

# Introduction to Statistical Method

---

## Descriptive Statics

---

- **population**: a large group of objects about which inferences are to be made.
- **sample**: a subset of the population
- at least one **random variable** whose behavior is to be studied relative to the population

## Random Samples

### Definition

Random sample of size  $n$  from the distribution of  $\mathbf{X}$  is a collection of  $n$  independent random variables  $X_1, \dots, X_n$ , each with the same distribution as  $\mathbf{X}$ .

$X_1, \dots, X_n$  are independent identically distributed random variables.

### Remark

In order to guarantee that the random variables in a random sample are indeed independently distributed, the size of random sample should be not exceed 5% of population.

## Sample Statistics

- **sample range**:  $\max_{1 \leq k \leq n} X_k - \min_{1 \leq k \leq n} X_k$
- **sample mean**:  $\bar{X} := \frac{1}{n} \sum_{k=1}^n X_k$
- **sample median**:  $\tilde{x} = \begin{cases} \frac{1}{2}(x_{n/2} + x_{n/2+1}) & \mathbf{n \text{ even}} \\ x_{(n+1)/2} & \mathbf{n \text{ odd}} \end{cases}$
- upper-case letters for statistics (random variables)
- lower-case letters for observed values of statistics (numbers)
- **sample variance**:  $S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$
- **sample standard deviation**:  $S := \sqrt{S^2}$

## Rounding of Statistics

- For the mean we give **one more decimal place** than the original data has.
- For the variance we give **two more decimal places** than the original data has.
- For the standard deviation we give **one more decimal place** than the original data has.
- The range and median are not rounded.

# Histogram

## Four Main Properties

- The number of categories should be suitable for the amount of data.  
According to Sturges's rule.
- Each datum should fall into exactly one category.
- The categories should have the same width.
- No datum should assume a boundary value.

## Histogram Creation Steps

- The desired number of categories (Sturges's rule)

Data Set Size	Number of Categories
$< 16$	Insufficient data
$16 - 31$	5
$2^{n-1}$ to $2^n - 1$	$n$

- Calculate **data (sample) range**.
- Divide data range by number categories  
Round up to the accuracy of the data or add a smallest decimal unit at accuracy of data.
- The **lower boundary for first category** lies **1/2 smallest decimal unit below** smallest datum.