TP d'ANOVA

Contents

Agence Nationale de la Statistique et de la Démographie (ANSD)

École Nationale de la Statistique et de l'Analyse Économique (ENSAE)

PRÉSENTATION DU PROJET – ANNÉE ACADÉMIQUE 2025-2026

ANOVA à mesure répétée

Travaux pratiques

Rédigé par

Célina NGUEMFOUO NGOUMTSA Mame Balla BOUSSO

Élèves Ingénieurs Statisticiens Economistes (ISE 2)

Sous la supervision de

M. Carlos AKAKPOVI

Ingénieur Statisticien Economiste

Document généré via R Bookdown.

Introduction

Introduction

L'Analyse de la Variance (ANOVA) constitue un outil fondamental en statistique pour étudier l'influence d'un ou plusieurs facteurs sur une variable quantitative. Elle permet de déterminer si les différences observées entre plusieurs groupes sont statistiquement significatives ou simplement dues au hasard.

Dans certaines situations expérimentales, les mesures ne sont pas indépendantes, notamment lorsque plusieurs observations sont réalisées dans un même cadre expérimental ou sur des unités similaires. L'ANOVA permet alors d'analyser simultanément les effets principaux des facteurs étudiés ainsi que leurs interactions éventuelles.

Le présent travail a pour objectif de présenter les fondements théoriques de l'ANOVA, puis d'appliquer ces outils à une base de données expérimentales réelles afin d'identifier les facteurs susceptibles d'influencer la croissance des plantes.

La base mise à disposition porte sur des données réelles recueillies dans le cadre d'une expérimentation agronomique.

Trois types de terreau ont été étudiés :

- Terreau de Manguier (Ma),
- Terreau de Caïlcédrat (Ca),
- Terreau d'Anacarde (An).

Ces terreaux ont été utilisés sur deux variétés de plantes :

- Variété 1 (Var1),
- Variété 2 (Var2).

L'expérience a été répétée quatre fois et, pour chaque combinaison terreau–variété, la hauteur des plantes a été mesurée.

L'objectif principal de l'étude est de déterminer s'il existe une association entre le type de terreau et le type de variété qui favorise significativement la croissance des plantes.

Le présent travail s'articule autour de trois parties principales :

- Une première partie consacrée à la présentation synthétique des fondements théoriques de l'Analyse de la Variance ;
- Une deuxième partie dédiée à l'analyse descriptive des variables d'étude et à l'exploration des données ;
- Une troisième partie portant sur l'analyse inferentielle, à travers l'application des modèles ANOVA à un ou deux facteurs (avec ou sans interaction), afin d'identifier les facteurs influençant la croissance des plantes.

I. Théorie sur l'ANOVA à mesure répétée

Théorie sur l'ANOVA à mesure répétée

L'Analyse de la Variance à Mesures Répétées (ANOVA-MR) est une extension de l'ANOVA classique adaptée aux situations où **les mêmes individus sont mesurés plusieurs fois**, au cours du temps ou sous différentes conditions expérimentales.

Elle tente de répondre à la question centrale : comment évaluer rigoureusement si les variations observées chez un même individu (dans le temps ou entre conditions) sont statistiquement significatives, tout en contrôlant la variabilité individuelle et les corrélations intra-sujets ?

Contrairement à l'ANOVA classique, qui suppose l'indépendance des observations, l'ANOVA à mesures répétées tient compte de la corrélation intra-sujet (les mesures provenant d'un même individu ne sont pas indépendantes).

C'est-à-dire :

$$\mathrm{Cov}(Y_{ij}, Y_{ik}) \neq 0 \quad \text{pour } j \neq k,$$

où j et k indexent deux mesures différentes effectuées sur un même individu.

Autrement dit, les observations répétées chez un même sujet ne sont pas indépendantes.

L'ANOVA à mesures répétées permet ainsi de modéliser explicitement cette dépendance intra-sujet.

1. Notations

On observe n sujets mesurés à t temps (ou conditions). Les données sont notées :

$$Y_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, t$$

où :

- i désigne le sujet,
- j désigne le temps ou la condition.

Chaque sujet est donc mesuré plusieurs fois, ce qui introduit une corrélation entre les observations d'un même individu.

Décomposition de la variance

La somme totale des carrés est :

$$SCT = \sum_{i=1}^{n} \sum_{j=1}^{t} (Y_{ij} - \bar{Y}_{..})^2$$

Elle se décompose en :

$$SCT = SS_{\text{sujets}} + SS_{\text{temps}} + SS_{\text{erreur}}.$$

- **Somme des carrés due aux sujets :**

$$SS_{\text{sujets}} = t \sum_{i=1}^{n} (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

Elle mesure la variabilité inter-individuelle.

- **Somme des carrés due au temps :**

$$SS_{\text{temps}} = n \sum_{j=1}^{t} (\bar{Y}_{.j} - \bar{Y}_{..})^2$$

Elle représente l'effet du facteur intra-sujet.

- **Somme des carrés résiduelle :**

$$SS_{\text{erreur}} = \sum_{i=1}^{n} \sum_{j=1}^{t} \left(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..} \right)^2$$

Elle correspond à l'interaction sujet $\times$ temps, c'est-à-dire à la variabilité non expliquée par les effets principaux.

