IZSAM G.CAPORALE TERAMO

OIE Collaborating Centre Veterinary Epidemiology

Sampling approaches in animal health studies

Al Ain 13-14 November 2017

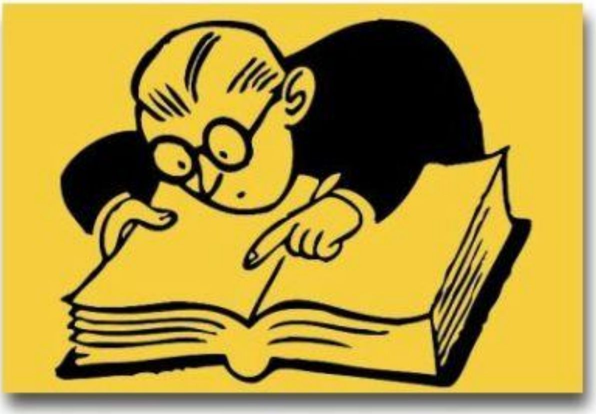
Definitions

Populations and Samples

From a **statistical** point of view:

✓ **Population**

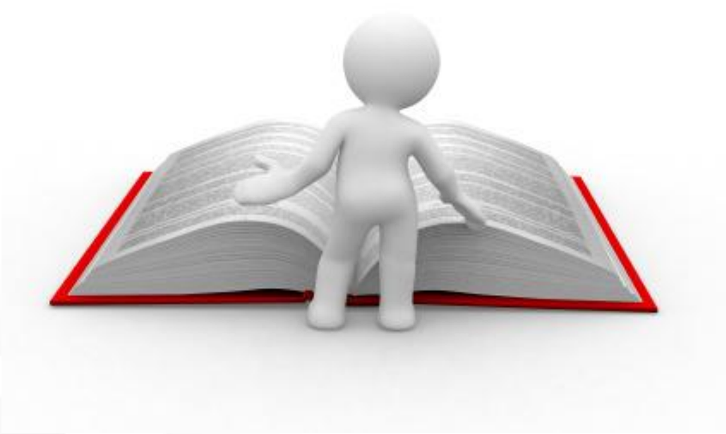
is a set of measurements, that may be finite or infinite, really existing or conceptual



✓ **Sample**

is a sub-set of measurements taken from the population and used to estimate population parameters

Definitions



Populations and Samples

From an **epidemiological** point of view:

- ✓ **Population**

  is a set of individuals (or a set of aggregates of individuals) that may be classified according to one or more homogeneous criteria (residence, production attitude, genetic structure, etc.)

- ✓ **Sample**

  is a sub-set taken from the population and used to gather information on the population itself

Populations and samples

Meanings and examples

STATISTIC.	EPIDEMIOL.	EXAMPLES
Population is a set of mea-surements ….	**Population** is a set of indi-viduals that may be classi-fied according to homoge-neous criteria	Antibody titers against a cer-tain disease in a cattle herd
that may be finite, ...		Animals in a herd
infinite, ...	(it exists only conceptually)	Italian cattle population

Populations and samples

Meanings and examples

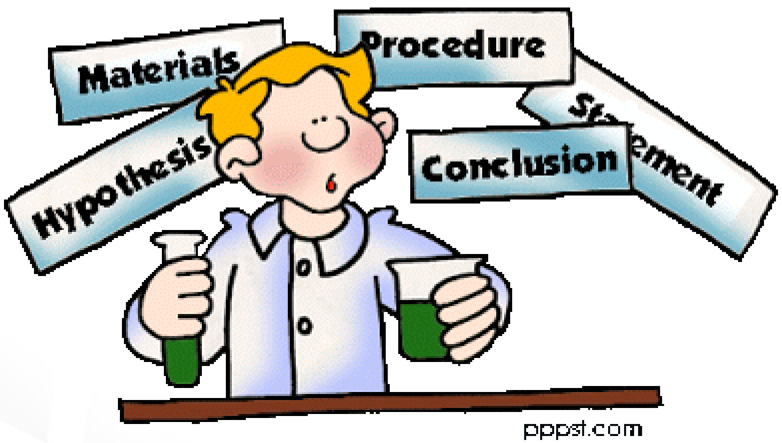
STATISTIC.	EPIDEMIOL.	EXAMPLES
really existing ...		The above mentioned cat-tle
or conceptual		The results of a series (poten-tially infinite) of administra-tions of a cer-tain drug to animals of a de-fined species

Populations and samples

Meanings and examples

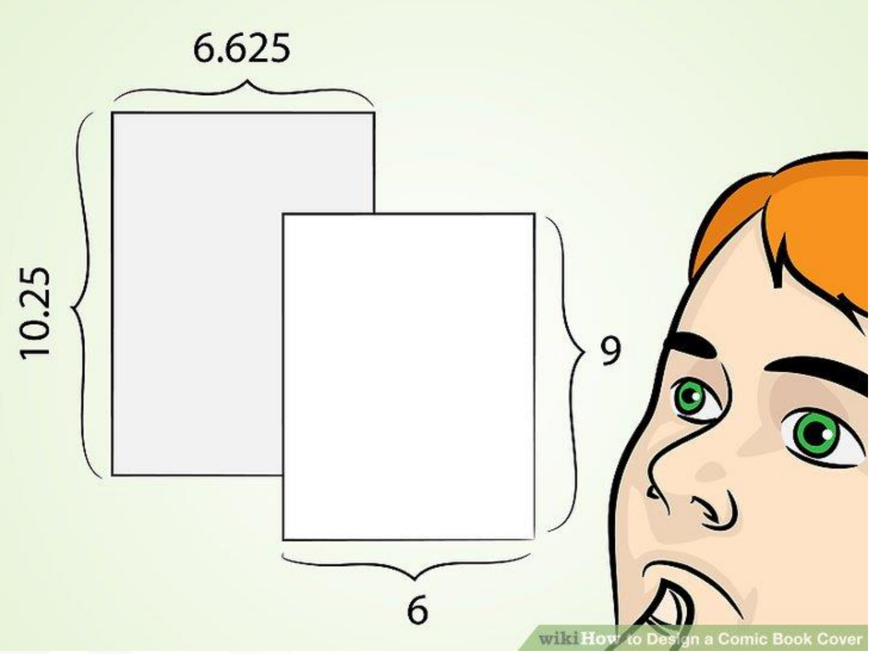
STATISTIC.	EPIDEMIOL.	EXAMPLES
Sample is a sub-set of measurements taken from the population and used to estimate population parameters	**Sample** is a sub-set taken from the population and used to gather information on the population itself	• 32 randomly selected animals • 10 females and 10 males randomly selected • Every tenth animal (arriving to an abattoir, a custom, etc.) • Reproductive stock (***)

Phases of sample design



- ✓ **Objectives** of sampling
- ✓ Target **population**
  - admission criteria
  - stratification criteria
- ✓ **Type** of sampling
- ✓ Sample **size**
- ✓ **Time** of sampling

Phases of sample design



- ✓ **Objectives** of sampling
- ✓ Target **population**
  - admission criteria
  - stratification criteria
- ✓ **Type** of sampling
- ✓ Sample **size**
- ✓ **Time** of sampling

Objectives

They should be:

