الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعـليم العـــالي والبحث العلمي
جـامعة محمد بوضياف – المسيلة

جامعة محمد بوضياف - المسيلة
niversité Mohamed Boudiaf - M'sila

كلية العلـوم
قسم علـوم الطبيعة و الحياة

الرقم: 01 /ق ع ط ح/ ل ع / 2026 المسيلة في: 2026/06/14

شهادة موافقة علمية على مطبوعة بيداغوجية

للأستاذ : عسلوم عبد المجيد يعقوب " أستاذ محاضر- ب- "

يشهد رئيس اللجنة العلمية لقسم علوم الطبيعة و الحياة،كلية العلوم بجامعة محمد بوضياف ،انه بعد الاطلاع على تقارير الخبرة الواردة من طرف الخبراء :

- الأستـاذ : شترة محمد أستاذ بجامعة محمد بو ضياف - المسيلة.
- الأستـاذة: راشدي منيرة أستاذ التعليم العالي بجامعة الشاذلي بن جديد - الطارف

و المعينين من طرف اللجنة العلمية لقسم علوم الطبيعة و الحياة ، في الإجتمـاع المنعقد في دورته العادية يوم : 2026/03/03 تحت رقم : ل ع/2026/01 لإجراء الخبرة للمطبوعة البيداغوجية الخاصة بالأستاذ: **عسلوم عبد المجيد يعقوب** " أستاذ محاضر- ب - " بقسم علوم الطبيعة و الحياة - كلية العلوم . جامعة محمد بوضياف – المسيلة و المقررة في برنامج التكوين الخاص بطلبة السنة السنة الثالثة " **بيوتكنولوجيا و تطوير النبات** " ميدان علوم الطبيعة و الحياة بكلية العلوم ، المعنونة بـ: ***Biostatistics*** .

تمت الموافقة عليها شكلا و مضمونا.

رئيس اللجنة العلمية لقسم علوم الطبيعة والحياة

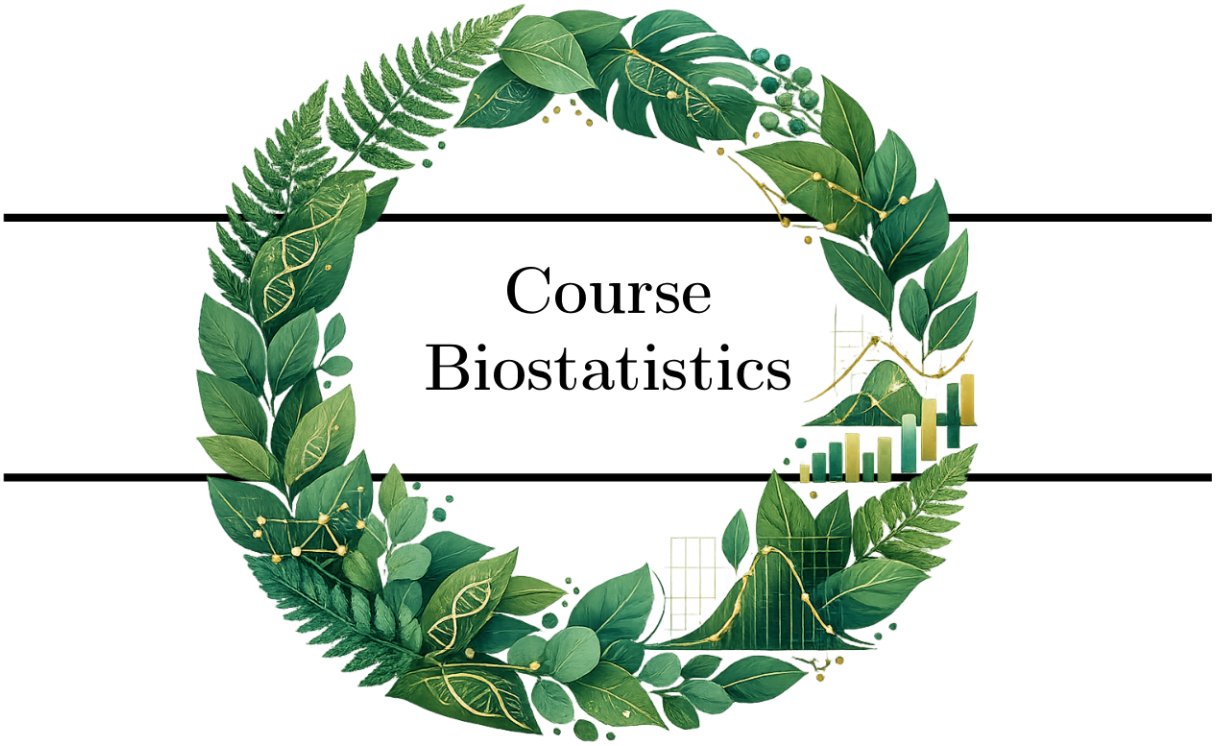
الأستاذ : خضور جمال

People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

MOHAMED BOUDIAF UNIVERSITY — M'SILA

Faculty of Science, Department of Nature and Life Science



Course
Biostatistics

Realized by:

Dr. ASLOUM Abdelmadjid Yagoub

Academic Year 2024–2025


بسم الله الرحمن الرحيم

Course Information

Course Instructor

Dr. ASLOUM Abdelmadjid Yagoub

This Course

Subject	Biostatistics
Credits	2
Coefficient	1
Teaching Unit	Discovery Teaching Unit 1: Biostatistics
Total hours	45 H00

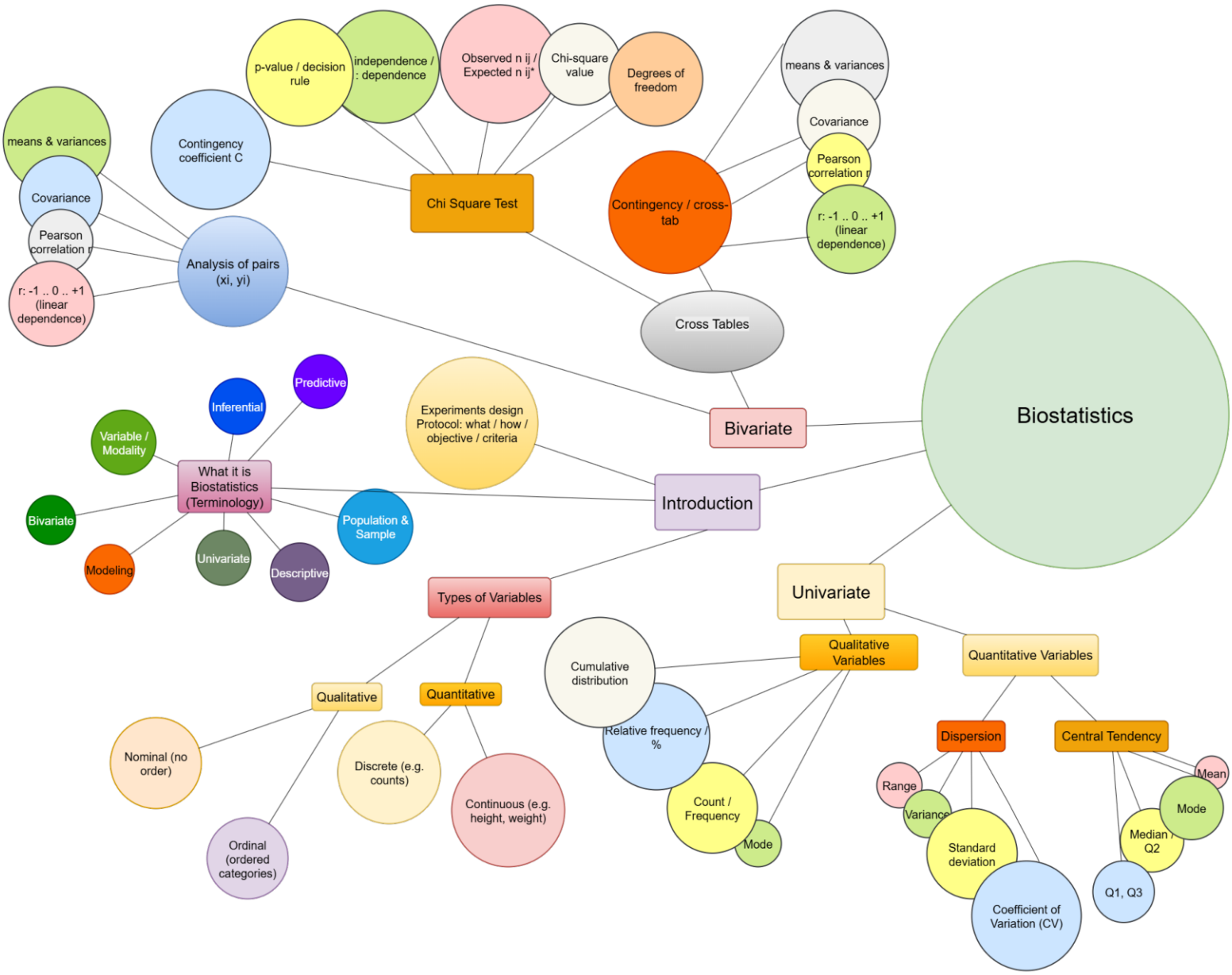
Course Objectives

Use of statistical tools in plant biotechnology.

