Statistical Concepts for Surveillance of Bee Pests and Diseases

***Daisy Salifu* (PhD)**

Biostatistics Unit - icipe

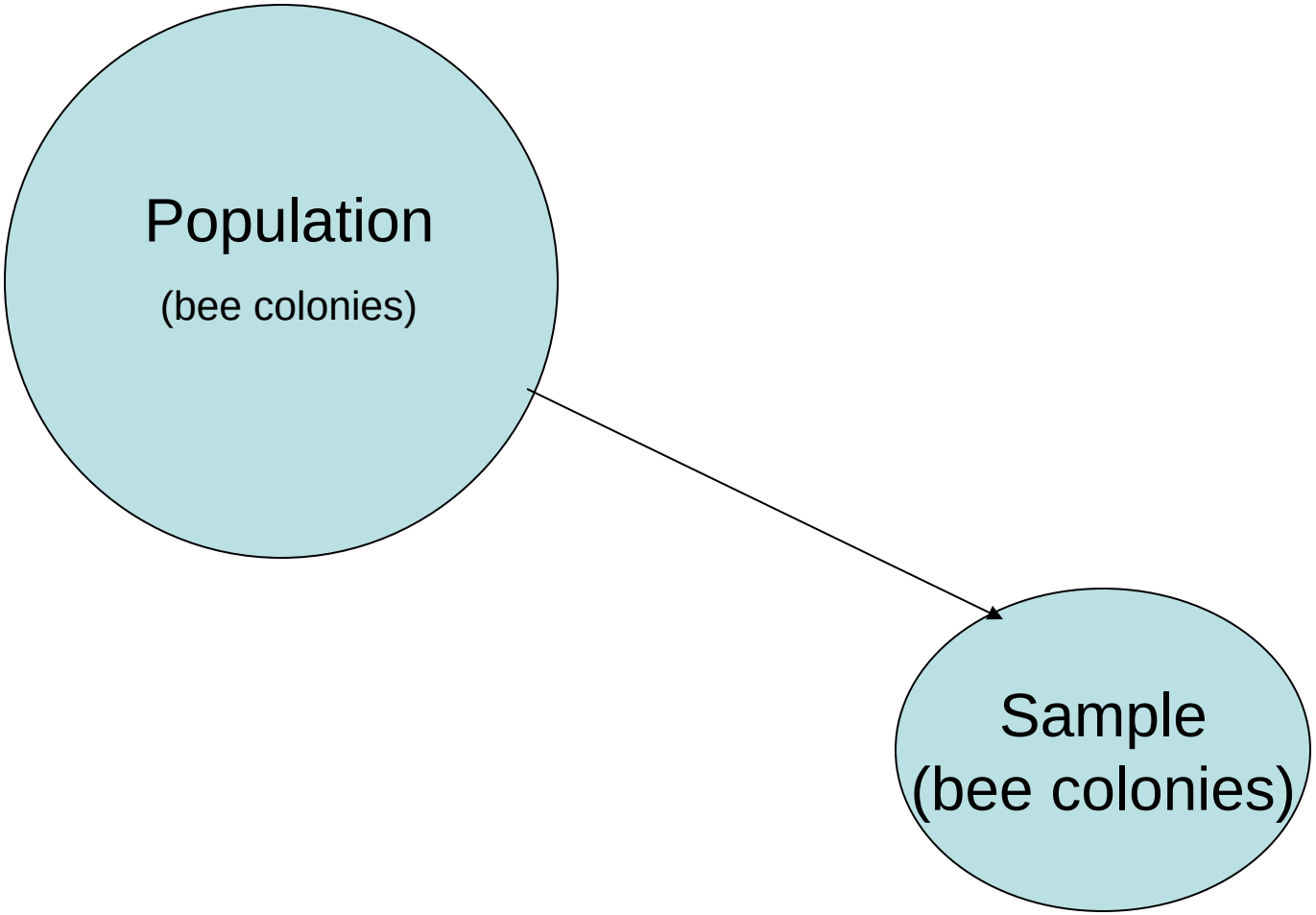
Outline

- Research questions
- Sampling of hives (observational units)
- Sample size determination
- Data to collect
- Data analysis

Research questions

- What are the current pests and diseases of honey bees in Africa?
- How bad are they?
- Are African honey bees more tolerant or resistant to certain diseases or pests
- What is the effect of environmental toxins - insecticide/pesticides in use in agricultural systems
- Can we link bee health with nutrition (pollen, nectar studies)
- Any variation due to seasons?

It all starts here….



How representative is the sample of the target population?

Know your situation well

- ❑ know bee keeping localities
- ❑ Number of apiaries in the localities
- ❑ Consider geographical distribution - are these locations similar in characteristics that we can just randomly select a number of apiaries to get colonies
- ❑ At apiary level, what type of hives are used – if different hives are used then we make sure each hive type is represented in our sample

Multi-stage sampling scheme…

Agro-ecological zones (several zones in a country)

select

Apiaries (in each agro-ecological zone)

select

Colonies (in each selected apiary)

Other features to consider

- Number of managed colonies in an apiary - apiaries with few colonies e.g <10 can be excluded (ref.”**USDA Honey Bee Pests and Diseases Survey Project Plan for 2013”)**

❑Scale of keeping – exclude hobbyst and consider semi or professional keepers

Representativeness

- If conclusions are to be valid for the whole study population, then the sample has to be representative of that population.
- The burden is to demostrate in our study report that the sample , regardless of how it was chosen, represents the ‘whole population’.

Sample size determination

How many colonies do we need?

- How big a sample do I require? – most frequently asked question
- It depends on what you want to estimate: examples
  - Prevalence of disease or pest (proportion)
  - Honey yield per colony (in kg – thus continuous data)
  - Density or counts of pests (e.g. varroa mites, small hive beetles)

Sample size for prevalence study

- **The following simple formula (Daniel, 1999) can be used:**

$$n = \frac{Z^2 P(1-P)}{d^2}$$

where n = sample size,
Z = Z statistic for a level of confidence (can be 95%, 99%),
P = expected prevalence or proportion, if 20%, P = 0.2
d = precision required, if 5%, d = 0.05

- **Z statistic** (values are obtained from the Z distribution tables)

Z value is 1.96 for the level of confidence of 95%

Z value is 2.58 for the level of confidence of 99%

That is, if investigators are to present their results with 95% or 99% confidence interval (CI), these Z values are used to calculate sample size

- **Expected proportion P** - this is the proportion the investigators are to estimate from the study

Can get this from previous studies or pilot study

- **Precision (d) –** how precise do we want to estimate P. e.g. we want to estimate P with a precision ±5%, thus d=0.05.

Notice then that (d) should be small for precise estimate of P.

- **We want a narrower CI and hence to obtain a narrower CI, we need to design a study with a smaller *d (good precision or smaller error estimate.***
- **Example:**

Prevalence = 30%, P = 0.3

d is 5% = 0.05 and choose Z=1.96

$n = [1.96^2 \times (0.3)(0.7)]/0.05^2 = 322.69$

Sample size *n* is 323

- We may want to over sample by 10% to 20% to cater for non-response or missing values.

Determining precision (d)

- **Guidelines (Naing et al.,2006)**

Use *d* = 5% if P is between 10% - 90%

If P <0.1 (10%), then *d* can be half of P.

If P >0.9 (90%), then d can be half of (1-P)

Reference

L. Naing T. Winn, B.N. Rusli (2006) Practical Issues in Calculating the Sample Size for Prevalence Studies. *Archives of Orofacial Sciences* Vol 1: 9-14.

Sampling plan – Kenyan case

Sites in Kenya	Agro-ecological Zone	No. of apiaries	No. of hives per apiary
1. Nairobi	Sub-humid	2	8
2. Mtwapa	Semi-humid	2	8
3. Taita	Semi-arid	2	8
4. Mwingi	Semi-arid	2	8
5. Kakamega	Humid	4	8
6. Isiolo	Arid	1	8
7. Mt. Kenya	Humid	4	8
8. Isiolo	Arid	1	8
9. Kinangop	Sub-humid	3	8

Total of 168 hives

Some conditions

- Kenya is divided into 5 agro-ecological zones and all zones are represented
- An apiary to be sampled must have at least ten hives
- An apiary must be easily accessible because of the repeated sampling over seasons

- Sample size software – there are lots on the internet.
- PS – Power and sample size software

http://softadvice.informer.com/Power_And_Sample_Size_Calculations.html

- Gpower 3.1

http://gpower.software.informer.com/3.1/

Data to be collected

Data to be collected

- Need to understand the nature or type of data to collect because that has implication on the type of analysis to be used
- Quantitative data such as:
  - Honey yield,
  - counts of small hive beetles per colony
  - counts of varroa mites per adult bee

- Categorical or qualitative data e.g.
  - Presence or absence of Amercian Foul Brood in hive
  - Presence or absence of Nosema in hive
  - Presence or absence of SHB, Varroa mites.

Weather data to support biological data: rainfall, temperature, relative humidity etc.

What to look for in the data

1. How are the data distributed?
2. What are the maximum and minimum values?
3. What is the mean?
4. What is the standard deviation (reflects variability in data)?
5. If data are categorical which category carries the most values?

Data description /summaries

- Categorical/Qualitative data are summarised in terms of frequencies – can be converted to percentage

- Quantitative data are summarised using numerical measures – mean, standard deviation, etc

- Make use of pivot table in MS Excel to get the story line of the data

Creating a pivot table in Excel:

- In MS Excel, click **insert** – **PivotTable** – click **ok,** choose the variables to create a table of summaries

What to look for in the data

- Most of the descriptive statistics : mean, frequencies, standard deviation etc can be obtained from ‘Pivot table’ in Excel.

- Beyond descriptive summaries then we need statistical software – there are several that require payment for licence: SPSS, SAS, GenStat, Minitab, Stata.

- *ICIPE uses R – a free software but not user friendly!!*

Data entry and analysis

- MS Excel is a very good spreadsheet for data entry
- Most of the descriptive statistics : mean, frequencies, standard deviation etc can be obtained from ‘Pivot table’ in Excel. ***Click on ‘Insert’ on menu bar in Excel to use Pivot table.***

- Beyond descriptive summaries (formal analysis) then we need statistical software – there are several that require payment for licence: SPSS, SAS, GenStat, Minitab, Stata

Thank you

icipe
African Insect Science for Food and Health