

Statistiques appliquées aux sciences

SCI 1018

Analyse de variance à un critère

Marc J. Mazerolle

Institut de recherche sur les forêts, Université du Québec en Abitibi-Témiscamingue



© TÉLUQ, 2013

Table des matières

1	Introduction	2
2	Comparaison de plusieurs groupes	2
3	Analyse de variance (ANOVA)	4
3.1	Décomposition de la variance	5
3.1.1	Sommes des carrés	5
3.1.2	Tableau de l'ANOVA	12
3.2	La distribution du F	13
3.3	Hypothèses statistiques	13
3.4	Suppositions	16
3.4.1	Résidus	16
3.4.2	Normalité des résidus et homogénéité de la variance	18
4	Comparaisons multiples : trouver les différences	25
4.1	Erreurs associées aux comparaisons multiples	26
4.2	Suppositions	29
4.3	Structure générale	29
4.4	Test de Tukey	30
4.5	Présentation des résultats	34
	Conclusion	37
	Index	39

1 Introduction

Dans la leçon précédente, nous avons vu l'analyse de fréquences, particulièrement le test du χ^2 d'ajustement et d'indépendance. Nous continuons ici avec l'analyse de données numériques continues, du même type que pour le test t . Dans la présente leçon, nous verrons un outil pour comparer plus de deux groupes, l'analyse de la variance.

2 Comparaison de plusieurs groupes

Nous avons vu que le test t permet de comparer deux groupes afin de déterminer s'ils proviennent de la même population. Dans certains cas, on s'intéresse à plus de deux groupes. Par exemple, on pourrait vouloir comparer la masse de poissons soumis à trois traitements différents. Rappelons qu'à chaque fois que nous réalisons un test d'hypothèse statistique, nous avons une probabilité α de commettre une erreur de type I (c.-à-d., déclarer une différence lorsqu'il n'y en a pas, un faux positif). Avec trois groupes (A, B, C), il existe trois comparaisons possibles :

$$A \text{ vs } B \quad A \text{ vs } C \quad B \text{ vs } C$$

À noter que la comparaison $A \text{ vs } B$ est identique à $B \text{ vs } A$. La probabilité de trouver une différence entre les trois moyennes est supérieure à α . Plus on fait de comparaisons avec un test t (e.g., $A \text{ vs } B$, $A \text{ vs } C$), plus on risque de conclure à tort qu'il y a une différence entre les moyennes des groupes. Si les tests sont indépendants, on peut calculer la probabilité de commettre au moins une erreur de type I à l'aide de l'équation :

$$P = 1 - (1 - \alpha)^k$$

où α est le seuil de signification pour chaque comparaison et k est le nombre de comparaisons effectuées. Pour un seuil fixé à $\alpha = 0.05$:

- avec 3 comparaisons, la probabilité de commettre une erreur de type I est de 0.14 ;
- avec 10 comparaisons, la probabilité de commettre une erreur de type I est de 0.40 ;
- avec 20 comparaisons, la probabilité de commettre une erreur de type I est de 0.64.

Toutefois, si les comparaisons ne sont pas indépendantes, la probabilité réelle de commettre une erreur de type I est inférieure ou égale à celle obtenue avec l'équation ci-haut. On comprend que l'augmentation du nombre de comparaisons risque d'entraîner des conclusions erronées.

Une alternative à ce problème consiste à corriger le seuil α par le nombre de comparaisons que l'on veut effectuer. C'est ce qu'on appelle l'**ajustement de Bonferroni** (*Bonferroni adjustment*).

Exemple 6.1 Dans une expérience où on étudie 5 traitements différents, nous avons cinq moyennes arithmétiques (5 groupes : A, B, C, D, E) et les comparaisons suivantes :

$$\begin{array}{ccccccc}
 A & vs & B & & & & \\
 A & vs & C & & B & vs & C \\
 A & vs & D & & B & vs & D & & C & vs & D \\
 A & vs & E & & B & vs & E & & C & vs & E & & D & vs & E .
 \end{array}$$

Puisqu'il y a 10 comparaisons, on devrait diviser le seuil α par 10 pour exécuter l'ajustement de Bonferroni. Si on utilise un seuil de 0.05, il faudra observer $P \leq 0.005$ pour déclarer une différence significative entre deux groupes ($\alpha/10 = 0.05/10 = 0.005$).