Contact: abdelmadjid-yagoub.asloum@univ-msila.dz

Assessment Method: Continuous assessment 40 % / Final exam 60 %

Mind Map of the Course



Semestre : 5 **Unité d'enseignement :** UEM [O/P]

Matière 1 : Biostatistique

Crédits : 2, Coefficient : 1

Objectifs de l'enseignement :

Utilisation des outils statistiques en biotechnologie végétale.

Connaissances préalables recommandées :

Mathématiques de base.

Contenu de la matière :

0. Rappels des statistiques descriptives
0. Théorie de l'estimation
0. Modèles linéaires
0. Distribution d'échantillonnage
0. Initiation à un logiciel de traitement statistique

Mode d'évaluation : Contrôle continu / Examen final

Contents

1. Introduction

Every scientific study, experimental or clinical, is based on a clear protocol with four key elements. These elements guide the study design, make the work reproducible, and ensure reliable and comparable results. The protocol answers four main questions:

1.1 What Was Studied?

Clearly define the study material: laboratory plant species, crop varieties, cell cultures, or genetically modified plant organisms analysed.

Example: In a plant biotechnology study on *Solanum tuberosum* (potato), the study material would be defined as *in vitro* cultured nodal explants from the cultivar 'Desiree', maintained on MS medium for 4 weeks.

1.2 Which Method Was Used?

Describe how the data were collected—either *prospectively* (planned in advance) or *retrospectively* (using existing data).

Example: Data were collected prospectively by measuring shoot length every 7 days for 28 days in *in vitro* banana (*Musa* spp.) cultures. Alternatively, retrospective data collection would involve analysing historical records of transformation efficiency from previous *Agrobacterium*-mediated transformation experiments in tomato.

1.3 What Was the Study Objective?

Specify what the researchers aimed to evaluate: for example, a biological test with reference values, the effectiveness of a treatment, or the impact of a risk factor.

Example: The objective was to evaluate the effectiveness of different cytokinin types (BAP, Kinetin, 2iP) on shoot multiplication rate in *Phalaenopsis* orchid micropropagation, or to test the impact of salt stress on callus growth in *Triticum aestivum* (wheat) tissue culture.

1.4 What Evaluation Criteria Were Applied?

List the criteria used, such as disease presence or absence, treatment effectiveness and toxicity, recurrence rate, or survival. Indicate the statistical methods used to analyse these criteria.

Example: Evaluation criteria included: callus induction percentage (%), callus fresh weight (g), number of shoots per explant, shoot length (cm), rooting percentage (%), and root number per shoot. Statistical analysis was performed using ANOVA followed by Duncan's multiple range test at $p < 0.05$.

These essential elements help structure a study rigorously and generate reliable data, ready for analysis and comparison. Within this framework, scientific studies are generally divided into two main types: *non-experimental studies* and *experimental studies*.

Non-experimental studies do not involve direct interventions on the subjects. They rely on observing subjects in their natural state. For example, in survey-type studies, researchers may follow wild populations of *Medicago truncatula* across different altitudes to analyse the relationship between elevation and flavonoid content without manipulating environmental conditions. Such studies can identify correlations or associations without manipulating variables.

Experimental studies, on the other hand, involve direct intervention or manipulation. For instance, to test the effectiveness of a plant growth regulator, researchers may divide *in vitro* cultures into groups: one receives the standard treatment, while others receive different concentrations of the test compound. This approach allows the effect of the treatment to be evaluated by comparing outcomes between the groups.

In short, whether experimental or non-experimental, all studies aim to generate analysable data that answer specific research questions. This classification highlights

how the choice of methodology is crucial for the quality and relevance of the results, and underlines the importance of following a well-structured protocol to ensure scientific validity and comparability.

Statistics is the science that deals with the collection, organisation, analysis, interpretation, and presentation of data in order to draw meaningful conclusions and support decision-making.

Biostatistics is a branch of statistics that applies these methods specifically to biological, medical, ecological, and plant biotechnology-related data, helping researchers understand variability in living systems and evaluate scientific hypotheses in life sciences, ultimately leading to a reliable conclusion.

2. Terminology in Biostatistics

2.1 Statistical Modelling

This step is essential for any study, whether experimental or not, and must be followed by statistical analysis. Statistical modelling is the process of applying statistical methods to interpret and predict phenomena based on collected data.

Example in Plant Biotechnology: Developing a predictive model for somatic embryo maturation in *Picea abies* (Norway spruce) based on culture medium composition (ABA concentration, osmolarity) and incubation time. The model uses multiple regression to predict the probability of reaching the cotyledonary stage.

2.2 Descriptive Statistics

Methods used to summarise and organise data. These methods include frequency tables, graphs, and measures of central tendency (such as the mean), etc.

Example: Summarising the distribution of shoot lengths in *Solanum tuberosum* (potato) micropropagation across 500 explants using frequency tables, histograms, and calculation of mean, median, and standard deviation.

2.3 Inferential Statistics

Methods that allow us to draw conclusions about a population from a sample. They rely on probability theory to estimate the characteristics of the population and test hypotheses (for example, surveys).

Example: Estimating the average transformation efficiency of *Agrobacterium tumefaciens* in tomato (*Lycopersicon esculentum*) leaf explants based on a sample of 50 experiments, and using this to infer the expected efficiency for the entire population of possible experiments under similar conditions.

2.4 Predictive Statistics

Predictive statistics is a branch of statistics that uses historical and current data to build models that can predict future events or outcomes.

Example: Predicting the optimal harvest time for maximum taxol production in *Taxus chinensis* (Chinese yew) cell suspension cultures based on growth curve data, nutrient consumption rates, and previous harvest data.

2.5 Univariate Statistics

A study focusing on a single variable, whether quantitative or qualitative. Univariate statistics is a branch of descriptive statistics and concentrates on the distribution and characteristics of a single variable.

Example: Analysing only the ploidy level (diploid vs. haploid) in *Datura stramonium* anther culture regenerants, without considering other variables like plant height or flower morphology.

2.6 Statistical Population

A complete set of individuals or objects on which a statistical study is based. It is the set of observable elements.

Example: All *in vitro* propagated plantlets of *Phalaenopsis* (moth orchid) produced in a laboratory during 2024, or all possible callus cultures initiated from immature embryos of a specific wheat cultivar.

2.7 Sample

A subset of the population selected for the study. The sample is used to draw conclusions about the population as a whole.

Example: 30 randomly selected plantlets from the above *Phalaenopsis* population for measurement of fresh weight and leaf number, or 50 randomly selected wheat embryos for callus induction experiments.

2.8 Individual

A unique element of the statistical population being studied. Each individual possesses characteristics or variables that will be measured.

Example: One specific *Phalaenopsis* plantlet with its unique ID number, or one specific wheat embryo explant cultured in a Petri dish.

2.9 Variable or Statistical Characteristic

A measurable property or characteristic of an individual in the studied population. The possible values of a variable are called *terms* or *modalities*.

