2 Collecting Data

MTH 3240 Environmental Statistics

Spring 2020

Topics

Objectives

Objectives:

- Recognize bias in non-random sampling schemes.
- Know how the two fundamental random sampling schemes, simple random sampling and systematic sampling, are carried out.

Introduction to Sampling

- In environmental studies, the **population** of interest is often a spatial region or a span of time.

Introduction to Sampling

- In environmental studies, the **population** of interest is often a spatial region or a span of time.
- In these cases sampling involves selecting spatial locations or time points at which a variable is measured.

- Samples can either be taken ***randomly*** or ***nonrandomly***.

- Samples can either be taken ***randomly*** or ***nonrandomly***.
- **Nonrandom** samples are often **biased**.

- Samples can either be taken ***randomly*** or ***nonrandomly***.
- **Nonrandom** samples are often **biased**.

  A sampling scheme is ***biased*** if it has a systematic tendency to favor certain groups within the population, leading to their ***over-representation*** in the sample.

- Samples can either be taken ***randomly*** or ***nonrandomly***.
- **Nonrandom** samples are often **biased**.

  A sampling scheme is ***biased*** if it has a systematic tendency to favor certain groups within the population, leading to their ***over-representation*** in the sample.

  It's ***unbiased*** if there's no such tendency.

- **Examples**: Here are some *nonrandom*, *biased* sampling schemes.

- **Examples**: Here are some *nonrandom*, *biased* sampling schemes.
  - Selecting sample locations because of their accessibility (e.g. close to a road or trail).

- **Examples**: Here are some *nonrandom*, *biased* sampling schemes.
  - Selecting sample locations because of their accessibility (e.g. close to a road or trail).

    This is an example of a ***convenience sample***.

- **Examples**: Here are some *nonrandom*, *biased* sampling schemes.
  - Selecting sample locations because of their accessibility (e.g. close to a road or trail).

    This is an example of a ***convenience sample***.
  - Hand-picking sample locations *thought* to be ”representative” of the study region.

- **Examples**: Here are some *nonrandom*, *biased* sampling schemes.
  - Selecting sample locations because of their accessibility (e.g. close to a road or trail).

    This is an example of a ***convenience sample***.
  - Hand-picking sample locations *thought* to be ”representative” of the study region.

    This is called ***judgmental sampling***.

- **Examples**: Here are some *nonrandom*, *biased* sampling schemes.
  - Selecting sample locations because of their accessibility (e.g. close to a road or trail).

    This is an example of a ***convenience sample***.
  - Hand-picking sample locations *thought* to be ”representative” of the study region.

    This is called ***judgmental sampling***.

    But our judgment can deceive us.

- **Random** samples, if selected carefully, are **unbiased**.

- **Random** samples, if selected carefully, are **unbiased**.
- **Examples**: Here are some types of *random*, *unbiased* sampling schemes.

- **Random** samples, if selected carefully, are **unbiased**.
- **Examples**: Here are some types of *random*, *unbiased* sampling schemes.
  1. ***Simple random sampling***

- **Random** samples, if selected carefully, are **unbiased**.
- **Examples**: Here are some types of *random*, *unbiased* sampling schemes.
  1. ***Simple random sampling***
  2. ***Systematic random sampling***

  (These are described in the slides ahead.)

Simple Random Sampling

- A ***simple random sample*** of size n is a sample that's selected in such a way that every possible size-n subset of the population has the same chance of being selected.

Simple Random Sampling

- A ***simple random sample*** of size n is a sample that's selected in such a way that every possible size-n subset of the population has the same chance of being selected.

  Think of "drawing names from a hat" or "throwing darts at a map."

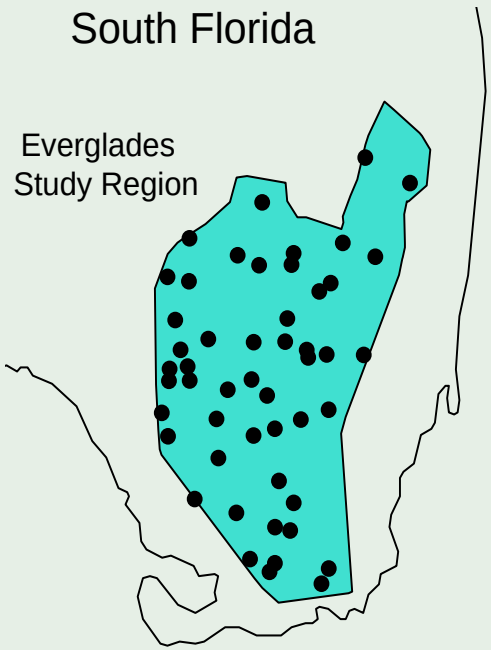
Example

To estimate the average density of sawgrass plants (plants per 1 m^2) in the Everglades region, we could take a **simple random sample** of locations, and count the number of sawgrass stems in a 1 m^2 *quadrat* at each location.