- ✓ Explicit
- ✓ Clearly defined
- ✓ Actually achievable

Examples of objectives

<table>
<tr><th>UNCORRECT</th><th>CORRECT</th></tr>
<tr><td>To know which diseases exist in UAE</td><td rowspan="2">To know whether or not «disease X» is present in a camel population of UAE at a prevalence rate higher than a predefined threshold</td></tr>
<tr><td>To know whether or not «disease X» is present in camels of UAE</td></tr>
</table>

Examples of objectives

<table>
<tr><th>UNCORRECT</th><th>CORRECT</th></tr>
<tr><td>To measure the occurrence of «disease X» in the population of camels of UAE</td><td rowspan="2">Considering an acceptable level of error (standard error), and an approximate value of expected prevalence, to measure the serological prevalence of «disease X» in the population of camels of UAE</td></tr>
<tr><td>To measure the prevalence of «disease X» in the population of camels of UAE</td></tr>
</table>

Definition of objectives

The definition of objectives in the case of a data collection on sample basis should be more precise than in the case of data collection on the whole population because the planning of qualitative and quantitative aspects of sampling is strictly dependent on chosen objectives

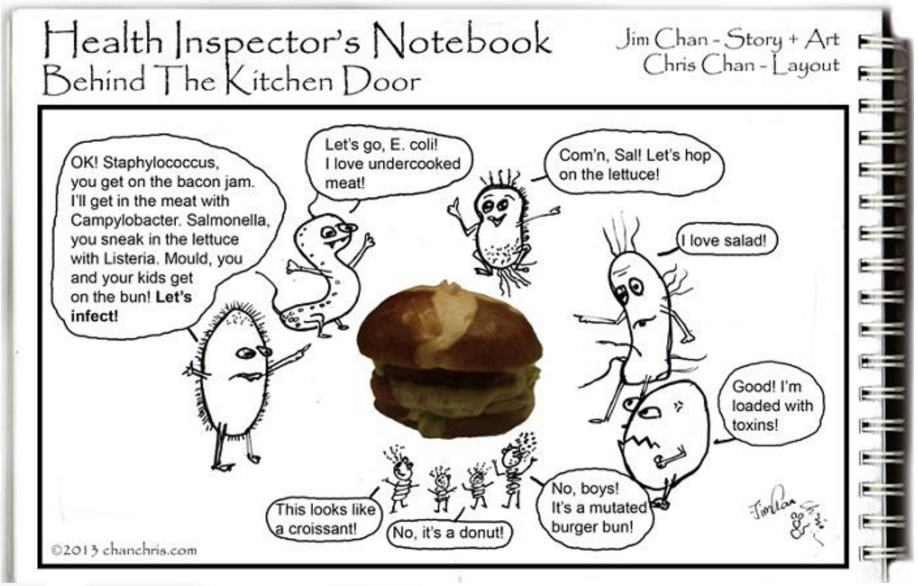
Definition of objectives

Examples

- Sample size is different in the case we need to detect whether a **disease is “not present”** in a certain population from the case we need to measure the **prevalence** of that disease in the same population

(continue)

Definition of objectives



Examples

- Target population is different in the case we need to investigate the **"risk for the consumer"** due to a certain disease/pollutant from the case we need to investigate the **"presence of that disease or pollutant"** in an animal population

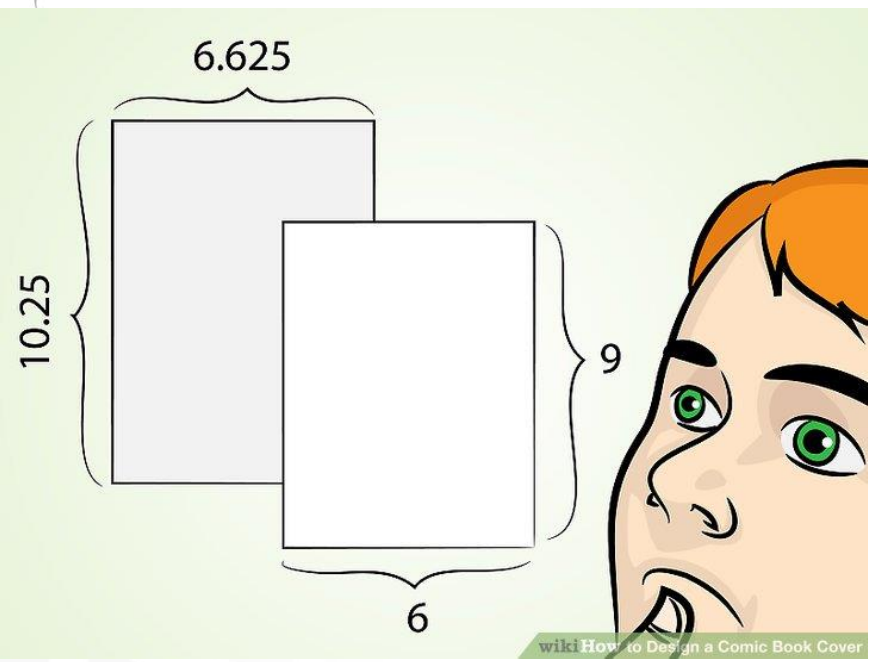
(continue)

Definition of objectives

Examples

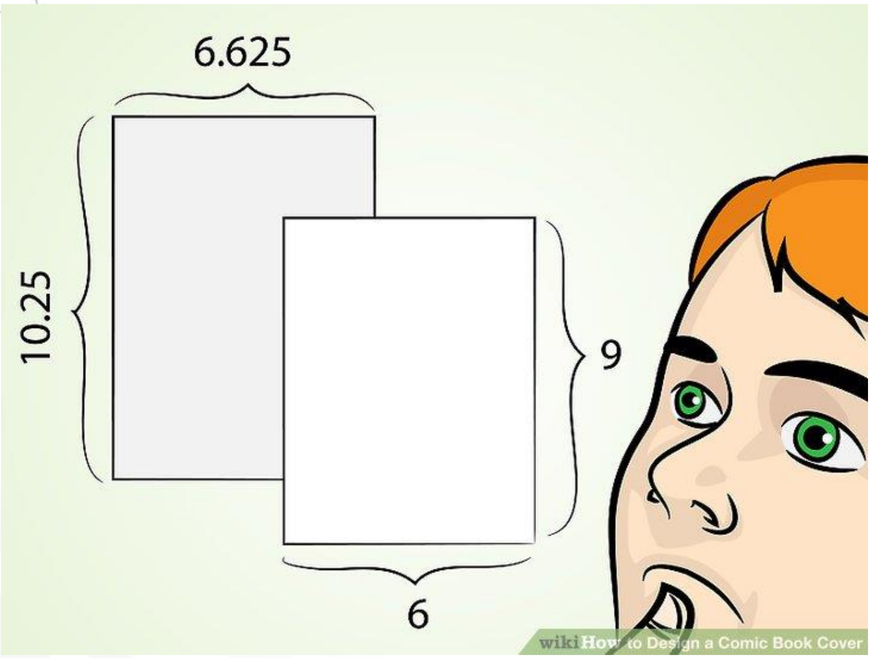
- The sizes of samples collected to estimate the prevalence of a certain disease differ, depending on the **expected prevalence** of that disease and on the **degree of precision** requested to the estimate itself

Phases of sample design



- ✓ **Objectives** of sampling
- ✓ Target **population**
  - admission criteria
  - stratification criteria
- ✓ **Type** of sampling
- ✓ Sample **size**
- ✓ **Time** of sampling

Phases of sample design



- ✓ **Objectives** of sampling
- ✓ Target **population**
  - admission criteria
  - stratification criteria
- ✓ **Type** of sampling
- ✓ Sample **size**
- ✓ **Time** of sampling

Target population

Admission criteria strictly depend on chosen objectives

...i.e....