Examples in Plant Tissue Culture:

- Number of shoots per explant (discrete quantitative)
- Callus colour: white, yellow, brown, green (nominal qualitative)
- Regeneration capacity: high, medium, low (ordinal qualitative)
- Fresh biomass in grams (continuous quantitative)
- Presence of contamination: yes/no (nominal qualitative)

2.10 Bivariate Statistics

Bivariate statistics is a branch of descriptive statistics that simultaneously analyses two variables to understand the relationship between them. The main objective is to explore and quantify how one variable can influence or be associated with another.

Example: Analysing the relationship between explant size (mm) and callus formation frequency (%) in *Gossypium hirsutum* (cotton) cotyledon culture, or between 2,4-D concentration and somatic embryo number in carrot (*Daucus carota*) suspension cultures.

2.11 The Statistical Process Flow

The diagram shows the progression:

3. Variable Types

3.1 Quantitative Variable

A variable expressed by numbers and representing a measurable quantity.

3.1.1 Continuous Quantitative Variable

A variable that can take an infinite number of values within a given range, measured with arbitrary precision.

Plant Biotechnology Examples:

- Plant height in cm (e.g. 15.3, 18.7, 12.5)
- Fresh weight of callus in grams (e.g. 2.45, 3.12, 1.89)
- Chlorophyll content in mg/g FW (e.g. 1.234, 1.456)
- pH of culture medium (e.g. 5.75, 5.80, 5.65)
- Temperature in culture room in °C (e.g. 24.5, 25.0)
- Osmolarity in mOsm/kg (e.g. 350.5, 400.2)

Specific Example: In a study of *in vitro* growth of *Eucalyptus* microshoots, plant height was measured as 15.3, 18.7, 12.5, 20.1, and 16.8 cm for five different plantlets. These are continuous values that can take any value within a range.

3.1.2 Discrete Quantitative Variable

A variable that takes on a finite or countable number of values (often integers).

Plant Biotechnology Examples:

- Number of shoots per explant (0, 1, 2, 3, . . .)
- Number of flowers on a regenerated plant (3, 5, 8, . . .)
- Number of seeds that germinated in a sample (25, 30, 28, . . .)
- Number of roots formed on a cutting (2, 5, 8, . . .)
- Days to callus initiation (7, 10, 14, 21)
- Number of somatic embryos per callus (10, 15, 20, 25)
- Number of nodes on a shoot (3, 4, 5, 6)

Specific Example: In *Musa* (banana) micropropagation, the number of shoots per explant was recorded as 1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5 for 12 explants. These are countable whole numbers.

3.2 Qualitative Variable

A variable that does not represent a numerical quantity but categories or qualities.

3.2.1 Nominal Qualitative Variable

A variable whose modalities are categories without any particular order.

Plant Biotechnology Examples:

- Flower colour in petunia (white, pink, purple, red)
- Place of origin of plant material (Algeria, France, USA, Japan)
- Explant type (leaf, stem, root, cotyledon, hypocotyl)
- Presence or absence of somaclonal variation (Yes, No)
- *Agrobacterium* strain used (EHA105, LBA4404, GV3101)
- Type of culture vessel (Petri dish, Erlenmeyer flask, Magenta box)
- Resistance to antibiotic (Resistant, Sensitive)

Specific Example: In a genetic transformation study of *Solanum lycopersicum* (tomato), the explant type used was categorised as: leaf (45 experiments), cotyledon

(30 experiments), and hypocotyl (25 experiments). These categories have no natural ordering.

3.2.2 Ordinal Qualitative Variable

A variable whose categories can be logically ordered.

Plant Biotechnology Examples:

- Callus quality (poor, fair, good, excellent)
- Level of hyperhydricity (none, mild, moderate, severe)
- Rooting efficiency (low, medium, high)
- Somatic embryo maturation stage (globular, heart, torpedo, cotyledonary)
- Tolerance to salinity in transgenic lines (sensitive, moderately tolerant, highly tolerant)
- Contamination level (clean, slightly contaminated, heavily contaminated)
- Shoot vigour (weak, moderate, strong)

Specific Example: In somatic embryogenesis of *Daucus carota* (carrot), embryo development stages were classified as: globular (50 embryos), heart (60), torpedo (70), and cotyledonary (20). These stages follow a logical developmental sequence.

4. Univariate Biostatistics

This part of the course covers statistical methods applied to a single variable at a time. The following two chapters detail the treatment of qualitative variables and quantitative variables separately.

4.1 Qualitative Variables

4.2 Definition

A qualitative variable is a variable that takes non-numerical values and describes a quality, category, or attribute.

Plant Biotechnology Examples:

- **Sex in dioecious plants:** Male, Female (in *Spinacia oleracea* — spinach)
- **Grafting compatibility:** Compatible, Incompatible, Partially compatible
- **Presence of contamination:** Yes, No (in *in vitro* cultures)
- **Flower colour:** Blue, Yellow, White (in *Ipomoea* species)
- **Explant response:** Callus formation, Direct organogenesis, No response
- **Transgene expression:** Positive, Negative (GUS assay)

4.3 Types of Qualitative Variables

4.3.1 Nominal Qualitative

The categories have no order or hierarchy.

Examples:

- Flower colours in transgenic *Rosa* (Blue, Yellow, Green, White)
- Types of growth regulators used (Auxin, Cytokinin, Gibberellin, Abscisic acid)
- Genetic transformation method (*Agrobacterium*-mediated, Biolistic, Electroporation, Protoplast fusion)
- Cultivar names (Grande Naine, Williams, Poyo in banana)

Specific Example: In a study of genetic diversity in *Olea europaea* (olive), the cultivars analysed were: Chemlal, Sigoise, Azeradj, and Limli. These are nominal categories with no inherent order.

4.3.2 Ordinal Qualitative

The categories can be classified in a logical order.

Examples:

- **Oxidative stress level:** Mild, Moderate, Severe, Lethal
- **Embryo development stage:** Globular → Heart → Torpedo → Cotyledonary
- **Acclimatisation success:** Failed → Surviving → Growing → Fully established
- **Callus texture:** Friable, Compact, Hard

Specific Example: In anther culture of *Hordeum vulgare* (barley), the response was classified as: no response (20 anthers), callus formation only (35), green spots (15), and plant regeneration (10). These represent increasing levels of successful response.

4.4 Descriptive Measures for a Qualitative Variable

4.4.1 Absolute Frequency

The number of observations in each category.

Example: Response of wheat (*Triticum aestivum*) immature embryos to tissue culture (100 explants):

Response Type	**Number of Explants**
Callus formation	40
Direct shoot regeneration	30
Root formation only	20
No response	10
Total	**100**

4.4.2 Relative Frequency (f)

Proportion of observations in each category:

$$f_i = \frac{n_i}{N}$$

For the wheat example:

- Callus formation: $f = 40/100 = 0.4$
- Direct shoot regeneration: $f = 30/100 = 0.3$
- Root formation only: $f = 20/100 = 0.2$
- No response: $f = 10/100 = 0.1$

4.4.3 Percentage Proportion ($P\%$)

Multiply the relative frequency by 100:

- Callus formation: $f = 40/100 = 0.4 \Rightarrow P = 40\,\%$

- Direct shoot regeneration: $f = 30/100 = 0.3 \Rightarrow P = 30\,\%$
- Root formation only: $f = 20/100 = 0.2 \Rightarrow P = 20\,\%$
- No response: $f = 10/100 = 0.1 \Rightarrow P = 10\,\%$

4.4.4 Cumulative Distribution

Especially useful for ordinal qualitative variables. It is the running total of the frequencies up to a certain category.

Example: Somatic embryo development stages in *Daucus carota* suspension culture (200 embryos):

Stage	**Count**	**Cumulative Count**	**Cumulative %**
Globular	50	50	25 %
Heart	60	110	55 %
Torpedo	70	180	90 %
Cotyledonary	20	200	100 %

Interpretation: 55 % of embryos have reached at least the heart stage, and 90 % have reached at least the torpedo stage.