On constate que l’ajustement de Bonferroni est une stratégie souvent trop conservatrice, qui augmente la probabilité de commettre une erreur de type II (un faux négatif). Une meilleure stratégie consiste à comparer tous les groupes simultanément. C’est d’ailleurs ce que permet de faire l’analyse de variance – on compare tous les groupes en une seule étape.

3 Analyse de variance (ANOVA)

L’**analyse de variance** (*analysis of variance*, ANOVA) est une approche statistique qui compare les moyennes de plusieurs groupes entre eux pour déterminer si au moins un des groupes diffère des autres. Le scénario typique de l’ANOVA est celui d’une expérience visant à comparer l’effet d’une **variable catégorique** sur une **variable réponse** d’intérêt. Une variable catégorique est une variable à partir de laquelle la variable réponse peut être classée dans une catégorie particulière. Par variable réponse, on entend une variable dépendante que l’on mesure et qui peut être influencée par la variable catégorique. Les termes « variable catégorique », « facteur » et « critère » sont souvent interchangeables dans la littérature. Nous utiliserons le terme « facteur » pour les besoins du cours.

Des exemples de variables catégoriques incluent les classes d’âge (jeune, mature, vieux), le traitement sylvicole (témoin, coupe partielle, coupe totale), et la concentration d’engrais (très faible, modérée, élevée, très élevée). On utilise aussi le terme **facteur** pour désigner une variable catégorique. Dans le contexte de l’ANOVA, la variable réponse est une variable numérique continue qui varie selon les niveaux du facteur. Comme nous le verrons plus loin, l’ANOVA est l’extension du test t pour des cas avec des variables catégoriques qui ont plus de 2 niveaux.

Étant donné que l’on s’intéresse aux moyennes, on pourrait se demander pourquoi appeler cette approche ANOVA. Lorsqu’on réalise une ANOVA, on détermine si la variabilité des données expliquée par le traitement est plus grande que la variabilité inexpliquée. Pour ce faire, nous calculons la **somme des carrés** (*sum of squares*). Le concept de la somme des

carrés a été présenté lors de la leçon *Statistiques descriptives*.

3.1 Décomposition de la variance

3.1.1 Sommes des carrés

Pendant l'ANOVA, on décortique la variance totale des données en différentes composantes : une partie expliquée et une autre inexpliquée. Puisque les sommes des carrés sont additives, on peut écrire :

$$SST = SSA + SSE ,$$

où SST correspond à la **somme des carrés totale**, SSA correspond à la **somme des carrés du facteur A** (aussi appelée somme des carrés du traitement) et SSE est la **somme des carrés des erreurs**. On peut diviser les sommes des carrés par les degrés de liberté appropriés pour obtenir le **carré moyen** (la variance) attribuable aux effets qui nous intéressent. Nous présentons dans les prochaines lignes le calcul de chacune des sommes des carrés discutées précédemment.

On calcule la somme des carrés totale (SST) :

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

où y_i est la $i^{\text{ième}}$ observation et \bar{y} représente la moyenne globale de toutes les observations (tous les groupes confondus). Cette somme des carrés compare chaque observation à la moyenne globale. Les degrés de liberté associés à SST sont obtenus avec $df = N - 1$, où N correspond au nombre total d'observations.

On calcule la somme des carrés des erreurs (SSE) :

$$SSE = \sum_{j=1}^k \sum_{i=1}^n (y_{ij} - \bar{y}_j)^2$$

où y_{ij} correspond à la $i^{\text{ième}}$ observation du groupe j et \bar{y}_j est la moyenne du groupe j . Cette

somme des carrés compare chaque observation à la moyenne de son groupe respectif. On obtient les degrés de liberté avec $df = k(n - 1)$, où k correspond au nombre de groupes et n correspond au nombre d'observations dans chacun des groupes.

