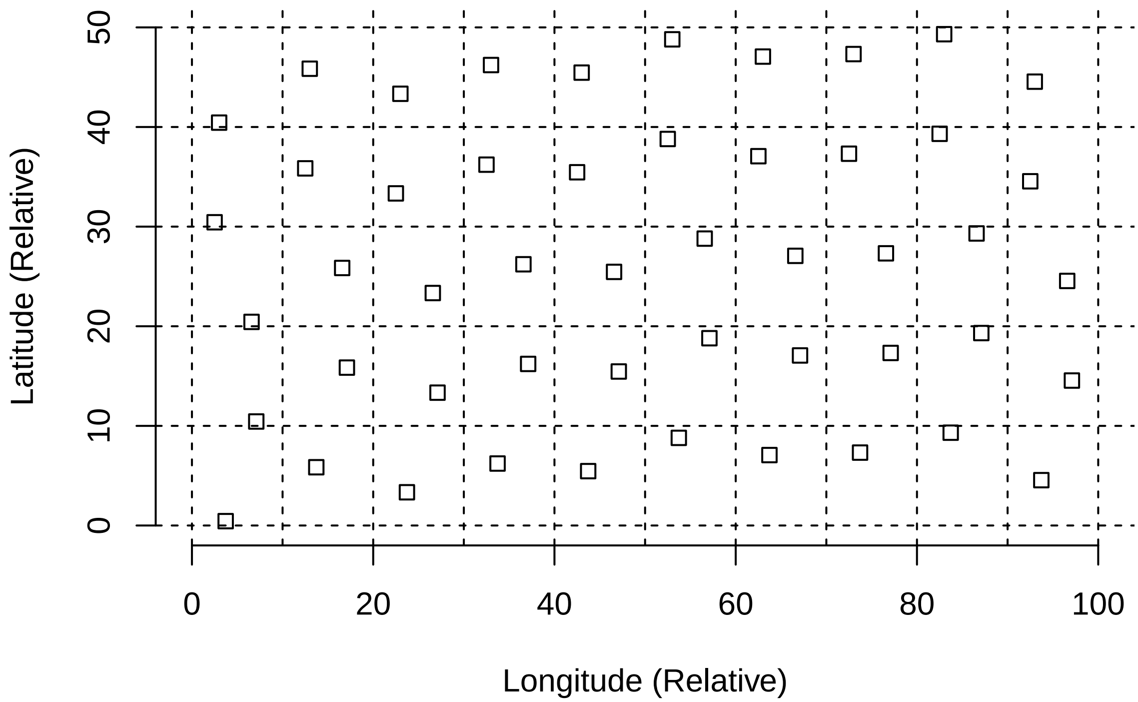
Example: Consider four designs for selecting 20 elements out of 100.