2. Tests spécifiques à l'ANOVA à mesures répétées

Contrairement à l'ANOVA classique, l'ANOVA à mesures répétées nécessite des vérifications supplémentaires liées à la dépendance intra-sujet.

a. Test de sphéricité (Mauchly)

Il vérifie l'hypothèse de **sphéricité**, condition fondamentale du modèle.

- **Hypothèses**

$$H_0 : \Sigma = \sigma^2 I \quad \text{(sphéricité respectée)}$$

$$H_1 : \text{sphéricité non respectée}$$

Cela revient à supposer que :

$$\mathrm{Var}(Y_i - Y_j) = \text{constante} \quad \forall (i, j).$$

- **Statistique** :

$$W = \frac{|S|}{\left(\dfrac{\mathrm{tr}(S)}{k-1}\right)^{k-1}},$$

avec S matrice de covariance des différences et k le nombre de conditions.

- **Approximation** :

$$\chi^2 = -(N-1)\ln(W) \sim \chi^2_{\frac{k(k-1)}{2}-1}.$$

Si p-value $< \alpha$, on applique une correction :

- Greenhouse–Geisser,
- Huynh–Feldt.

b. Tests intra-sujets

Ils analysent les variations à l'intérieur d'un même individu (évolution dans le temps ou entre conditions).

- **Effet principal du temps** :

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_k \quad H_1 : \text{au moins deux temps diffèrent.}$$

- **Statistique de test** :

$$F = \frac{SCE_{\text{temps}}/(k-1)}{SCR_{\text{intra}}/[(k-1)(N-1)]}.$$

Sous H_0 (et si sphéricité respectée) :

$$F \sim F(k-1, (k-1)(N-1)).$$

- **Comparaisons complémentaires**

Si l'effet principal est significatif :

- Contrastes (polynomial, Helmert, différences),
- Comparaisons post-hoc (Bonferroni, Tukey).

c. Interaction temps et groupe (design mixte)

On étudie l'interaction entre un facteur intra-sujet (temps) et un facteur inter-sujets (groupe).

$$H_0 : \text{profils d'évolution parallèles entre groupes}$$

$$H_1 : \text{au moins un groupe a un profil différent.}$$

Si l'interaction est significative :

- **Analyse des effets simples** :
  - effet du temps dans chaque groupe,
  - effet du groupe à chaque temps ;
- **Comparaisons multiples** (Bonferroni, Tukey).

3. Modélisation et tests en ANOVA à mesures répétées

a. ANOVA à un facteur intra-sujet

On observe n sujets mesurés à t temps (ou conditions).

- **Modèle**

$$Y_{is} = \mu + \tau_s + \alpha_i + \varepsilon_{is}$$

où :

- μ : moyenne générale,
- τ_s : effet fixe du facteur intra-sujet,
- $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$: effet aléatoire sujet,
- $\varepsilon_{is} \sim \mathcal{N}(0, \sigma^2)$: erreur.
- **Hypothèse testée**

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_t$$

$$H_1 : \text{au moins deux niveaux diffèrent}$$

- **Statistique de test**

$$F = \frac{CM_{\text{temps}}}{CM_{\text{erreur}}} \sim F(t-1, (t-1)(n-1))$$

- **Règle de décision**

Si p-value $< \alpha$, on rejette H_0 : le facteur intra-sujet a un effet significatif.

Si la sphéricité est violée, on utilise les ddl corrigés (Greenhouse–Geisser ou Huynh–Feldt).

b. ANOVA à deux facteurs à mesures répétées (double intra-sujet)

Chaque sujet est mesuré pour toutes les combinaisons des niveaux des deux facteurs A et B.

- **Modèle**

$$Y_{ijs} = \mu + \alpha_i + \beta_j + \gamma_s + (\alpha\beta)_{ij} + (\alpha\gamma)_{is} + (\beta\gamma)_{js} + \varepsilon_{ijs}$$

où γ_s représente l'effet aléatoire sujet.

- **Hypothèses testées**

Effet principal A :

$$H_0 : \alpha_1 = \cdots = \alpha_I$$

Effet principal B :

$$H_0 : \beta_1 = \cdots = \beta_J$$

Interaction :

$$H_0 : (\alpha\beta)_{ij} = 0$$

- **Statistiques de test**

$$F_A = \frac{CM_A}{CM_{A\times Sujet}}$$

$$F_B = \frac{CM_B}{CM_{B\times Sujet}}$$

$$F_{AB} = \frac{CM_{AB}}{CM_{\text{erreur}}}$$

- **Règle de décision**

Si p-value $< \alpha$, l'effet est significatif.

En cas d'effet principal significatif : comparaisons post-hoc ou contrastes.

- **Contraintes d'identifiabilité**

$$\sum_{i=1}^{I} \alpha_i = 0 \qquad \sum_{j=1}^{J} \beta_j = 0$$

$$\sum_{i=1}^{I} (\alpha\beta)_{ij} = 0 \qquad \sum_{j=1}^{J} (\alpha\beta)_{ij} = 0$$

$$\sum_{s} \rho_s = 0$$

Les effets aléatoires sont centrés et indépendants.