On calcule la somme des carrés du facteur A (le traitement qui définit les différents groupes) (SSA) :

$$SSA = \sum_{j=1}^k \sum_{i=1}^n (\bar{y}_j - \bar{y})^2 = n \sum_{j=1}^k (\bar{y}_j - \bar{y})^2$$

où \bar{y}_j correspond à la moyenne du groupe j , \bar{y} est la moyenne globale et n correspond au nombre d'observations dans chacun des groupes. On peut aussi la calculer par soustraction à l'aide de $SSA = SST - SSE$. La somme des carrés du traitement compare la moyenne de chaque groupe à la moyenne globale. Les degrés de liberté s'obtiennent avec $df = k - 1$ où k est le nombre de groupes dans le facteur A .

Exemple 6.2 On veut déterminer si la concentration d'ozone en parties par million (ppm) diffère entre trois jardins d'une ville industrielle. Le jeu de données `jardins.txt` contient les observations suivantes :

```
> ##importation du jeu de données
> ozone <- read.table("jardins.txt", header = TRUE)
> ozone
```

	Ozone	Jardin
1	14.82	A
2	20.59	A
3	6.18	A
4	7.61	A
5	31.35	A
6	11.62	A
7	32.70	A

8	26.18	A
9	19.91	A
10	10.96	A
11	12.60	A
12	17.09	A
13	5.97	A
14	17.95	A
15	9.60	A
16	20.46	B
17	18.26	B
18	28.63	B
19	14.81	B
20	14.22	B
21	13.97	B
22	20.19	B
23	16.21	B
24	23.63	B
25	21.02	B
26	20.41	B
27	18.59	B
28	13.19	B
29	18.28	B
30	11.15	B
31	10.03	C
32	11.58	C
33	26.68	C
34	5.61	C

35	12.21	C
36	24.87	C
37	18.60	C
38	34.78	C
39	14.63	C
40	17.09	C
41	5.05	C
42	20.53	C
43	28.67	C
44	12.19	C
45	28.65	C

On peut visualiser les données à l'aide d'un diagramme de boîtes et moustaches (fig. 1). On remarque que la variabilité est similaire d'un jardin à l'autre.

```
> ##diagramme de boîtes et moustaches
> boxplot(Ozone ~ Jardin, data = ozone,
          ylab = "Concentration d'ozone (ppm)",
          xlab = "Jardin", cex.lab = 1.2)
```

À partir des données, on calcule les sommes des carrés totales,

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad \bar{y} = 17.54$$

$$SST = (14.82 - 17.54)^2 + (20.59 - 17.54)^2 + \dots + (12.19 - 17.54)^2 + (28.65 - 17.54)^2$$

$$SST = 2483.9$$

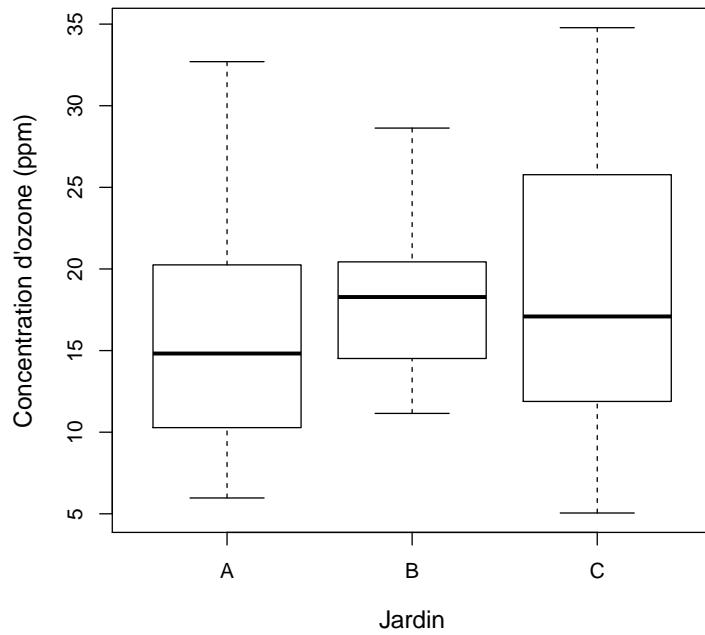


FIGURE 1 – Diagramme de boîtes et moustaches de la concentration d’ozone dans les trois jardins.

