GEOG 312 – DATA ANALYSIS IN GEOGRAPHY
ASSIGNMENT #1: DATA COLLECTION IN GEOGRAPHY

Most studies draw a sample from a population, which is ultimately of greatest interest in our research. As we have discussed in class, your final results are only as good as how you sampled the population and then applied your statistics. In this respect, it is critical that prior to starting a study, you decide on what to measure, where, when and how often.

TYPES OF SAMPLES

In general, we can obtain an **unbiased sample** by giving every member of a population an equal chance of being included in the sample.

- *Simple random sample*: every individual has an equal chance of being included in a sample. Random samples are, however, prone to error. For example, quite by chance a random sample might contain a disproportionately large number of individuals with specific characteristics. A list of random numbers is provided on the following page- it doesn't matter how you use it, as long as you do it consistently.

- *Systematic sampling*: every *n*th individual is selected or individuals are sampled every *n*th minute/hour etc.

- *Stratified random sample*: the population is divided into strata and samples are drawn randomly (from the above methods) from each strata.

- *Cluster sampling*: groups of individuals or samples within specific areas, with the individuals in each cluster drawn at random. A strength and a weakness of cluster sampling is that individuals from each cluster tend to be homogeneous.

The opposite of unbiased sampling is to collect a **sample of convenience**, in which the investigator chooses the individuals to sample out of convenience.

- *Purposive sampling*: individuals are selected that are believed to give us the best information.

It is important to remember that sampling by its nature is biased and that you do not place too much confidence in your results- do not assume your results have concreteness.

73735	45963	78134	63873
02965	58303	90708	20025
98859	23851	27965	62394
33666	62570	64775	78428
81666	26440	20422	05720
15838	47174	76866	14330
89793	34378	08730	56522
78155	22466	81978	57323
16381	66207	11698	99314
75002	80827	53867	37797
99982	27601	62686	44711
84543	87442	50033	14021
77757	54043	46176	42391
80871	32792	87989	72248
30500	28220	12444	71840

SPATIAL SAMPLING

Data will be collected through space rather than repeated measurement in the laboratory or at a location through time. The starting point for spatial sampling is a map of the area to studied, such as provided on the next page of this handout. As discussed in class, spatial sampling can involve:

- *Traverses*: line samples in which data is collected from subjects that fall on or within a certain distance of the line
- *Point*: subjects at or within a certain distance of a point are sampled
- *Quadrats*: subjects that fall within an area are sampled. The quadrats are usually distributed randomly

Figure 6.6 Types of Spatial Sampling

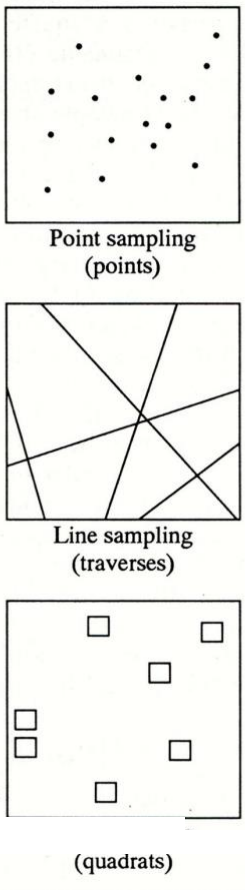
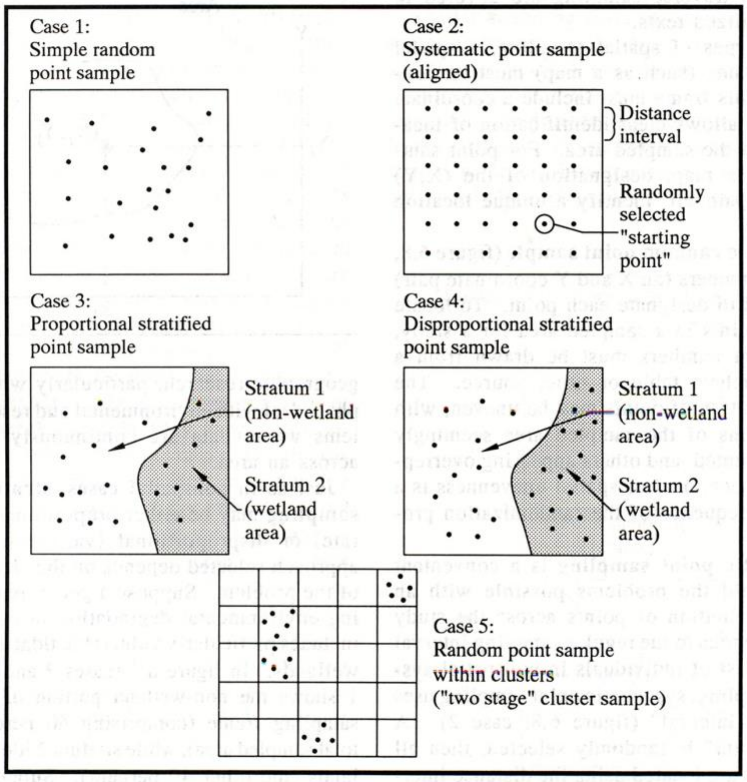


Figure 6.8 Types of Spatial Point Sampling



DATA COLLECTION PROCESS

Once you are ready to start collecting data it is important to organize the collection process, including the data collection forms and data coding.