Tableau récapitulatif : ANOVA à 2 facteurs

Table 1: Résumé des informations : ANOVA à 2 facteurs (MR sur A et B)

Facteur A	SC_A	$I-1$	$CM_A = \frac{SC_A}{I-1}$	$F_A = \frac{CM_A}{CM_{A\times S}}$
Facteur B	SC_B	$J-1$	$CM_B = \frac{SC_B}{J-1}$	$F_B = \frac{CM_B}{CM_{B\times S}}$
Interaction $A \times B$	SC_{AB}	$(I-1)(J-1)$	$CM_{AB} = \frac{SC_{AB}}{(I-1)(J-1)}$	$F_{AB} = \frac{CM_{AB}}{CM_{AB\times S}}$
Sujet	SC_S	$n-1$	$CM_S = \frac{SC_S}{n-1}$	
Interaction $A \times Sujet$	$SC_{A\times S}$	$(I-1)(n-1)$	$CM_{A\times S} = \frac{SC_{A\times S}}{(I-1)(n-1)}$	
Interaction $B \times Sujet$	$SC_{B\times S}$	$(J-1)(n-1)$	$CM_{B\times S} = \frac{SC_{B\times S}}{(J-1)(n-1)}$	
Interaction $A\times B\times Sujet$	$SC_{AB\times S}$	$(I-1)(J-1)(n-1)$	$CM_{AB\times S} = \frac{SC_{AB\times S}}{(I-1)(J-1)(n-1)}$	
Totale	SC_T	$nIJ-1$		

d. Modélisation d'une ANOVA avec trois facteurs

On considère trois facteurs fixes A, B et C ayant respectivement I, J et K niveaux. Selon le protocole expérimental, on peut avoir un, deux ou trois facteurs mesurés de manière répétée (intra-sujet).

- **Formes usuelles du modèle**
- **Cas 1** : *mesures répétées sur un seul facteur*

On suppose que le facteur A est répété (intra-sujet) et que B et C sont inter-sujets. Un modèle possible est :

$$Y_{ijk\ell} = \mu + \alpha_i + \delta_{\ell(i)} + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijk\ell},$$

où $\delta_{\ell(i)}$ représente l'effet aléatoire du sujet ℓ (imbriqué dans la modalité répétée), et $\varepsilon_{ijk\ell}$ l'erreur.

- **Cas 2** : *mesures répétées sur deux facteurs*

On suppose que A et B sont répétés (intra-sujet), et C inter-sujets. Un modèle s'écrit alors :

$$Y_{ijk\ell} = \mu + \alpha_i + \beta_j + \delta_\ell + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijk\ell}.$$

Ici, δ_ℓ capture la variabilité propre à chaque sujet.

- **Cas 3** : *mesures répétées sur les trois facteurs*

Lorsque chaque sujet passe toutes les combinaisons des niveaux de A, B et C (plan entièrement intra-sujet), on peut écrire :

$$Y_{ijk\ell} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_\ell + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijk\ell}.$$

0.0.1 Contraintes d'identifiabilité

Pour assurer l'unicité des paramètres (effets centrés), on impose classiquement :

$$\sum_{i=1}^{I} \alpha_i = 0, \qquad \sum_{j=1}^{J} \beta_j = 0, \qquad \sum_{k=1}^{K} \gamma_k = 0.$$

Pour les interactions, les sommes sont nulles sur chacun des indices, par exemple :

$$\sum_{i=1}^{I} (\alpha\beta)_{ij} = 0 \ \forall j, \qquad \sum_{j=1}^{J} (\alpha\beta)_{ij} = 0 \ \forall i,$$

et de même pour $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$ et $(\alpha\beta\gamma)_{ijk}$ (somme nulle sur chaque dimension).

Tableau récapitulatif

Table 2: Synthèse des tests : ANOVA à trois facteurs

Facteur A	SC_A	$I-1$	$CM_A = \frac{SC_A}{I-1}$	$F_A = \frac{CM_A}{CM_R}$
Facteur B	SC_B	$J-1$	$CM_B = \frac{SC_B}{J-1}$	$F_B = \frac{CM_B}{CM_R}$
Facteur C	SC_C	$K-1$	$CM_C = \frac{SC_C}{K-1}$	$F_C = \frac{CM_C}{CM_R}$
Interaction AB	SC_{AB}	$(I-1)(J-1)$	$CM_{AB} = \frac{SC_{AB}}{(I-1)(J-1)}$	$F_{AB} = \frac{CM_{AB}}{CM_R}$
Interaction AC	SC_{AC}	$(I-1)(K-1)$	$CM_{AC} = \frac{SC_{AC}}{(I-1)(K-1)}$	$F_{AC} = \frac{CM_{AC}}{CM_R}$
Interaction BC	SC_{BC}	$(J-1)(K-1)$	$CM_{BC} = \frac{SC_{BC}}{(J-1)(K-1)}$	$F_{BC} = \frac{CM_{BC}}{CM_R}$
Interaction ABC	SC_{ABC}	$(I-1)(J-1)(K-1)$	$CM_{ABC} = \frac{SC_{ABC}}{(I-1)(J-1)(K-1)}$	$F_{ABC} = \frac{CM_{ABC}}{CM_R}$
Résiduelle	SC_R	$n-I-J-K-1$	$CM_R = \frac{SC_R}{n-I-J-K-1}$	
Totale	SC_T	$n-1$		