4.4.5 Complete Statistical Table

Response Type	**Count**	**Cumulative**	**Frequency**	**%**
Callus formation	40	40	0.4	40
Direct shoot regen.	30	70	0.3	30
Root formation only	20	90	0.2	20
No response	10	100	0.1	10
Total	**100**		**1.0**	**100**

4.4.6 Mode

The most frequent category (the one with the highest frequency).

Example: If 40 explants showed callus formation out of 100, then **Callus formation** is the mode.

Additional Examples:

Example 1: Contamination types in 150 *Prunus* (cherry) cultures:

- Bacterial contamination: 45 cases
- Fungal contamination: 60 cases (**Mode**)
- Yeast contamination: 30 cases
- No contamination: 15 cases

Example 2: Graft compatibility in 80 citrus combinations:

- Compatible: 50 cases (**Mode**)
- Partially compatible: 20 cases
- Incompatible: 10 cases

Example 3: Ploidy level in 100 *Datura stramonium* regenerants:

- Haploid: 25 plants
- Diploid: 60 plants (**Mode**)
- Tetraploid: 15 plants

4.5 Quantitative Variables

A quantitative variable is a variable that takes numerical values. They are measurable and allow for mathematical or statistical calculations. They differ from qualitative variables in that they express a quantity or a magnitude.

4.6 Characteristics of a Quantitative Variable

- It allows you to measure or count.

- Mathematical operations (addition, subtraction, average, standard deviation, etc.) are possible.
- The difference between two values has meaning (for example, a 5 g difference between two callus samples).

4.7 Types of Quantitative Variables

4.7.1 Continuous Quantitative

Takes values within a range (decimals allowed), measured with arbitrary precision.

Plant Biotechnology Examples:

- Height of *in vitro* plantlets in cm: 15.3, 18.7, 12.5, 20.1
- Fresh weight of hairy roots in g: 2.45, 3.12, 1.89, 4.56
- Chlorophyll *a* content in mg/g FW: 1.234, 1.456, 0.987, 1.678
- Medium pH: 5.75, 5.80, 5.65, 5.90
- Osmolarity in mOsm/kg: 350.5, 400.2, 375.8, 425.3
- Leaf area in cm^2: 3.45, 4.12, 5.67, 6.89

Specific Example: Fresh weight measurements of *Brassica napus* (rapeseed) callus cultures after 4 weeks: 2.45, 3.12, 1.89, 4.56, 3.78, 2.91, 3.45, 4.01 g.

4.7.2 Discrete Quantitative

Takes finite or countable values (often integers).

Plant Biotechnology Examples:

- Number of shoots per explant: 0, 1, 2, 3, 4, 5, 6, 7
- Number of somatic embryos per callus: 10, 15, 20, 25, 30
- Days to root initiation: 7, 10, 14, 21, 28
- Number of nodes on a shoot: 3, 4, 5, 6, 7, 8
- Number of true leaves: 2, 4, 6, 8, 10

- Number of anthers cultured per Petri dish: 20, 25, 30

Specific Example: Number of adventitious shoots formed on *Petunia* leaf explants after 6 weeks of culture: 3, 5, 2, 8, 6, 4, 7, 5, 6, 4, 3, 9 shoots per explant.

4.8 Descriptive Measures for a Quantitative Variable

4.8.1 Measures of Central Tendency

Arithmetic Mean

The value that each individual should have if the total were shared equally.

For a discrete quantitative variable (frequency table):

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

For a continuous quantitative variable (grouped data):

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{k} c_i \, n_i$$

where c_i is the class centre.

Example 1: Mean number of shoots per explant in *Musa* (banana) micropropagation:

Shoots per explant (x_i)	**Frequency** (n_i)
1	5
2	12
3	20
4	15
5	8
Total	$N = 60$

$$\bar{x} = \frac{(15) + (212) + (320) + (415) + (58)}{60} = \frac{5 + 24 + 60 + 60 + 40}{60} = \frac{189}{60} = 3.15$$

Interpretation: On average, each banana explant produces 3.15 shoots.

Example 2: Fresh weight of callus (g) from 50 *Brassica napus* explants (grouped data):

Weight Class (g)	**Class Centre** (c_i)	**Frequency** (n_i)
[0.5, 1.5[	1.0	8
[1.5, 2.5[	2.0	15
[2.5, 3.5[	3.0	18
[3.5, 4.5[	4.0	9
Total		$N = 50$

$$\bar{x} = \frac{(1.08) + (2.015) + (3.018) + (4.09)}{50} = \frac{8 + 30 + 54 + 36}{50} = \frac{128}{50} = 2.56 \text{ g}$$

Median

The value that separates the lower and upper halves of the data.

For a discrete quantitative variable:

- If N is even ($N = 2P$): $Me = \dfrac{X_P + X_{P+1}}{2}$
- If N is odd ($N = 2P + 1$): $Me = X_{P+1}$

Example: Root length (mm) in 7 *Arabidopsis thaliana* plantlets: 12, 15, 18, 20, 22, 25, 30.

Ordered data: 12, 15, 18, **20**, 22, 25, 30.

$N = 7$ (odd), $P = 3$, so $Me = X_4 = 20\,\text{mm}$.

For a continuous quantitative variable (grouped data):

$$Me = B_I + (B_S - B_I)\,\frac{\dfrac{N}{2} - n_{ic-1}}{n_i c - n_{ic-1}}$$

where:

- B_I = lower bound of the median class
- B_S = upper bound of the median class
- n_{ic-1} = cumulative effective before the median class
- n_ic = cumulative effective up to the median class

Example (callus weight data):

- $N/2 = 25$
- Cumulative effective: 8, 23, 41, 50
- Median class: $[2.5,\ 3.5[$ (since $23 < 25 \leq 41$)
- $B_I = 2.5, \quad B_S = 3.5, \quad n_{ic-1} = 23, \quad n_ic = 41$

$$Me = 2.5 + (3.5 - 2.5)\frac{25 - 23}{41 - 23} = 2.5 + 1\frac{2}{18} = 2.61 \text{ g}$$

Quartiles

A quartile is one of three values Q_1, Q_2 (median), and Q_3 that divide the data into four equal parts.

For a discrete variable:

$$\text{Rank}_{Q_1} = \frac{N}{4}, \qquad \text{Rank}_{Q_3} = \frac{3N}{4}$$

For a continuous variable (grouped data):

$$Q_1 = B_I + (B_S - B_I)\,\frac{\frac{N}{4} - n_{ic-1}}{n_ic - n_{ic-1}}, \qquad Q_3 = B_I + (B_S - B_I)\,\frac{\frac{3N}{4} - n_{ic-1}}{n_ic - n_{ic-1}}$$

Example: Days to flowering in 40 transgenic *Nicotiana tabacum* (tobacco) plants:

- $Q_1 = 45$ days (25 % of plants flowered by day 45)
- $Q_2 = 52$ days (50 % by day 52)
- $Q_3 = 58$ days (75 % by day 58)

Interpretation: The interquartile range IQR $= Q_3 - Q_1 = 13$ days indicates moderate variability. The middle 50 % of plants flowered between days 45 and 58.

Mode

The most frequent value, i.e. the modality x_i with the largest number of occurrences.

For a continuous variable (grouped data):

$$\text{Mod} = B_I + a\,\frac{n_{\text{co}} - n_{\text{co}-1}}{(n_{\text{co}} - n_{\text{co}-1}) + (n_{\text{co}} - n_{\text{co}+1})}$$

where:

- B_I = lower bound of the modal class
- a = class width
- n_{co} = effective of the modal class
- $n_{\text{co}-1}$ = effective of the preceding class
- $n_{\text{co}+1}$ = effective of the following class

Example (callus weight): The modal class is $[2.5,\ 3.5[$ with frequency 18.

$$\text{Mod} = 2.5 + 1\frac{18 - 15}{(18 - 15) + (18 - 9)} = 2.5 + \frac{3}{12} = 2.75 \text{ g}$$

4.8.2 Measures of Dispersion

Range (Extent)

Difference between the largest and smallest values:

$$e = x_{\max} - x_{\min}$$

Example: Shoot length in *Petunia* micropropagation: $x_{\min} = 2.3$ cm, $x_{\max} = 8.7$ cm, so $e = 6.4$ cm.