Example

To estimate the average density of sawgrass plants (plants per 1 m^2) in the Everglades region, we could take a **simple random sample** of locations, and count the number of sawgrass stems in a 1 m^2 *quadrat* at each location.

The next slide shows a computer-generated **simple random sample** of $n = 50$ locations in the Everglades.


South Florida
Everglades
Study Region

- An advantage of simple random sampling is that it's always **unbiased**.

- An advantage of simple random sampling is that it's always **unbiased**.

  A disadvantage is that there's no guarantee that the sample of locations will be evenly spread over the study region (especially if n is small).

Systematic Random Sampling

- A ***systematic random sample*** is one in which the sample of locations lie on a **regular grid**.

Systematic Random Sampling

- A ***systematic random sample*** is one in which the sample of locations lie on a **regular grid**.

  (When sampling time points, the sample of time points lie at regular intervals).

Systematic Random Sampling

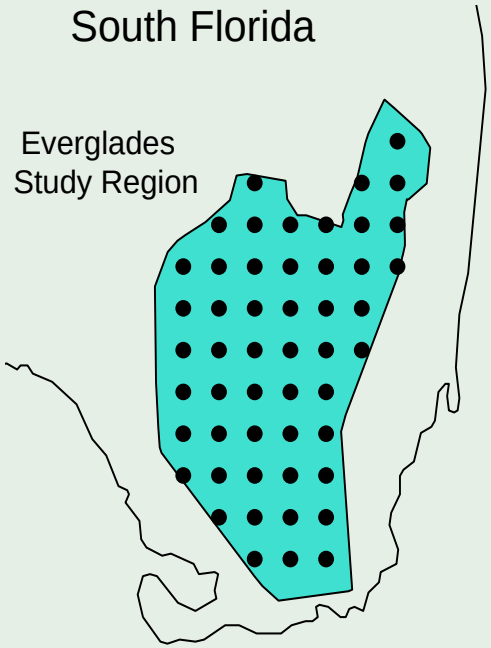
- A ***systematic random sample*** is one in which the sample of locations lie on a **regular grid**.

  (When sampling time points, the sample of time points lie at regular intervals).

- To incorporate chance into the sample selection process, the grid is initialized at a single, randomly selected starting location.

Example

The next slide shows a **systematic random sample** of $n = 51$ locations in the Everglades.


South Florida
Everglades
Study Region

- Two advantages of **systematic random sampling** are:
  - It's **unbiased**.
  - The sample of locations is guaranteed to be evenly spread over the study region.

- Two advantages of **systematic random sampling** are:
  - It's **unbiased**.
  - The sample of locations is guaranteed to be evenly spread over the study region.
- A disadvantage (particularly when sampling **time points**) is that if there's a **cyclical pattern** in the variable being measured, and the interval between sampled time points coincides with the **period** of the **cyclical pattern** (e.g. diurnal, weekly, seasonal etc.), the sample **won't** be representative of the population.

- Note that an unbiased sampling scheme doesn't *guarantee* that the sample will be representative of the population, ...

- Note that an unbiased sampling scheme doesn't *guarantee* that the sample will be representative of the population, ...

  ... just that there won't be any *systematic tendency* for it to be unrepresentative.

- Note that an unbiased sampling scheme doesn't *guarantee* that the sample will be representative of the population, ...

  ... just that there won't be any *systematic tendency* for it to be unrepresentative.

  In particular, an unbiased sampling scheme using only a small sample size can, just by chance, lead to samples that are uncharacteristic of the population.

Other Sampling Schemes

- Other sampling schemes (used less often) include:
  1. ***Stratified random sampling*** – **Partition** the study region into **sub-regions** defined by a categorical variable (e.g. habitat types), then take a simple random sample from each sub-region.

Other Sampling Schemes

- Other sampling schemes (used less often) include:
  1. ***Stratified random sampling*** – **Partition** the study region into **sub-regions** defined by a categorical variable (e.g. habitat types), then take a simple random sample from each sub-region.
  2. ***Two-stage random sampling*** – Take a simple random sample of **sites**, and then, from each site, take another simple random sample of **specimens** (soil, water, etc.).