- **Remarque** : le dénominateur exact (erreur appropriée) dépend du caractère inter/intra-sujet des facteurs. Si un ou plusieurs facteurs sont répétés, certaines erreurs doivent être remplacées par des interactions avec le sujet (ex. $A \times Sujet$, $AB \times Sujet$) et des corrections de ddl peuvent être nécessaires en cas de violation de sphéricité.

Après avoir présenté les fondements théoriques de l'Analyse de la Variance et les différents modèles envisageables selon la structure expérimentale, nous passons désormais à l'application pratique sur les données issues de l'expérimentation relative aux types de terreau et aux variétés de plantes.

II. Analyse descriptive des variables d'étude

Analyse descriptive des variables d'étude

La base utilisée porte sur des données réelles recueillies dans le cadre d'une expérimentation.Dans le cadre de cette expérimentation, trois types de terreau : Manguier (Ma), Caïlcédrat (Ca) et Anacarde (An) ont été utilisées sur deux types de variétés de plantes (Var1 et Var2). L'expérience a été répété quatre fois et chaque fois on mesure la hauteur des plantes.

Voici un aperçu de la base :

```
# A tibble: 6 x 5
  Sujet Répétition Terreau variété Hauteur
  <dbl> <fct>      <fct>   <fct>     <dbl>
1     1 T1         Ma      Var1         43
2     2 T1         Ma      Var1         42
3     3 T1         Ma      Var1         40
4     4 T1         Ma      Var1         46
5     5 T1         Ma      Var1         42
6     6 T1         Ma      Var1         40
```

1. Description de la base de données

```
[1] 253   5
```

La base contient 5 variables et 253 lignes. Comme il y'a des répétitions, le nombre de ligne n'est pas égal au nombre d'individus. La variable de répetition se présente comme suit :

Table 3: Répartition des données selon la répétition

Répétition	Effectif (n)
T1	78
T2	57
T3	63
T4	55

Valeurs manquantes

On constate des valeurs manquantes, car le nombre d'observations pour chaque répétition n'est pas le même.

L'ANOVA à mesures répétées classique repose sur un plan complet et ne considère que les sujets disposant de l'ensemble des mesures. Les observations incomplètes sont donc exclues de l'analyse.

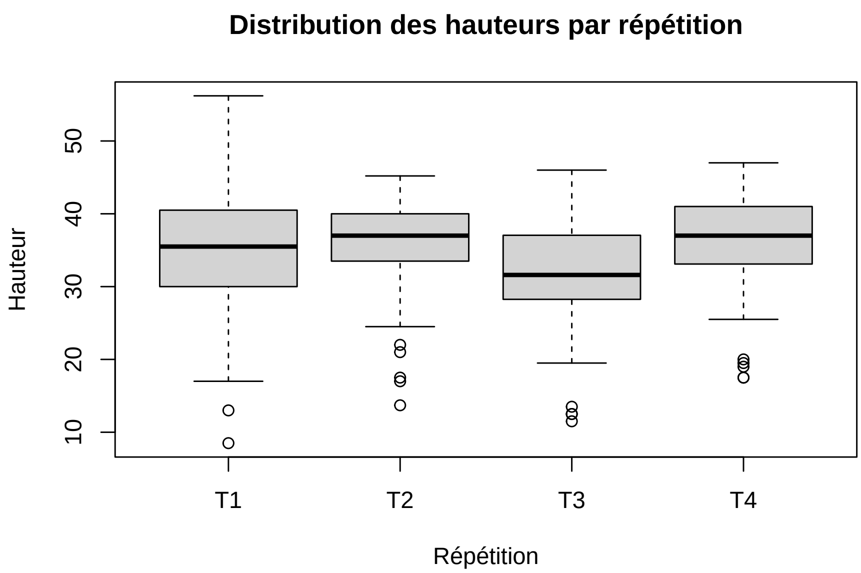
Sous l'hypothèse que ces données manquantes sont aléatoires (MCAR), cette exclusion n'introduit pas de biais systématique mais peut réduire la puissance statistique.

Valeurs abberantes

Les données analysées proviennent d'une expérimentation agronomique contrôlée. Les hauteurs observées correspondent à des mesures physiques réelles et non à des déclarations ou des données administratives. Ainsi, les valeurs extrêmes potentielles reflètent davantage la variabilité biologique naturelle que des anomalies statistiques.