Variance and Standard Deviation

They indicate the variability of the data around the mean.

Variance:

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

Standard deviation:

$$\sigma = \sqrt{\sigma^2}$$

Example 1: Number of roots in 5 *Ficus* (fig) microcuttings: 3, 5, 4, 6, 2.

1. Mean: $\bar{x} = (3+5+4+6+2)/5 = 4$
2. Squared deviations: $(3-4)^2 = 1$, $(5-4)^2 = 1$, $(4-4)^2 = 0$, $(6-4)^2 = 4$, $(2-4)^2 = 4$
3. Variance: $\sigma^2 = (1+1+0+4+4)/5 = 2$
4. Standard deviation: $\sigma = \sqrt{2} \approx 1.41$

Interpretation: The average number of roots is 4 ± 1.41.

Example 2: Shoot length (cm) in 10 *Solanum tuberosum* plantlets: 5.2, 6.1, 4.8, 5.5, 6.3, 5.0, 5.8, 6.0, 5.3, 5.9.

- $\bar{x} = 5.59\,\text{cm}$
- $\sum(x_i - \bar{x})^2 = 2.009$
- $\sigma^2 = 2.009/10 = 0.2009\,\text{cm}^2$
- $\sigma = 0.448\,\text{cm}$

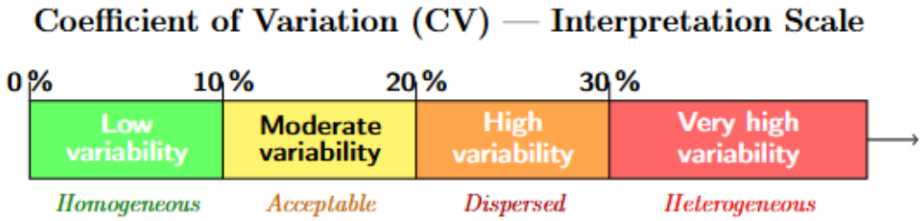
Coefficient of Variation (CV)

The coefficient of variation expresses the standard deviation as a proportion of the mean, allowing comparison across data sets with different units or scales:

$$CV = \frac{\sigma}{\bar{x}} 100\,\%$$

Common categorisation:

- $CV < 10\,\%$: Low variability (homogeneous data)
- $CV = 10$–$20\,\%$: Moderate variability
- $CV = 20$–$30\,\%$: High variability
- $CV > 30\,\%$: Very high variability (heterogeneous data)



Example — Comparing two genotypes:

Genotype A (Wheat): Mean $= 65\,\%$, SD $= 8\,\%$ $\Rightarrow$ $CV_A = 8/65100 = 12.3\,\%$

Genotype B (Barley): Mean $= 42\,\%$, SD $= 12\,\%$ $\Rightarrow$ $CV_B = 12/42100 = 28.6\,\%$

Conclusion: Genotype B shows higher relative variability in regeneration capacity despite lower absolute standard deviation.

Comparing shoot multiplication across species:

- *Musa* spp.: Mean $= 4.2$, SD $= 0.8$, $CV = 19.0\,\%$
- *Solanum*: Mean $= 3.1$, SD $= 0.6$, $CV = 19.4\,\%$
- *Phalaenopsis*: Mean $= 2.8$, SD $= 0.9$, $CV = 32.1\,\%$

Phalaenopsis shows the highest relative variability in multiplication rate.

5. Bivariate Biostatistics

5.1 Introduction

The variables x and y can be analysed separately. We can calculate all parameters including means and variances.

Note: These parameters are called *marginal parameters*: marginal variances, marginal means, marginal standard deviations, marginal quantiles, etc.

Plant Biotechnology Context: We often need to analyse relationships between two variables, such as:

- Explant size and regeneration efficiency
- Hormone concentration and callus growth
- Culture duration and secondary metabolite accumulation
- Gene expression level and phenotypic trait
- Fresh weight and chlorophyll content

5.2 Analysis of Pairs (x_i, y_i)

5.2.1 Relationship Between X and Y

The main objective is to determine whether X and Y are:

- **Independent:** no significant link between the two variables.
- **Correlated:** a linear or non-linear relationship exists.

5.2.2 Data Organisation

A statistical table groups the n observations in the form of pairs.

Example: Relationship between 2,4-D concentration (μM) and callus fresh weight (g) in *Oryza sativa* (rice) anther culture:

2,4-D Concentration (X)	**Callus Weight in g (Y)**
0.5	0.12
1.0	0.28
1.5	0.45
2.0	0.62
2.5	0.75
3.0	0.82
3.5	0.78
4.0	0.65

5.2.3 Calculations on the Double Series

Mean and Variance for Each Variable

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad V(x) = \overline{x^2} - \bar{x}^2$$

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i, \qquad V(y) = \overline{y^2} - \bar{y}^2$$

Example: Using the rice anther culture data:

$\bar{x} = (0.5 + 1.0 + 1.5 + 2.0 + 2.5 + 3.0 + 3.5 + 4.0)/8 = 2.25\ \mu$M

$\bar{y} = (0.12 + 0.28 + 0.45 + 0.62 + 0.75 + 0.82 + 0.78 + 0.65)/8 = 0.559$ g

Covariance Between X and Y

$$\mathrm{Cov}(x, y) = \frac{1}{N}\sum_{i=1}^{N} x_i\, y_i - \bar{x}\,\bar{y}$$

Interpretation:

- $\text{Cov}(x, y) > 0$: positive relationship
- $\text{Cov}(x, y) < 0$: negative relationship
- $\text{Cov}(x, y) = 0$: linear independence

Calculation:

$$\sum x_i y_i = (0.50.12) + (1.00.28) + \cdots + (4.00.65) = 11.235$$

$$\text{Cov}(x, y) = \frac{11.235}{8} - (2.250.559) = 1.404 - 1.258 = 0.146$$

Conclusion: Positive covariance indicates that as 2,4-D concentration increases, callus weight tends to increase (up to a point, then decreases).

Second Example: Temperature (X, °C) vs. shoot length (Y, cm) in *Brassica oleracea* culture:

Temperature	**Shoot Length**
20	3.2
22	4.1
24	5.5
26	6.8
28	7.2
30	6.5
32	5.1

$\text{Cov}(x, y) = 2.34$ (positive relationship up to optimum, then negative).

5.3 Analysis of Cross-Tabulations

Contingency table analysis (cross-tabulation) is a statistical method used to examine the relationships between two qualitative or categorical variables.

5.3.1 Cross-Tables of Two Quantitative Variables

A cross-table presents two statistical series simultaneously. The $n_{i.}$ and $n_{.j}$ are called *marginal frequencies.*

	y_1	y_2	$\cdots$	y_p	Total
x_1	n_{11}	n_{12}	$\cdots$	n_{1p}	$n_{1\cdot}$
x_2	n_{21}	n_{22}	$\cdots$	n_{2p}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}	$\cdots$	n_{kp}	$n_{k\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot p}$	N

Where:

- $n_{i\cdot} = \sum_{j=1}^{p} n_{ij}$: total for row i
- $n_{\cdot j} = \sum_{i=1}^{k} n_{ij}$: total for column j
- n_{ij} : frequency for cell (i, j)
- $N = n_{\cdot\cdot}$: grand total

Example: Explant type vs. regeneration response in *Solanum tuberosum* (potato):

Explant Type	**Direct Shoot**	**Callus**	**No Response**	**Total**
Leaf	25	15	10	50
Stem	30	12	8	50
Petiole	20	20	10	50
Total	75	47	28	150

Means of Marginal Distributions

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{k} n_{i\cdot}\, x_i, \qquad \bar{y} = \frac{1}{N} \sum_{j=1}^{p} n_{\cdot j}\, y_j$$

Covariance

$$\mathrm{Cov}(x, y) = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{p} n_{ij}\, x_i\, y_j \; - \; \bar{x}\, \bar{y}$$

Pearson's Linear Correlation Coefficient (r)

$$r(x, y) = \frac{\mathrm{Cov}(x, y)}{\sigma(x)\, \sigma(y)}$$

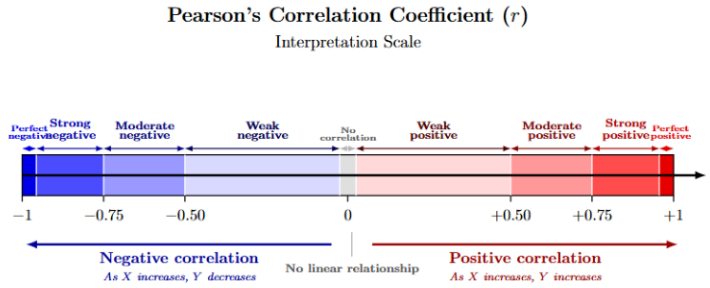
Example: Fresh weight vs. chlorophyll content in *in vitro* grapevine (*Vitis vinifera*) plantlets:

- $\mathrm{Cov}(x, y) = 0.45$
- $\sigma(x) = 0.8\,\mathrm{g}, \quad \sigma(y) = 0.6\,\mathrm{mg/g}$
- $r = 0.45/(0.80.6) = 0.9375$

Interpretation: $r \approx 0.94$ indicates a very strong positive correlation.

