

## SCIU4T4: Central tendency and measures of spread

## Descriptive versus inferential statistics

### **Descriptive Statistics**

- ▶ Summarise observations
- ▶ E.g., average monthly temperature

### **Inferential Statistics**

- ▶ Make estimates or predictions
- ▶ E.g., predict temperature from latitude

# Descriptive statistics in jamovi

### Descriptives

Variables: Soil organic carbon (g C / kg soil)

Descriptives | Variables across columns  Frequency tables

Statistics

**Sample Size**  
 N  Missing

**Percentile Values**  
 Cut points for 4 equal groups  
 Percentiles 25,50,75

**Dispersion**  
 Std. deviation  Minimum  
 Variance  Maximum  
 Range  IQR

**Mean Dispersion**  
 Std. error of Mean  
 Confidence interval for Mean 95 %

**Central Tendency**  
 Mean  
 Median  
 Mode  
 Sum

**Distribution**  
 Skewness  
 Kurtosis

**Normality**  
 Shapiro-Wilk

**Outliers**  
 Most extreme 5 values

### Results

#### Descriptives

|                     | Soil organic carbon (g C / kg soil) |
|---------------------|-------------------------------------|
| N                   | 34                                  |
| Missing             | 0                                   |
| Mean                | 6.52353                             |
| Median              | 5.80000                             |
| Mode                | 2.40000 *                           |
| Standard deviation  | 4.49701                             |
| Variance            | 20.22307                            |
| IQR                 | 7.27500                             |
| Range               | 15.60000                            |
| Minimum             | 0.60000                             |
| Maximum             | 16.20000                            |
| Skewness            | 0.55655                             |
| Std. error skewness | 0.40305                             |
| Kurtosis            | -0.73034                            |
| Std. error kurtosis | 0.78790                             |
| 25th percentile     | 2.42500                             |
| 50th percentile     | 5.80000                             |
| 75th percentile     | 9.70000                             |

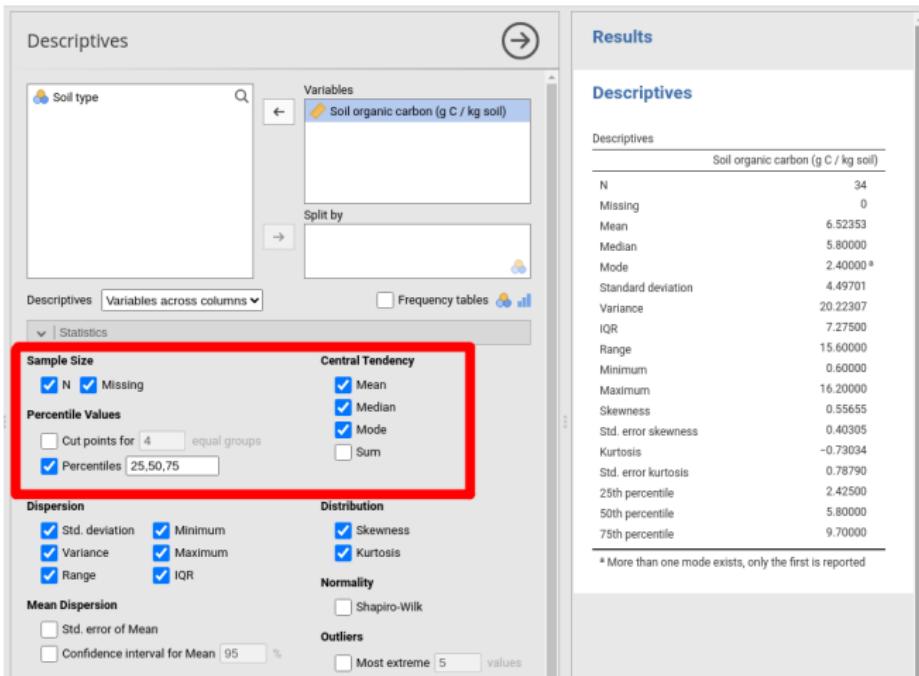
\* More than one mode exists, only the first is reported

# Properties of distributions

- ▶ Central tendency
- ▶ Spread
- ▶ Skew & Kurtosis

**Will focus on *samples*  
rather than populations**

# Descriptive statistics: Central tendency



The screenshot shows the SPSS 'Descriptives' dialog box and the resulting 'Results' table.

**Descriptives Dialog Box:**

- Variables:** Soil organic carbon (g C / kg soil)
- Descriptives:** Selected
- Central Tendency (highlighted in red):**
  - N
  - Missing
  - Mean
  - Median
  - Mode
  - Sum
- Percentile Values:**
  - Cut points for 4 equal groups
  - Percentiles 25,50,75
- Dispersion:**
  - Std. deviation
  - Variance
  - Range
  - Minimum
  - Maximum
  - IQR
- Distribution:**
  - Skewness
  - Kurtosis
- Normality:**
  - Shapiro-Wilk
- Outliers:**
  - Most extreme 5 values

**Results Table:**

| Descriptives        |                                     |
|---------------------|-------------------------------------|
|                     | Soil organic carbon (g C / kg soil) |
| N                   | 34                                  |
| Missing             | 0                                   |
| Mean                | 6.52353                             |
| Median              | 5.80000                             |
| Mode                | 2.40000 <sup>a</sup>                |
| Standard deviation  | 4.49701                             |
| Variance            | 20.22307                            |
| IQR                 | 7.27500                             |
| Range               | 15.60000                            |
| Minimum             | 0.60000                             |
| Maximum             | 16.20000                            |
| Skewness            | 0.55655                             |
| Std. error skewness | 0.40305                             |
| Kurtosis            | -0.73034                            |
| Std. error kurtosis | 0.78790                             |
| 25th percentile     | 2.42500                             |
| 50th percentile     | 5.80000                             |
| 75th percentile     | 9.70000                             |

<sup>a</sup> More than one mode exists, only the first is reported

# Mean, median, and mode

## Arithmetic mean

Add values, divide by number ( $N$ )

**For example,  $N = 3$  temperatures:**

- ▶  $12.5 \text{ } ^\circ\text{C}$
- ▶  $13.4 \text{ } ^\circ\text{C}$
- ▶  $14.0 \text{ } ^\circ\text{C}$

$$\bar{x} = \frac{12.5 + 13.4 + 14.0}{3} = 13.3$$

## Calculating the mean of 7 temperatures ( $^{\circ}\text{C}$ )

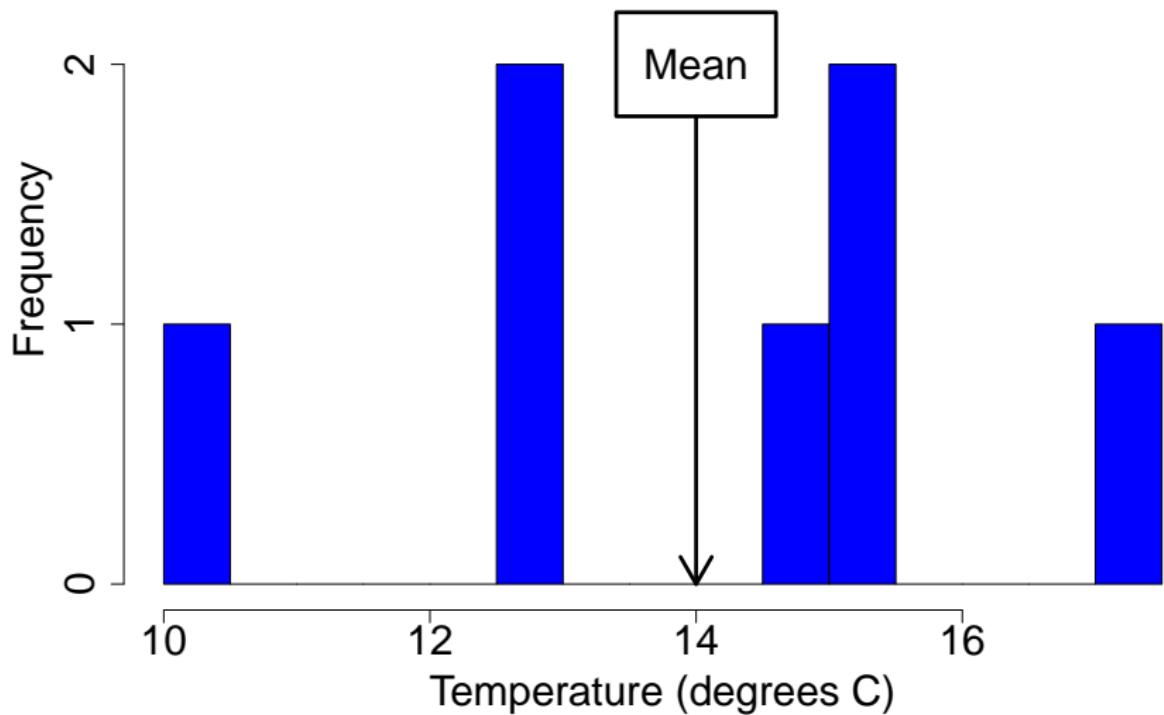
Table 1: Seven values ( $x$ ) of soil temperature ( $^{\circ}\text{C}$ ) at a site