La somme des carrés des erreurs s’obtient ainsi :

$$SSE = \sum_{j=1}^k \sum_{i=1}^n (y_{ij} - \bar{y}_i)^2 \quad \bar{y}_A = 16.34 \quad \bar{y}_B = 18.20 \quad \bar{y}_C = 18.08$$

$$SSE = (14.82 - 16.34)^2 + \dots + (9.60 - 16.34)^2 + (20.46 - 18.20)^2 + \dots + (11.15 - 18.20)^2 + \\ (10.03 - 18.08)^2 + \dots + (28.65 - 18.08)^2$$

$$SSE = 2451.5$$

On calcule la somme des carrés des traitements avec l’équation complète ou par

soustraction :

$$SSA = \sum_{j=1}^k \sum_{i=1}^n (\bar{y}_i - \bar{y})^2 = n \sum_{j=1}^k (\bar{y}_j - \bar{y})^2 \quad n_A = n_B = n_C = 15$$

$$SSA = 15 \cdot ((16.34 - 17.54)^2 + (18.20 - 17.54)^2 + (18.08 - 17.54)^2)$$

$$SSA = 32.43$$

On effectue ces calculs dans R :

```
> ##SST
> SST <- sum((ozone$Ozone - mean(ozone$Ozone))^2)
> SST
[1] 2483.9
> ##degrés de liberté de SST
> df.tot <- nrow(ozone) - 1
> df.tot
[1] 44
> ##SSE
> ##sous-jeu de données
> ozoneA <- ozone[ozone$Jardin == "A", ]
> ozoneB <- ozone[ozone$Jardin == "B", ]
> ozoneC <- ozone[ozone$Jardin == "C", ]
> ##SSE de A
> SSE.A <- sum((ozoneA$Ozone-mean(ozoneA$Ozone))^2)
> ##SSE de B
> SSE.B <- sum((ozoneB$Ozone-mean(ozoneB$Ozone))^2)
> ##SSE de C
> SSE.C <- sum((ozoneC$Ozone-mean(ozoneC$Ozone))^2)
```

```

> SSE <- SSE.A + SSE.B + SSE.C
> SSE

[1] 2451.5

> ##degrés de liberté de SSE
> df.erreur <- 3 * (15 - 1)
> df.erreur

[1] 42

> ##SSA
> SSA <- SST - SSE
> SSA

[1] 32.43

> ##autre façon de la calculer
> SSA.alt <- (15 * (mean(ozoneA$Ozone) - mean(ozone$Ozone))^2) +
  (15 * (mean(ozoneB$Ozone) - mean(ozone$Ozone))^2) +
  (15 * (mean(ozoneC$Ozone) - mean(ozone$Ozone))^2)
> SSA.alt

[1] 32.43

> ##degrés de liberté du traitement
> df.trait <- 3 - 1
> df.trait

[1] 2

```

3.1.2 Tableau de l'ANOVA

Après avoir calculé les sommes des carrés des différents termes ainsi que leurs degrés de liberté respectifs, on peut les assembler dans un tableau d'ANOVA (tableau 1). Ce tableau constitue une partie importante des résultats de l'analyse et doit figurer dans les rapports ou au moins être résumé dans par écrit.

Tableau 1 – Calcul des différents éléments du tableau d'ANOVA.

Source	Somme des carrés (SS)	Degrés de liberté (df)	Carré moyen (MS)	ratio F
Traitement	SSA	$k - 1$	SSA/df_A	MSA/MSE
Erreur	SSE	$k(n - 1)$	SSE/df_{erreur}	
Total	SST	$N - 1$		

Le carré moyen du traitement (SSA/df_A) estime la variance due au traitement, alors que le carré moyen des erreurs (SSE/df_{erreur}) estime la variance inexpliquée aussi appelée **variance résiduelle** (*residual variance*). La variance résiduelle est la meilleure estimation de la variance (σ^2) commune à tous les k groupes comparés. C'est une des raisons pour lesquelles la méthode est appelée « analyse de variance » : on estime en même temps la variance commune à tous les groupes.

