Sampling: Lecture Notes

You may want to know something or quantitative.

Introduction: the question may be

- "How many elephants in Sapo?"
- "How often do elephants come to Gola?"
- "How many raphia trees per hectare are there in a typical swamp?"
- How heavy is the average zebra duiker in the region?
- How old is the average *XYZ* tree?

► Get everyone to think of a meaningful biological or sociological question.

We cannot count them all. So, we **sample.**

SAMPLING means to count the number of things or events using a pre-determined sampling methodology.

This can provide an accurate **ESTIMATE** of **all** of the things you want to measure in the study area.

YOU HAVE TO DEVELOP A ROBUST SAMPLING DESIGN…this may be unique for each new study.

If you are sampling **density** of plants or animals, the bigger area to which your final estimates apply is called the ***sampling area***. E.g a national park…or the area within 5km of a village…all habitat types or one…

Check that everyone knows what density means.

► Get everyone to name at least one sampling area that applies to their own question .

If it is a village of people or a specific group of animals, **population** is used, e.g., all the villages around the park, or just the elephants using the area around Gola…

Get everyone to name at least one population that could apply to their own question, or to one of the other people's questions. . animals, trees, humans

Do not change the design half way through…you can only draw firm conclusions from the sample, and you have to extrapolate your results to the rest.

SAMPLING UNITS are smaller sections within which you count the plants, animals, humans, or events. These can be plots or transects (area, length) or fixed amounts of time per hour, or units such as households.

Think of sampling units appropriate to people's own questions.

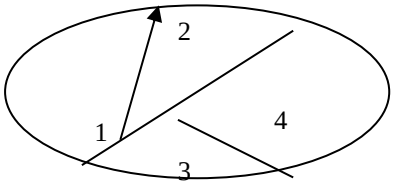
PLOTS AND TRANSECTS

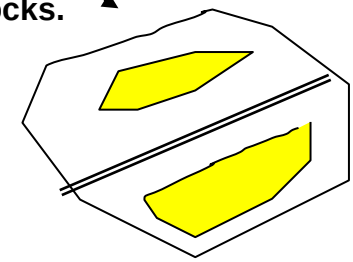
Plots- you count everything. Immovable things (botany)
Transects- have no fixed width: count what you detect from the centre line. (moving things: animals)

REPRESENTATIVE SAMPLING.
All sampling units should be representative of the sampling area or population. The proportion of the different things in the sample, or how they are dispersed should reflect the reality in the wider area.

Examples: show the duiker density design suggested in the manual (what happens if you sample close to roads- you miss a lot of duikers).

AVOID BIAS. Prelocate your sample units within blocks.





Discussion section

People to contribute

What kinds of bias should we consider?- **SPATIAL…TEMPORAL…SOCIAL….**

Ask people to show how rivers might affect the distribution of water chevrotain. If we want to know the number of water chevrotain per km2 in NNNP how should we sample? Along rivers only? All over the Park? Along rivers but keeping in mind that this is only representative of riversides?

Ask how we might select people for interviews about the use of forest products. Do we select only houses near the river? Do we select houses near the market? Do we ask people we meet in the village hangar? DISCUSS

How might season affect our results? Time of day?

Random sampling.
This means all the points in your study area have an equal probability of being chosen. So it is unlikely that any systematic bias will affect the sample.

Construct a numbered system which represents all possible sample units. (random numbers).

To select 10 sample units from within 100, number the units first. Then do a random numbers exercise: numbers 0-9 in a hat, choose 20 times whilst putting it back each time.,

write them down in order, pair them up. This gives 10 double digit numbers (including 02 etc).

Try the stop watch system

Systematic sampling.

Things are not usually systematic in space in nature, so a random start point can be the corner for a systematic design (plots or transects).

If you deliberately choose plots/ transects for convenience, you BIAS your sample

Explain with a map: roads are usually on dry flat ridges or floodplains, and have human influence along them in a band. NOT representative of anything except roads!

CHOOSING SAMPLING UNITS IN SPACE

Transects should cross major topographic features like the drainage and/or mountain ridges. Show people how this is representative of the vegetation.
Use a baseline preselected on a map. Choose –at random- a start point for the transect series. Why should you do this before going to the terrain? How might you choose the start point?

Transects should be at 90º to this.

Plots
Superimpose a grid on your map.

Number the grid, and choose 20 of these squares as your random sample plots.

EXERCISE

Need: 4 maps, 4 rulers. Divide the area into 100 (10 x 10 grid)

Choose random plots (doing the random numbers as before)
Choose systematic plots (every xth so that it is well distributed- need a random start, though).

Do you think that some of these plots will be very different from the others?
Why? (near and far from the river and village.)

So- maybe you need to **stratify**
Dividing into blocks

This ensures no area is left unsampled. This helps to ensure good coverage across the whole area: then you choose your plots within each block.

STRATIFICATION
This means dividing into categories. (strata) This is when you suspect a variable affects your results

Think of different variables with their effects

(like vegtype, hunting pressure, even time of day (morning/ afternoon, or season, or moonlit nights for nocturnal species…)

If there are no differences in density between strata, they can be pooled- this increases sample size and reduces variablity.

POSTSTRATIFICATION

If you have collected environmental data whilst carrying out a survey, but you did not know that one of the variables made a difference, you may come to suspect it during the survey or when you look at the results.

So you can take the variable, split the sample into two strata, and test (statistically) if there is a difference.

Example?

SAMPLE SIZE
…is important because the bigger your sample size, the better your precision. This is because your variance goes down.

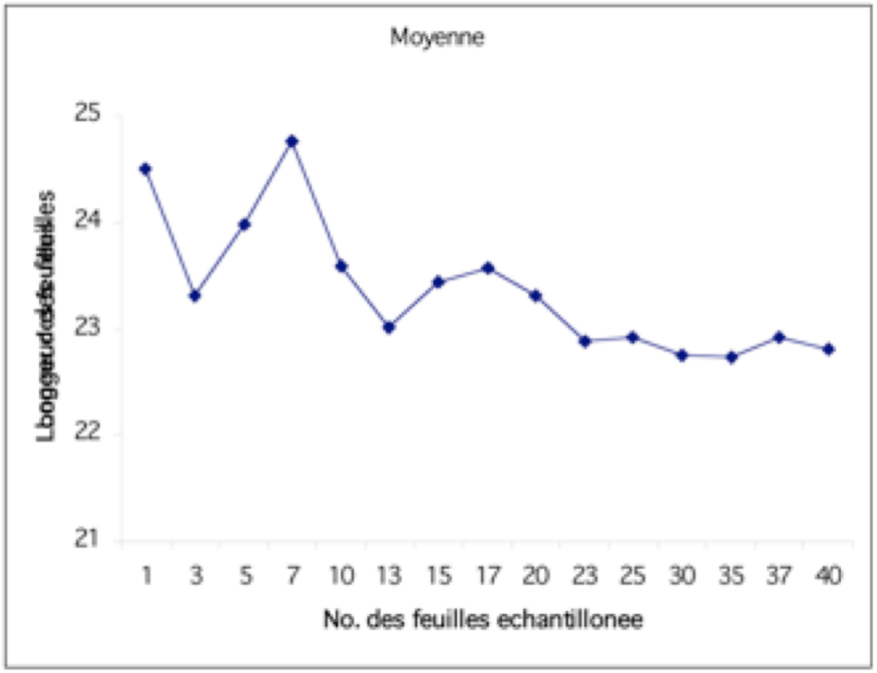
Look at the figure in the manual (3.2)

What sample size is reached before the mean estimate of sighting frequency stabilises?

Would this sample size be the same for all kinds of data?

Exercise
Collect 100 things (leaves?) and take a sample of 5 each time. Take the mean of the measurements. Repeat: mean of 3, 5, 7, 10; 13, 15 etc; see how the line flattens off (running mean)

1	24.5
2	18.8
3	26.6
4	26.5
5	23.5
6	23.2
7	30.2
8	21.7
9	21.7
10	19.1
11	23.1
12	21.2
13	19.0
14	31.7
15	20.7
16	23.6
17	25.7
18	22.8
19	19.4
20	23.0
21	21.6
22	21.7
23	16.9
24	22.6
25	24.0
26	28.9
27	19.1
28	24.5
29	20.2
30	17.1
31	26.1
32	22.6
33	22.5
34	24.1
35	17.7
36	29.3
37	22.9
38	19.8
39	22.5
40	22.0



AFTERNOON:
Read and Discuss Richard Barnes's work: distance to roads affects elephant density.