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ |
|-------|-------|-------|-------|-------|-------|-------|
| 17.1  | 15.2  | 14.9  | 12.6  | 15.2  | 10.3  | 12.7  |

$$\bar{x} = \frac{17.1 + 15.2 + 14.9 + 12.6 + 15.2 + 10.3 + 12.7}{7}$$

$$\bar{x} = 14$$

## Arithmetic mean visualisation (histogram)



## General formula for arithmetic mean

- ▶ Sample mean:  $\bar{x}$  (or  $\hat{\mu}_x$ )
- ▶ Sample size:  $N$

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_{N-1} + x_N}{N}$$

## General formula for arithmetic mean

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_{N-1} + x_N}{N}$$

$$\sum_{i=1}^N x_i = x_1 + x_2 + \dots + x_{N-1} + x_N$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

## The mode

Most frequently occurring observation

---

| $x_1$ | $x_2$       | $x_3$ | $x_4$ | $x_5$       | $x_6$ | $x_7$ |
|-------|-------------|-------|-------|-------------|-------|-------|
| 17.1  | <b>15.2</b> | 14.9  | 12.6  | <b>15.2</b> | 10.3  | 12.7  |

---

Also applies to categorical data

---

| $x_1$ | $x_2$      | $x_3$ | $x_4$      | $x_5$      | $x_6$ |
|-------|------------|-------|------------|------------|-------|
| dog   | <b>cat</b> | bird  | <b>cat</b> | <b>cat</b> | dog   |

---

## Visualising the mode

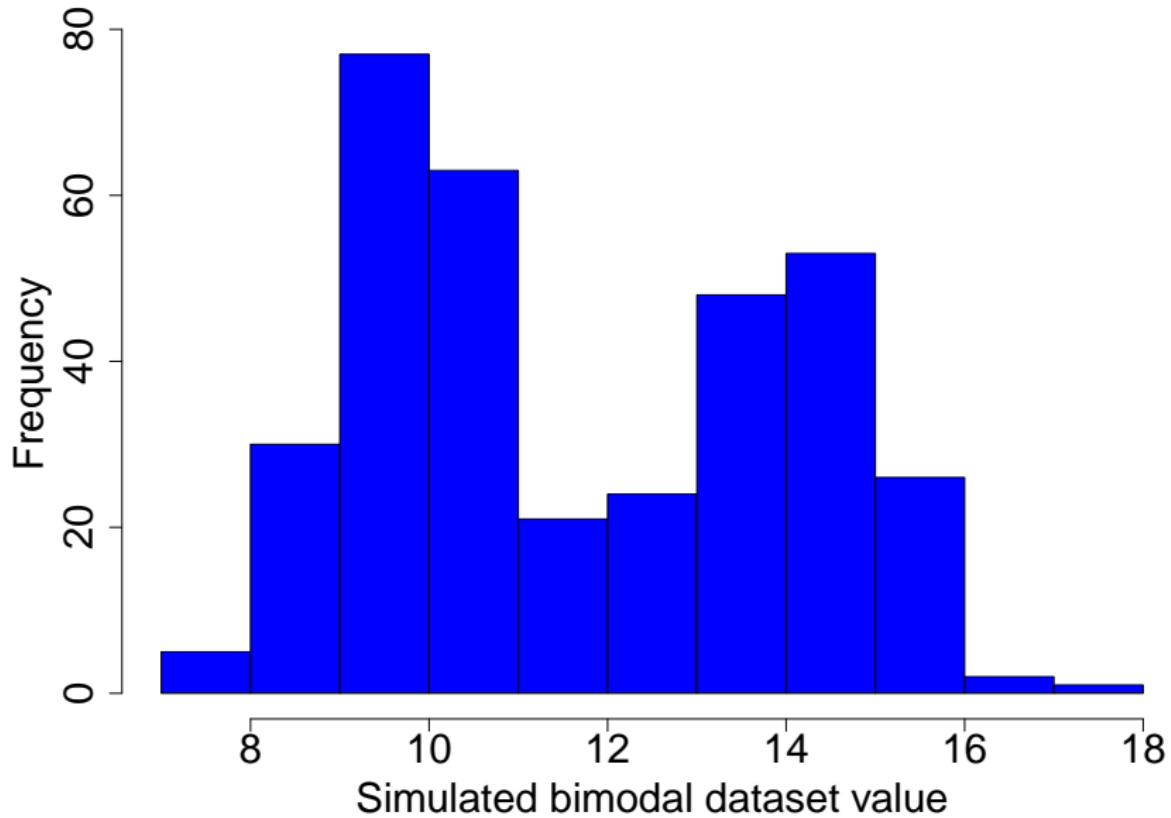


Figure 1: Hypothetical dataset that has a bimodal distribution.

## The median

- ▶ Observation in the middle when the observations are arranged in ascending order
- ▶ There are an equal number of observations lower and higher than the median

## The median

---

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ |
|-------|-------|-------|-------|-------|-------|-------|
| 17.1  | 15.2  | 14.9  | 12.6  | 15.2  | 10.3  | 12.7  |

---

Sorting the data:

---

| $x_6$ | $x_4$ | $x_7$ | $x_3$       | $x_2$ | $x_5$ | $x_1$ |
|-------|-------|-------|-------------|-------|-------|-------|
| 10.3  | 12.6  | 12.7  | <b>14.9</b> | 15.2  | 15.2  | 17.1  |

---

## The median

Median is a type of **quantile** (50%)

- ▶ Can break distribution into other quantiles
  - ▶ First **quartile** (25% quantile)
  - ▶ Third **quartile** (75% quantile)
- ▶ Quantiles also called 'percentiles'

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|-------|-------|-------|-------|-------|
| 2     | 4     | 5     | 6     | 8     |

## The median

If there is no middle value

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|-------|-------|-------|-------|-------|-------|
| 3.1   | 3.5   | 3.8   | 4.0   | 4.2   | 4.2   |

