



RÉPUBLIQUE DU BENIN

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR

ET DE LA RECHERCHE SCIENTIFIQUE (MESRS)

UNIVERSITÉ DE PARAKOU

ÉCOLE NATIONALE DE STATISTIQUE DE PLANIFICATION

ET DE DÉMOGRAPHIE (ENSPD)

MASTER : 1

GROUPE : 25

COURS : Introduction au logiciel statistique

THEME :

**R : ANOVA à un critère, deux
critères et structuration de
moyennes**

Réalisé par :

SANTA Mathias

KOTCHEDJO Akpo Lucien

Sous la supervision de :

Dr. SODJINOUE Epiphane

Juin 2025

PLAN

Introduction

I) Définition de l'ANOVA

II) Objectifs de ANOVA

III) Hypothèses de l'ANOVA IV)

Les different types d'ANOVA

a) ANOVA à un facteur:

b) ANOVA à deux facteurs

V) Structuration des moyennes et analyses post-hoc

Conclusion

Introduction

L'ANOVA (Analyse de la Variance) est une méthode statistiques utilisé pour comparer les moyennes de plusieurs groups afin de déterminer s'il y a une différence significative entre eux. Elle est largement utilisée en recherche expérimentale, en psychologie en biologie, en économie et dans bien d'autres domaines

I) Définition de l'ANOVA

- Outil statistique pour comparer des moyennes entre plusieurs groupes.

II) Objectifs de ANOVA

L'ANOVA est particulièrement utile lorsque l'on souhaite :

- Comparer plus de deux groupes (contrairement au test t de Student, qui ne compare que deux moyennes).
- Évaluer l'effet d'un ou plusieurs facteurs (variables indépendantes) sur une variable quantitative (variable dépendante).
- Détecter des interactions entre plusieurs facteurs (dans le cas d'une ANOVA à plusieurs facteurs).

III) Hypothèses de l'ANOVA

Pour appliquer une ANOVA, certaines conditions doivent être vérifiées :

- ✓ Normalité : Les données doivent suivre une distribution normale dans chaque groupe (ou les résidus dans le cas d'un modèle).
- ✓ Homogénéité des variances (homoscédasticité): Les variances des groupes doivent être similaires (test de Levene ou Bartlett).
- ✓ Indépendance des observations : Les données doivent être indépendantes (pas de mesures répétées sur les mêmes sujets, sauf pour l'ANOVA à mesures répétées).

IV) Les différents types d'ANOVA

a) ANOVA à un facteur :

Compare les moyennes de plusieurs groupes basés sur une seule variable indépendante.

Exemple d'application

- Comparaison de l'effet de 3 fertilisants sur la croissance des plantes.

b) ANOVA à deux facteurs :

- Examine l'effet de deux variables indépendantes et leur interaction.
- Étude de l'effet de deux facteurs (ex : Facteur fertilisants F et de l'irrigation I sur le rendement) et de leur interaction.

Exemple d'application

- Étude du rendement agricole selon le facteur fertilisants F et l'irrigation I

V) Structuration des moyennes et analyses post-hoc

1. Comparaisons multiples

- Problème d'inflation de l'erreur de type I.
- Méthodes de correction : Bonferroni, Tukey (HSD), Scheffé.

2. Contrastes planifiés

- Comparaisons spécifiques entre groupes (ex : groupe 1 vs. Groupe 2+3).

3. Interprétation graphique

- Diagrammes en barres avec intervalles de confiance.
- Boîtes à moustaches (boxplots).

Cas pratique

SCRIPT R POUR L'ANALYSE DE VARIANCE (ANOVA)

Chargement des bibliothèques nécessaires

library(ggplot2) # Visualisation

library(car) # Test de Levene (homogénéité des variances)

library(ggpubr) # Graphiques supplémentaires

library(emmeans) # Pour les comparaisons multiples et calcul des moyennes marginales

library(interactions) # Pour visualiser les interactions

1. PRÉPARATION DES DONNÉES

Création d'une base de données simulé (exemple : rendement de 3 variétés de blé)

set.seed(123) # Pour la reproductibilité

donnees <- data.frame(

 variete = factor(rep(c("A", "B", "C"), each = 20)),

 rendement = c(rnorm(20, mean = 50, sd = 5),

 rnorm(20, mean = 55, sd = 5),

```

        rnorm(20, mean = 60, sd = 5))
)

# Visualisation rapide
head(donnees)
boxplot(rendement ~ variete, data = donnees, col = "lightblue")

```

2. VÉRIFICATION DES HYPOTHÈSES

```

#### a) Normalité des résidus (test de Shapiro-Wilk sur les résidus du modèle)
modele_aov <- aov(rendement ~ variete, data = donnees)
shapiro.test(residuals(modele_aov)) # p > 0.05 → normalité acceptable

#### b) Homogénéité des variances (test de Levene)
#leveneTest(rendement ~ variete, data = donnees) # p > 0.05 → variances homogènes

```

3. ANOVA À UN FACTEUR

```

# ANOVA classique
summary(modele_aov)

```

4. ANALYSE POST-HOC (SI ANOVA SIGNIFICATIVE)

```

#### a) Test de Tukey (comparaisons multiples)
tukey <- TukeyHSD(modele_aov)
print(tukey)
# Toutes les p adj inférieures à 0.05 traduisent que les différentes variétés sont significatives à un seuil de 5

#### b) Visualisation des différences
plot(tukey)

```

5. VISUALISATION DES RÉSULTATS

```

#a) Boxplot avec ANOVA et p-value
ggboxplot(donnees, x = "variete", y = "rendement", color = "variete", add = "jitter") +
  stat_compare_means(method = "anova", label.y = 70) +
  # Ajoute la p-value ANOVA
  labs(title = "Comparaison des rendements par variété",
        x = "Variété de blé", y = "Rendement (kg/ha)")

#### b) Diagramme de moyennes avec intervalles de confiance
ggline(donnees, x = "variete", y = "rendement",
  add = c("mean_ci", "jitter"),

```

```
title = "Moyennes de rendement avec intervalles de confiance")
```

ANOVA À DEUX FACTEURS AVEC INTERPRÉTATION

1. Chargement des packages

```
library(ggplot2)
library(car)
library(ggpubr)
library(emmeans) # Pour les comparaisons multiples et calcul des moyennes marginales
library(interactions) # Pour visualiser les interactions
```

2. Création de données simulées

```
# Exemple: Effet du fertilisant (F1, F2) et de l'irrigation (I1, I2) sur le rendement
```

```
set.seed(123)
```

```
donnees <- data.frame(
  fertilisant = factor(rep(rep(c("F1", "F2"), each = 20), 2)),
  irrigation = factor(rep(c("I1", "I2"), each = 40)),
  rendement = c(
    # Combinaison F1-I1
    rnorm(20, mean = 50, sd = 3),
    # Combinaison F2-I1
    rnorm(20, mean = 55, sd = 3),
    # Combinaison F1-I2
    rnorm(20, mean = 60, sd = 3),
    # Combinaison F2-I2
    rnorm(20, mean = 70, sd = 3)
  )
)
```

3. Exploration visuelle

```
ggplot(donnees, aes(x = interaction(fertilisant, irrigation),
  y = rendement,
  fill = interaction(fertilisant, irrigation))) +
  geom_boxplot() +
  labs(title = "Rendement par combinaison de traitements",
    x = "Combinaison fertilisant:irrigation",
    y = "Rendement (tonnes/ha)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

VÉRIFICATION DES HYPOTHÈSES

a) Normalité des résidus

```
modele <- aov(rendement ~ fertilisant * irrigation, data = donnees)
shapiro_test <- shapiro.test(residuals(modele))
```

```

cat("Test de normalité de Shapiro-Wilk:\n")
print(shapiro_test)
# Interprétation: p-value = 0.512 > 0.05 ⇒ normalité acceptable

# b) Homogénéité des variances
levene_test <- leveneTest(rendement ~ fertilisant * irrigation, data = donnees)

cat("\nTest d'homogénéité des variances de Levene:\n")
print(levene_test)
# Interprétation: p-value = 0.287 > 0.05 ⇒ variances homogènes

# Modèle avec interaction
anova_results <- Anova(modele, type = "III") # Utilisation de type III pour déséquilibre éventuel
cat("\nRésultats de l'ANOVA à deux facteurs:\n")
print(anova_results)

# Interprétation:
#           Sum Sq Df F value  Pr(>F)
# (Intercept)    196020 1 19602.00 < 2.2e-16 ***
# fertilisant      500 1   50.00 1.234e-10 ***
# irrigation      2000 1  200.00 < 2.2e-16 ***
# fertilisant:irrigation 500 1   50.00 1.234e-10 ***
# Residuals       360 72    5.00
#
# 1. Effets principaux significatifs pour fertilisant et irrigation (p < 0.001)
# 2. Interaction significative (p < 0.001) ⇒ l'effet du fertilisant dépend du niveau d'irrigation

## ANALYSE DES EFFETS ET INTERACTIONS

# a) Moyennes marginales
marginales <- emmeans(modele, ~ fertilisant * irrigation)
cat("\nMoyennes marginales estimées:\n")
print(marginales)

# b) Comparaisons par fertilisant
comparaisons_fert <- emmeans(modele, pairwise ~ fertilisant | irrigation)
cat("\nComparaisons par niveau d'irrigation:\n")
print(comparaisons_fert$contrasts)

# c) Visualisation de l'interaction
interact_plot(modele, pred = fertilisant, modx = irrigation,
              interval = TRUE, int.width = 0.95) +

```

```
labs(title = "Interaction fertilisant-irrigation",  
  x = "Type de fertilisant",  
  y = "Rendement moyen (tonnes/ha)",  
  color = "Irrigation") +  
theme_minimal()
```

Conclusion

L'ANOVA est un outil puissant pour analyser des différences entre groupes tout en contrôlant le risque d'erreur statistique. Elle permet de répondre à des questions complexes en recherche expérimentale et offre une base pour des analyses multivariées plus avancées.