Le ratio F consiste à comparer la variabilité expliquée et la variabilité inexpliquée. Si le ratio est supérieur à 1, c'est parce que la variance expliquée par le traitement excède la variance résiduelle : il y a un effet du traitement sur la variable réponse. Un ratio inférieur ou égal à 1 suggère qu'il n'y a pas d'effet du traitement. L'interprétation du ratio F dépend aussi de la taille de l'échantillon. Ce ratio de variances est comparé à une distribution théorique, celle du F .

3.2 La distribution du F

La distribution du F est une distribution continue qui est définie par la fonction de densité :

$$f(x|\nu_1, \nu_2) = \frac{\sqrt{\frac{(\nu_1 x)^{\nu_1} \nu_2^{\nu_2}}{(\nu_1 x + \nu_2)^{\nu_1 + \nu_2}}}}{xB\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)},$$

où ν_1 et ν_2 sont les degrés de liberté et B est la fonction beta qui implique des intégrales (voir `?beta` dans R). En ce qui concerne l'application de cette distribution à l'ANOVA, les degrés de liberté ν_1 et ν_2 sont ceux associés au numérateur et au dénominateur du ratio F , respectivement. La Figure 2 présente la forme de la distribution du F selon différents degrés de liberté. Comme d'habitude, nous comparons la statistique obtenue à partir des données de l'échantillon à celle de la distribution du F selon les degrés de liberté et le seuil de signification spécifiés lorsque l'hypothèse nulle est vraie. Si la valeur du F observée est plus grande que celle déterminée par la distribution du F , nous rejetterons l'hypothèse nulle.

3.3 Hypothèses statistiques

Comme nous l'avons vu dans les leçons précédentes, l'hypothèse nulle (bilatérale) avec le test t de Student sur deux groupes indépendants implique l'égalité des moyennes des deux groupes :

$$H_0 : \mu_A = \mu_B \text{ (non-différence)}$$

$$H_a : \mu_A \neq \mu_B \text{ (différence)}$$

En ce qui a trait à l'ANOVA, l'hypothèse nulle a la même forme. Par exemple, pour un facteur qui compte quatre niveaux (p. ex., faible, modéré, élevé, très élevé), nous aurions l'hypothèse nulle suivante :

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D \text{ (non-différence)}$$