Une inspection graphique par boîtes à moustaches, réalisée par répétition expérimentale, ne met pas en évidence de valeurs aberrantes isolées susceptibles d'indiquer une erreur de mesure ou de saisie. Les éventuelles valeurs extrêmes observées sont cohérentes avec la dispersion attendue dans un contexte biologique réel et ont donc été conservées dans l'analyse.

```r
boxplot(Hauteur ~ Répétition, data = data,
        main = "Distribution des hauteurs par répétition",
        ylab = "Hauteur")
```



2. Analyse univariée

Variable hauteur

- **Résumé de la variable hauteur**

Le tableau suivant présente quelques statistiques descriptives de la variable à expliquer hauteur, suivant les périodes.

Table 4: Statistiques descriptives de la Hauteur par Répétition

T1	8.5	30.00	35.5	34.32	40.50	56.2
T2	13.7	33.50	37.0	35.41	40.00	45.2
T3	11.5	28.25	31.6	31.42	37.05	46.0
T4	17.5	33.10	37.0	36.27	41.00	47.0

Les plantes semblent globalement connaître une nette évolution entre la période une et deux, puis une baisse de croissance entre la prériode 2 et 3, et enfin une hausse entre les deux dernières périodes.

Variable terreau

- **Résumé de la variable terreau**

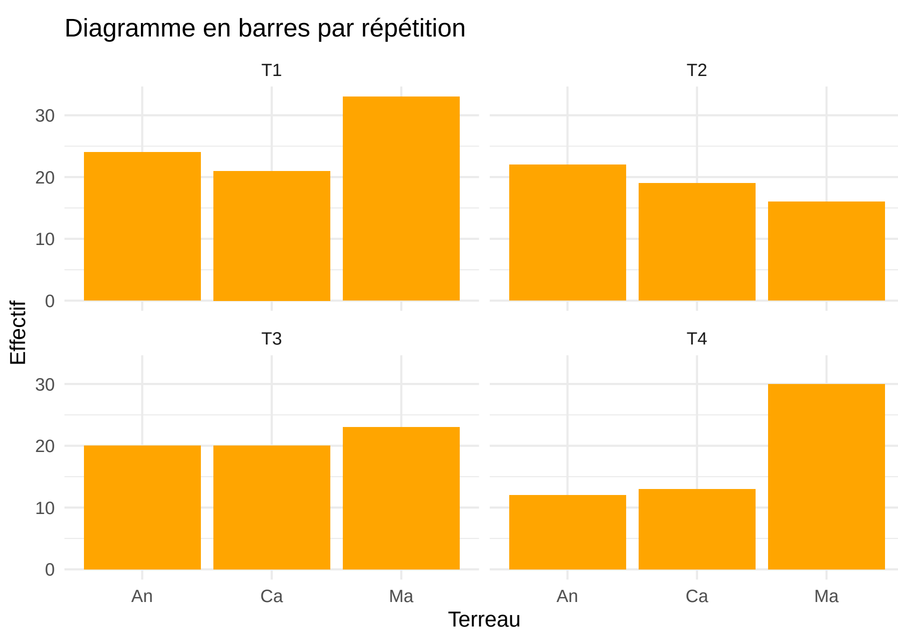
Le tableau suivante présente les effectifs de chaque type de terreau par période.

Table 5: Répartition du nombre d'observations par Répétition et par Terreau

	Types de Terreau		
Répétition	**An**	**Ca**	**Ma**
T1	24	21	33
T2	22	19	16
T3	20	20	23
T4	12	13	30

Ce tableau est illustré par les diagrammes en barre suivant :

- **Diagramme en barre de la variable terreau**



Variable variété

- **Résumé de la variable variété**

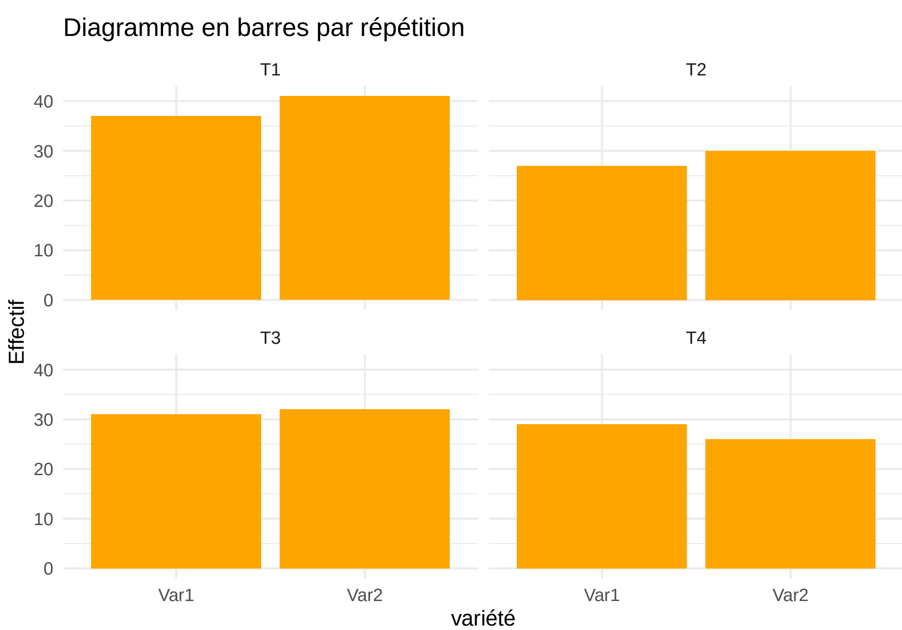
Le tableau suivante présente les effectifs de chaque type de variété par période