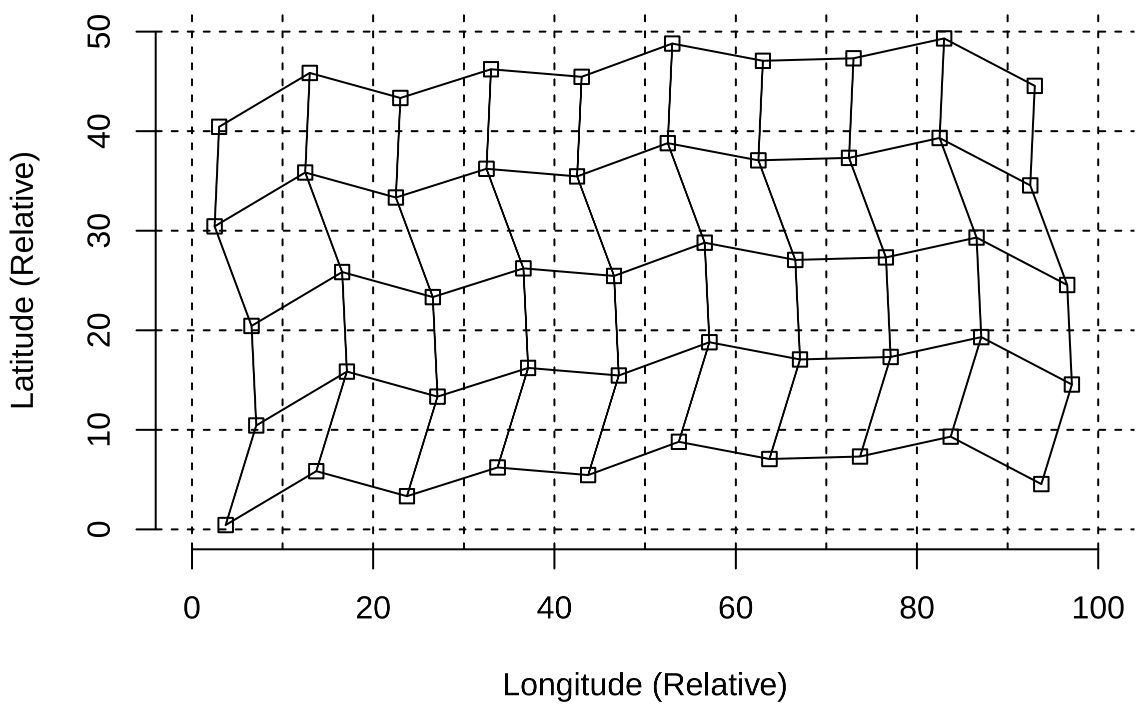


Aligned Grid Systematic Sampling

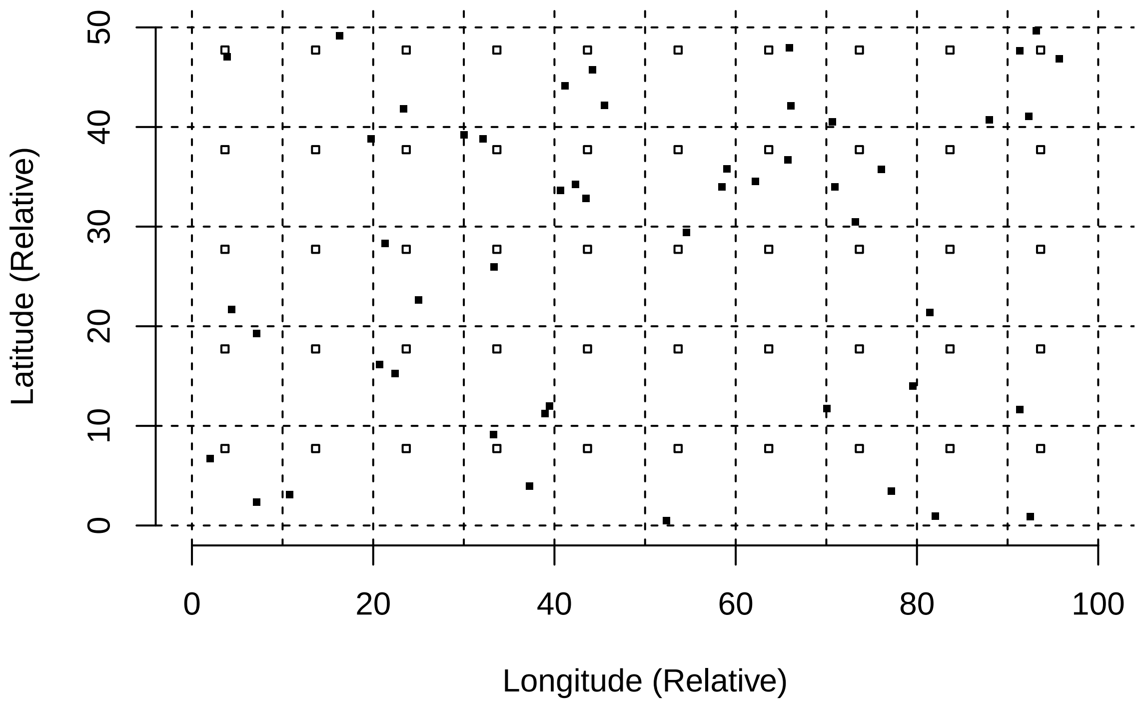


Unaligned Grid Systematic Sample

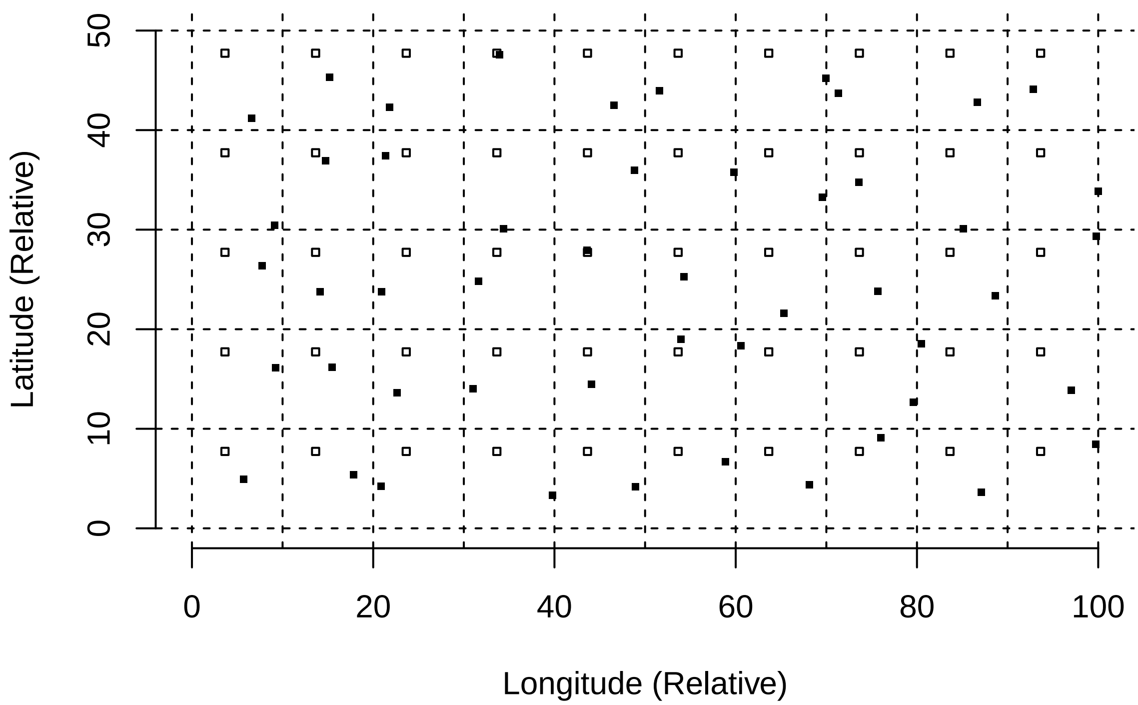




Grid Systematic Sampling Versus Simple Random Sampling



Grid Systematic Sampling Versus Stratified Random Sampling



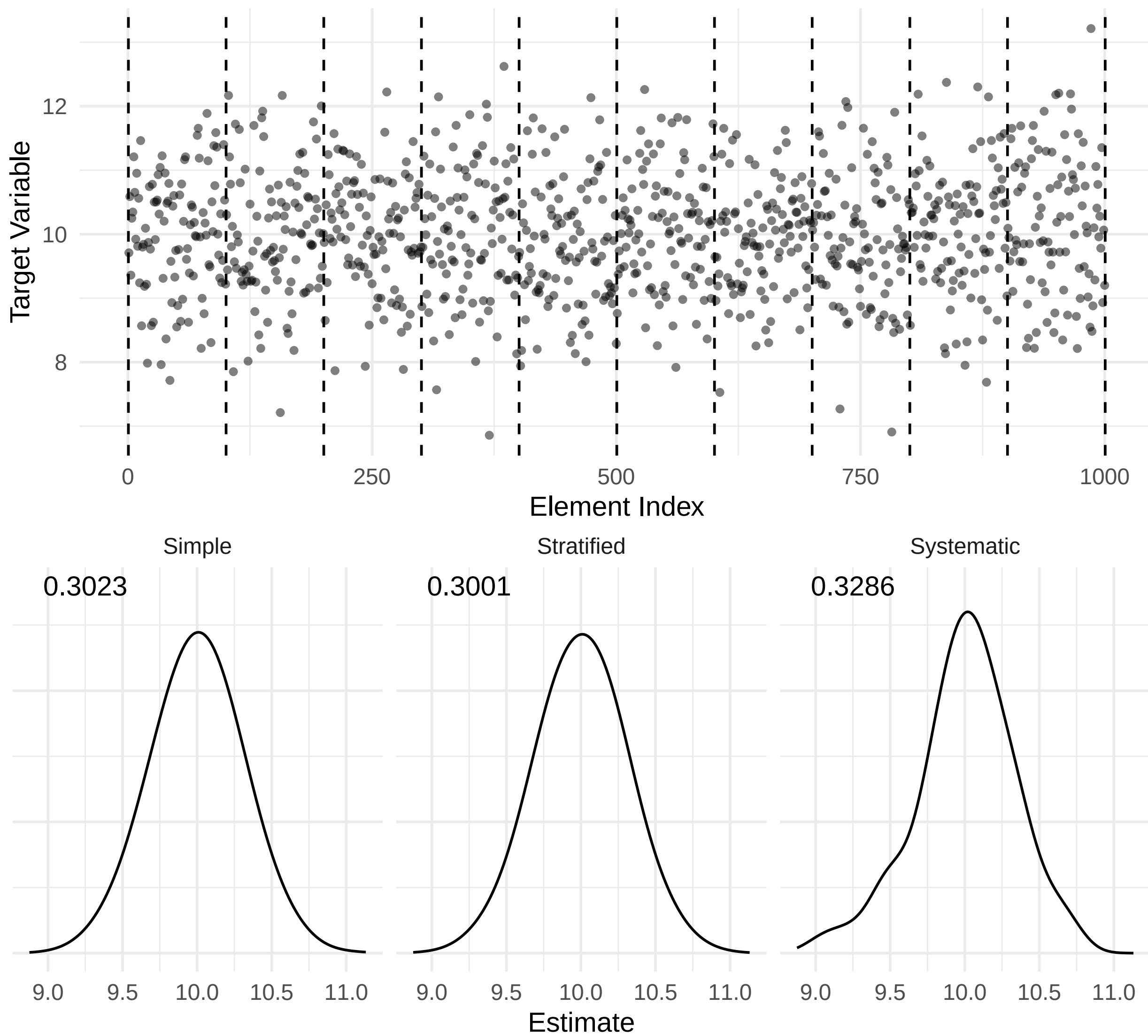
Performance of Systematic Sampling

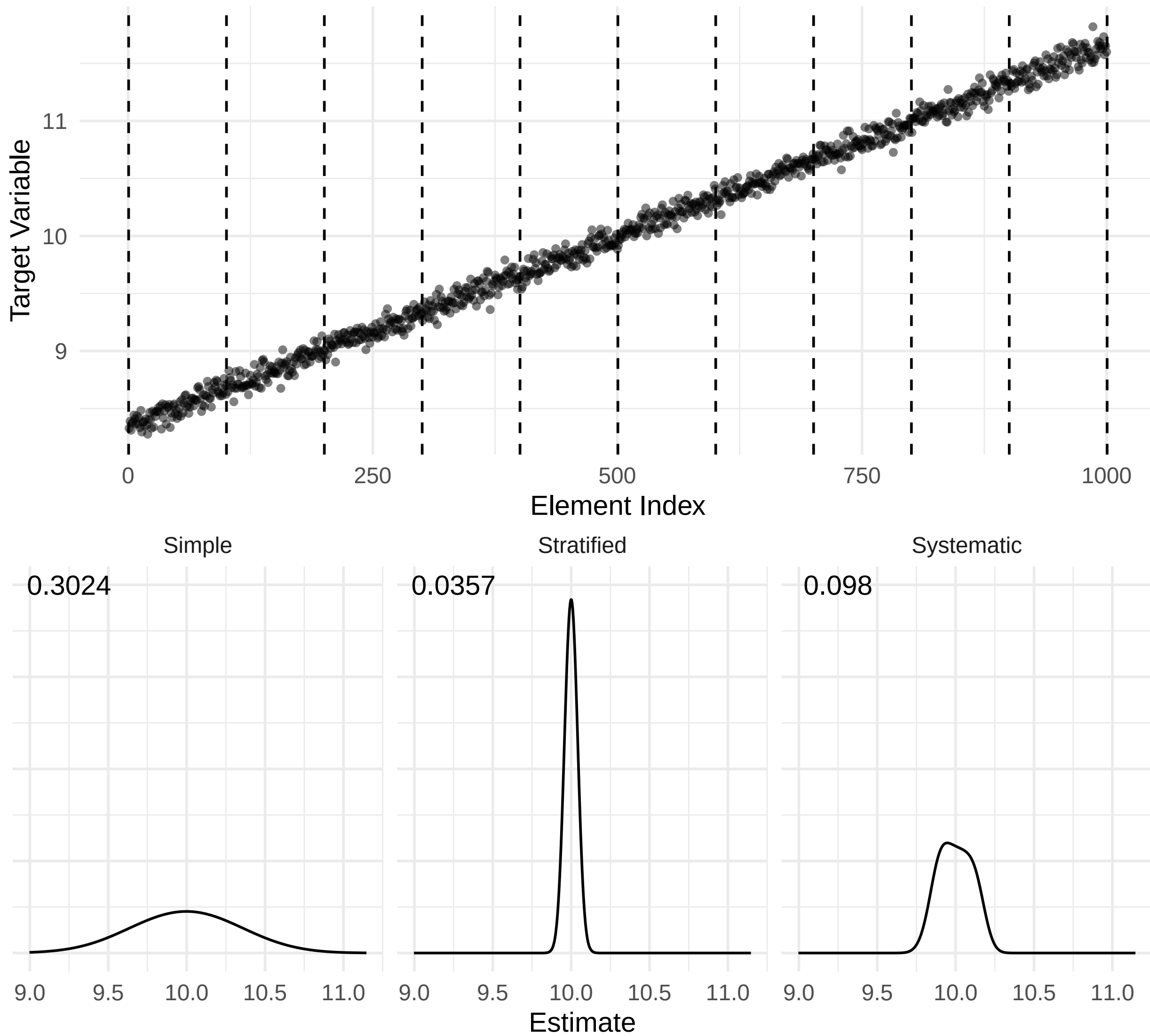
Example: Consider the following populations of 1000 ordered elements and three different designs.

1. Simple random sampling.

2. Stratified random sampling by treating groups of 100 adjacent elements as strata.
3. 1-in-100 ordered systematic sampling.

In each case 10 elements were selected and averaged to estimate μ.





Note: The numbers shown in each plot are the standard errors (i.e., the square root of the variance of the estimator).