

# The Online R-FETPV 1st Module : Basic Epidemiology and Surveillance Data Analysis

5 April -28 May 2021



**USAID**  
FROM THE AMERICAN PEOPLE



**Food and Agriculture  
Organization of the  
United Nations**



# Session Title: Descriptive statistics

**Sangeeta Rao**, BVSc, MVSc, PhD  
Colorado State University



**USAID**  
FROM THE AMERICAN PEOPLE



**Food and Agriculture  
Organization of the  
United Nations**



# Outline:

- Epidemiological Data
- Subjects and Variables
- Data Types
- Precision and Validity of data
- Measures to describe continuous epidemiological data
- Measures to describe categorical epidemiological data

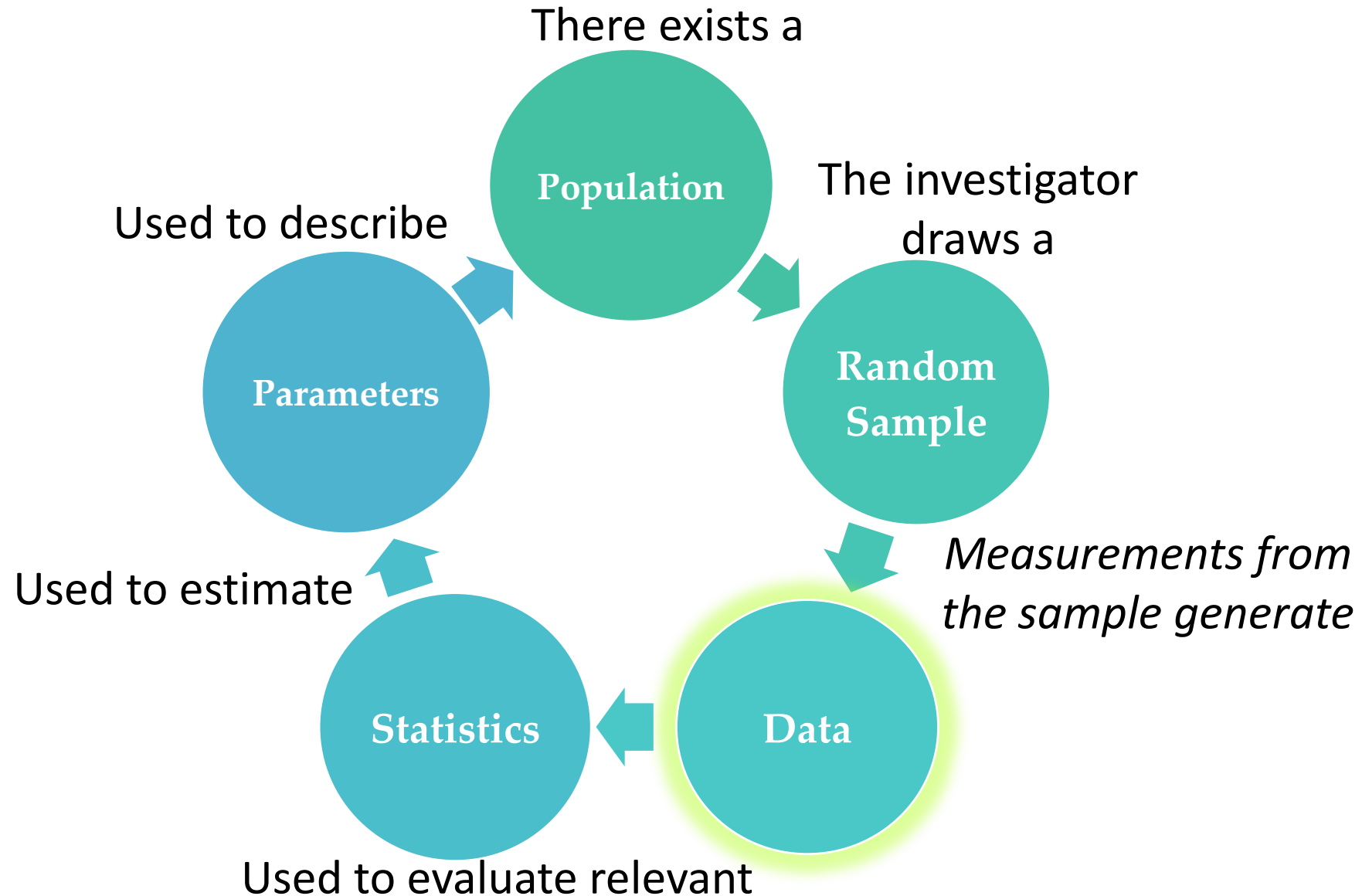
# Epidemiological field scenario:

You have been noticing a higher occurrence of cows with abortions during their last term of pregnancy. You know that some new animals were purchased in your jurisdiction including bulls many months ago. You sample milk and serum from few of the affected animals to detect the cause. As you suspected, the laboratory results (with ELISA test) were positive for *Brucella abortus*.



123rf.com

# Epidemiological study summary



# Outline:

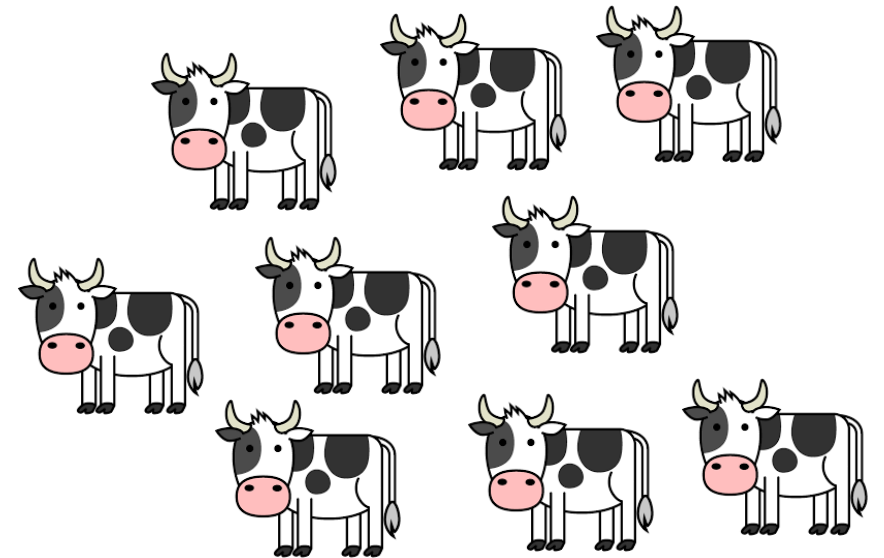
- Epidemiological Data
- Subjects and Variables
- Data Types
- Precision and Validity of data
- Measures to describe continuous epidemiological data
- Measures to describe categorical epidemiological data

# Data

- Data
  - *measurements* or *values* of a characteristic of interest.

## Data ≠ Information

- Information is generated by “working” with data
  - Group data
  - Summarize data
- Information, not data, is used to make decisions.



# Individuals and variables

(= observations)

*Subjects/ Individuals* are the objects described by a set of data.

Subjects may be animals, people, or things.

- Jersey cows, Golden retrievers, fields of corn, bacterial cells etc.

A *variable* is any characteristic of an individual. A characteristic/variable can take different values for different individuals.

- Age, gender, disease status, blood pressure, blood type, temperature, eye color



Association

Risk factor/  
predictor

Disease outcome  
in the population

Requires Collection  
of data

Description of data

# Outline:

- Epidemiological Data
- Subjects and Variables
- **Data Types**
- Precision and Validity of data
- Measures to describe continuous epidemiological data
- Measures to describe categorical epidemiological data

Disease outcomes or factors [Data] → Quantitative

→ Qualitative/  
Categorical

## Quantitative data

### Examples

Blood counts

Body weight

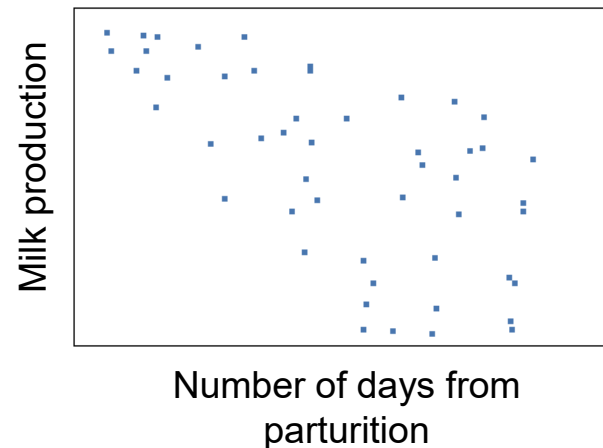
Blood pressure

Cfu

Body temperature

Age

Litter size

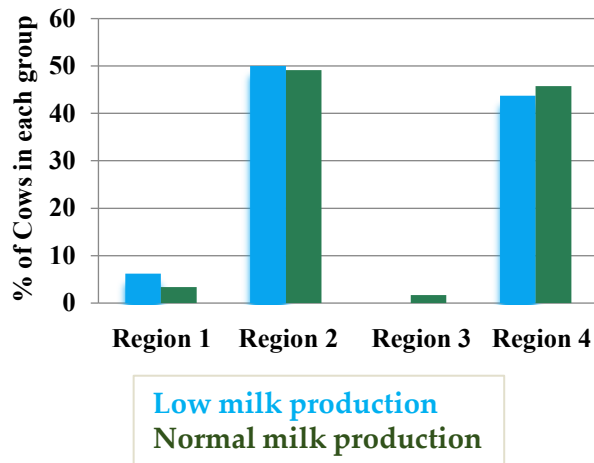


# Disease outcomes or factors[Data]

## Qualitative/ Categorical data

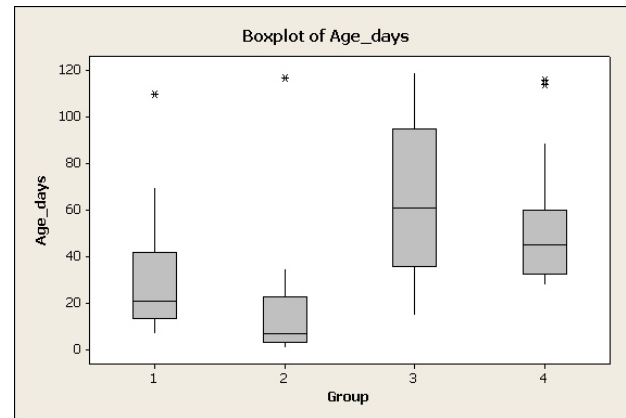
### Binary

Fever yes/ no  
Seropositive yes/ no  
Disease yes/no



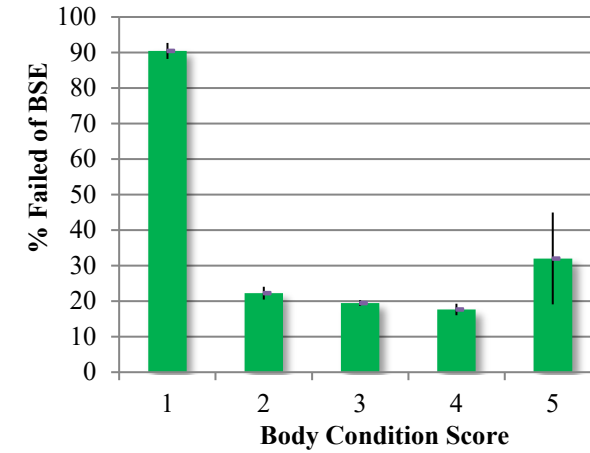
### Nominal

Tx groups  
Color of urine  
Color of mucus membrane  
Gender



### Ordinal

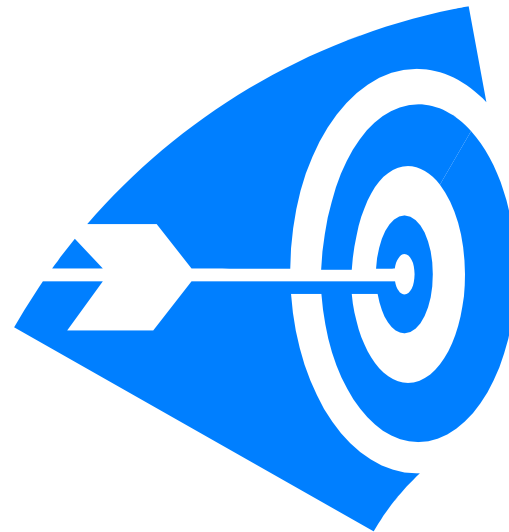
Stages of cancer  
Clinical Score  
Body condition score



# Outline:

- Epidemiological Data
- Subjects and Variables
- Data Types
- **Precision and Validity of data**
- Measures to describe continuous epidemiological data
- Measures to describe categorical epidemiological data

# Precision and validity of the data



# Example:

- You measure the body temperature of a cow 3 times consecutively using the same thermometer and get slightly different values: 38, 38.2, 38.3°C.

What could be the reason?



<https://www.twi.co.uk>

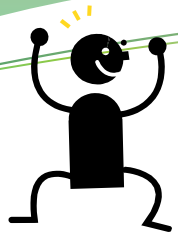
## Quality of Measurements

### Accuracy (or Validity)

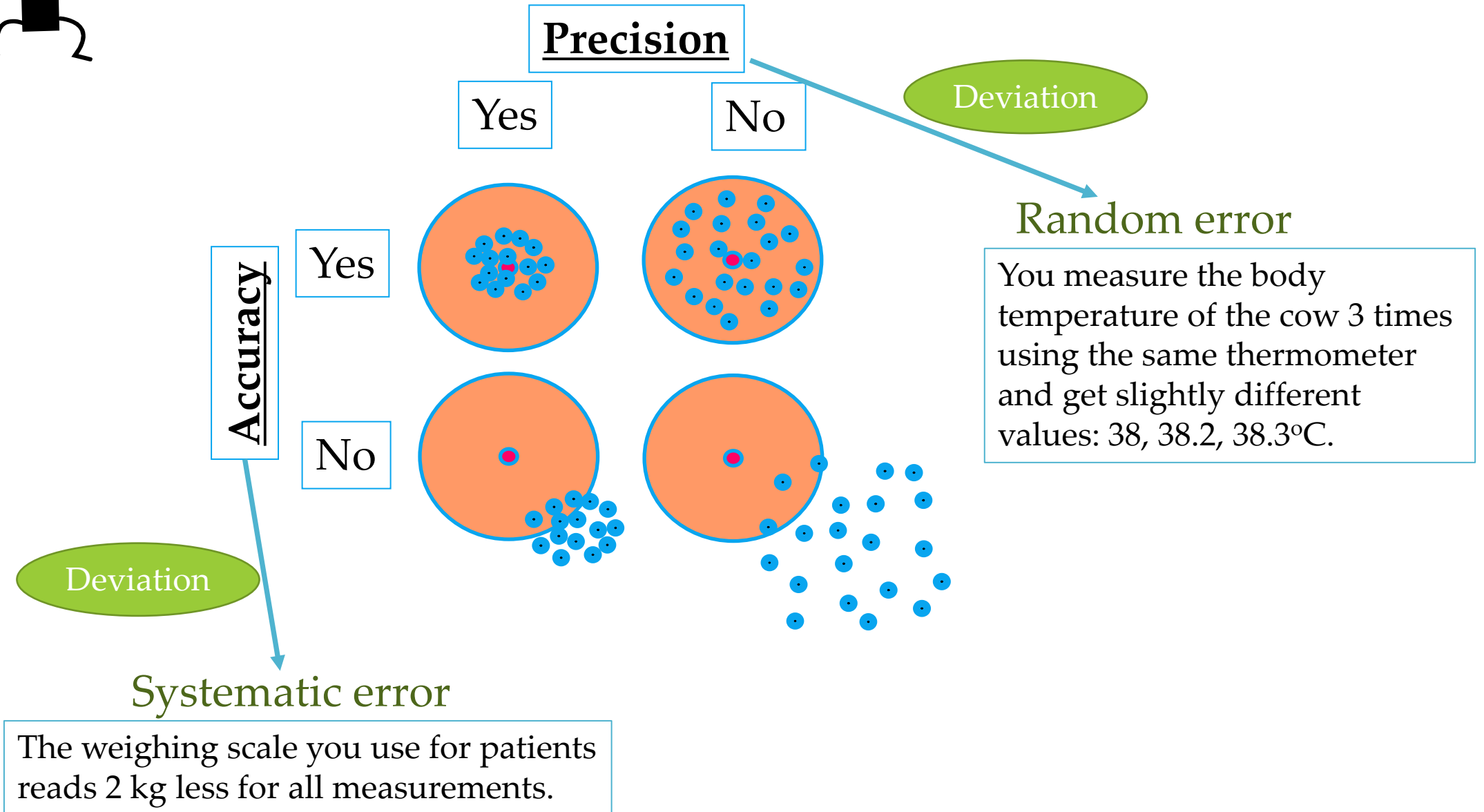
Observations or measurements represent the true status of characteristic

### Precision (or repeatability, reliability)

Repeated measurements yield similar results



# Quality of Measurements





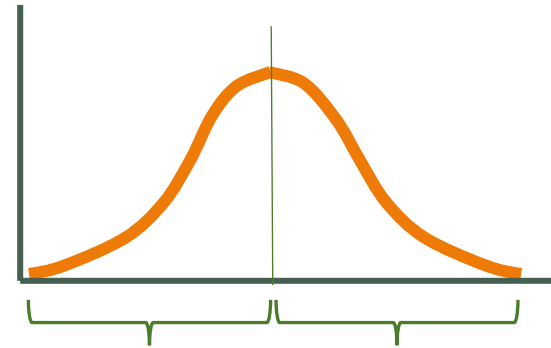
# Outline:

- Epidemiological Data
- Subjects and Variables
- Data Types
- Precision and Validity of data
- **Measures to describe continuous epidemiological data**
- Measures to describe categorical epidemiological data

# Describing quantitative data (Continuous)

---

- Measures of the center of data:
  - Mean
  - Median
  - Mode
- Measures of dispersion
  - The range
  - The interquartile range
  - The variance
  - The standard deviation



# Measures of Central Tendency

## Mean:

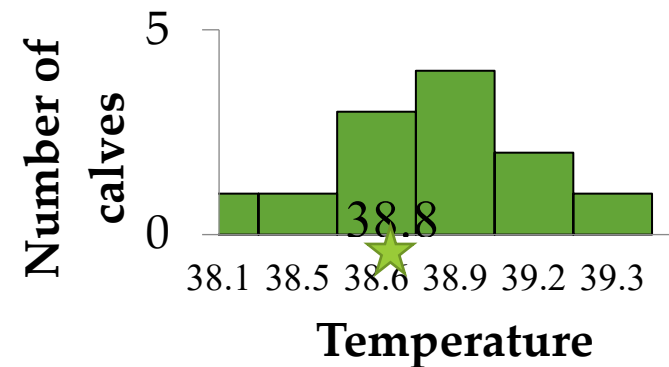
To calculate add all values, then divide by the number of individuals.

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Temperature (° C) in 12 calves

CALF	TEMP °C
1	39.2
2	39.3
3	38.1
4	38.6
5	38.6
6	38.9
7	38.9
8	38.9
9	38.9
10	39.2
11	38.6
12	38.5

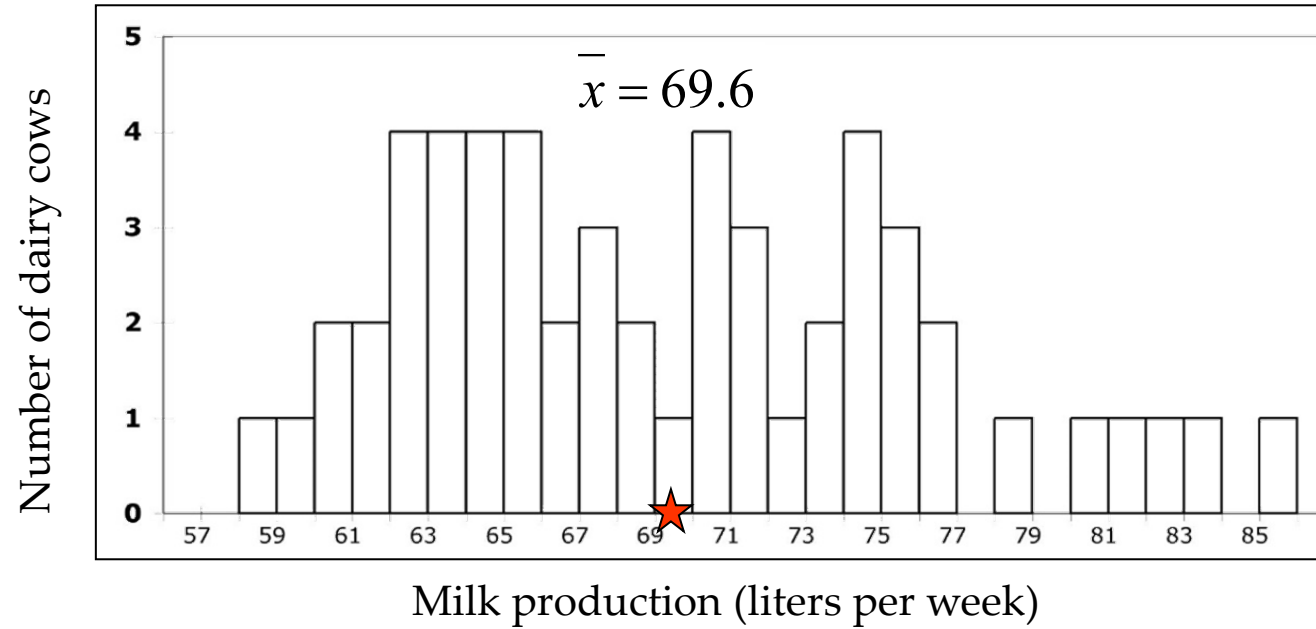
$\Sigma=465.7$



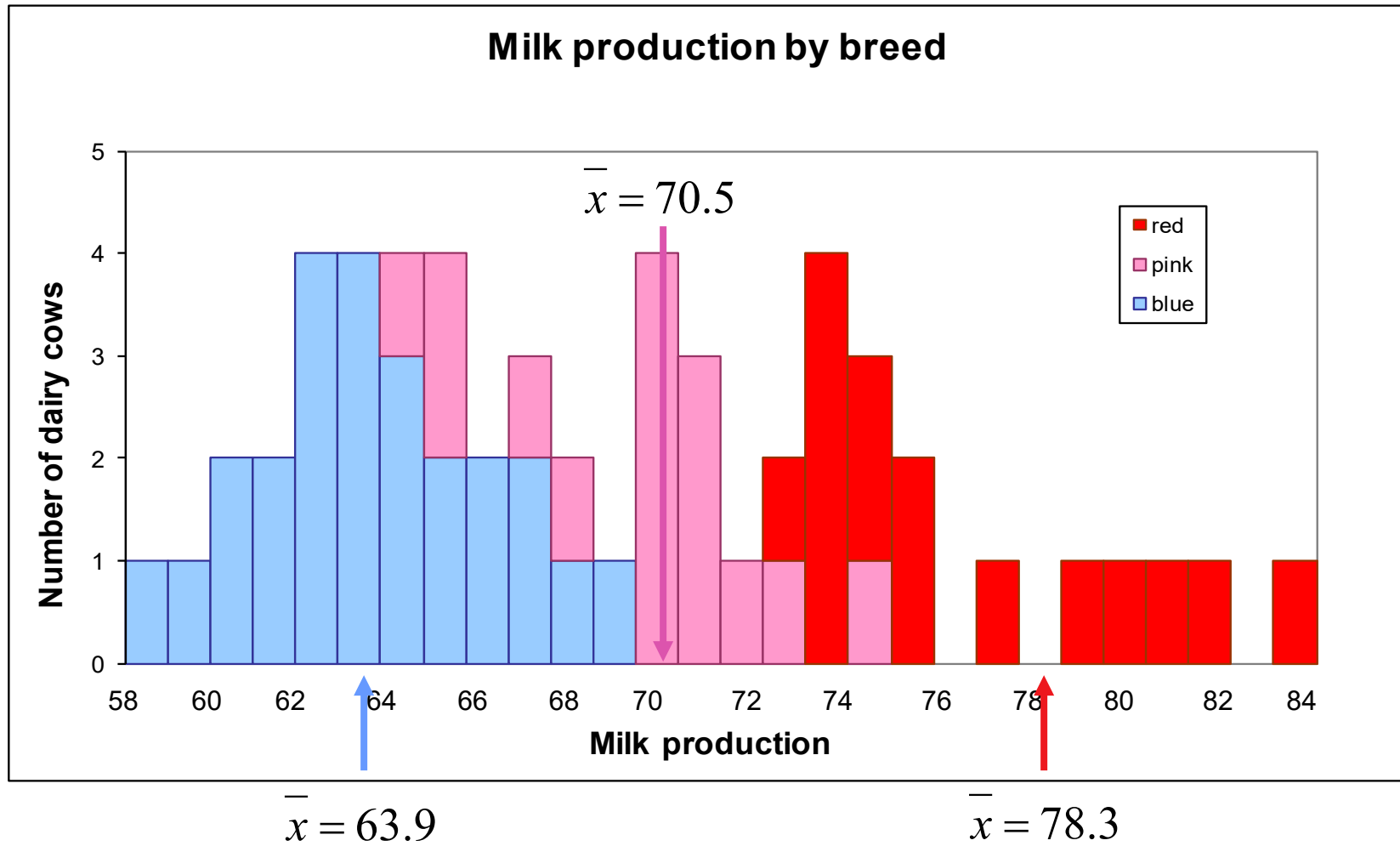
The mean is a good numerical summary.

$$\bar{X} = \frac{465.7}{12} = 38.8$$

Another example:



Here the shape of the distribution is wildly irregular. Why?



A single numerical summary here would not make sense.



# Measures of Central Tendency

- **Mode** = Most frequent value

1

3

4

6

8

8

9

9

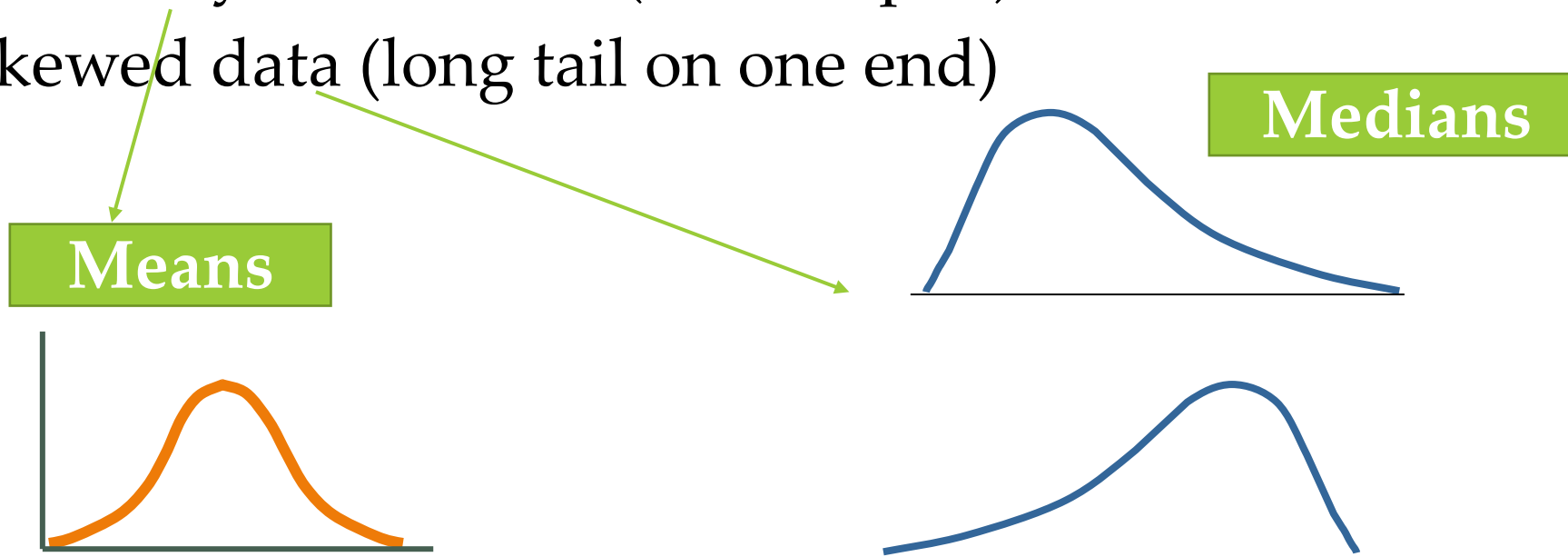
9

Mode = 9

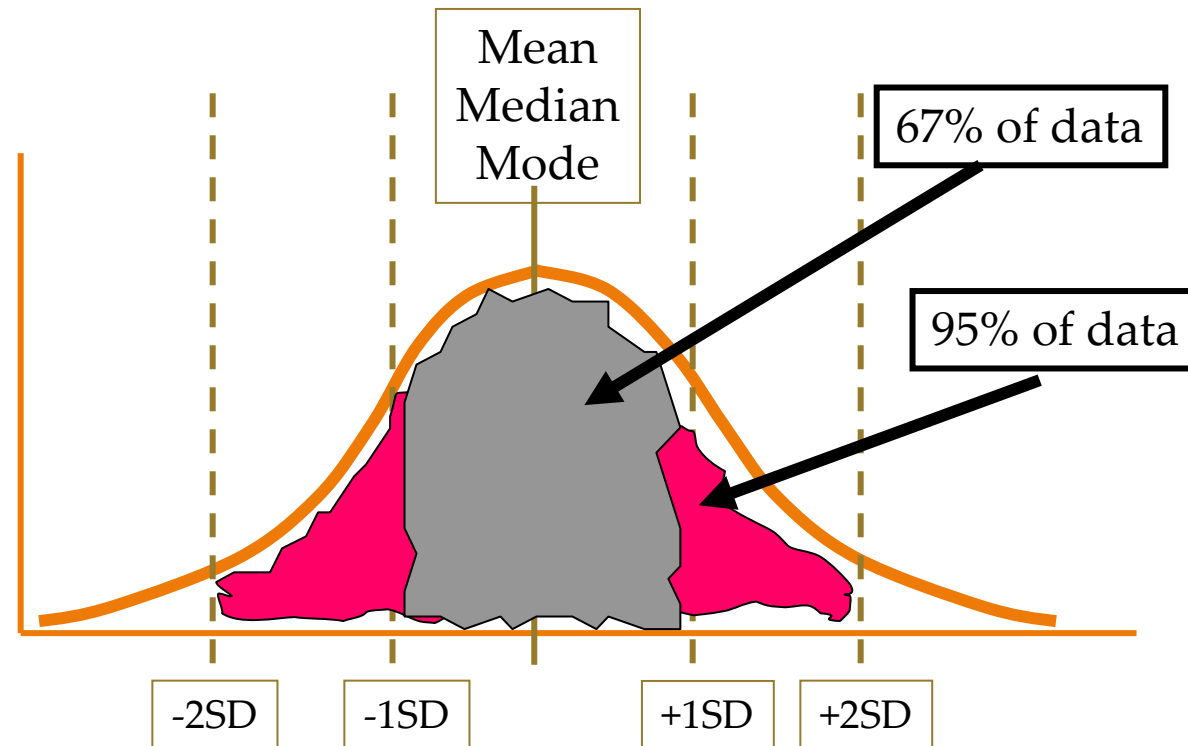
Example: Litter size in 9 Sows

# Which one is best to use?

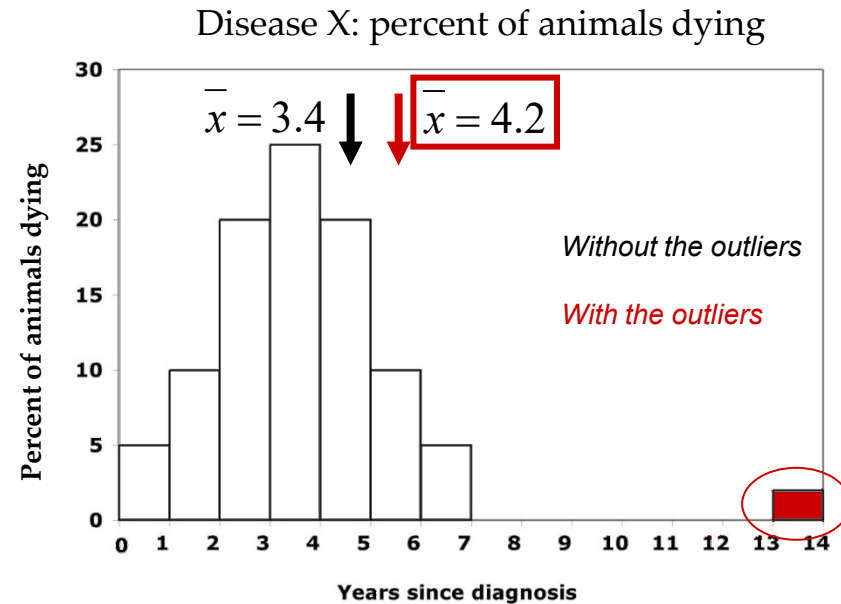
- It depends
  - What kind of data do you have?
    - Normally-distributed (bell-shaped) data
    - Skewed data (long tail on one end)



# The Power of Normal Data



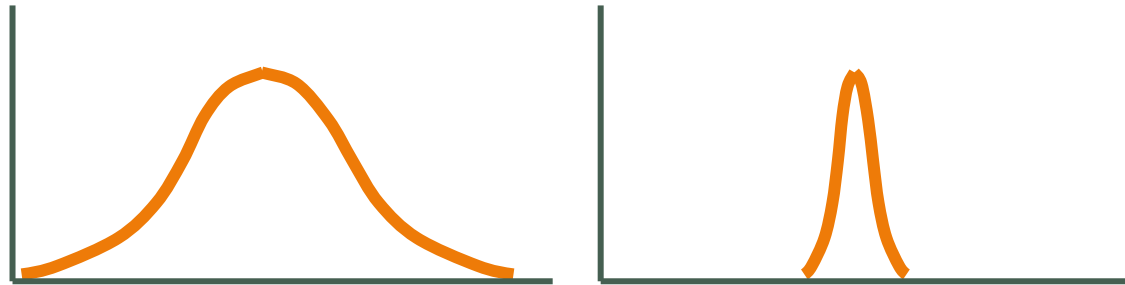
# Mean and median of a distribution with outliers



The mean is pulled to the right a lot by the outliers (from 3.4 to 4.2).

The median, on the other hand, is only slightly pulled to the right by the outliers (from 3.4 to 3.6).

# Measures of dispersion (spread) in continuous data



- Variance
- Standard deviation
- Standard error
- Ranges

# Measures of dispersion (spread) of continuous data

When your data is summarized using the “mean”:

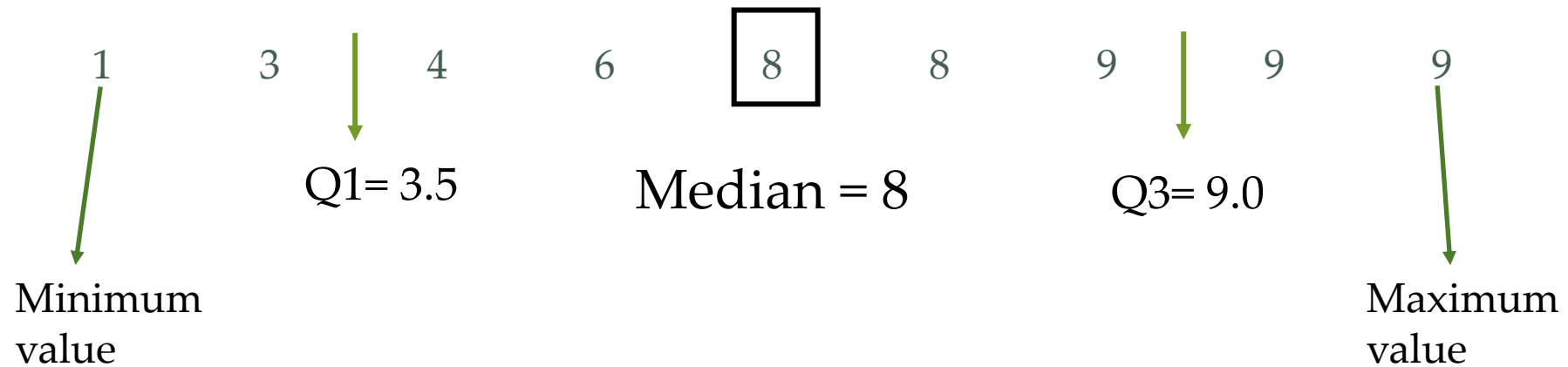
- Variance ( $S^2$ )
  - Average of [(difference of each value from mean)<sup>2</sup>]
  - Measure of variability from the mean
- Standard Deviation (SD)
  - Square root of variance
  - Measure of individual variability from the mean
- Standard Error of the Mean (SE or SEM)
  - $SD / \sqrt{n}$
  - Measure of population or group variability

# Measures of dispersion (spread) of continuous data

When your data is summarized using the “median”:

- Measure of individual variability from “center”
- Range
  - Maximum – Minimum value
- Interquartile range
  - $Q3 - Q1$

# Range and Inter-quartile Range:



$$\text{Range} = \text{Max} - \text{Min} = 9.0 - 1.0 = 8.0$$

$$\text{IQR} = \text{Q3} - \text{Q1} = 9.0 - 3.5 = 5.5$$



# Outline:

- Epidemiological Data
- Subjects and Variables
- Data Types
- Precision and Validity of data
- Measures to describe continuous epidemiological data
- Measures to describe categorical epidemiological data



## Describing Categorical health data (Qualitative)

---

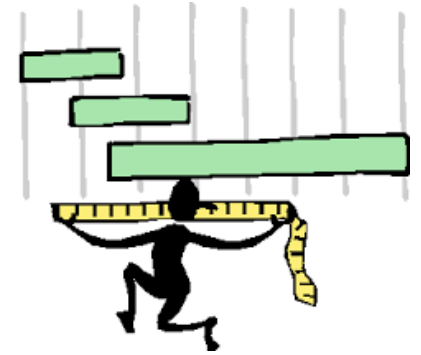
- Frequency (counts)
- Percentages

# Measures of health indices



# Measuring Disease (health event):

- Crude vs. specific measure
  - Crude
    - Measure in entire study population
    - Eg. 10/100 of entire study population sick
  - Specific
    - Measure in a subgroup of study population
    - Age, sex, exposure group, risk factor
    - E.g. 40/100 were females



- Count

- Number of diseased, infected, dead
  - 16 goats had foot injuries

- Population-at-risk (PAR)

- Number “capable” of experiencing disease, infection or death
  - Watch your denominators: Bulls aren’t at risk of abortion, exclude bulls from PAR if studying uterine diseases
  - Recently recovered individuals (in many cases) aren’t at risk of developing the same disease



- Morbidity

- Total number of illnesses during a given time period/total population

- Mortality

- Total number of deaths during a given time period/total population in which the deaths occurred

- Case-fatality rate

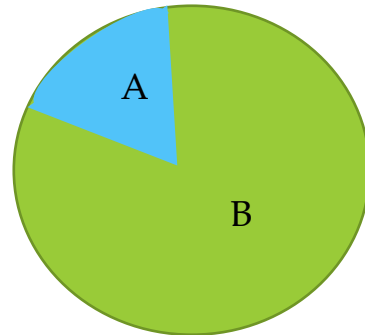
- Number of deaths during a given time period from a particular disease/number of individuals **with the specific disease**

# Disease (health event) frequency measures:

## 1) Proportions

- Values: 0 to 1 or 0%-100%
- *Example: In a region with 50 veterinary hospitals, 10 were multi-specialty facilities and 40 were not. The proportion of hospitals that were multi-specialty facility are  $10/50 = 0.2$  or 20%*

$$\frac{A}{A + B}$$

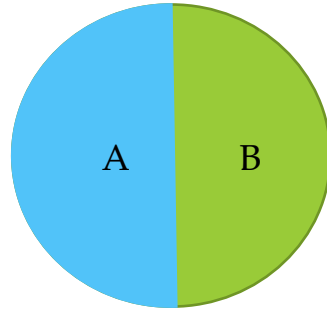


*Note: Numerator is a part of the denominator*

## 2) Ratios

- Purpose: Compare 2 mutually exclusive groups
- Values: Any positive value
- Example: In a region with 50 veterinary hospitals, 10 were multi-specialty facilities and 40 were not. The ratio of hospitals that were multi-specialty to that were not is 10/40 or 1:4

$$\frac{A}{B} \text{ or } A : B$$



*Note: Denominator does not include Numerator*

### 3) Rates

- Purpose: measure change per unit-time.
- Tells us how fast disease is occurring
- Values: Any value per time-unit
- *Example: In a region with 50 dairy farms, 10 reported abortions last year and 40 did not. The rate of abortions in the region was 0.2 per year.*

# Measures of Health Indices:

- “Risk” of disease
  - The probability that an individual will be diseased
  - A proportion
  - Example: prevalence or incidence risk

# Individuals with disease

---

# Individuals @ risk

- “Rate” of disease
  - The speed of disease through the population
  - A rate
  - Example: incidence rate

New cases of disease

---

Individual-time @ risk

# Measures of Health Indices:

## Prevalence

Proportion of cases in the population-at-risk at a GIVEN point in time.

$$\frac{\text{Number of cases observed at time } t}{\text{Total number of individuals at risk at time } t}$$

Prevalence is a **proportion** not a rate  
(prevalence does not involve “speed” of disease occurrence or spread)

- Point Prevalence – proportion at a specific point in time
- Period prevalence – proportion during a certain time period

# Incidence

Incidence =  $\frac{\text{Number of "new" cases observed}}{\text{Population-at-risk}}$  during time interval T

- Incidence is a rate
  - Measures the flow of individuals from the disease-free state to the disease state over time in the population-at-risk
- Measure of how quickly disease is spreading
- What is happening moment to moment or year to year

# Epidemiological field scenario:

You have been noticing a higher occurrence of cows with abortions during their last term of pregnancy. You know that some new animals were purchased in your jurisdiction including bulls many months ago. You sample milk and serum from few of the affected animals to detect the cause. As you suspected, the laboratory results (with ELISA test) were positive for *Brucella abortus*.