Admission criteria are the criteria according to which each individual may or may not be included in the sample. They are a direct consequence of the objectives of sampling

Admission criteria

Example #1

Objective	Criterion
To estimate the risk for the consumer get brucellosis drinking raw camel milk	This objective needs to be split into three sub-objectives: 1. To estimate the average amount of raw camel milk in the diet 2. To estimate the prevalence of brucellosis in camel herds 3. To estimate the level of *Brucella* concentration in camel milk

Admission criteria

Example #1

Objective	Criterion
1. To estimate the average amount of raw camel milk in the diet	The whole human population of the country, subdivided by age groups, gender and other possible relevant conditions (such as rural vs urban population, etc.)

Objective	Criterion
2. To estimate the prevalence of brucellosis in camel herds	All adult female camels during the milking periods

Objective	Criterion
3. To estimate the level of *Brucella* concentration in camel milk	Milk produced in infected herds

Admission criteria

Example #2

Objective	Criterion
To determine the probability that RVF infection is introduced into EAU through imported camels	Two objectives: 1. Probability that one or more RVF infected camels are introduced into EAU in the unit of time 2. Probability that a RVF infected camel will be not detected by the import controls

Admission criteria

Example #2

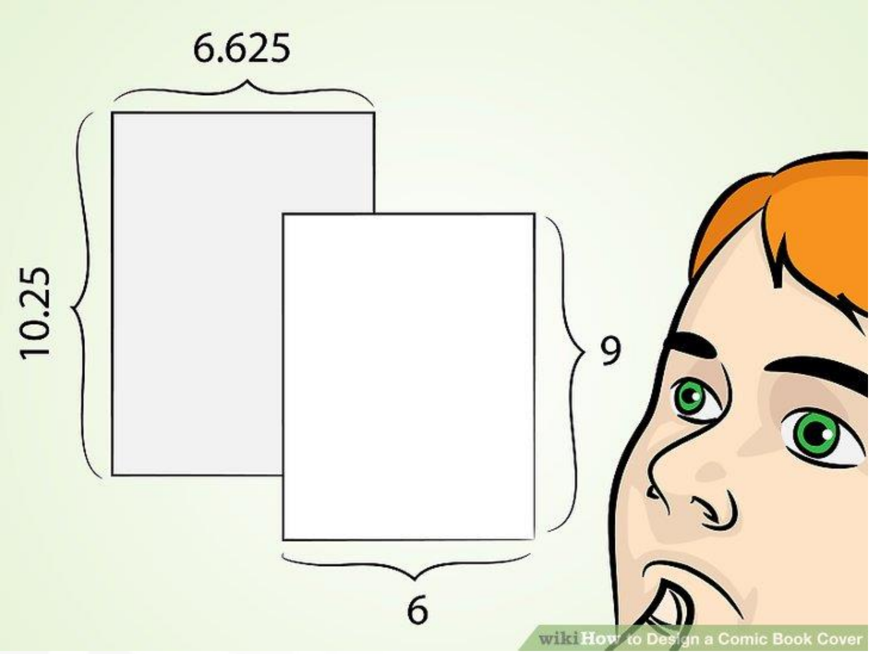
Objective	Criterion
1. Probability that one or more RVF infected camels are introduced into EAU in the unit of time	All (not vaccinated against RVF) camels imported from RVF infected countries during the vector season

Admission criteria

Example #2

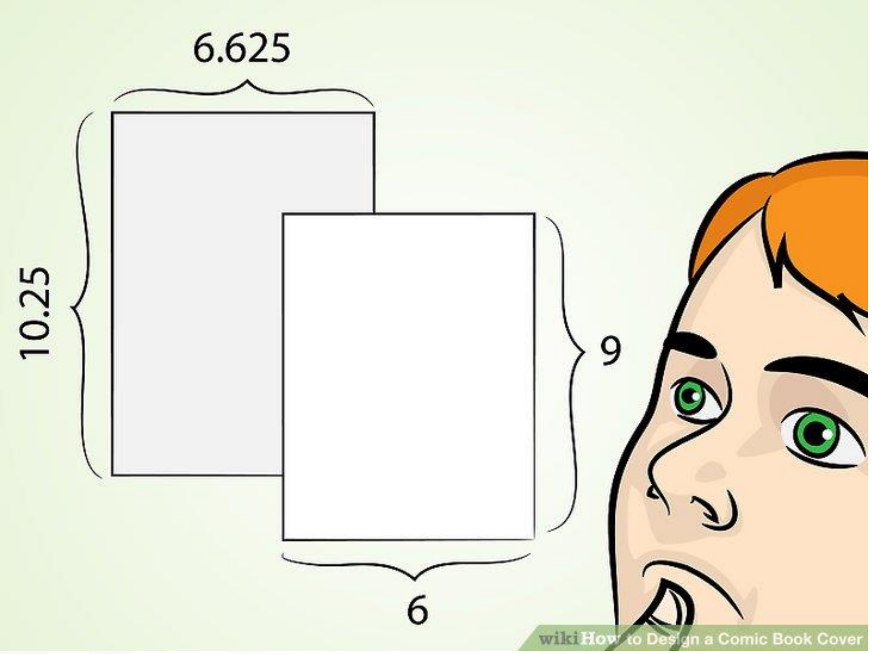
Objective	Criterion
2. Probability that a RVF infected camel will be not detected by the import controls	All imported RVF infected camels subjected to import controls

Phases of sample design



- ✓ **Objectives** of sampling
- ✓ Target **population**
  - admission criteria
  - stratification criteria
- ✓ **Type** of sampling
- ✓ Sample **size**
- ✓ **Time** of sampling

Phases of sample design



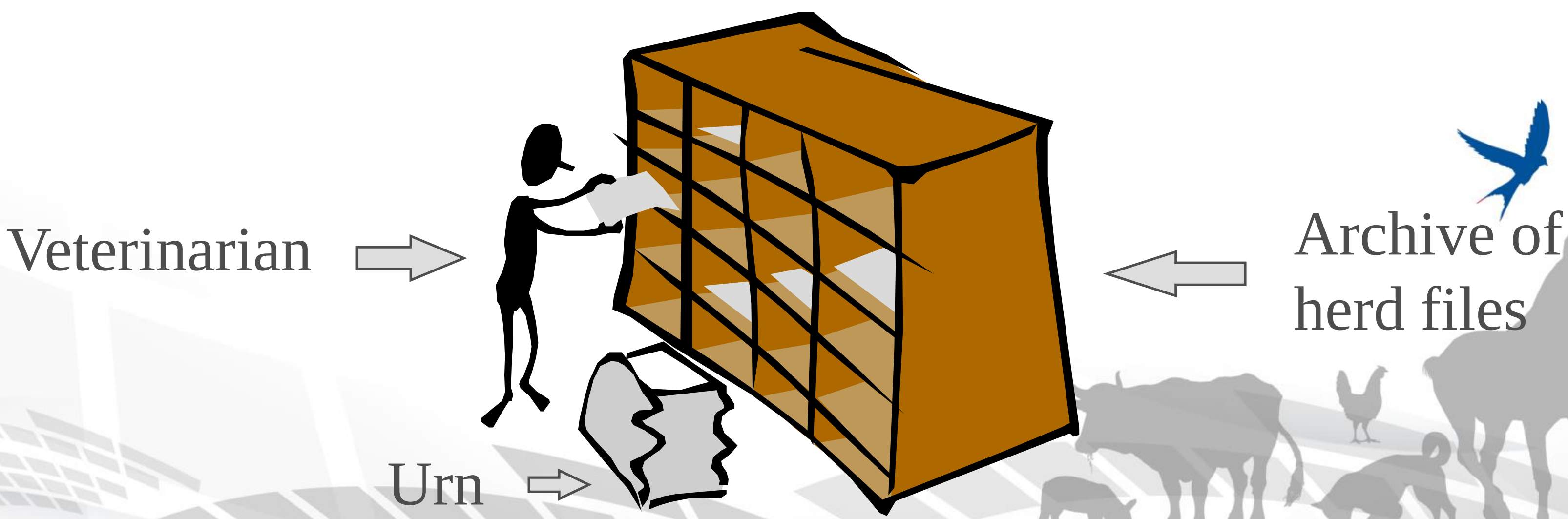
- ✓ **Objectives** of sampling
- ✓ Target **population**
  - admission criteria
  - stratification criteria
- ✓ **Type** of sampling
- ✓ Sample **size**
- ✓ **Time** of sampling

Type of sampling

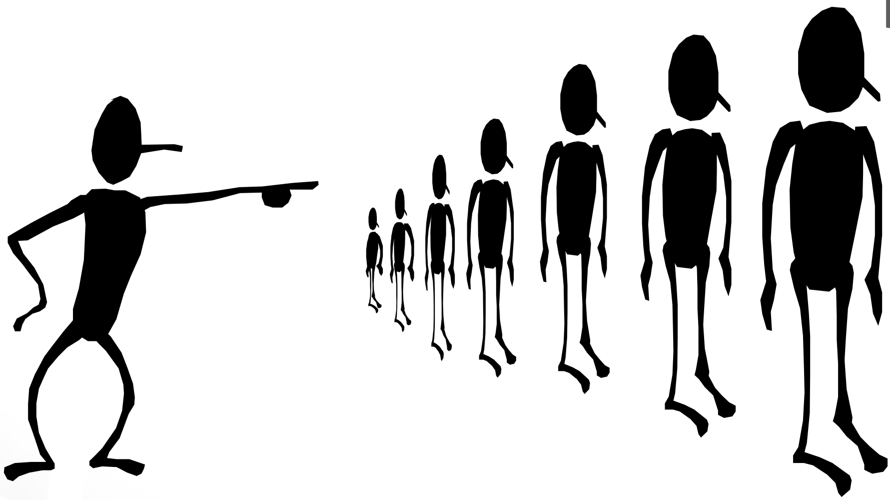
- ✓ Simple random sampling
- ✓ Systematic random sampling
- ✓ Stratified sampling
- ✓ Multistage sampling

Systematic random sampling

It is obtained by putting in a urn a set of tags (one for each element of the population) with the ID numbers of each element (e.g. eartag number) and taking from the urn the desired number of tags. It may also be obtained using random number tables or using a computer



Systematic random sampling



We need to extract a sample of **n** elements from a population of size **N**:

- a sampling step is defined: **k=N/n**
- the first animal to be sampled is randomly selected between 1 and k
- and every **k**$^{\mathbf{th}}$ animal starting from the first selected animal is introduced in the sample

IZSAM G.CAPORALE TERAMO

OIE Collaborating Centre Veterinary Epidemiology

Systematic random sampling

Advantages

- We do not need to know each element of the population, the only needed piece of information is population size
- It is suitable, for example, in quality control or when population units manifest themselves in a time sequence (e.g. in a laboratory, slaughterhouse, custom office, etc.)

Stratified sampling

- Population is subdivided into strata (sub-groups) and a simple or systematic random sample is taken from each stratum
- Characters chosen as strata (age, sex, breed, geographical origin, herd size, etc.) must be relevant for the investigation

Stratified sampling

Advantages

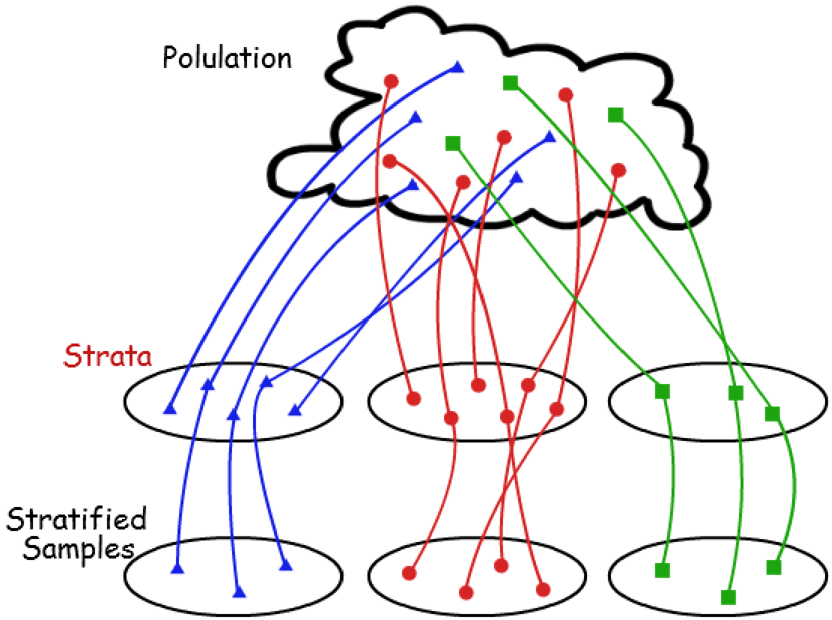
- It allows to perform a sampling that is representative of each stratum in the population, even in the case of very small sub-groups
- It allows (in the case of non proportional stratified sampling) to draw also inferences on single strata of the population
- It reduces sampling variance (increase in precision of estimates)

Stratified sampling

Disadvantages

Effects of stratification features need to be known in advance

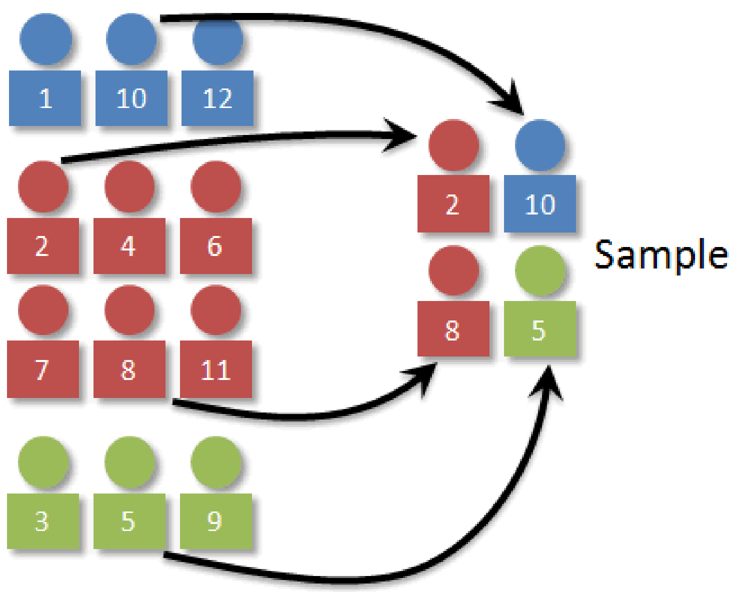
Stratified sampling



Two types of Stratified sampling exist:

- Proportional s.s.
- Non-proportional s.s.

Stratified sampling



Examples

We need to sample 60 sheep from a population of 7800, subdivided in 4 flocks.

✓ **Proportional stratified sampling**:

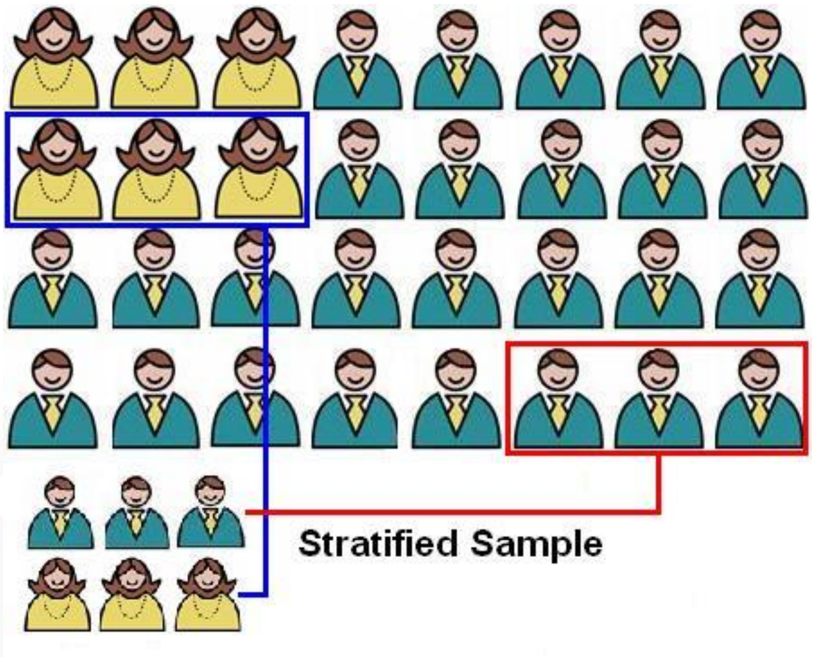
from each flock a sample is taken proportionally to the weight of that flock relative to total population

Stratified sampling

Examples

I flock	3000	60*3000:7800=	23
II flock	800		6
III flock	2500		19
IV flock	1500		12
TOTAL	7800		60

Stratified sampling



Examples

We need to sample 60 sheep from a population of 7800, subdivided in 4 flocks.

- ✓ **Non-proportional stratified sampling:**

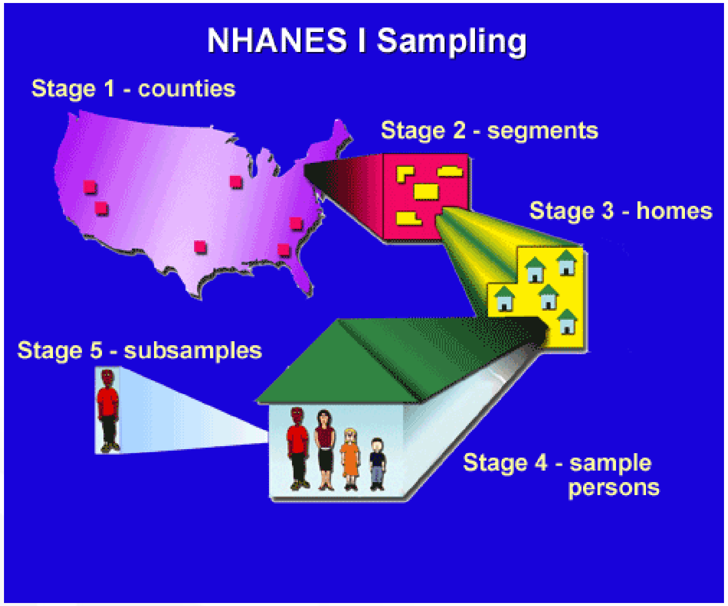
  from each flock an equal sample is taken (e.g. to have a pre-defined and equal precision of the estimates for each stratum of the population)

Stratified sampling

Examples

I flock	3000	15
II flock	800	15
III flock	2500	15
IV flock	1500	15
TOTAL	7800	60

Multi-stage sampling



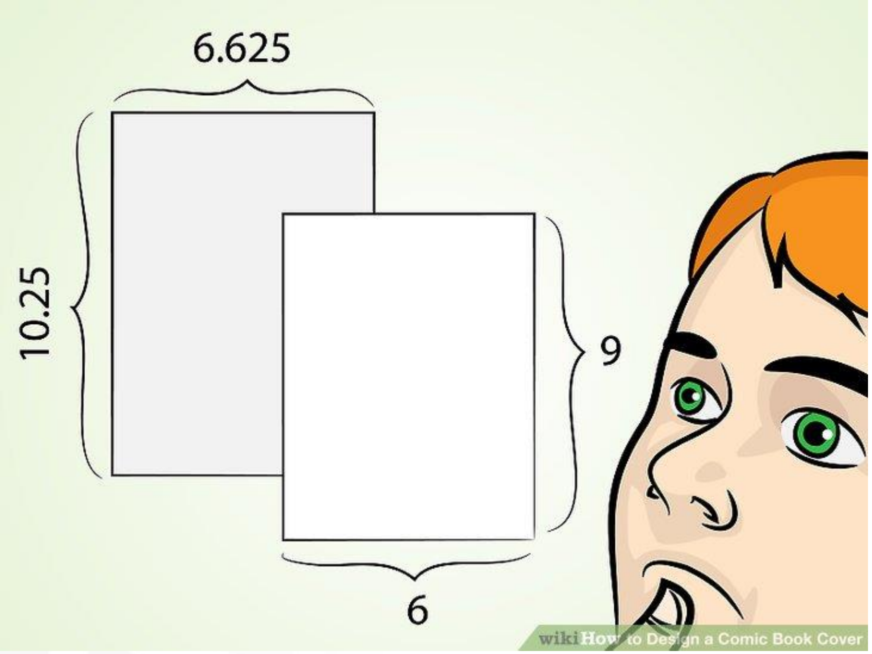
It is a combination of the previous methods

The sample is selected in two or more phases

Example:

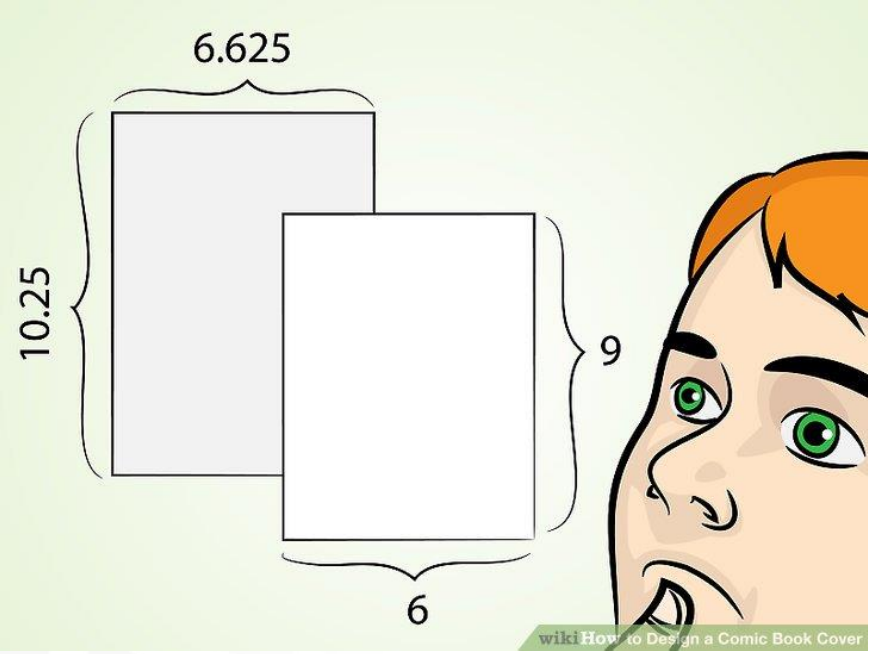
- a sample of dairy herds is randomly selected and then
- a sample of heifers is randomly selected in each selected herd

Phases of sample design



- ✓ **Objectives** of sampling
- ✓ Target **population**
  - admission criteria
  - stratification criteria
- ✓ **Type** of sampling
- ✓ Sample **size**
- ✓ **Time** of sampling

Phases of sample design



- ✓ **Objectives** of sampling
- ✓ Target **population**
  - admission criteria
  - stratification criteria
- ✓ **Type** of sampling
- ✓ Sample **size**
- ✓ **Time** of sampling

Sample size

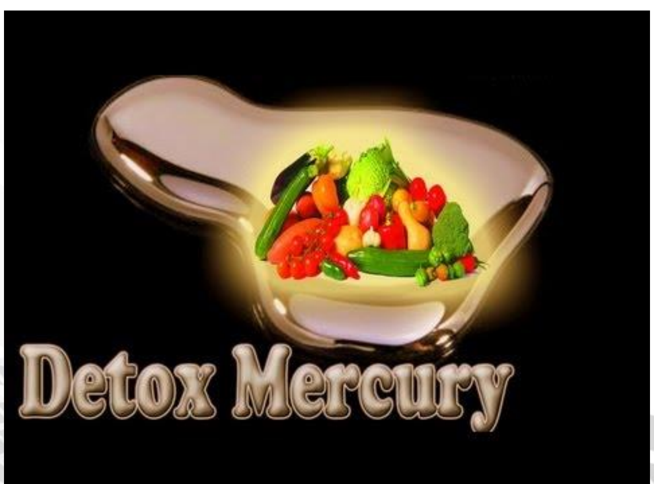
We need to consider two different types of variables to be studied:

✓ **Dichotomous variables**

- Presence/absence of a disease
- Presence/absence of residues or pollutants

IZSAM G.CAPORALE TERAMO

Sample size

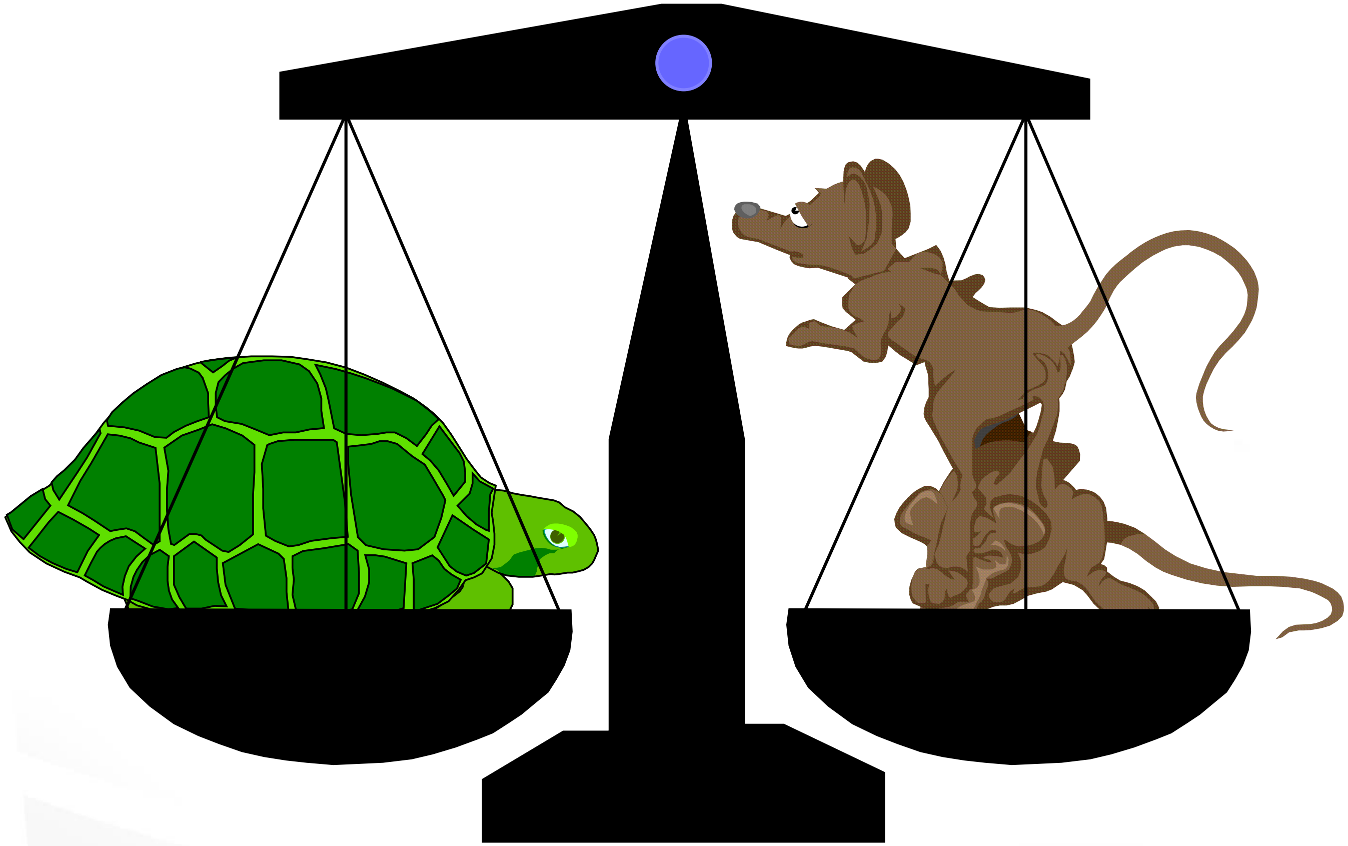


✓ **Continuous variables**

- Intake of mercury through the diet
- Concentration of anti-microbial drugs in food

NOTE: in both cases, dichotomization may be performed based on defined thresholds, but it would produce a loss of information

Dichotomous variables



Sample size to detect a disease in a population with a threshold prevalence P

POPULATION SIZE	0.50%	1.00%	2.00%	5.00%	10.00%
20	20	20	20	19	15
50	50	50	48	34	22
100	100	95	77	44	25
250	227	174	112	52	27
500	349	224	128	55	28
1000	450	258	138	57	28
2000	517	277	143	58	28
10000	580	294	147	58	28
INFINITE	598	298	148	58	28

Confidence level = 95%

Sample size to detect a disease in a population with a threshold prevalence P

POPULATION SIZE	0.50%	1.00%	2.00%	5.00%	10.00%
20	20	20	20	20	18
50	50	50	50	41	29
100	100	99	90	59	35
250	244	210	149	75	40
500	420	300	183	82	42
1000	601	367	204	86	43
2000	736	409	215	88	43
10000	878	448	225	89	44
INFINITE	919	458	228	90	44

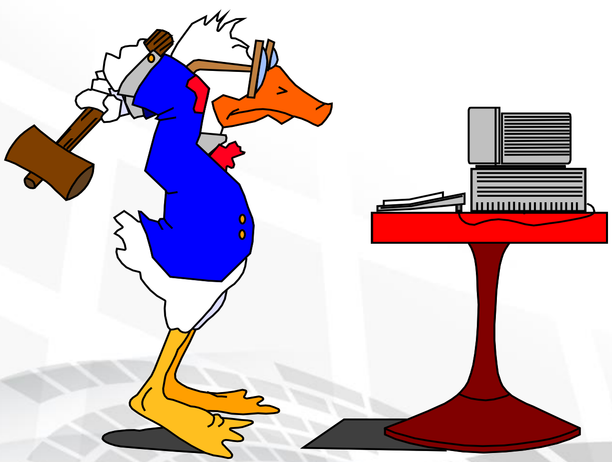
Confidence level = 99%

Finite populations:

$$n = \left[1-(1-\alpha)^{\frac{1}{D}}\right] * \left[N - \frac{(D-1)}{2}\right]$$

D=number of animals infected in the population
N=total number of animals in the population
α=confidence level

Infinite populations:



$$n = \frac{\log(1-\alpha)}{\log(1-p)}$$

α=confidence level
p=prevalence of infection

Calculation of the sample size

IZSAM G.CAPORALE TERAMO

OIE Collaborating Centre Veterinary Epidemiology

- For infinite populations in may be easily calculated using a spreadsheet like Excel:
- given the threshold prevalence p, for example 1%=0.01
- and the confidence level, for example 95%=0.95
- sample size may be calculated as

  =log(1-0.95)/log(1-0.01) and the result is 298

Collaborating Centre
Veterinary Epidemiology

Sample size needed to estimate a prevalence of infection with confidence level = 95%

Infinite population

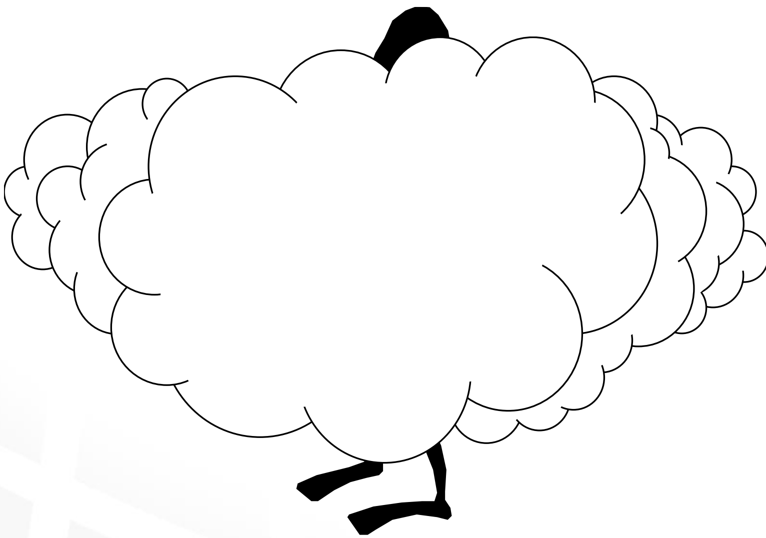
EXPECTED PREVALENCE IN THE POPULATION	DESIRED LEVEL OF PRECISION				
	1,00%	2,00%	5,00%	8,00%	10,00%
5%	1825	456	73		
10%	3457	864	138	54	35
15%	4898	1225	196	77	49
20%	6147	1537	246	96	61
40%	9220	2305	369	144	92
50%	9604	2401	384	150	96

Sample size needed to estimate a prevalence of infection with confidence level = 99%

Infinite population

EXPECTED PREVALENCE IN THE POPULATION	DESIRED LEVEL OF PRECISION 1,00%	2,00%	5,00%	8,00%	10,00%
5%	3150	787	126		
10%	5968	1492	239	93	60
15%	8454	2114	338	132	85
20%	10609	2652	424	166	106
40%	15914	3978	637	249	159
50%	16577	4144	663	259	166

Calculation of the sample size



From: $$n = \frac{z\alpha^2 * p * (1-p)}{E^2}$$

p=expected prevalence

E=accepted level of error

za=1.96 for 95% c.l.

za=2.575 for 99% c.l.

Calculation of the sample size

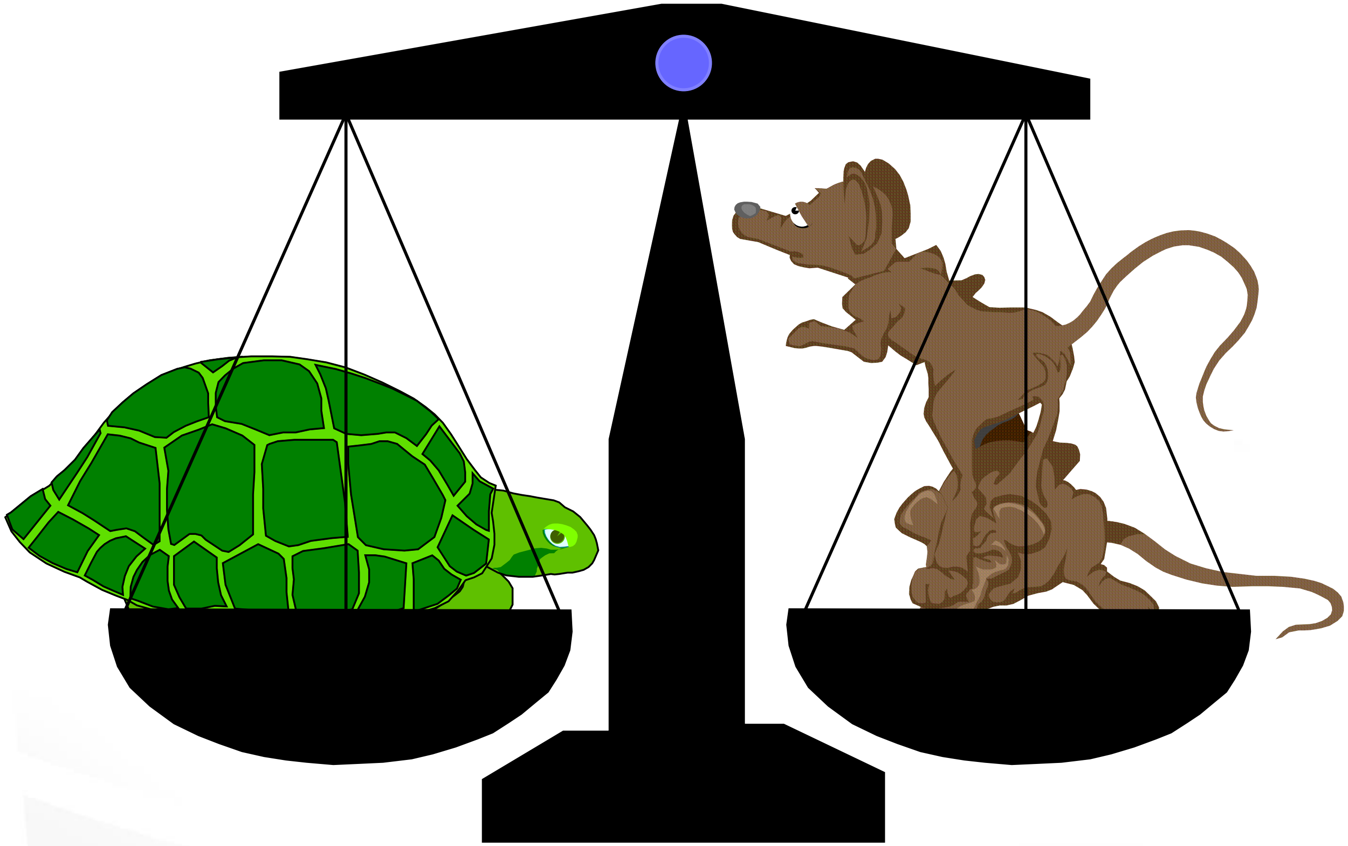
For finite populations, calculate the sample size using the following formula:

$$n_f=(n_i*N)/(n_i+N)$$

Where:

- n_f = sample size for **finite** population
- n_i = sample size for **infinite** population (from the table)
- N = size of the total population

Continuous variables



Calculation of the sample size

Sample size for estimating mean value of a continuous variable					
Confidence level	Precision of estimation				
	±10%	±20%	±30%	±50%	±100%
95%	384	96	43	15	4
99%	697	174	77	28	7

$$n = \left(\frac{1,96 * \sigma}{E} \right)^2$$

Confidence level = 95%

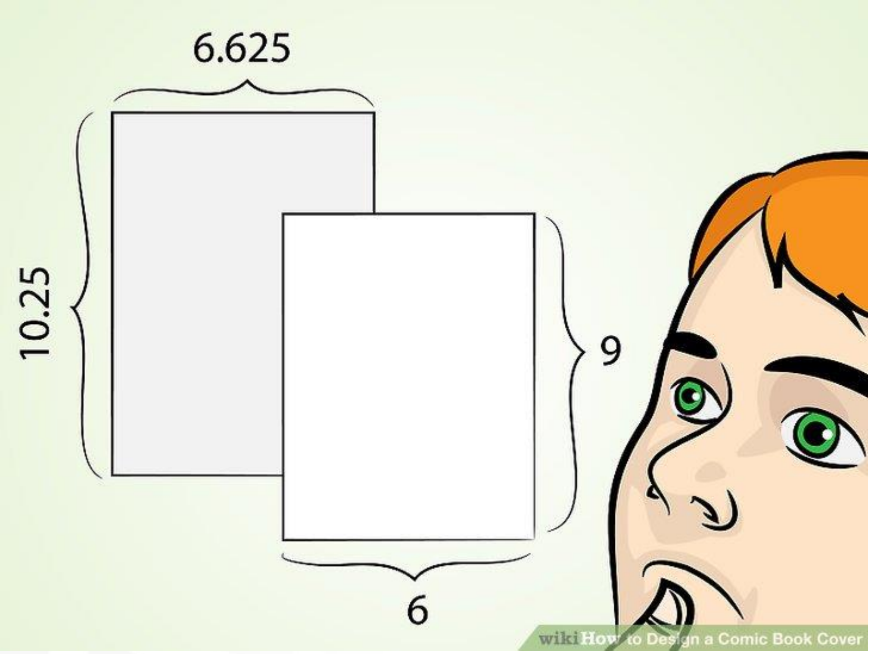
$$n = \left(\frac{2,64 * \sigma}{E} \right)^2$$

Confidence level = 99%

E=accepted level of error

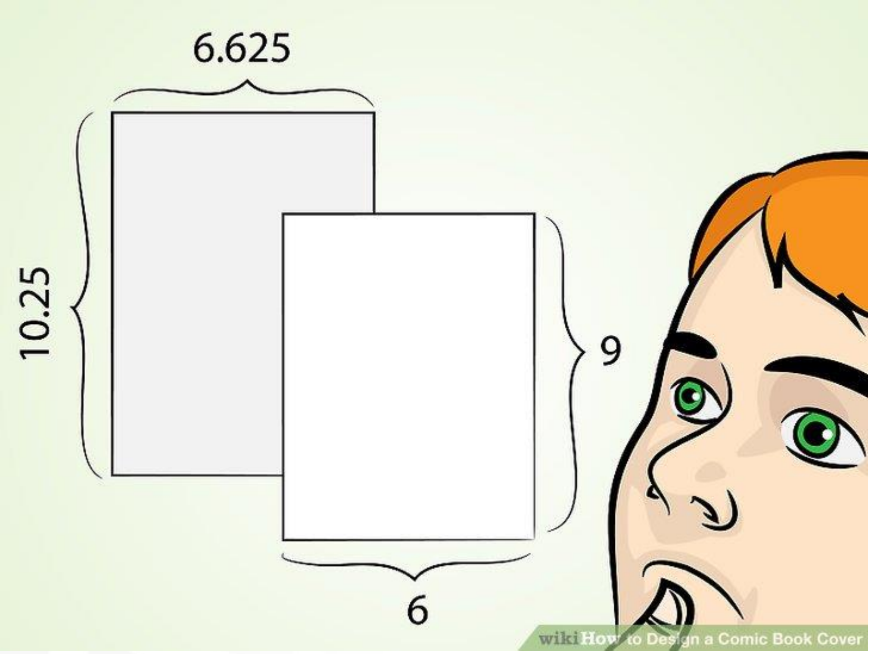
σ=standard deviation of variable's distribution

Phases of sample design



- ✓ **Objectives** of sampling
- ✓ Target **population**
  - admission criteria
  - stratification criteria
- ✓ **Type** of sampling
- ✓ Sample **size**
- ✓ **Time** of sampling

Phases of sample design



- ✓ **Objectives** of sampling
- ✓ Target **population**
  - admission criteria
  - stratification criteria
- ✓ **Type** of sampling
- ✓ Sample **size**
- ✓ **Time** of sampling

Time and duration of sampling

They depend on:

- Objective of sampling
- Type of study
  - transversal
  - longitudinal
- Seasonality of the studied phenomenon, which depends on:
  - host biology
  - parasite biology
  - husbandry habits and traditions
- Logistic constraints

Final remarks

In case of rare events sample surveys are not suitable

Examples of rare events are:

- *Exotic diseases or newly introduced ones*
- *Diseases in final phases of eradication programs, when prevalence approaches zero*
- *Diseases rare for their own nature (e.g. BSE)*

For this reason, early warning systems cannot be based on random sampling

Final remarks

- In all these cases, clinical ("passive") surveillance is more effective
- Even if "passive" surveillance has a low sensitivity, this low sensitivity is compensated by the large portion of the population that it can check
- Therefore, in case of rare events, the most cost-effective way to deal with problems is:
  - *Passive surveillance*
  - *Risk-based surveillance which is a targeted (not random) active surveillance targetting those population strata with higher probability of being exposed and infected*

IZSAM G.CAPORALE TERAMO

OIE Collaborating Centre Veterinary Epidemiology

Thank you for your attention!