Table 6: Répartition du nombre d'observations par Répétition et par variété

	Types de variété	
Répétition	**Var1**	**Var2**
T1	37	41
T2	27	30
T3	31	32
T4	29	26

Ce tableau est illustré par les diagrammes suivants :

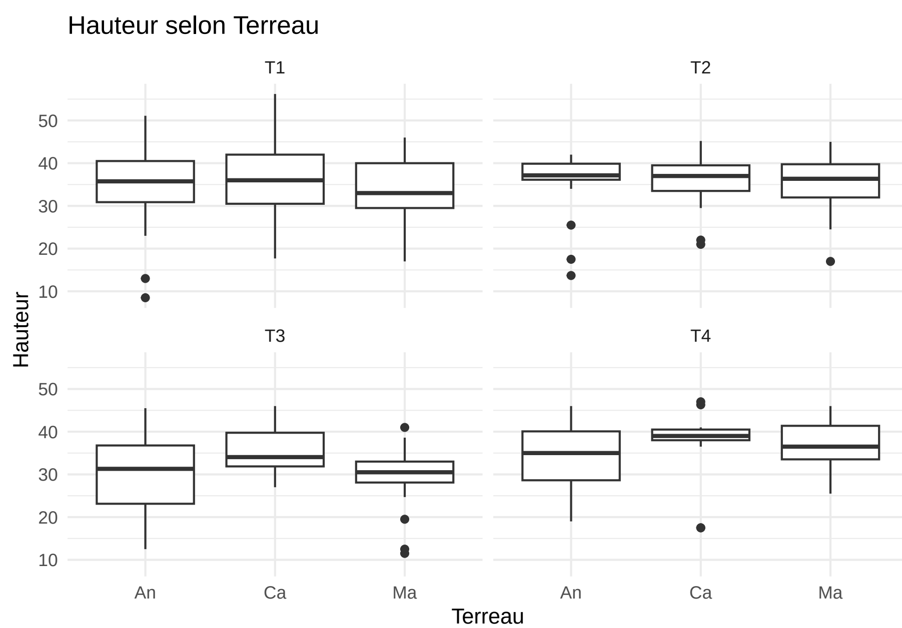
- **Diagramme en barre de la variable variété**



3. Analyse bivariée

Variable terreau et hauteur

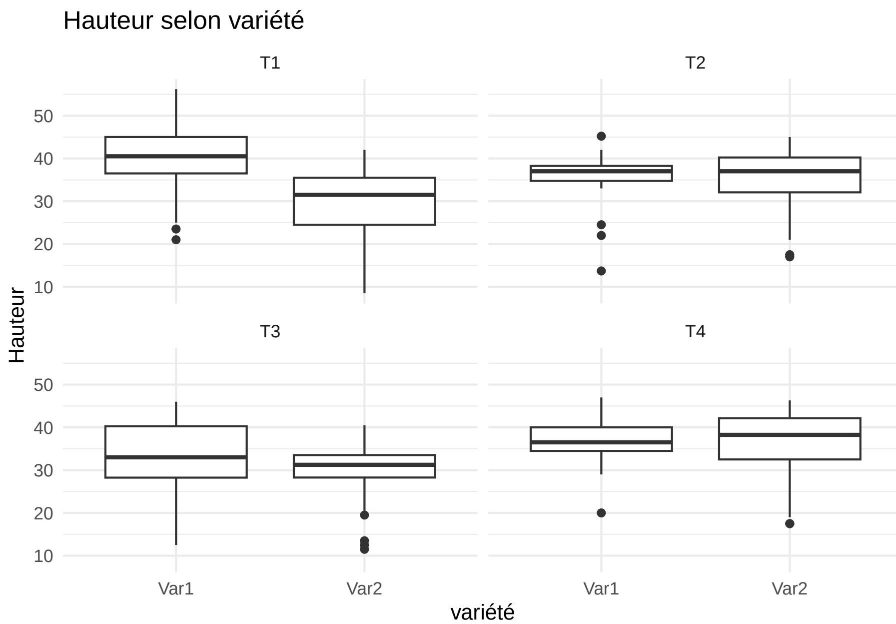
La figure suivante présente les boxplots de la variable hauteur selon les modalités de la variable terreau et suivant les périodes.



D'après ces boxplots, nous pouvons constater qu'au cours des périodes 1 et 2, le type de terreau utilisé ne semble pas influencer la hauteur des plantes. Durant les périodes 3 et 4 cependant, le type de terreau semble avoir un effet sur les plantes. En effet, les plantes ayant utilisées le terreau de type Ca semblent plus grandes que les autres.

Variable variété et hauteur

La figure suivante présente les boxplots de la variable hauteur selon les modalités de la variable variété et suivant les périodes.