Important Notes:

- $-1 \leq r \leq 1$.
- When $x_i = y_i$ for all i, the covariance equals the variance.
- $r = 0$ implies no *linear* dependence; a non-linear relationship may still exist.
- $r > 0$: points aligned along an increasing line.
- $r < 0$: points aligned along a decreasing line.



Additional Examples:

- IBA concentration vs. rooting percentage in *Olea europaea*: $r = 0.85$ (strong positive).
- Culture duration vs. remaining sucrose in *Taxus* cell suspension: $r = -0.92$ (strong negative).

5.3.2 The χ^2 Test

Test Objective

The χ^2 test analyses relationships between two qualitative variables. It tests whether the two variables are independent (no association) or dependent (significant association).

Applications: Testing whether regeneration response depends on explant type, contamination rate depends on sterilisation method, or somaclonal variation depends on genotype.

Hypotheses

H_0**:** The two variables are independent.

H_1**:** The two variables are dependent.

Observed frequencies (n_{ij}): actual counts from the cross-tabulation.

Expected frequencies (n^*_{ij}):

$$n^*_{ij} = \frac{n_{i\cdot} n_{\cdot j}}{N}$$

Example: Expected frequency for Leaf/Direct Shoot in the potato table: $n^*_{11} = 5075/150 = 25$.

Test Statistic

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - n^*_{ij})^2}{n^*_{ij}}$$

Degrees of freedom: $df = (L-1)(C-1)$, where L = number of rows, C = number of columns.

Decision rule: Compare χ^2_{calc} with χ^2_{crit} at significance level α and df degrees of freedom.

- $\chi^2_{\text{calc}} > \chi^2_{\text{crit}} \Rightarrow$ reject H_0 (variables are dependent).
- Otherwise: do not reject H_0.

Complete Example: Association between genotype and somaclonal variation in banana (*Musa* spp.):

Genotype	**Normal**	**Variant**	**Total**
Grande Naine	45	5	50
Williams	40	10	50
Poyo	48	2	50
Total	133	17	150

Expected frequencies (all rows share the same marginal totals):

$$n^*_{\text{Normal}} = \frac{50133}{150} = 44.33, \qquad n^*_{\text{Variant}} = \frac{5017}{150} = 5.67$$

$$\begin{aligned}\chi^2 &= \frac{(45-44.33)^2}{44.33} + \frac{(5-5.67)^2}{5.67} + \frac{(40-44.33)^2}{44.33} \\ &\quad + \frac{(10-5.67)^2}{5.67} + \frac{(48-44.33)^2}{44.33} + \frac{(2-5.67)^2}{5.67} \\ &= 0.01 + 0.08 + 0.42 + 3.30 + 0.30 + 2.37 = 6.48\end{aligned}$$

$df = (3-1)(2-1) = 2$. Critical value at $\alpha = 0.05$, $df = 2$: $\chi^2_{\text{crit}} = 5.99$.

Conclusion: $6.48 > 5.99 \Rightarrow$ reject H_0. There is a significant association between genotype and somaclonal variation.

Second Example: Sterilisation method vs. contamination in *Prunus* cultures:

Method	**Contaminated**	**Clean**	**Total**
A ($HgCl_2$)	8	42	50
B (NaOCl)	15	35	50
C (Heat)	25	25	50
Total	48	102	150

$\chi^2 = 12.35, \quad df = 2, \quad \chi^2_{\text{crit}} = 5.99.$

Conclusion: $12.35 > 5.99 \Rightarrow$ significant association. Sterilisation method affects contamination rate.

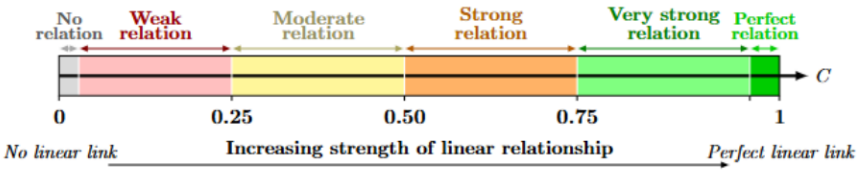
Contingency Coefficient (C)

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

Properties:

- $0 \leq C < 1$.
- $C = 0$: complete independence.
- C close to 1: strong association.
- C can never reach 1 unless the table is square.

Interpretation of Pearson's Correlation Coefficient (C)



Banana example:

$$C = \sqrt{\frac{6.48}{6.48 + 150}} = \sqrt{\frac{6.48}{156.48}} = \sqrt{0.041} = 0.203$$

Interpretation: $C = 0.20$ indicates a weak to moderate association between genotype and somaclonal variation.

Sterilisation example:

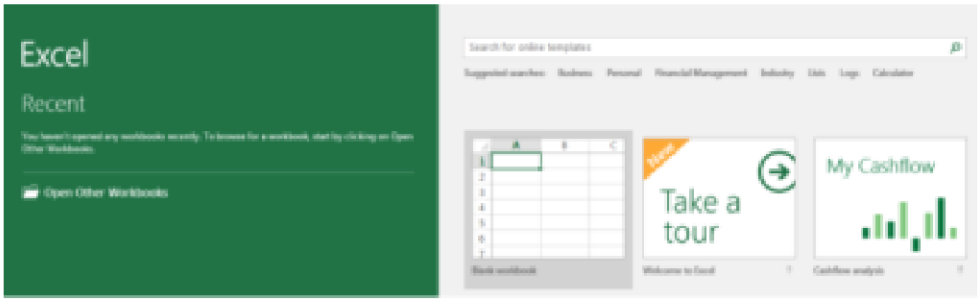
$$C = \sqrt{\frac{12.35}{12.35 + 150}} = \sqrt{\frac{12.35}{162.35}} = \sqrt{0.076} = 0.276$$

Moderate association between sterilisation method and contamination.

6. Graphical Presentation

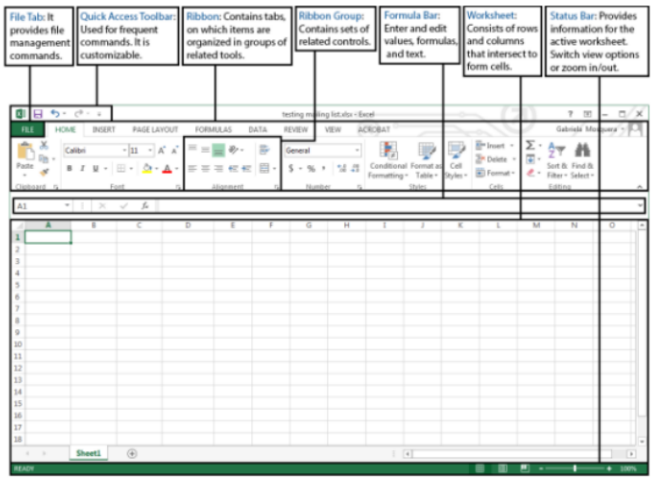
6.1 Microsoft Excel 2016

Introduction: Excel (full name: Microsoft Excel) is a spreadsheet application developed by Microsoft. It is part of the Microsoft Office suite (now Microsoft 365). Excel allows users to organize, format, and calculate data using formulas, functions, charts, pivot tables, and macros (via Visual Basic for Applications). Common uses include data analysis, accounting, budgeting, project management, and creating visualizations of numerical information.In biology, Excel is widely used for everything from basic data entry to complex statistical tests. Interface: The Excel interface is where you see and use the tools in Excel on the screen. This includes the way the tools are organised and presented to you, the software user.