Take mean of middle values:

$$\frac{3.8 + 4.0}{2} = 3.9$$

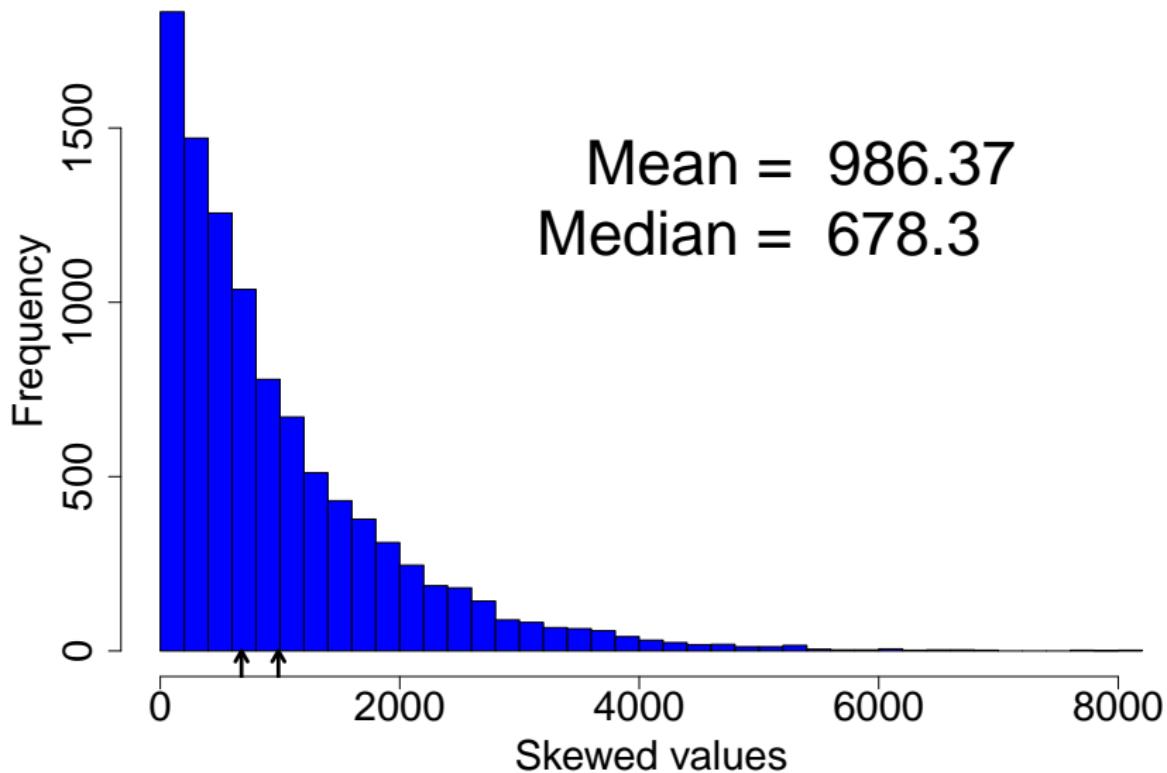
## The median

- ▶ Multiple valid ways to calculate quantiles<sup>1</sup>
- ▶ No one ‘right’ way
- ▶ Jamovi’s approach might differ from other software

---

<sup>1</sup>Hyndman, RJ, & Y Fan. 1996. American Statistician 50:361–65.

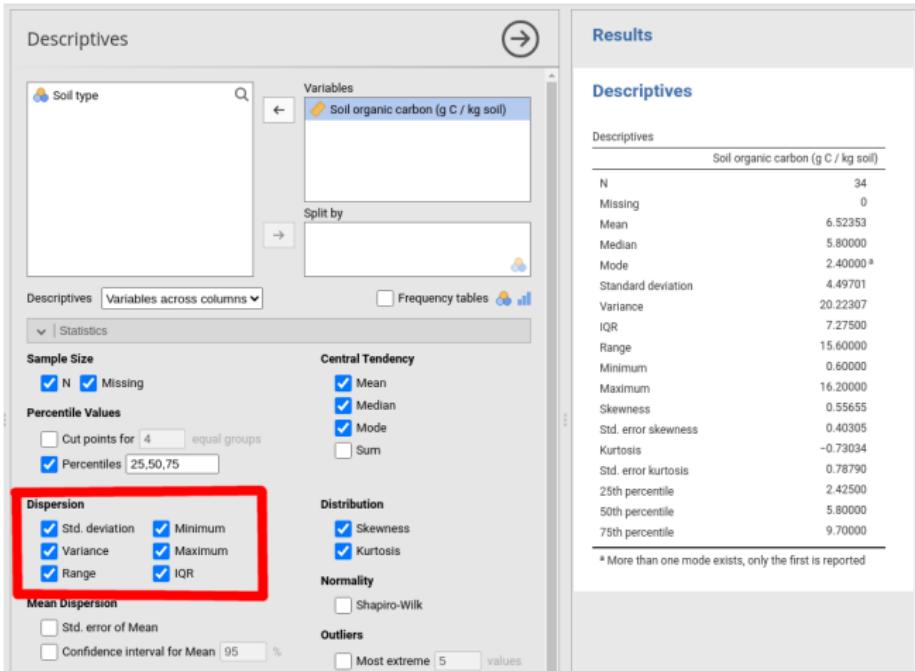
## Median more robust to outliers



## Measures of spread

- ▶ Range
- ▶ Interquartile range (IQR)
- ▶ Variance ( $s^2$ )
- ▶ Standard deviation ( $s$ )
- ▶ Coefficient of variation ( $CV$ )

# Measures of spread

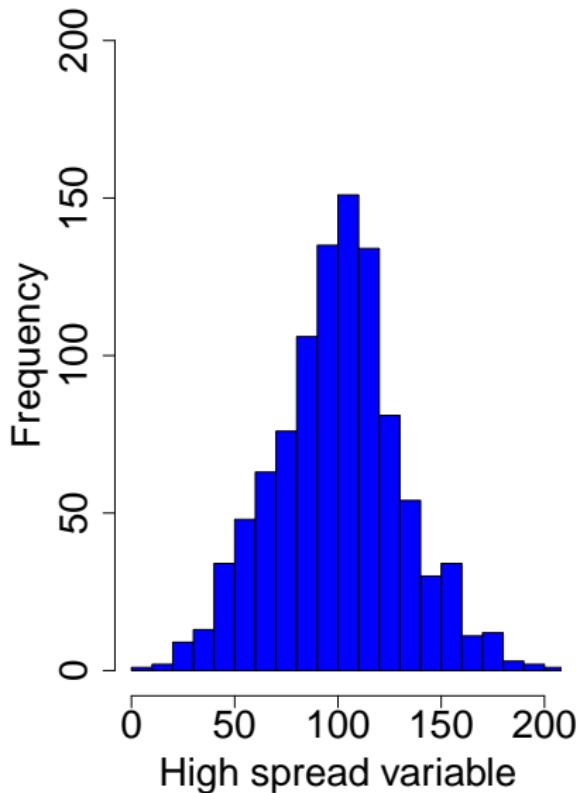
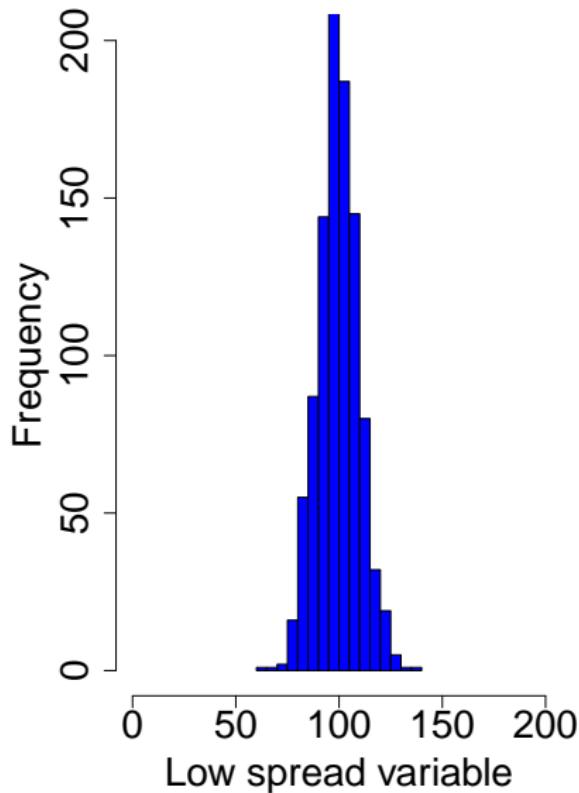


The screenshot shows the SPSS 'Descriptives' dialog box. On the left, under 'Descriptives', 'Soil type' is selected as the split variable for 'Soil organic carbon (g C / kg soil)'. The 'Dispersion' section is highlighted with a red box, showing checked options for Std. deviation, Variance, Range, Minimum, Maximum, and IQR. The 'Central Tendency' section includes Mean, Median, Mode, and Sum. The 'Distribution' section includes Skewness and Kurtosis. The 'Normality' section includes Shapiro-Wilk. The 'Outliers' section includes Most extreme values. On the right, the 'Results' pane displays descriptive statistics for 'Soil organic carbon (g C / kg soil)' across 34 observations, including N (34), Mean (6.52353), Median (5.80000), Mode (2.40000\*), Standard deviation (4.49701), Variance (20.22307), IQR (7.27500), Range (15.60000), Minimum (0.60000), Maximum (16.20000), Skewness (0.55655), Std. error skewness (0.40305), Kurtosis (-0.73034), Std. error kurtosis (0.78790), 25th percentile (2.42500), 50th percentile (5.80000), and 75th percentile (9.70000). A note at the bottom states: '\* More than one mode exists, only the first is reported.'