D'après ce graphique, de manière générale, le type de variété semé semble influencer la hauteur des plantes. On constate par exemple que dans les périodes 1 et 3, les hauteurs des plantes de la variété 2 semblent inférieures à celle des plantes de la variété 1.

Toutes ces observations vont être testées de façon plus rigoureuses par l'ANOVA à mesure répétées.

III. Pratique

Pratique

1. Tests d'hypothèses de l'ANOVA

a. Indépendance des sujets

L'hypothèse d'indépendance des observations stipule que les mesures d'un sujet ne doivent pas influencer celles d'un autre. Dans notre expérience :

Chaque plante a été sélectionnée indépendamment des autres. Les mesures répétées concernent la même plante, mais les plantes entre elles sont distinctes.

Ainsi, on considère que l'hypothèse d'indépendance est respectée.

b. Normalité des résidus

Concernant l'analyse de la variance, la validité des tests F repose sur l'hypothèse de normalité des résidus. Toutefois, conformément au Théorème Central Limite (TCL), la distribution des moyennes d'échantillons tend vers la normalité dès lors que l'effectif est suffisamment important (typiquement $n > 30$).Dans le cadre de cette étude, notre effectif total dépasse ce seuil critique. En vertu de la robustesse de l'ANOVA face aux déviations modérées de la normalité pour de tels échantillons, nous considérons cette condition comme satisfaite. Cette approche permet de garantir la fiabilité des conclusions statistiques sans nécessiter de transformations de variables, préservant ainsi l'interprétabilité directe des mesures de hauteur.

c. Homogénéité des variances et sphéricité

Nos deux facteurs variété de la plante et type de terreau sont unique à une plante, donc il s'agit de facteurs inter-sujet. Ainsi le test qui sera utilisé sur les deux facteurs est celui de l'homogéneité des variances.

- **Facteur Variété**

```
leveneTest(Hauteur ~ variété, data = data)
```

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   1  1.8743 0.1722
      251
```

Le test de Levene pour le facteur variété donne une p-value de 0,1722, supérieure au seuil classique de 0,05. Cela indique que les variances entre les groupes ne sont pas significativement différentes, donc l'hypothèse d'homogénéité des variances est respectée pour ce facteur.

- **Facteur terreau**

```
leveneTest(Hauteur ~ Terreau, data = data)
```

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   2  1.3192 0.2692
      250
```

Là aussi, l'hypothèse d'homogénité des variances est respectée.

De manière générale, les hypothèses de l'ANOVA à mesure répétée sont respectées pour notre jeu de données. Passons à présent à L'ANOVA-MR en elle même.

2. ANOVA à un facteur

Facteur terreau

```
anova_terreau <- aov(
  Hauteur ~ Terreau + Error(Sujet/Répétition),
  data = data
)
summary(anova_terreau)
```

```
Error: Sujet
        Df Sum Sq Mean Sq
Terreau  1    104     104

Error: Sujet:Répétition
          Df Sum Sq Mean Sq F value Pr(>F)
Terreau    2  131.7   65.86   0.224  0.831
Residuals  1  294.1  294.11

Error: Within
           Df Sum Sq Mean Sq F value  Pr(>F)
Terreau     2    721   360.5   5.705 0.00379 **
Residuals 246  15544    63.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

L'analyse montre que :

- **Effet inter-sujet** (entre les plantes) : F = 0,224, p = 0,831, ainsi, pas de différence significative entre les plantes selon le type de terreau.
- **Effet intra-sujet** (répétitions sur chaque plante) : F = 5,705, p = 0,00379, donc il y'a un effet significatif du type de terreau sur la hauteur des plantes à travers les répétitions.

Conclusion : le type de terreau influence la croissance des plantes lorsqu'on considère les mesures répétées sur chaque plante, mais il n'explique pas de différences globales entre les plantes individuellement.

Variable variété

```r
anova_terreau <- aov(
  Hauteur ~ variété + Error(Sujet/Répétition),
  data = data
)
summary(anova_terreau)
```

```
Error: Sujet
        Df Sum Sq Mean Sq
variété  1    104     104
```

```
Error: Sujet:Répétition
          Df Sum Sq Mean Sq F value Pr(>F)
variété    1    7.2    7.16   0.034   0.87
Residuals  2  418.7  209.33

Error: Within
           Df Sum Sq Mean Sq F value   Pr(>F)
variété     1   1223  1222.6   20.08 1.14e-05 ***
Residuals 247  15042    60.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

L'analyse montre que :

- **Effet inter-sujet** (entre les plantes) : F = 0,034, p = 0,87, donc pas de différence significative entre les plantes selon la variété.
- **Effet intra-sujet** (répétitions sur chaque plante) : F = 20,08, $p < 0{,}001$, donc il y'a un effet significatif de la variété sur la hauteur des plantes à travers les répétitions.

Conclusion : la variété influence la croissance des plantes lorsque l'on considère les mesures répétées sur chaque plante, mais elle n'explique pas de différences globales entre les plantes individuellement.