You will learn about:

- The Welcome Page
- The Ribbon
- Quick Access Toolbar
- File Tab



6.1.1 The Welcome Page

When you first open Excel 2016, you will see the Welcome Page. Take a moment to browse the many templates available for specific uses. Notice the Search Box near the top-centre where you can search for templates.

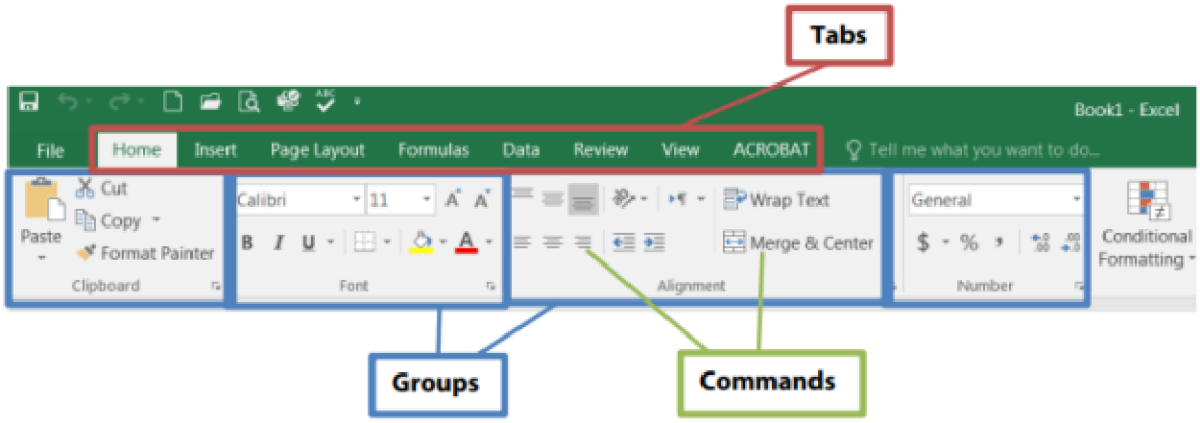
Tip: If you have a Microsoft account (such as Outlook, Hotmail, or Office 365), you can log in near the upper-right corner of the Welcome Page where it reads, "Sign in to get the most out of Office". Doing this allows you to save your work to "OneDrive", and then access it from any other computer with Internet connection. For example, you start a document in the library, save it to your Microsoft account, and retrieve it at home or work later. This is known as "saving to the cloud," or more formally "cloud computing".

Try: Open a new "Blank workbook" from the Welcome Page. Simply click the template with that name. You will be working in this blank workbook.

6.1.2 The Ribbon

The Ribbon is a toolbox at the top of the screen. It is organised into three main parts:

1. **Tabs** — Tabs represent a general activity area. For example, the "Home" tab has the tools most often used, and the "Insert" tab has the tools to "put objects into" the work area.
2. **Groups** — Groups show related "tools" together more specifically, like "Font" or "Alignment".
3. **Commands** — A command is one of the actual "tools", which can be a button, expandable menu, or a box for entering information.

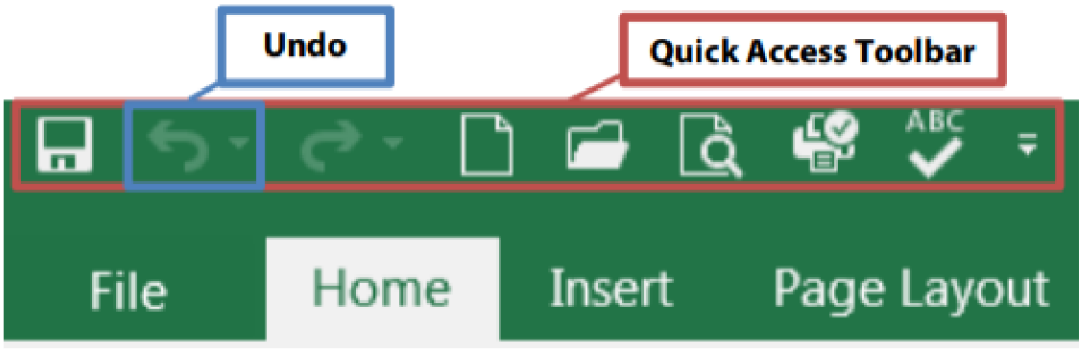


Try it: Click the various tabs and observe how the Groups and Commands change. Let the mouse arrow rest over a command (icon) and the name and description of that command will appear.

Tip: If you have a question about Excel, type it into the box labelled "Tell me what you want to do. . . "

6.1.3 Quick Access Toolbar

The Quick Access Toolbar is above the Tabs and has the commands used most often, such as Save, Undo, and Redo.



Try it: Click into any of the blank rectangles (cells) in the large white grid area (worksheet). Type your name, then click the "Undo" button. The Undo command "takes back" any changes made in the document. Use it when you want to "go back" a few steps.

Important: It is important to "save early and save often".

6.1.4 File Tab

The File Tab is where you can:

- Create a New document
- Open an existing one
- Save changes
- Save As a different file with a different name
- Print the current workbook

- And many other options

Try it: Click the File Tab and observe the options described. This area is called the "backstage view". Click the "back" arrow at the top to return to the main Excel work area.

6.2 Basic Formatting

6.2.1 Labels

The words you type into a cell are called "labels". Excel has many formatting tools to make labels look better and easier to read. For example, the label "Number Sold" is too long to fit into a cell B2.

Try it: Select cell range A1:E1 (click and drag from cell A1 to E1). Click the "Merge and Center" command in the Alignment Group of the Home Tab. Now add a little style: click the "Good" command in the Styles Group of the Home Tab.

	A	B	C	D	E
1	My Pet Store Earnings				
2	Type of Pet	Number Sold	Price	Total per type	
3	Dogs				
4	Cats				
5	Fish				
6	Birds				
7	Rodents				
8	Reptiles				
9	Arachnids				
10					

Formatting Tips:

- Select cell range A1:E1, click "Merge and Center" in the Alignment Group
- Use "Wrap Text" for long labels
- Adjust column width by double-clicking between column letters

Try: Place the mouse pointer on the thin line between column letters "B" and "C". Double click, and the columns will automatically adjust to fit the text in column "B".

Make the text in every column easy to read. You may also click and drag to adjust the width of the column.

6.2.2 Values

The numbers you type into a cell are called "values". Working with values in Excel will begin to show you the power of the software.

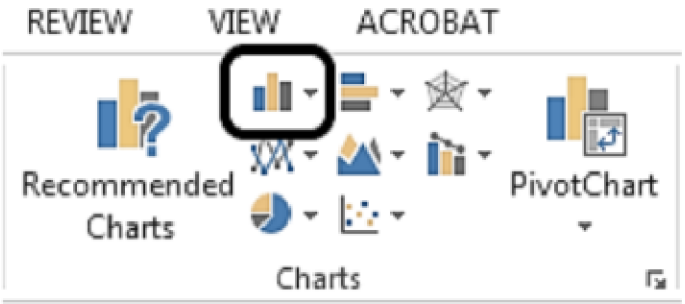
Formulas: Start with = sign

- `=AVERAGE(B2:B10)` — calculates the mean
- `=STDEV(B2:B10)` — calculates standard deviation
- `=SUM(B2:B10)` — calculates sum
- `=COUNT(B2:B10)` — counts cells with numbers

6.3 Charts

There are many types of charts to choose from when you want to share your data with others. Sometimes it can be difficult to know which chart type to use for the data one wants to present. Excel 2016 has an excellent new feature called "Recommended Charts", which shows a live preview of selected data to be charted. It is found in both Insert Chart dialog as well as Quick Analysis.