Data Collection Forms
When you collect data, it is important that you have an organized collection form, especially when working in the field. Think of your **raw data** as the pieces to a jigsaw puzzle that you will use to put it all together. The data you collect and the strategy that you use to collect the data should match the analysis that you will be using to test your research question/hypothesis. Since this is the main link between your field work and analysis it is important that the form be easy to understand and easy to work with.

Important things to remember when collecting data:
- Begin thinking about the data you are going to collect early in the research design process
- Think about where and when you will be sampling your data
- Make sure that the collection method is clear and easy to use, but also avoids bias and in-field decision making
- Make a clear and concise data form to facilitate data collection and the analysis
- Copy all handwritten completed data forms and enter your data into a spreadsheet program immediately (also make photographic backups)
- Plan a schedule for when and where you will be collecting your data and stick to it
- Do a test sample (of at least 30) to determine the number of samples you will need to be confident on your estimates
- NEVER discard original data

Coding Data
It is critical that you keep clean notes to avoid transcription error and confusion when you get to the analysis stage. It is also a good idea to draw a well-labeled map of your study site so that you can remember where you took your samples. The availability of GPS on your smartphone gives you no excuse to forget a map!

EXAMPLE FIELD DATA SHEET

Names:_____
Sampling Approach:_____

Notes (e.g. weather conditions, your own conditions, etc.):

Sample	Lat.	Long.	Circum.	Diameter
1				
2				
3				

TASK

You have been hired by the University to determine the average size of trees on east campus as measured by the diameter at breast height (DBH). You do not need to measure every tree on east campus, but rather need to estimate the average size of the trees on campus. Use an engineering tape to measure the DBH. Working in groups of up to 3, measure 30 trees using each of the sampling strategies below and take Lat/Long using a GPS or Google Maps/Google Earth (ask Dr. Allen if you can't access this on your phone).

- Approach 1: Systematic survey
- Approach 2: Random survey
- Approach 3: Random cluster (10 clusters of 3 adjacent trees)

Create a data collection field for the field sampling and enter your data in Excel (or another program) using a separate sheet for each sampling approach.

DELIVERABLES

1. (30% of points) Using R, create one or more aesthetically-pleasing histograms of tree diameter collected using the three sampling approaches listed above. Tip: Save your data table as a .CSV file format. Then read the table into R using the function read.csv(), specifically:

   table1 = read.csv("The/Path/To/Your/CSV/File.csv", header=T).

Provide a clear description of the data presented in your histograms(s) along with a description of the units of measurement, sampling dates, etc. The title should allow the figure to stand-alone from the main body of the assignment, project report or paper. Your histogram(s) for this assignment should include axes titles, legend (inside of the graph(s)) and units. Your histogram(s) can be included directly in your report by using the Export button on the RStudio Plots window. We will go over how to make histograms in R in class next week.

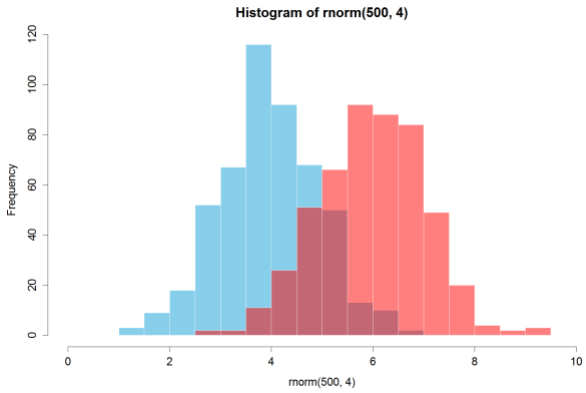


FIG. 1. …

2. (20% of points) Using the R functions that will be discussed in class next week, calculate the following for each sampling strategy:
    a. mean
    b. standard deviation
    c. skewness
    d. kurtosis
    e. max
    f. min
    g. range

    Use the following table structure to present your data including a heading:

Table 1: Provide a clear description of the data presented in your table along with a description of the units of measurement, sampling dates, etc. The title should allow the table to stand-alone from the main body of the assignment, project report or paper.

	Systematic Sample	**Random Sample**	**Random Cluster**
Mean			
Standard Deviation			
Skewness			
Kurtosis			
Maximum			
Minimum			
Range			

3. (20% of points) Based on the histograms and descriptive statistics, describe the differences between the datasets. Don't try to explain the differences, but rather *describe* how they are different and what that tells us.

4. (30% of points) Provide a description of how you completed your systematic, random and random cluster sampling. A map of campus (from Google Maps, etc.) would be very helpful to explain your sampling strategy

SUBMIT THE FOLLOWING ITEMS (ONLY ONE PER GROUP)

1. Cover page with assignment title, unique 6-digit ID, and names of the group members (select a unique 6-digit ID for your group, e.g., last 2 digits of phone numbers).
2. Digital Data Collection Tables (scaled so that the data for each sampling strategy is on 1 page only).
3. Neatly presented histogram(s) of your data that is properly labeled,
4. Neatly presented table of your statistical results.
5. Typed answers to the above questions.