3. ANOVA à deux facteurs sans interaction

```r
anova_2facteurs <- aov(
  Hauteur ~ Terreau + variété + Error(Sujet/Répétition),
  data = data
)
summary(anova_2facteurs)
```

```
Error: Sujet
        Df Sum Sq Mean Sq
Terreau  1    104     104

Error: Sujet:Répétition
        Df Sum Sq Mean Sq
Terreau  2  131.7   65.86
variété  1  294.1  294.11
```

```
Error: Within
           Df Sum Sq Mean Sq F value   Pr(>F)
Terreau     2    721   360.5    6.01 0.002831 **
variété     1    848   848.1   14.14 0.000212 ***
Residuals 245  14695    60.0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

L'analyse montre que :

- **Effet inter-sujet** (entre les plantes) : pas de différences significatives (F = 104 ou 131,7 pour les blocs Sujet et Sujet:Répétition), donc les plantes individuelles ne présentent pas de variations globales importantes selon le type de terreau ou la variété.
- **Effet intra-sujet** (répétitions sur chaque plante) :
  - **Terreau** : F = 6,01, p = 0,00283 : effet significatif du type de terreau sur la hauteur des plantes à travers les répétitions.
  - **Variété** : F = 14,14, p = 0,000212 : effet significatif de la variété sur la hauteur des plantes à travers les répétitions.

Conclusion : à travers les mesures répétées, le type de terreau et la variété influencent significativement la croissance des plantes, même si les différences globales entre les plantes individuelles ne sont pas significatives.

4. ANOVA à deux facteurs avec interaction

```r
anova_2facteurs_int <- aov(
  Hauteur ~ Terreau * variété + Error(Sujet/Répétition),
  data = data
)
summary(anova_2facteurs_int)
```

```
Error: Sujet
        Df Sum Sq Mean Sq
Terreau  1    104     104

Error: Sujet:Répétition
        Df Sum Sq Mean Sq
```

```
Terreau  2  131.7   65.86
variété  1  294.1  294.11

Error: Within
                 Df Sum Sq Mean Sq F value   Pr(>F)
Terreau           2    721   360.5   5.961 0.002969 **
variété           1    848   848.1  14.025 0.000225 ***
Terreau:variété   2      1     0.7   0.012 0.988462
Residuals       243  14694    60.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

L'analyse montre que :

- **Effets principaux intra-sujets** (répétitions sur chaque plante) :
  - **Terreau** : F = 5,961, p = 0,00297 : effet significatif du type de terreau sur la hauteur des plantes à travers les répétitions.
  - **Variété** : F = 14,025, p = 0,000225 → effet significatif de la variété sur la hauteur des plantes à travers les répétitions.
- **Interaction Terreau × Variété** : F = 0,012, p = 0,988 → interaction non significative, ce qui indique que l'effet du type de terreau est similaire pour toutes les variétés, et inversement.

Conclusion : la croissance des plantes est influencée indépendamment par le type de terreau et par la variété, mais il n'existe pas d'effet combiné supplémentaire entre ces deux facteurs.

5. Modèle final

Après analyse, le modèle à deux facteurs avec interaction a montré que l'interaction entre le type de terreau et la variété n'était pas significative (p = 0,988). Les effets principaux de Terreau et de Variété sont en revanche significatifs. Pour simplifier l'interprétation et respecter le principe de parcimonie, nous retenons donc le modèle final sans interaction :

$$\text{Hauteur} \sim \text{Terreau} + \text{Variété} + \text{Error(Sujet/Répétition)}$$

Ce modèle permet de mettre en évidence l'effet indépendant du type de terreau et de la variété sur la croissance des plantes, tout en conservant la structure des mesures répétées.

Conclusion

Conclusion

Dans cet exposé, nous avons étudié l'effet du type de terreau et de la variété sur la croissance des plantes en utilisant des ANOVA à mesures répétées. Avant d'effectuer les analyses, nous avons vérifié les hypothèses fondamentales du modèle. L'indépendance des sujets a été respectée, chaque plante constituant une unité expérimentale distincte. La normalité des résidus, hypothèse clé pour la validité de l'ANOVA, a été prise en compte et l'homogénéité des variances entre les groupes inter-sujets a été confirmée par le test de Levene.

Les résultats des analyses montrent que, lorsqu'on considère chaque facteur séparément, le type de terreau et la variété ont tous deux un effet significatif sur la hauteur des plantes à travers les répétitions. Les modèles à deux facteurs ont permis de confirmer ces effets principaux, tandis que l'interaction entre le type de terreau et la variété n'était pas significative, ce qui indique que l'influence du terreau est similaire pour toutes les variétés et que l'effet de la variété est indépendant du type de terreau. Sur cette base, le modèle final retenu est l'ANOVA à deux facteurs sans interaction, qui met en évidence les effets indépendants du type de terreau et de la variété tout en conservant la structure des mesures répétées.

Ainsi, les résultats suggèrent que la croissance des plantes est affectée à la fois par la variété et par le type de terreau, mais que ces deux facteurs agissent de manière indépendante. Ces conclusions fournissent des indications précieuses pour le choix des combinaisons optimales de variété et de terreau afin de maximiser la croissance des plantes, tout en respectant la rigueur statistique nécessaire pour des mesures répetées sur les mêmes sujets.