GEOG 413/613

LECTURE 2

1

Sampling

- When a phenomenon occurs, it may be due to
  - a random process
  - a systematic process
    - Statistical analysis helps to determine the nature of the process
- Processes may be linear, non-linear
- Sometimes processes may be acting in combination

2

2

Sampling

- Spatial Sampling
  - A biogeographer selects locations to examine environmental change in a national park
  - Medical geographer chooses to examine hospital use patterns in certain neighborhoods
- Non-spatial sampling
  - a geographer conducting a study on attitudes towards landslides may choose a nonspatial list of households in the area
- A study on housing quality may be taken from nonspatial list e.g. tax rolls or a spatial source a map

3

3

Sampling

- Advantages
  - It often a necessity in geography because population is too large
  - Efficient and cost effecting methods of data collection
  - Can provide highly detailed information
  - Can provide highly accurate information
  - Allows repeated collection of data (e.g. at specific times of the year)

4

4

Process, Pattern and Sampling

- In ecological studies the match between pattern and process is far from perfect
  - Changes in process intensity can create different patterns
  - Different process can create the same pattern signature
- A spatial pattern is a snapshot in time reflecting a single process or multiple processes at the given time
- Ecological data are a result of embedded and confounded processes evidenced as
  - Trends at larger scales
  - Patchiness at intermediate scales
  - Randomness/noise at small scales

5

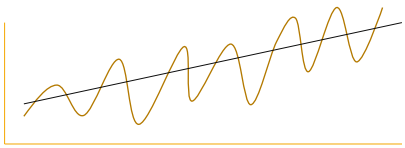
5

Spatial Dependence: Topography, Drainage, Soil

Spatial Autocorrelation: Vegetation a, Vegetation b, Vegetation c, Vegetation d, Vegetation e	Spatial Autocorrelation Vegetation v	Spatial Autocorrelation Vegetation a, Vegetation b, Vegetation c, Vegetation d, Vegetation e
Fire	Fire, Drought, Grazing, Insect outbreak	Fire, Drought, Grazing, Insect outbreak
Vegetation a: trend Vegetation b: patchy Vegetation c: random Vegetation d: patchy Vegetation e: patchy	Vegetation v: patchy	Vegetation a: trend + patchy Vegetation b: repeated patches Vegetation c: patches at two scales Vegetation d: patchy Vegetation e: repeated patches
One process resulting in several patterns	**Several processes resulting in the same pattern**	**Several processes resulting in several patterns**

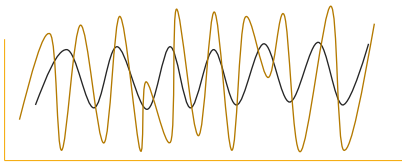
6

6



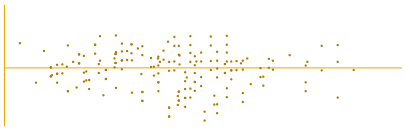
Large scale trends (gradients)

If data are gathered along a temperature gradient, tree height can increase in a linear fashion at a large scale



Cycle patches (hot spots)

Topography and spatial dispersal processes can generate patchy patterns at intermediate (landscape) scale

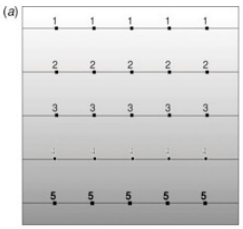


Noise at local scale

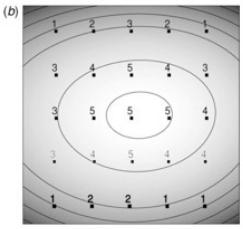
There is only random noise at a micro (local) scale

7

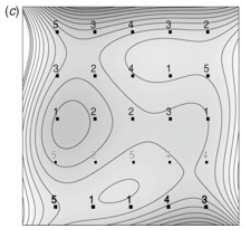
7



gradient



Patch



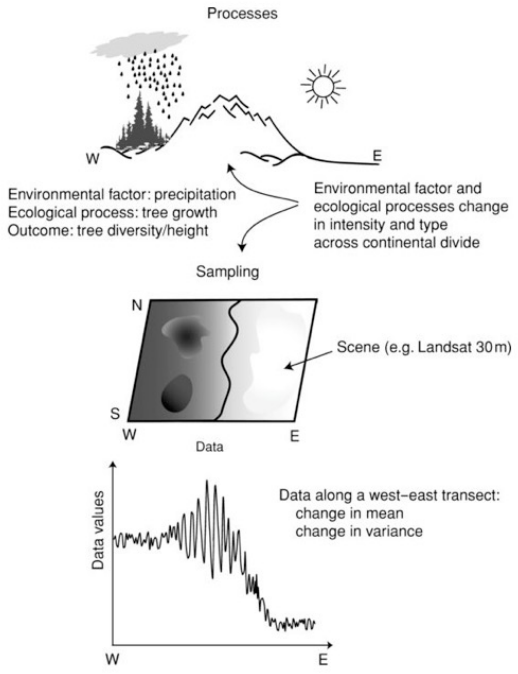
Random

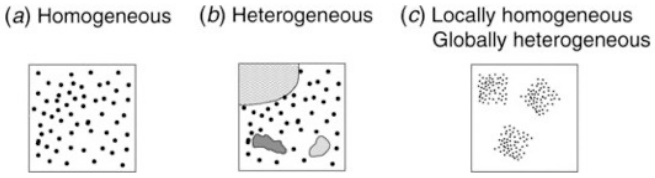
- Exogenous (induced) pattern generation is based on factors independent of the process of the variable of interest
  - For example, species abundance can respond to an exogenous process such as a moisture gradient
- Endogenous(inherent) process such as dispersal, competition or inhibition is an inherent property of the variable of interest
- Most data have some degree of spatial structure
  - "Everything is related to everything else, but near things are more related than distant things" Tobler 1970

Fortin, Marie-Josée,Dale, Mark R. T. Spatial Analysis: A Guide for Ecologists. Cambridge University Press.

8

8





Homogeneous (a) versus heterogeneous landscapes (b, c). In (b), the study area includes subregions where plants cannot grow, such as a lake, on rock or in inappropriate soil types. In (c) the plants can grow only in subregions where the soil type is appropriate.

The measurement of spatial statistics is based on the assumption of of stationarity (homogeneity) in that pattern is independent of the location and direction.

9

9

Sampling Methods

- Any sampling design for studying processes, imposes an arbitrary
  - template
  - filter
  - segmentation
- Selecting a representative part of a population for statistical analysis;
  - various designs of sampling can be applied
    - random sampling
    - systematic sampling
    - stratified sampling
  - Within the above designs, one may decide on
    - point
    - line
    - or area method

10

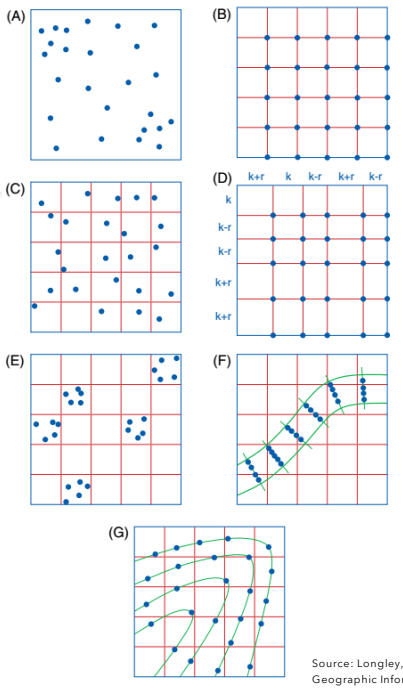
10

Sampling

- Provides knowledge about a whole population
  - i.e. make inference about a population from the sample data
- Larger sample sizes are more accurate representations of the whole
  - Large samples are costly: time, labour
    - Can be wasteful since we can statistically infer from appropriate samples
- A sampling strategy with the minimum bias is the most statistically valid

11

11



Sampling Methods

Spatial sample designs: (A) simple random sampling, (B) systematic sampling, (C) stratified random sampling, (D) stratified sampling with random variation in grid spacing, (E) clustered sampling, (F) transect sampling, and (G) contour sampling.

Source: Longley, Paul A.; Goodchild, Michael F.; Maguire, David J.; Rhind, David W.. Geographic Information Science and Systems, 4th Edition. Wiley.

12

12

Random Sampling

- Random sampling: each member of the population has an equal chance of being selected
  - Advantages:
    - Can be used with large sample populations
    - Avoids bias
  - Disadvantages:
    - Can disproportionately represent some parts of the population at the expense of others

13

13

Systematic Sampling

- Systematic Sampling: Samples are chosen at regular intervals
  - Sample locations are evenly distributed for example every two metres along a transect line
    - systematic sampling implies a regularly spaced grid
      - Advantages:
        - It is more straight-forward than random sampling
        - Provides a good coverage of the study area
      - Disadvantages:
        - It is more biased: not all points have an equal chance of being selected
        - It may lead to over or under representation if there is periodicity in the data (e.g. sampling at the same interval as the location of erosion barriers along a beach. Or a city road grid)

14

14

Stratified sampling

- Stratified sampling: used when the parent population is made up of sub-groups that of interest.
  - Divide the sampling design into strata(classes), and then select a sample from each stratum
  - The strata are defined so that individuals inside each class are similar based on the characteristic believed to influence the phenomena

15

15

Stratified sampling

- Advantages:
  - If the proportions of the subgroups are known, the results are representative of the whole population
  - Correlations and comparisons can be made between subgroups
- Disadvantages:
  - The proportions of the subgroups must be known

16

16

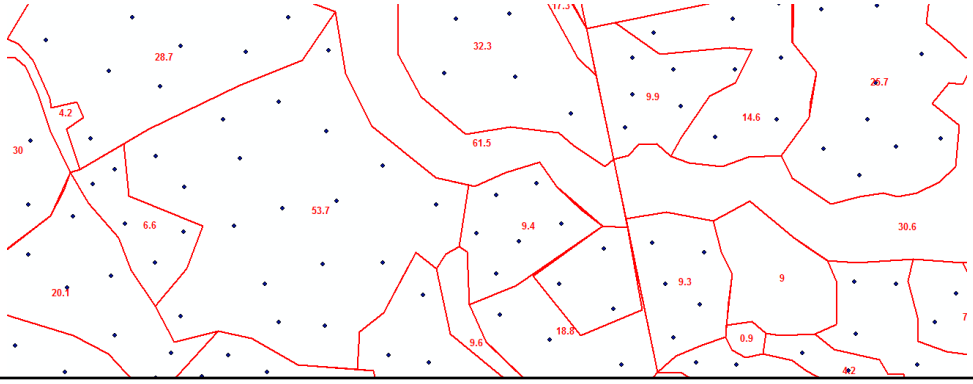
Air Photos for Stratified Sampling

- Looking for distinct, uniform areas
  - Crown size (age), harvest history
  - Hardwoods (gray) and softwoods (green)

17

Stratified Sampling

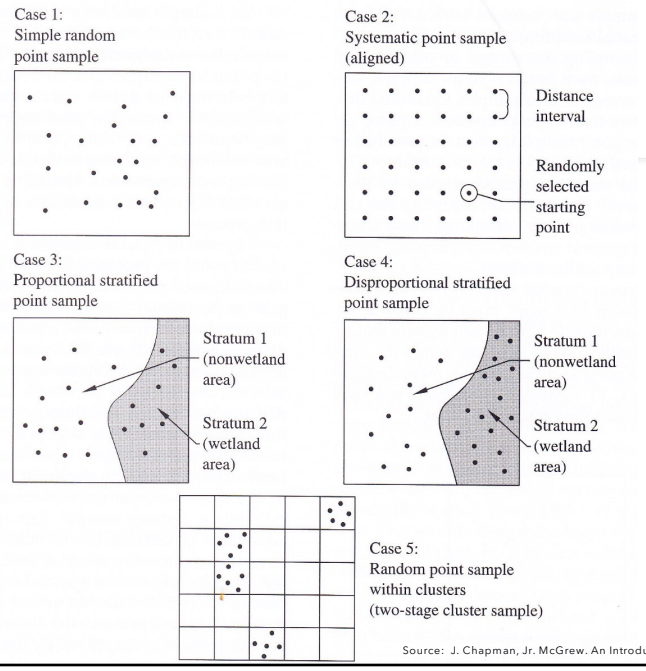
- Generate sample points randomly
  - X points per area, e.g. 1 point every 3 hectares
  - Each point tied to polygon = unique stand



18

Stratified Sampling: Population

19

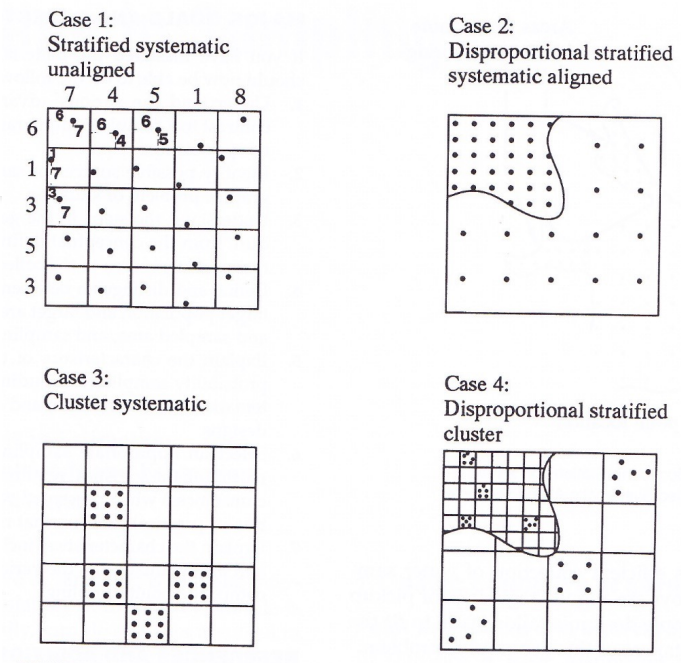


Point Sampling Methods

20

Source: J. Chapman, Jr. McGrew. An Introduction to Statistical Problem Solving in Geography

20



Hybrid Point Sampling Methods

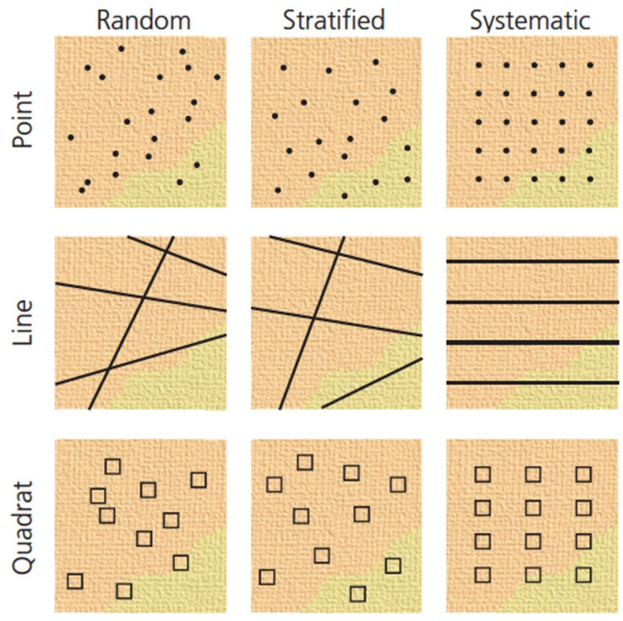
Consider a study on the distribution and intensity of nitrogen and phosphorus distribution in a bay receiving agricultural run off.

- What would be the appropriate sampling design?

Source: J. Chapman, Jr. McGrew. An Introduction to Statistical Problem Solving in Geography

21

21



Source: https://www.geography-fieldwork.org/a-level/before-starting/methods/sampling/

22

22

Sampling: some considerations

- the number of observations (sample size) at least 30. The more the better
- the size of the study area (the extent);
- the size (granularity) and shape of the sampling units
- the sampling strategy or spatial layout of the sampling units used to collect the data
- the spatial lag (spatial distance) among sampling units

23

23

Declustering

- Point samples may be unduly clustered spatially, for a variety of reasons
  - samples from boreholes and wells may provide the basis for a chemical analysis of groundwater sources. The distribution of boreholes is often clustered
  - Geological and hydrographic data collection is often in localized areas, with sparse sampling elsewhere.
  - Practical constraints, such as access in built-up or secure zones may also dictate sampling schemes that exhibit strong clustering

24

24

Declustering

- Declustering is the removal of the known or estimated adverse effects of clustering in order to obtain a better representation of population data
- Procedures involving adjusting the sample before futther analysis
  - Defining a grid over sampled points
  - Defining a grid over sampled points and use point density as weighting function
  - Use of voronoid regions

25

25

Levels or Scales of Measurement

- Nominal
  - Categorical data e.g. land use type, religious affiliation
- Ordinal
  - Ranked data , e.g. main, secondary, minor roads
- Interval:
  - Interval between any two units can be measured on scale. Zero value is assigned arbitrarily e.g. Celsius and Fahrenheit scales (80^{o}F is not twice as hot as 40^{o}F)
- Ratio:
  - interval data with an absolute zero value

26

26

Summary of Levels of Measurement

Level of measurement	Brief description
Nominal	Each value or unit of data is assigned to one of at least two categories or qualitative classes; no assumptions are made about relationships between categories—only that they are "different."
Ordinal	Values themselves are placed in some rank order.
Strongly ordered	Each value or unit of data is given a particular position in a rank-order sequence; that is, each value is assigned its own particular rank.
Weakly ordered	Each value or unit of data is assigned to a category, and the categories are then rank ordered.
Interval	Each value or unit of data is placed on a measurement scale, and the interval between any two units of data on this scale can be measured; origin or zero starting point is assigned arbitrarily (i.e., origin does not have a "natural" or "real" meaning).
Ratio	Each value or unit of data is placed on a measurement scale, and the interval between any two units of data on this scale can be measured; origin or zero starting point is "natural" or non-arbitrary, making it possible to determine the ratio between values.

Source: J. Chapman, Jr. McGrew. An Introduction to Statistical Problem Solving in Geography

OK to compute....	Nominal	Ordinal	Interval	Ratio
frequency distribution.	Yes	Yes	Yes	Yes
median and percentiles.	No	Yes	Yes	Yes
add or subtract.	No	No	Yes	Yes
mean, standard deviation, standard error of the mean.	No	No	Yes	Yes
ratio, or coefficient of variation.	No	No	No	Yes

27

27

References (strongly recommended)

- Dale, M. R. T., & Fortin, M.-J. (2014). *Spatial analysis : a guide for ecologists* (Second edition). Cambridge University Press.
  - Chapter 1

28

28