| Descriptives        | Soil organic carbon (g C / kg soil) |
|---------------------|-------------------------------------|
| N                   | 34                                  |
| Missing             | 0                                   |
| Mean                | 6.52353                             |
| Median              | 5.80000                             |
| Mode                | 2.40000*                            |
| Standard deviation  | 4.49701                             |
| Variance            | 20.22307                            |
| IQR                 | 7.27500                             |
| Range               | 15.60000                            |
| Minimum             | 0.60000                             |
| Maximum             | 16.20000                            |
| Skewness            | 0.55655                             |
| Std. error skewness | 0.40305                             |
| Kurtosis            | -0.73034                            |
| Std. error kurtosis | 0.78790                             |
| 25th percentile     | 2.42500                             |
| 50th percentile     | 5.80000                             |
| 75th percentile     | 9.70000                             |

Range, IQR,  $s^2$ ,  $s$ , CV

## Measures of spread



## Measures of spread: Range

$$\text{Range}(X) = \text{Maximum}(X) - \text{Minimum}(X)$$

---

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ |
|-------|-------|-------|-------|-------|-------|-------|
| 17.1  | 15.2  | 14.9  | 12.6  | 15.2  | 10.3  | 12.7  |

---

$$\text{Range}(X) = 17.1 - 10.3 = 6.8$$

## Measures of spread: Interquartile Range

$$IQR(X) = Q_3(X) - Q_1(X)$$

---

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|-------|-------|-------|-------|-------|
| 2     | 4     | 5     | 6     | 8     |

---

$$IQR(X) = 6 - 4 = 2$$

## Measures of spread: Variance ( $s^2$ )

- ▶ Expected squared deviation from mean
- ▶ More useful than range or IQR
- ▶ Less intuitive than range or IQR<sup>1</sup>
- ▶ Jamovi will calculate this for us

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2.$$

We can break this down step by step!

---

<sup>1</sup><https://bradduthie.github.io/stats/app/forest/>

## Measures of spread: Variance ( $s^2$ )

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2.$$

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ |
|-------|-------|-------|-------|-------|-------|-------|
| 17.1  | 15.2  | 14.9  | 12.6  | 15.2  | 10.3  | 12.7  |

1. Take  $x_1$  minus mean, squared  $(17.1 - 14)^2 = 9.61$
2. Repeat step 1 for  $x_2, x_3, \dots, x_N$
3. Sum up all these  $(x_i - \bar{x})^2$  values
4. Multiply the sum by  $1/(N - 1)$

## Measures of spread: Variance ( $s^2$ )

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2.$$

---

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ |
|-------|-------|-------|-------|-------|-------|-------|
| 17.1  | 15.2  | 14.9  | 12.6  | 15.2  | 10.3  | 12.7  |

---

$$\begin{aligned}SS &= (17.1 - 14)^2 + (15.2 - 14)^2 + \cdots + (12.7 - 14)^2 \\&= (3.1)^2 + (1.2)^2 + \cdots + (-1.3)^2 \\&= 30.64\end{aligned}$$

$$s^2 = \frac{1}{7-1} \times 30.64 = 5.1067 \text{ } {}^{\circ}\text{C}^2$$

## Measures of spread: Standard deviation ( $s$ )

- ▶ Mean deviation from the mean
- ▶ Square-root of the variance
- ▶ Gets back to original units

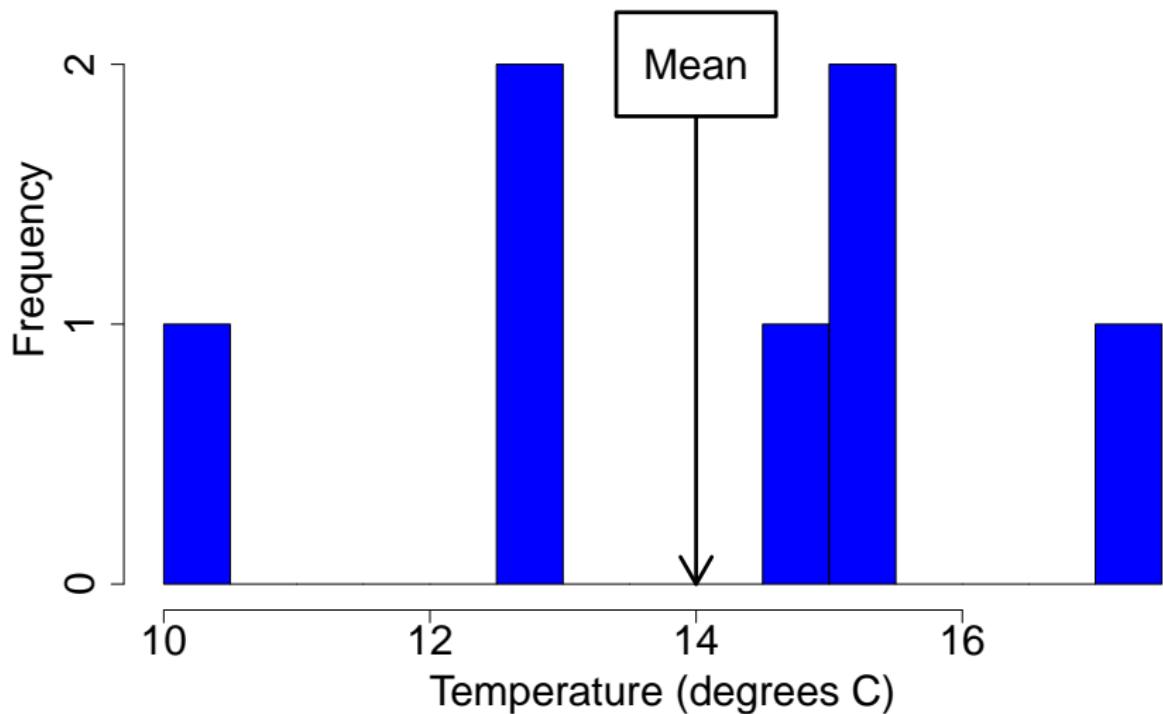
$$s^2 = 5.1067 \text{ } ^\circ C^2$$

$$s = \sqrt{5.1067} = 2.2598 \text{ } ^\circ C$$

---

<sup>2</sup>[https://bradduthie.github.io/stats/app/normal\\_pos\\_neg/](https://bradduthie.github.io/stats/app/normal_pos_neg/)

## Standard deviation of the mean: does it look right?



## Standard deviation of the mean

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}.$$

- ▶ One checkbox in jamovi
- ▶ Spread of a variable

## Coefficient of variation (CV)

Standard deviation divided by the mean

---

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ |
|-------|-------|-------|-------|-------|-------|-------|
| 17.1  | 15.2  | 14.9  | 12.6  | 15.2  | 10.3  | 12.7  |

---

$$CV = \frac{s}{\bar{x}} = \frac{2.2598 \text{ } ^\circ C}{14 \text{ } ^\circ C} = 0.1614$$

Note that the units cancel out.

## Coefficient of variation (CV)

Often expressed as a percentage

---

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ |
|-------|-------|-------|-------|-------|-------|-------|
| 17.1  | 15.2  | 14.9  | 12.6  | 15.2  | 10.3  | 12.7  |

---

$$CV = \frac{2.2598 \text{ } ^\circ\text{C}}{14 \text{ } ^\circ\text{C}} \times 100\% = 16.14\%$$

Useful for comparing variation across categories (e.g., species)

# Descriptive statistics: Skew and kurtosis

Descriptives

Variables: Soil organic carbon (g C / kg soil)

Split by:

Descriptives Variables across columns  Frequency tables

Statistics

Sample Size:  N  Missing

Percentile Values:  Cut points for 4 equal groups  Percentiles 25,50,75

Dispersion:  Std. deviation  Variance  Range  Std. error of Mean  Confidence interval for Mean 95 %

Central Tendency:  Mean  Median  Mode  Sum

Distribution:  Skewness  Kurtosis

Normality:  Shapiro-Wilk

Outliers:  Most extreme 5 values

Results

### Descriptives

| Soil organic carbon (g C / kg soil) |           |
|-------------------------------------|-----------|
| N                                   | 34        |
| Missing                             | 0         |
| Mean                                | 6.52353   |
| Median                              | 5.80000   |
| Mode                                | 2.40000 * |
| Standard deviation                  | 4.49701   |
| Variance                            | 20.22307  |
| IQR                                 | 7.27500   |
| Range                               | 15.60000  |
| Minimum                             | 0.60000   |
| Maximum                             | 16.20000  |
| Skewness                            | 0.55655   |
| Std. error skewness                 | 0.40305   |
| Kurtosis                            | -0.73034  |
| Std. error kurtosis                 | 0.78790   |
| 25th percentile                     | 2.42500   |
| 50th percentile                     | 5.80000   |
| 75th percentile                     | 9.70000   |

\* More than one mode exists, only the first is reported

## Skew is the asymmetry of a distribution

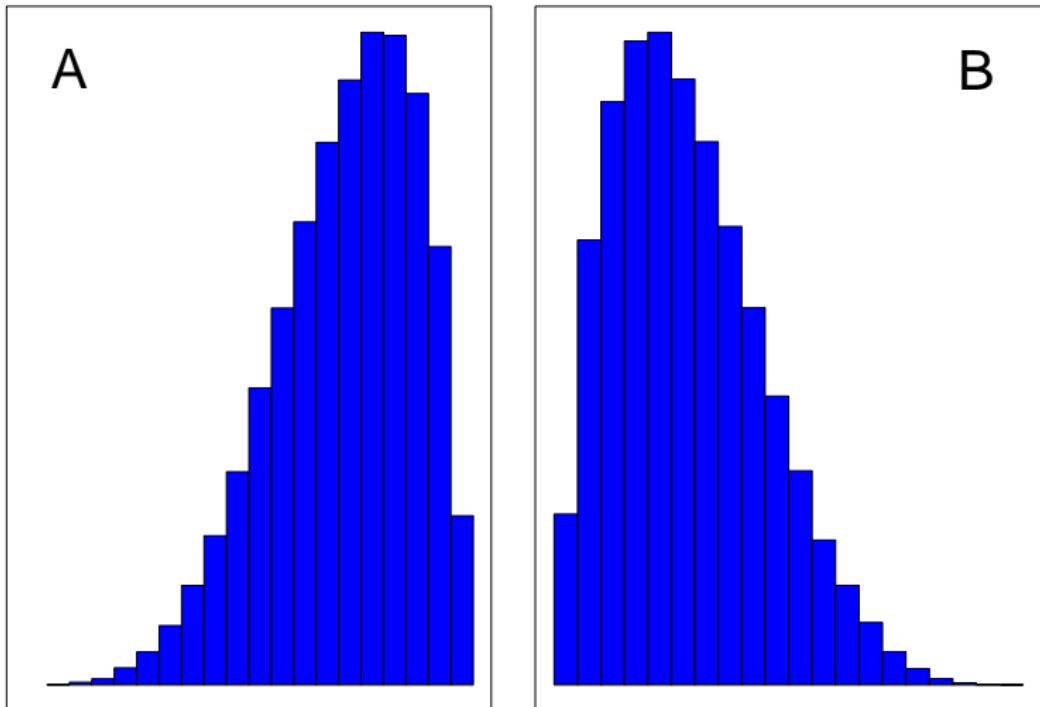


Figure 2: Histograms showing a (A) distribution that has a negative (i.e., 'left') skew and (B) distribution that has a positive (i.e., 'right') skew.

Kurtosis is the flatness of a distribution

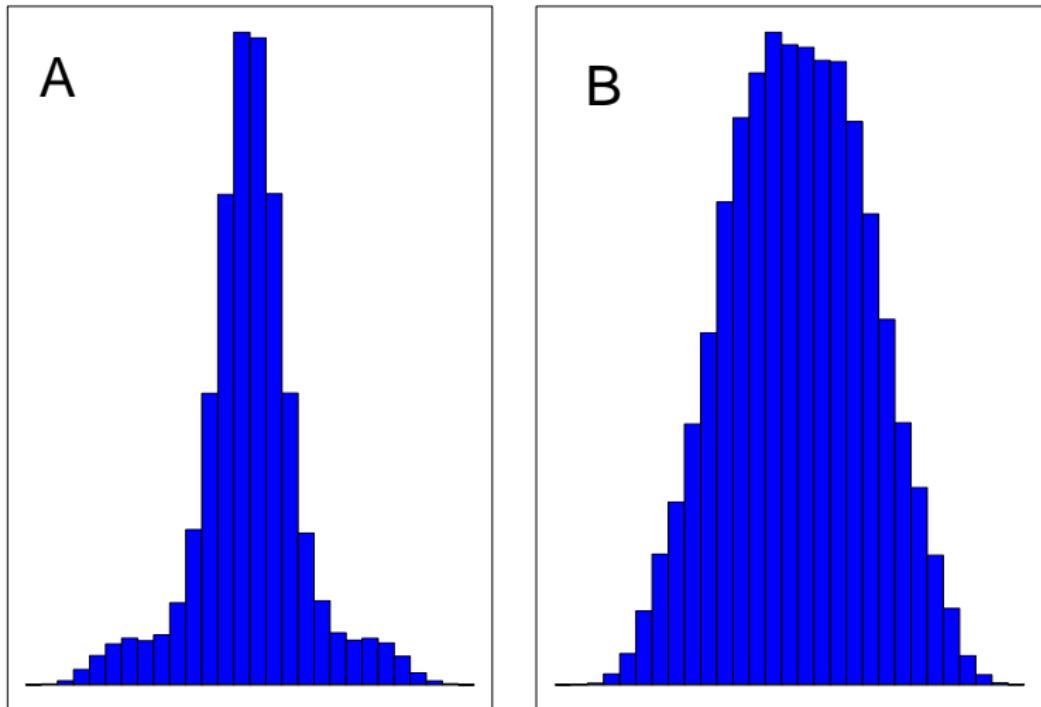


Figure 3: Histograms showing a (A) leptokurtic distribution and (B) platykurtic distribution.

## Statistical moments

1. Mean
2. Variance
3. Skew
4. Kurtosis

Mathematically, deviations from mean raised to some power give the shape of a distribution.

# Descriptive statistics in jamovi

### Descriptives

Variables: Soil organic carbon (g C / kg soil)

Descriptives | Variables across columns  Frequency tables

Statistics

**Sample Size**  
 N  Missing

**Percentile Values**  
 Cut points for 4 equal groups  
 Percentiles 25,50,75

**Dispersion**  
 Std. deviation  Minimum  
 Variance  Maximum  
 Range  IQR

**Mean Dispersion**  
 Std. error of Mean  
 Confidence interval for Mean 95 %

**Central Tendency**  
 Mean  
 Median  
 Mode  
 Sum

**Distribution**  
 Skewness  
 Kurtosis

**Normality**  
 Shapiro-Wilk

**Outliers**  
 Most extreme 5 values

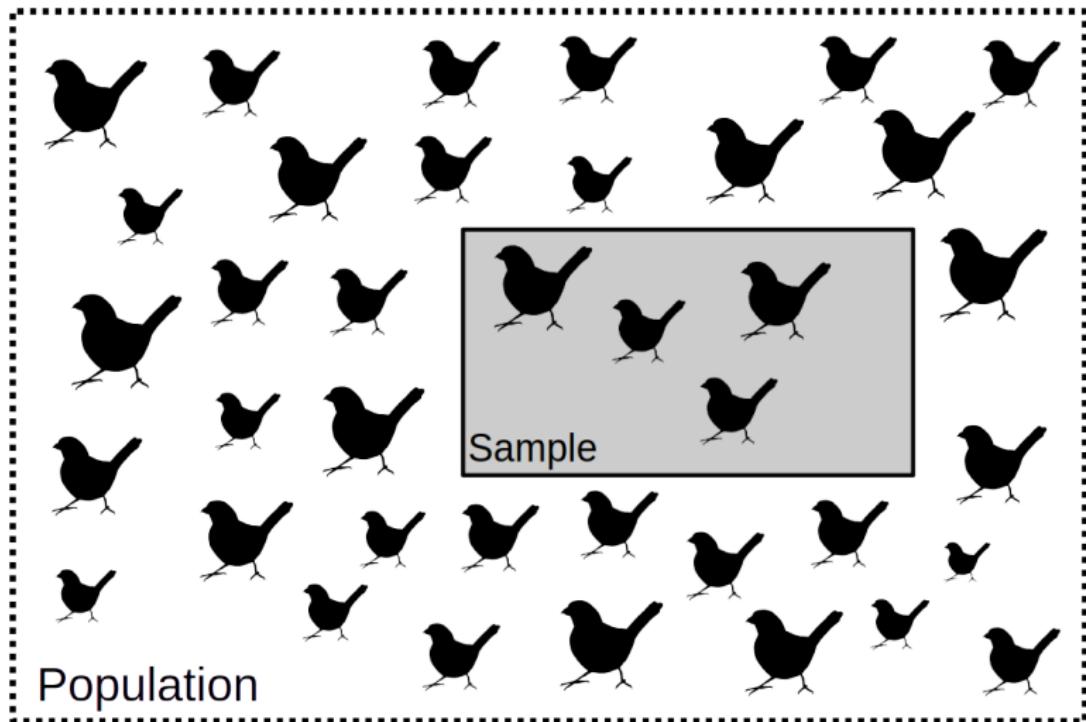
### Results

#### Descriptives

|                     | Soil organic carbon (g C / kg soil) |
|---------------------|-------------------------------------|
| N                   | 34                                  |
| Missing             | 0                                   |
| Mean                | 6.52353                             |
| Median              | 5.80000                             |
| Mode                | 2.40000 *                           |
| Standard deviation  | 4.49701                             |
| Variance            | 20.22307                            |
| IQR                 | 7.27500                             |
| Range               | 15.60000                            |
| Minimum             | 0.60000                             |
| Maximum             | 16.20000                            |
| Skewness            | 0.55655                             |
| Std. error skewness | 0.40305                             |
| Kurtosis            | -0.73034                            |
| Std. error kurtosis | 0.78790                             |
| 25th percentile     | 2.42500                             |
| 50th percentile     | 5.80000                             |
| 75th percentile     | 9.70000                             |

\* More than one mode exists, only the first is reported

## The standard error: Sample means and population mean



## The standard error: Sample means and population mean

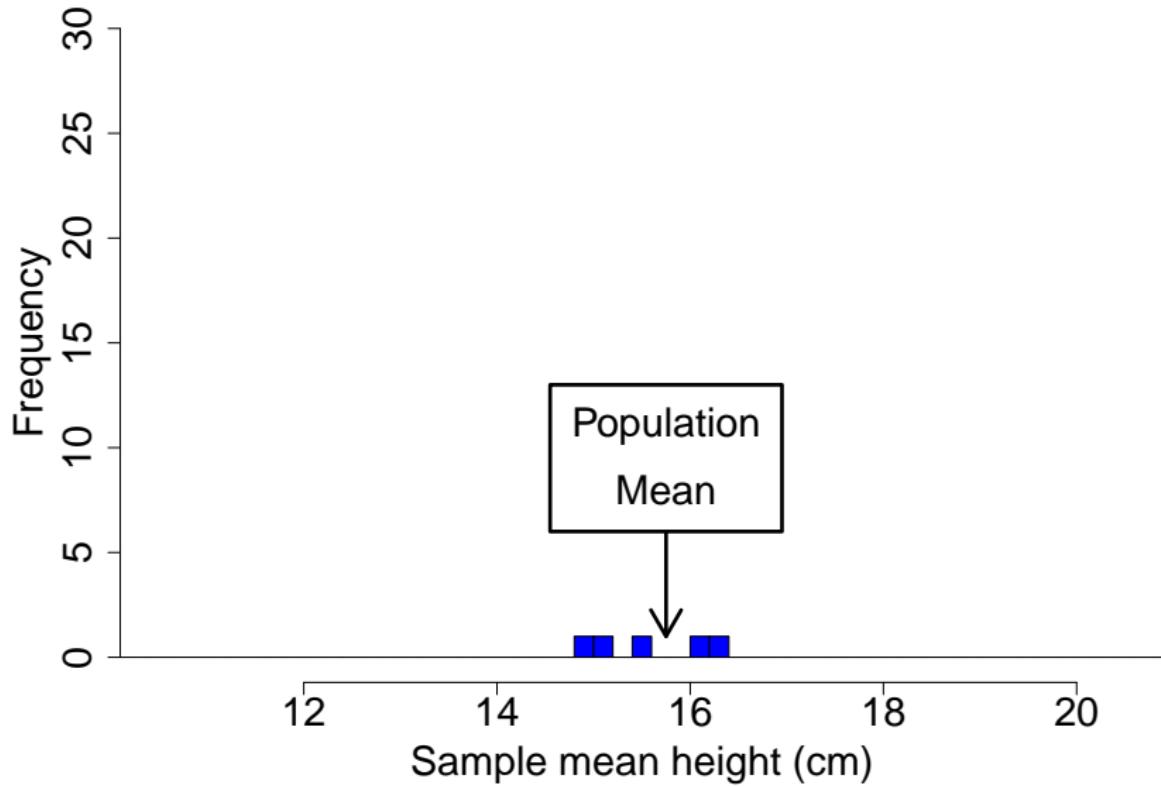
| <u>Samples</u>  | <u>Average height</u> |
|---|-----------------------|
|  | 15.53 cm              |
|  | 16.21 cm              |
|  | 16.09 cm              |
|  | 14.88 cm              |
|  | 15.06 cm              |

## Repeated re-sampling

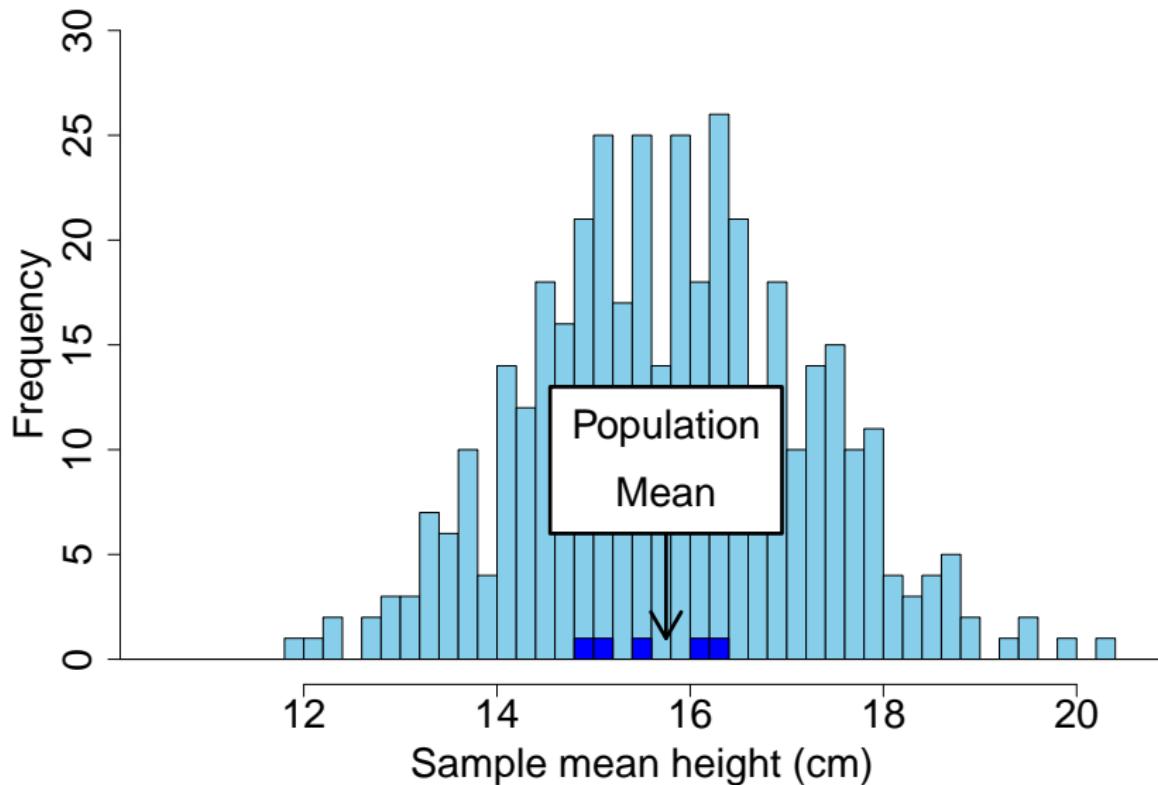
- ▶ What is our expectation?
- ▶ What is our uncertainty?

What is the **distribution** of the *sample mean* ( $\bar{x}$ ) around the *population mean* ( $\mu_x$ )?

## The standard error: Sample means and population mean



## The standard error: Sample means and population mean



## The standard error: Sample means and population mean

- ▶ What is the **distribution** of the *sample mean* ( $\bar{x}$ ) around the *population mean* ( $\mu_x$ )?
- ▶ **Standard error:** The standard deviation of sample means around the population mean

## The standard error: estimation

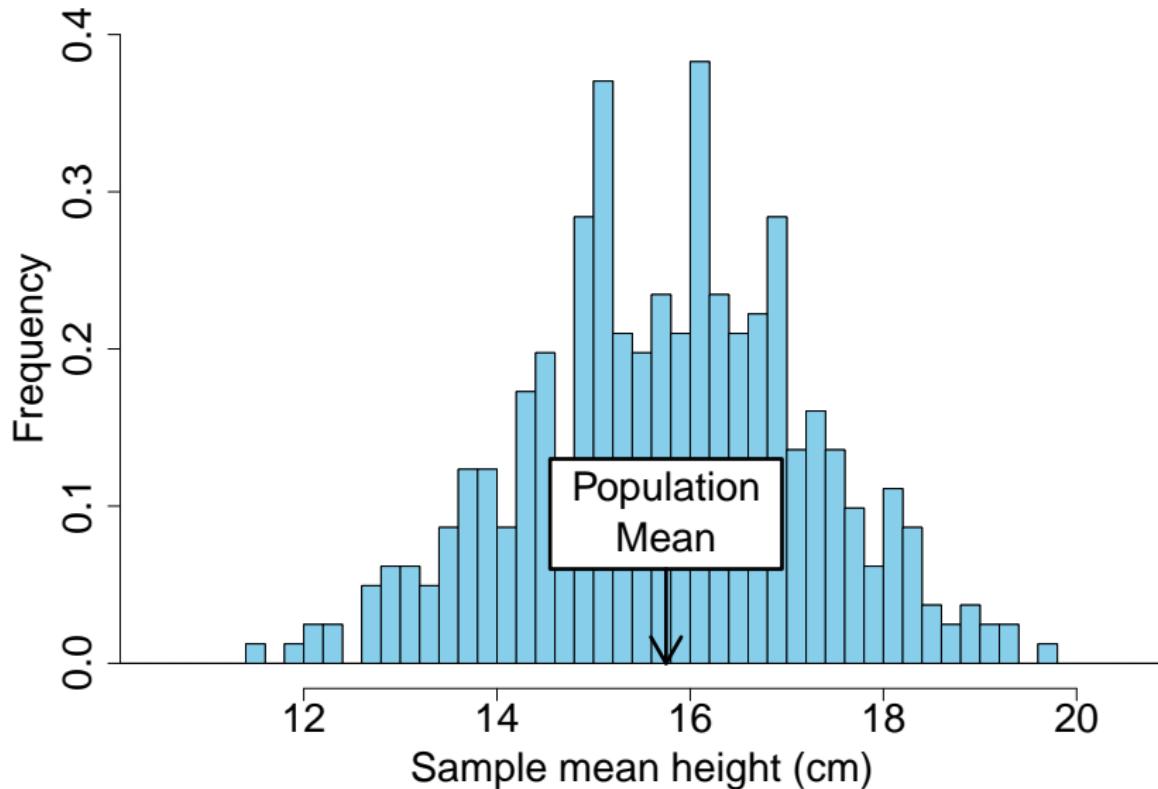
$$SE = \frac{\text{Sample standard deviation}}{\sqrt{\text{Sample size}}}$$

## The standard error: estimation

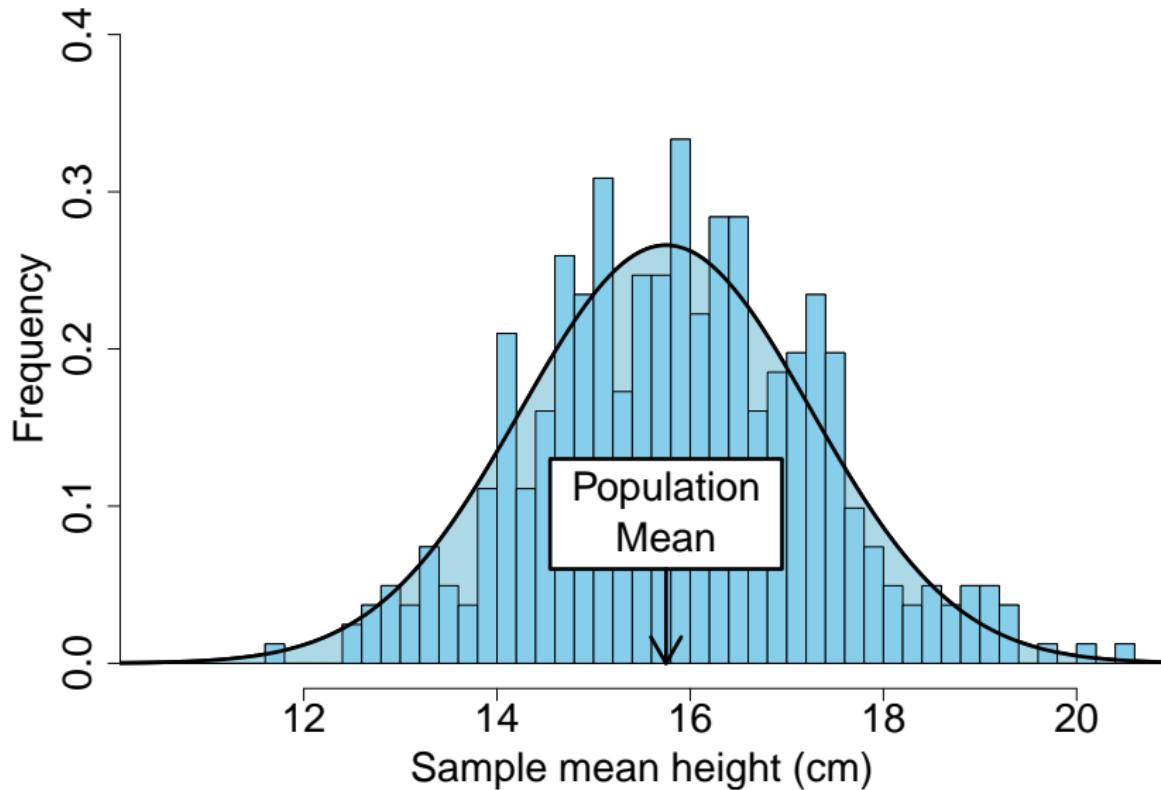
$$SE = \frac{\text{Sample standard deviation}}{\sqrt{\text{Sample size}}}$$

$$SE = \frac{s}{\sqrt{N}}$$

## The standard error



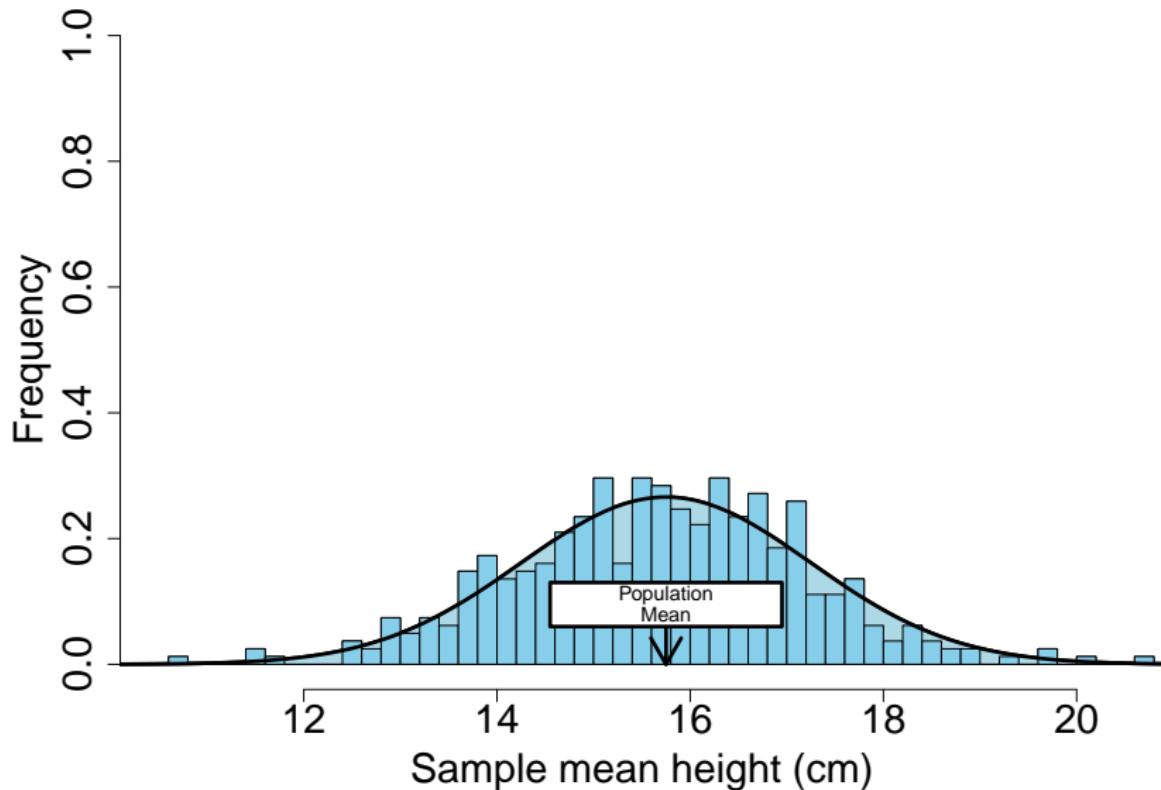
## The standard error



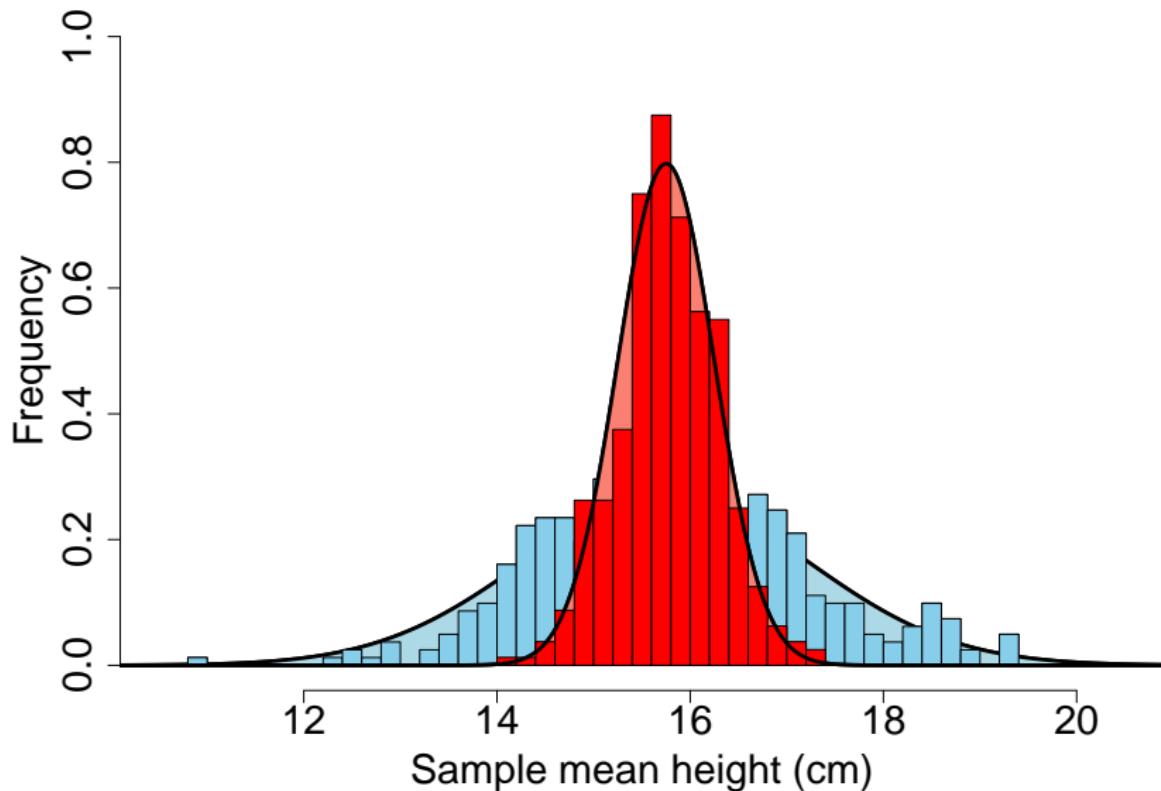
## The standard error: Sample means and population mean

- ▶ **Standard error:** The standard deviation of sample means around the population mean
- ▶ Standard error measures the uncertainty of the sample mean

## The standard error



## The standard error



# The standard error in jamovi

The screenshot shows the jamovi interface with the 'Descriptives' dialog on the left and the 'Results' table on the right.

**Descriptives Dialog:**

- Variables:** Soil organic carbon (g C / kg soil) is selected.
- Descriptives:** The 'Variables across columns' tab is selected.
- Statistics:** The 'Central Tendency' section is highlighted with a red box, containing:
  - N
  - Missing
  - Mean
  - Median
  - Mode
  - Sum
- Percentile Values:** Contains:
  - Cut points for 4 equal groups
  - Percentiles 25,50,75
- Dispersion:** Contains:
  - Std. deviation
  - Variance
  - Range
  - Std. error of Mean
  - Confidence interval for Mean 95 %
- Distribution:** Contains:
  - Minimum
  - Maximum
  - IQR
  - Skewness
  - Kurtosis
  - Shapiro-Wilk
- Normality:** Contains:
  - Shapiro-Wilk
- Outliers:** Contains:
  - Most extreme 5 values

**Results Table:**

### Descriptives

|                     | Soil organic carbon (g C / kg soil) |
|---------------------|-------------------------------------|
| N                   | 34                                  |
| Missing             | 0                                   |
| Mean                | 6.52353                             |
| Median              | 5.80000                             |
| Mode                | 2.40000 *                           |
| Standard deviation  | 4.49701                             |
| Variance            | 20.22307                            |
| IQR                 | 7.27500                             |
| Range               | 15.60000                            |
| Minimum             | 0.60000                             |
| Maximum             | 16.20000                            |
| Skewness            | 0.55655                             |
| Std. error skewness | 0.40305                             |
| Kurtosis            | -0.73034                            |
| Std. error kurtosis | 0.78790                             |
| 25th percentile     | 2.42500                             |
| 50th percentile     | 5.80000                             |
| 75th percentile     | 9.70000                             |

\* More than one mode exists, only the first is reported