Measures of spread

STATISTICAL TECHNIQUES IN TABLEAU

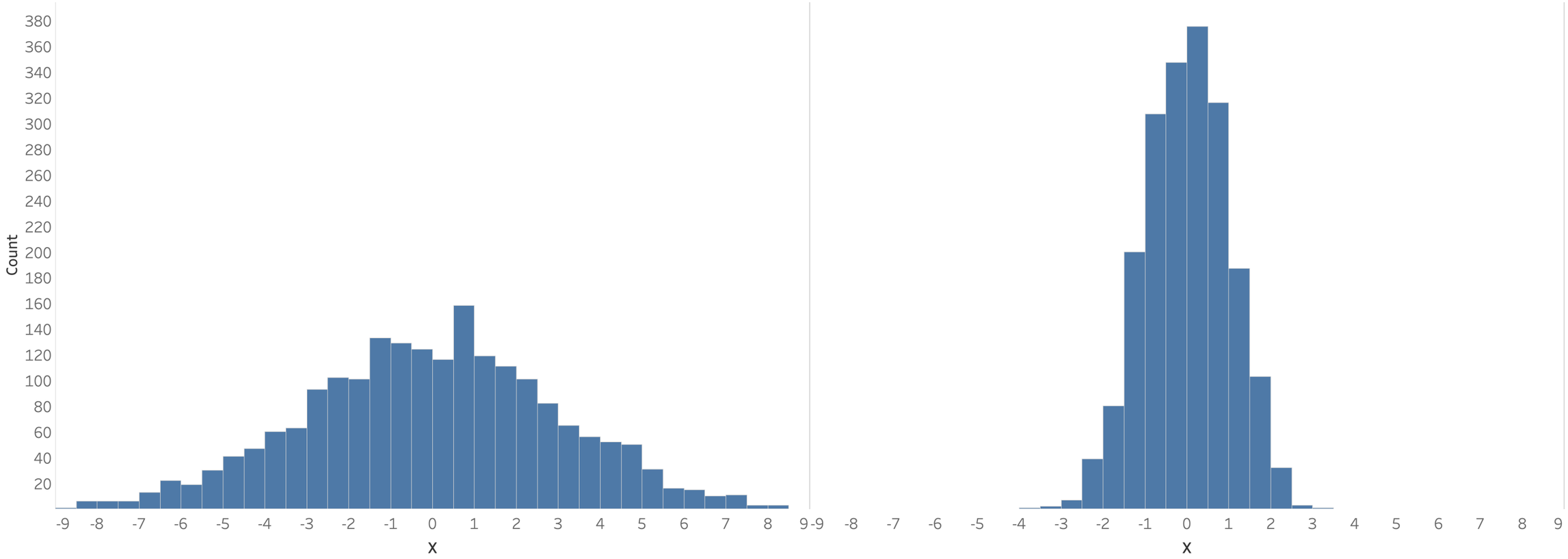
Maarten Van den Broeck
Content Developer at DataCamp

datacamp

Statistics for describing a variable

Statistic	Description
Count	number of observations
Median	midpoint of your observations
Average	mean value of your observations
Min/Max	lowest and highest value
Quartile/IQR	25th and 75th percentile / spread of the 50% of your middlemost observations
Modality/Mode	number of modes / most occurring value
Skewness	(a)symmetry of the distribution
Kurtosis	distribution of extreme values

Measures of spread



- Spread is affected by kurtosis (outliers) and skewness (asymmetry)
- Typically, spread around the mean is only useful for normal distributions

Variance

$x_i - \overline{x}$

$(x_i - \overline{x})^2$

$\sum(x_i - \overline{x})^2$

$\frac{\sum(x_i - \overline{x})^2}{n-1}$

x_i = individual data point, $\overline{x}$ = sample mean

n = number of observations

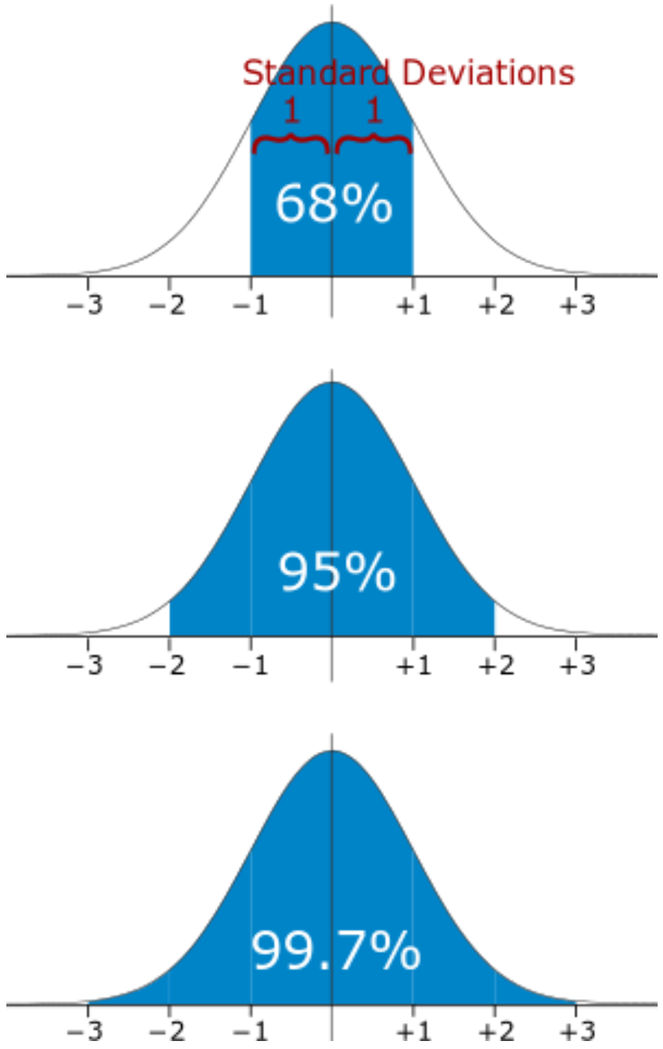
- Variance is the average of the squared differences from the mean
- Higher variance means higher spread of the data
- Unit of variance is squared

[1] Note: you don't need to memorize the formulas. They unveil the black box of Tableau's calculations.

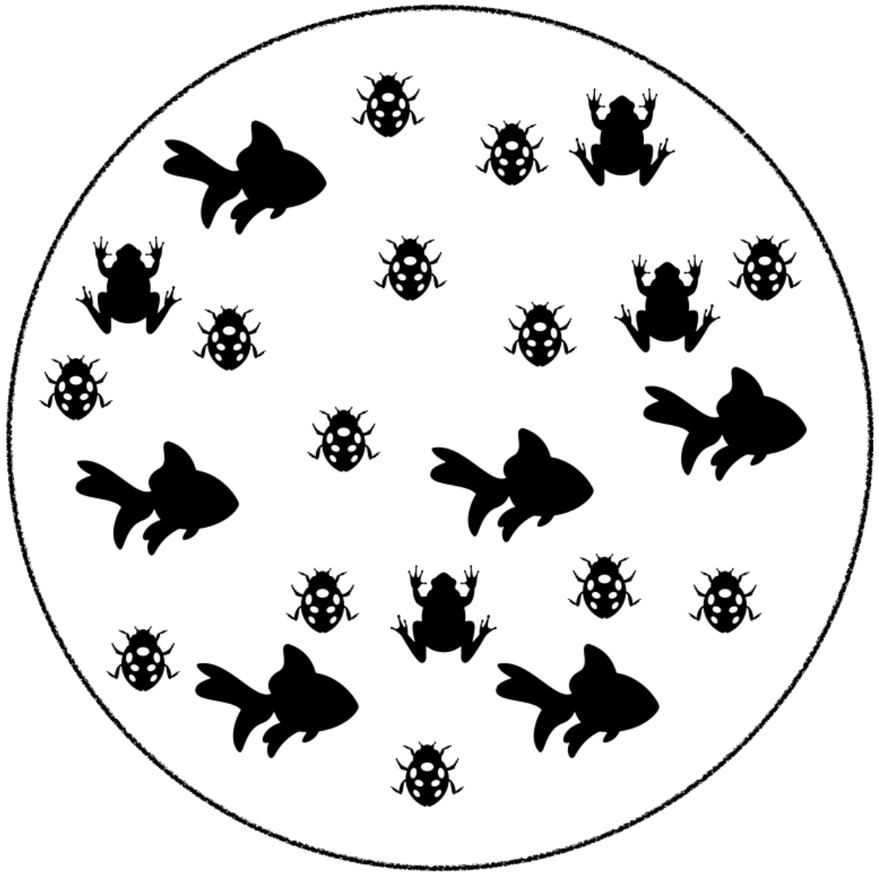
Standard deviation (SD or s)

$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$ or $s = \sqrt{variance}$

- Unit of standard deviation is same as the variable
- How far on average lie the data points from the mean
- 68% of the observations lies within $[-1s, 1s]$ range if data is normally distributed
- Number of standard deviations can be used as a threshold to pinpoint unusual values

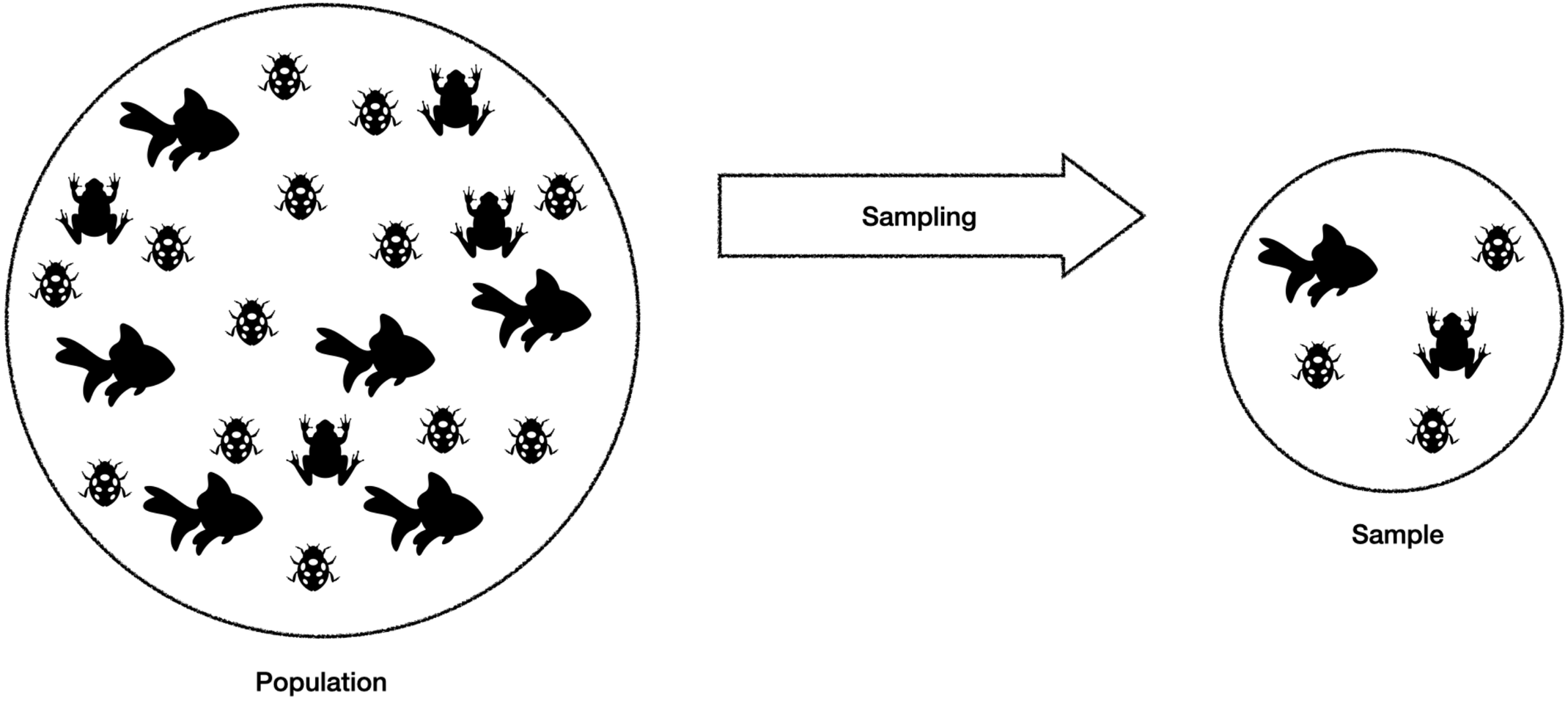


Population vs. sample

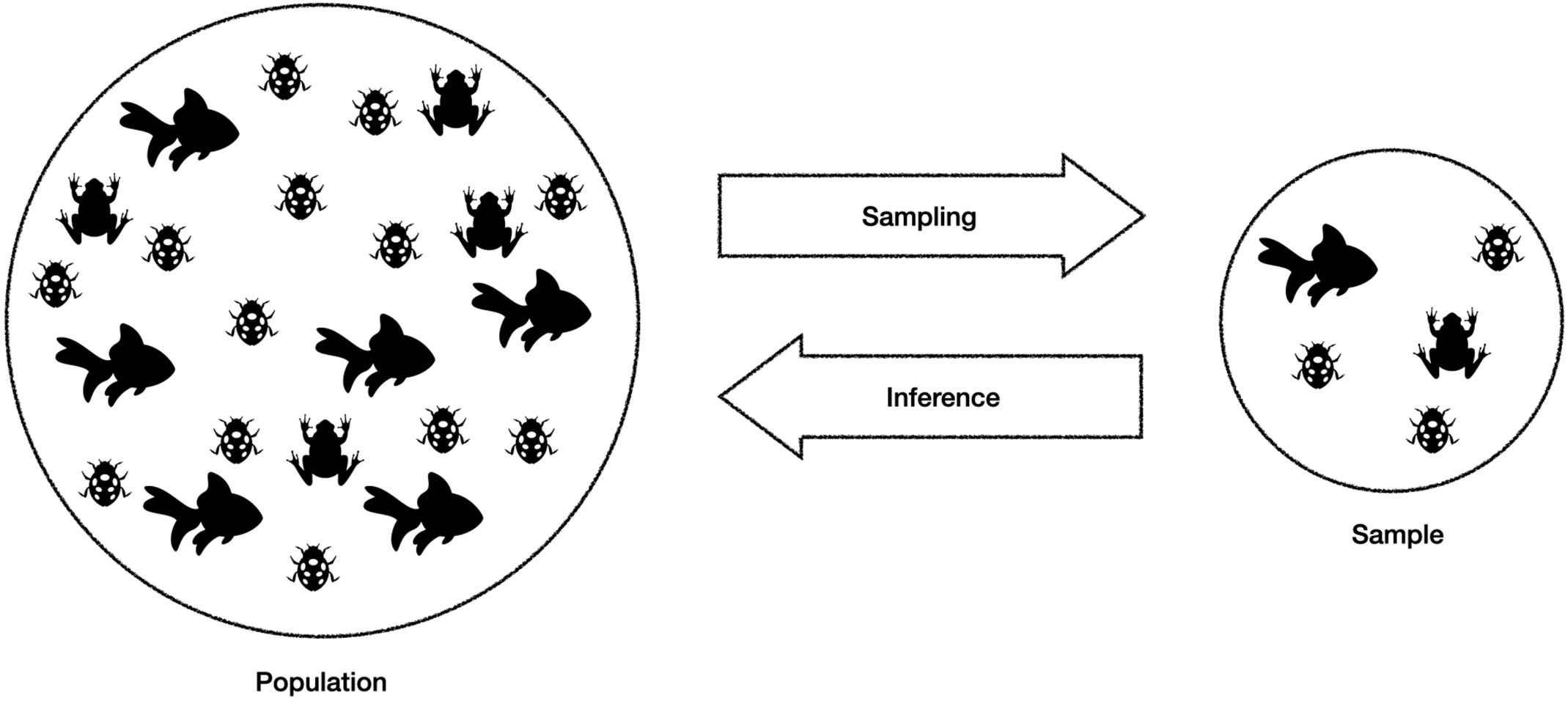


Population

Population vs. sample



Population vs. sample



Calculating spread in sample vs. population

Sample variance s^2

$$s^2 = \frac{\sum(x_i - \overline{x})^2}{n-1}$$

data per country (sample)
generalize for Europe (population)

Sample standard deviation s

$$s = \sqrt{\frac{\sum(x_i - \overline{x})^2}{n-1}}$$

$\overline{x}$ = sample mean

n = sample size

Population variance σ

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$$

data of your university (population)
no need for generalizing

Population standard deviation σ^2

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

μ = population mean

N = population size

[1] Note: you don't need to memorize the formulas. They unveil the black box of Tableau's calculations.

Calculating spread in sample vs. population

Sample variance s^2

$$s^2 = \frac{\sum(x_i - \overline{x})^2}{\mathbf{n - 1}}$$

data per country (sample)
generalize for Europe (population)

Sample standard deviation s

$$s = \sqrt{\frac{\sum(x_i - \overline{x})^2}{\mathbf{n - 1}}}$$

$\overline{x}$ = sample mean

n = sample size

Population variance σ^2

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{\mathbf{N}}$$

data of your university (population)
no need for generalizing

Population standard deviation σ

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{\mathbf{N}}}$$

μ = population mean

N = population size

[1] Note: you don't need to memorize the formulas. They unveil the black box of Tableau's calculations.

Let's practice!

STATISTICAL TECHNIQUES IN TABLEAU

Tableau: summary cards and spread

STATISTICAL TECHNIQUES IN TABLEAU

Maarten Van den Broeck
Content Developer at DataCamp

datacamp

Let's practice!

STATISTICAL TECHNIQUES IN TABLEAU

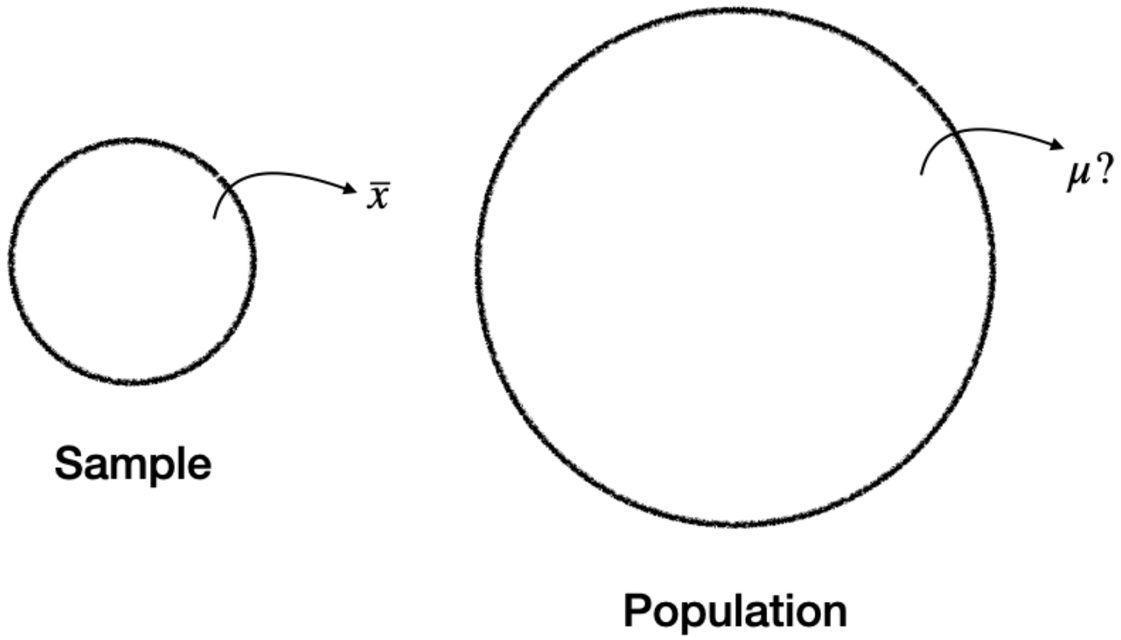
Standard error and confidence intervals

STATISTICAL TECHNIQUES IN TABLEAU

Maarten Van den Broeck
Content Developer at DataCamp

datacamp

The standard error (SE)

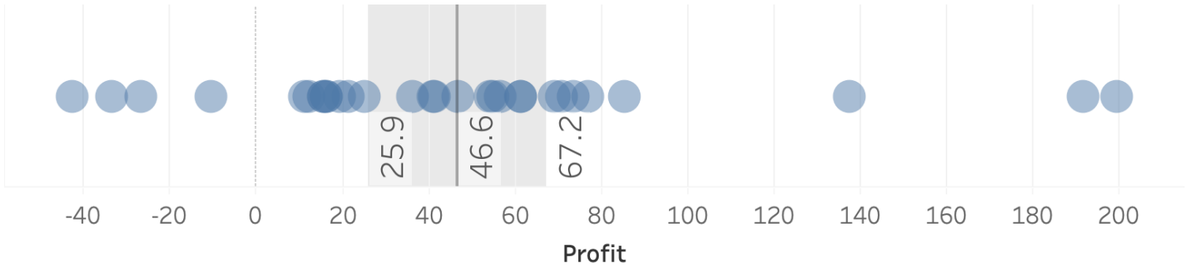


- $SE = \frac{SD}{\sqrt{n}}$ SD = standard deviation
  n = sample size
- Larger sample size means smaller SE

- Standard error of the mean: how much the sample mean and population mean deviate

The confidence interval (CI)

"I'm 95 percent confident that the true population mean lies within the ranges of my confidence interval."



- When taking more samples, 95% of the confidence intervals will contain the true population mean
- Sample $n = 30$
  - Sample mean $\overline{x}$ = `46.6`
  - Lower bound = `25.9`
  - Upper bound = `67.2`
- Population mean = `44.4`
- $CI = \overline{x} \pm level * SE$
- $level = 1.96$ for 95% CI

Let's practice!

STATISTICAL TECHNIQUES IN TABLEAU

Tableau: adding lines and distribution bands

STATISTICAL TECHNIQUES IN TABLEAU

Maarten Van den Broeck
Content Developer at DataCamp

datacamp

Let's practice!

STATISTICAL TECHNIQUES IN TABLEAU