6.3.1 Column Charts



Creating a Column Chart:

1. Highlight the range of cells
2. Click the INSERT tab
3. In the Charts group, click the Insert Column Chart drop-down menu
4. Click the desired column chart

Choosing a Different Layout:

1. Click the chart
2. Click the DESIGN tab
3. In the Chart Layouts group, click the Quick Layout drop-down menu
4. Select the desired layout

Change the Orientation of the Data:

1. Click the chart
2. Click the DESIGN tab
3. In the Data group, click Switch Row/Column
4. This is a toggle, clicking it again will switch the row and column again

Adding a Chart Title:

1. Click the title placeholder
2. To link the title to a cell:
   - Click the formula bar and type the equals sign (=)
   - Click the cell
   - Press Enter
3. Or simply type the title in the formula bar and press Enter

Changing the Chart Subtype:

1. Click the chart
2. Click the DESIGN tab
3. In the Type group, click Change Chart Type
4. Click the desired chart type

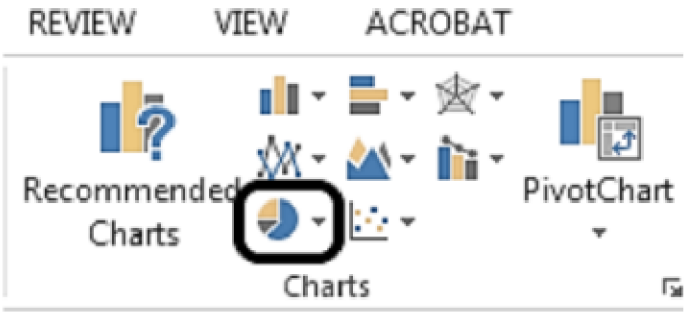
Displaying a Data Table:

1. Click the chart

2. Click the Plus sign that appears to the right
3. In the Chart Elements window, check the Data Table

Plant Biotechnology Example: Comparing shoot regeneration rates across different hormone treatments in *Musa* micropropagation:

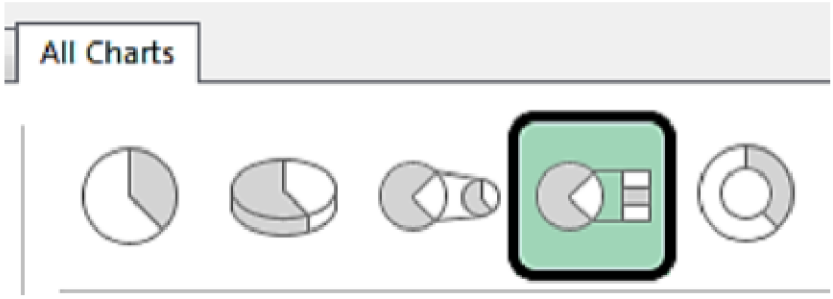
6.3.2 Pie Charts



Creating a Pie Chart:

1. Highlight the range of cells
2. Click the INSERT tab
3. In the Charts group, click the Pie dropdown menu
4. Click the desired pie chart

Adding Data Labels:



1. Click the chart
2. Click the Plus sign
3. Select Data Labels

Adding a Chart Title:

1. Select the chart
2. Click the Plus sign that appears to the right
3. In the Chart Elements window, check Chart Title

Adding Sparklines to a Worksheet:

1. Highlight the range of cells
2. Click the INSERT tab
3. In the Sparklines group there are three different options, click the desired option (Line, Column, or Win/Loss)

4. The Create Sparklines dialog box appears
5. Click the icon to the right of the Location Range field
6. Highlight the range of cells
7. Click the icon to the right of the field
8. Click OK

Plant Biotechnology Example: Distribution of explant response types in *Solanum tuberosum*:

- Callus Formation: 40 %
- Direct Regeneration: 30 %
- Rooting Only: 20 %
- No Response: 10 %

6.3.3 Line Charts

Creating a Line Chart:

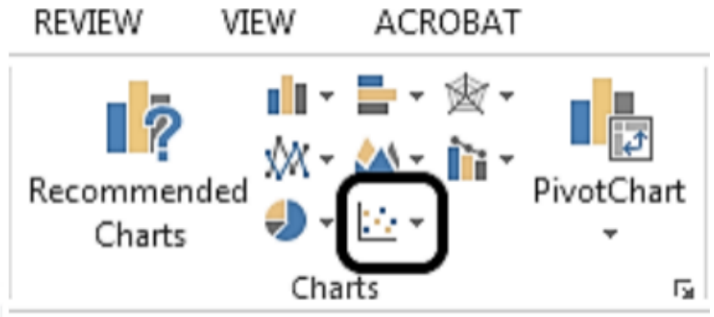
1. Highlight the range of cells
2. Click the INSERT tab
3. In the Charts group, click the Line dropdown menu
4. Click the desired line chart

Add a Data Series to a Line Chart:

1. Click the chart
2. Click the DESIGN tab
3. In the Data group, click Select Data
4. The Select Data Source dialog box appears

5. Click Add
6. Click the cell that is the Series Name
7. Press Tab
8. Highlight the values in the series (vertical axis)
9. Click OK
10. Under Horizontal (Category) Axis Labels, Click Edit
11. Highlight the values in the category (horizontal axis)
12. Press Enter
13. Click OK

6.3.4 Scatter Charts



Creating a Scatter Chart:

1. Highlight the range of cells
2. Click the INSERT tab
3. In the Charts group, click the Scatter drop-down menu
4. Click the desired scatter chart

Adding a Trendline:

1. Click the chart
2. Click the Plus sign that appears to the right
3. In the Chart Elements window, place the pointer over Trendline and click the arrow that appears to the right
4. Choose Linear, Exponential, or other trendline types

Plant Biotechnology Example: Relationship between explant size and regeneration efficiency in *Gossypium hirsutum*:

6.3.5 Bar Charts

Creating a Bar Chart:

1. Highlight the range of cells
2. Click the INSERT tab
3. In the Chart group, click the Bar dropdown menu
4. Click the desired type of bar chart

Plant Biotechnology Example: Comparison of secondary metabolite content across different elicitor treatments in *Artemisia annua*:

6.3.6 Applying a Chart Style

1. Click the chart
2. Click the DESIGN tab
3. In the Chart Styles group, click the desired style

Exercise 2

A researcher studied the contamination sources in 180 *in vitro* cultures of *Olea europaea* (olive). The results were:

Contamination Source	**Number of Cultures**
Bacterial	55
Fungal	65
Yeast	30
Mixed (bacterial + fungal)	10
No contamination	20
Total	**180**

1. What type of qualitative variable is this?
2. Calculate all descriptive measures (absolute frequency, relative frequency, percentage).
3. What is the mode?
4. What percentage of cultures showed some form of contamination?
5. Represent the data graphically. Which type of chart would be most appropriate and why?

Exercise 3

In an acclimatisation experiment, 120 micropropagated plantlets of *Fragaria ananassa* (strawberry) were evaluated after 8 weeks. Their survival status was classified as:

Status	**Number of Plantlets**
Dead	15
Surviving (no growth)	25
Moderate growth	40
Vigorous growth	30
Fully established	10
Total	**120**

1. Is this a nominal or ordinal variable? Justify your answer.
2. Construct the complete statistical table with cumulative frequencies.
3. What percentage of plantlets showed at least moderate growth?
4. Determine the mode and interpret the result.
5. If you were to represent the cumulative distribution graphically, what type of chart would you use?

Exercise 4

A genetic transformation experiment was conducted on *Solanum lycopersicum* (tomato) using three different methods. A total of 150 transformation attempts were classified by method:

Transformation Method	**Number of Attempts**
Agrobacterium-mediated	65
Biolistic (gene gun)	50
Electroporation	25
Protoplast fusion	10
Total	**150**

1. What type of qualitative variable is this?
2. Calculate the relative frequency and percentage for each method.
3. Determine the mode.
4. What percentage of transformations used a non-biological vector method (electroporation or protoplast fusion)?
5. Which graphical representation would best display these data? Justify your choice.

Exercise 5

In a somatic embryogenesis study on *Coffea arabica* (coffee), 250 embryos were classified by their developmental stage: