Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences
Institutionen för skoglig resurshushållning

The benefits of balanced sampling versus mapping efforts

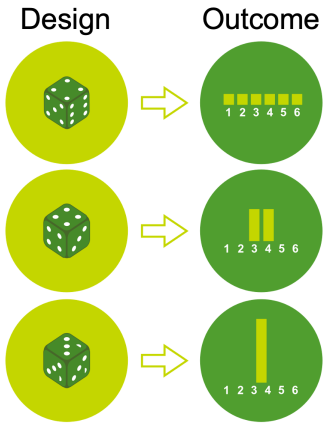
Anton Grafström
Swedish University of Agricultural Sciences

6th Natura 2000 Monitoring Workshop, Umeå, 2025-09-30

What is design-based statistics?

- The population is considered fixed (during a data collection).
- We use a known design to select a probability sample.
- We collect the data in the field.
- We use the sample data to make inference about the population.
- It is the ONLY approach that can provide unbiased estimates of population quantities, such as means and totals. Including change over time.

Design-based statistics



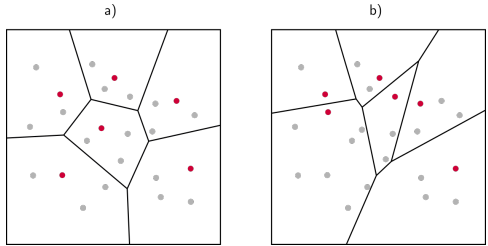
Tobler’s first law of geography

“Everything is related to everything else, but near things are more related than distant things.” - Waldo Tobler.

What is spatially balanced sampling?

- It builds on a generalization of Tobler’s first law of geography.
- The generalization is in distance. Similarity in auxiliary variables imply similarity in target values.
- Spatial balance means that the sample is well spread in some auxiliary variables.

What is spatially balanced sampling?



- More precisely, spatial balance means that the empirical multivariate distribution of auxiliary variables in the sample is close to the sampling distribution.
- If the auxiliary variables have explanatory power for the target variable(s), then we also come close to the sampling distribution for the target variable(s).

What does it achieve?

If every sample captures the sampling distribution well, we minimize the between sample variance which is the cause of sampling variance.

Some methods to select spatially balanced samples

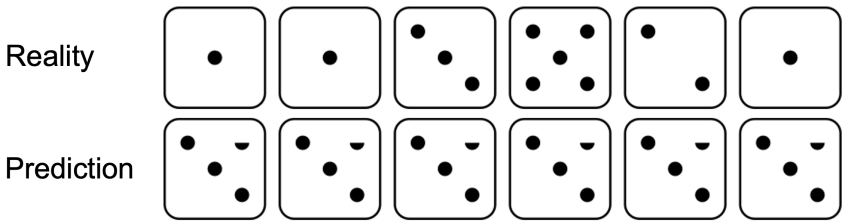
- The local pivotal method
- Spatially correlated Poisson sampling
- Locally correlated Poisson sampling
- The local cube method (double balance, spatial and marginal)

All are available in the `BalancedSampling` R-package.

What is model-based statistics?

Assumption: Reality is generated from the model. The current reality is an outcome from the model.

Model-based statistics



What do we estimate?

Model-based Statistics

- Does not care about reality.
- Cannot estimate without systematic errors (built into the method).
- Works exceptionally well when every realization lies close to the prediction (when the errors are ALWAYS small relative to the prediction).

Thus, it is not enough for the model to be correct. It must also provide a very precise picture of reality.

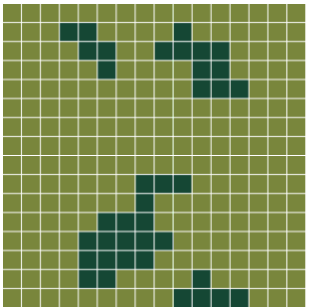
The strength of a map: Answers "Where?"

Maps are needed for:

- **Planning:** Where should we place actions?
- **Prioritizing:** Where do we have hotspots?
- **Visualizations:** Communicate spatial patterns and distributions.

They provide a qualitative picture, but do in general not answer “How much?".

Maps - Do NOT sum pixels (pixel-values)!



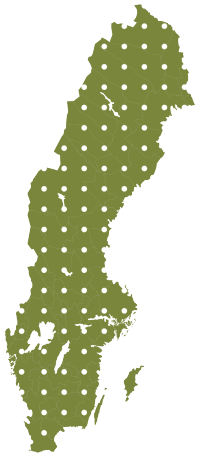
If you need unbiased estimates or scientifically solid evidence for your claims, dot NOT sum pixels as your primary method.

The Strength of Design-based: Answers "How much?"

Design-based statistics are best for quantitative answers:

- **Reporting:** Establishing status in relation to environmental goals (e.g., means, totals).
- **Trend analysis:** Measuring changes over time.
- **Compliance:** Producing legally and scientifically robust evidence.

The method provides a reliable figure with an honestly calculated uncertainty.



Key Takeaways!

- Correct statistical analyses require appropriate data collection. There are no shortcuts.
- A model (map) cannot replace field data collection – on the contrary, it requires data in order to be validated locally.
- Design-based and model-assisted methods are robust choices. They use collected data in a transparent and statistically defensible way.
- Models (maps) should be used to improve sample surveys, both in the design phase and in the estimation phase.