$$H_a : \text{au moins une moyenne diffère des autres moyennes}$$

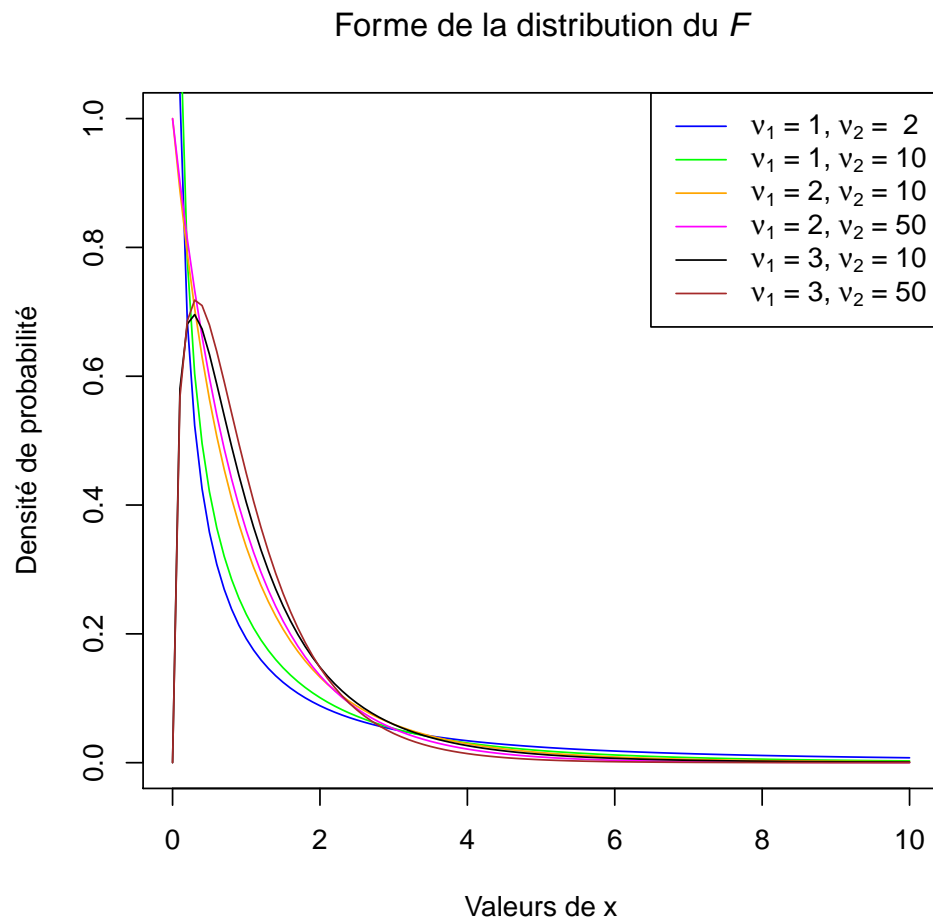


FIGURE 2 – Distribution du F selon différentes valeurs de degrés de liberté.

Exemple 6.3 Reprenons notre exemple sur la concentration d’ozone dans les trois jardins d’une ville industrielle. Nous pouvons émettre les hypothèses statistiques suivantes :

$$H_0 : \mu_{\text{jardin A}} = \mu_{\text{jardin B}} = \mu_{\text{jardin C}} \text{ (non-différence)}$$

$$H_a : \text{au moins une moyenne diffère des autres moyennes}$$

$$\alpha = 0.05$$

Puisque nous avons déjà calculé les sommes des carrés dans l’exemple 6.2, nous pouvons construire le tableau d’ANOVA qui nous permettra de tester l’hypothèse nulle (tableau 2).

Tableau 2 – Tableau d’ANOVA concernant la concentration d’ozone dans les jardins d’une ville industrielle.

Source	Somme des carrés (<i>SS</i>)	Degrés de liberté (<i>df</i>)	Carré moyen (<i>MS</i>)	ratio <i>F</i>
Jardin	32.43	2	16.22	0.28
Residuals	2451.45	42	58.37	

On constate que le ratio F est de 0.28 (16.22/58.37). Étant donné que $P(F_{2,42} \geq 0.28) = 0.7588$ ¹, on ne rejette pas H_0 et on conclut que les moyennes des groupes ne diffèrent pas les unes des autres. La fonction `aov()` permet de réaliser une ANOVA dans R lorsque chaque groupe contient le même nombre d’observations².

On peut extraire le tableau de l’ANOVA à l’aide de la fonction `summary()`. Cette dernière fonction est souvent utilisée pour obtenir un résumé des résultats d’une

1. Rappel : Ici, la valeur de P correspond à la probabilité d’obtenir une valeur de F supérieure ou égale à celle qu’on a observée (dans les données originales) lorsqu’on répète l’échantillonnage avec la même taille d’échantillon dans la même population et que l’hypothèse nulle est vraie. On peut écrire plus succinctement : $P = 0.7588$.

2. Dans les cas où le nombre d’observations varie d’un groupe à l’autre, il est préférable d’utiliser la fonction plus générale, `lm()` qui permet de réaliser plusieurs types de modèles linéaires.

analyse statistique.

```
> ##exécuter l'ANOVA
> aov1 <- aov(Ozone ~ Jardin, data = ozone)
> ##on extrait les résultats
> summary(aov1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Jardin	2	32	16.2	0.28	0.76
Residuals	42	2451	58.4		

3.4 Suppositions

L'analyse de variance est une extension du test t de Student lorsqu'on souhaite comparer les moyennes entre plus de deux groupes. Il n'est donc pas surprenant que l'analyse de variance doive répondre aux mêmes conditions d'utilisation que le test t . Pour effectuer l'ANOVA, les erreurs doivent être indépendantes. L'échantillonnage aléatoire et l'utilisation d'un bon dispositif expérimental aident à respecter cette condition. L'ANOVA requiert également que les variances des groupes soient égales (homoscédasticité) : $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 \dots \sigma_k^2$. Finalement, on suppose que chacun des groupes est issu d'une population normale (c.-à-d., résidus ou erreurs sont distribués normalement). Autrement dit, afin de pouvoir comparer des groupes dont la moyenne peut potentiellement différer d'un groupe à l'autre, il faut que ces groupes aient la même variance. Cette variance commune est estimée par le carré moyen de l'erreur (*residual mean square, error mean square*).

3.4.1 Résidus

À l'instar du test t , nous utilisons les résidus pour diagnostiquer des problèmes liés à la normalité et à l'homogénéité des variances. Les résidus sont obtenus en calculant la différence

entre les valeurs observées y_i et les valeurs qui sont prédites \hat{y}_i par notre modèle d'ANOVA. Dans le cas de l'ANOVA à un critère, les valeurs prédites (\hat{y}_i) correspondent à la moyenne arithmétique de chaque groupe. Nous obtenons les résidus à l'aide de $y_i - \hat{y}_i$. Les fonctions `residuals()` et `fitted()` dans R permettent d'extraire les résidus et les valeurs prédites de plusieurs types d'analyses statistiques. Le prochain exemple illustre l'application directe de ces fonctions à l'objet qui contient le résultat de l'ANOVA.

Exemple 6.4 On peut vérifier les suppositions de l'ANOVA que nous avons exécutée dans l'exemple précédent. On obtient les résidus en calculant la différence entre les valeurs observées et les valeurs prédites, $y_i - \hat{y}_i$. On commence par extraire les valeurs observées et les valeurs prédites :

```
> ##valeurs observées
> obs <- ozone$Ozone
> obs[1:10]

[1] 14.82 20.59  6.18  7.61 31.35 11.62 32.70 26.18 19.91
[10] 10.96

> ##extraire les valeurs prédites
> pred <- fitted(aov1)
> pred[1:10]

      1      2      3      4      5      6      7      8
16.342 16.342 16.342 16.342 16.342 16.342 16.342 16.342
      9     10
16.342 16.342

> ##moyenne arithmétique de chaque groupe
> tapply(X = ozone$Ozone, INDEX = ozone$Jardin, FUN = mean)
```

	A	B	C
	16.342	18.201	18.078

On remarque que les valeurs prédites par l'ANOVA à un critère correspondent à la moyenne arithmétique de chaque groupe, comme l'indique par la fonction `tapply()`³. Les résidus s'obtiennent facilement à l'aide du calcul $y_i - \hat{y}_i$ ou avec la fonction `residuals()` :

```
> ##résidus
> obs[1:10] - pred[1:10]
```

1	2	3	4	5	6	7
-1.522	4.248	-10.162	-8.732	15.008	-4.722	16.358
8	9	10				
9.838	3.568	-5.382				

```
> ##extraire résidus du modèle
> res <- residuals(aov1)
> res[1:10] #valeurs identiques au calcul obs - pred
```

1	2	3	4	5	6	7
-1.522	4.248	-10.162	-8.732	15.008	-4.722	16.358
8	9	10				
9.838	3.568	-5.382				

3.4.2 Normalité des résidus et homogénéité de la variance

De la même façon qu'on utilise le graphique quantile-quantile lorsqu'on effectue un test t , on utilise ce même graphique pour diagnostiquer des déviations par rapport à la supposition

3. La fonction `tapply()` applique une fonction donnée (ici, `mean`) aux valeurs d'une variable `X` (`ozone$Ozone`) regroupées selon les différentes catégories de `INDEX` (`ozone$Jardin`).

de normalité. La fonction `qqnorm()` permet d'obtenir ce graphique et la fonction `qqline()` ajoute une droite théorique de $1 : 1$, c'est-à-dire une droite où chaque point de l'abscisse (axe des valeurs x) est égal à celui de l'ordonnée à l'origine (axe des valeurs y).

Pour évaluer l'homogénéité des variances des différents groupes, on peut utiliser le diagramme de boîtes et moustaches (`boxplot()`) des résidus en fonction de chaque groupe. Un autre graphique utile pour diagnostiquer des problèmes de variances hétérogènes est le graphique des résidus en fonction des valeurs prédites. Ce dernier devrait montrer un patron nul, c'est-à-dire des points distribués uniformément de part et d'autre de 0 sur l'axe des y sans patron apparent (fig. 3a). L'hétérogénéité des variances se traduit parfois par l'apparition d'un patron en forme d'entonnoir, indiquant que les variances augmentent avec les valeurs prédites (fig. 3b).

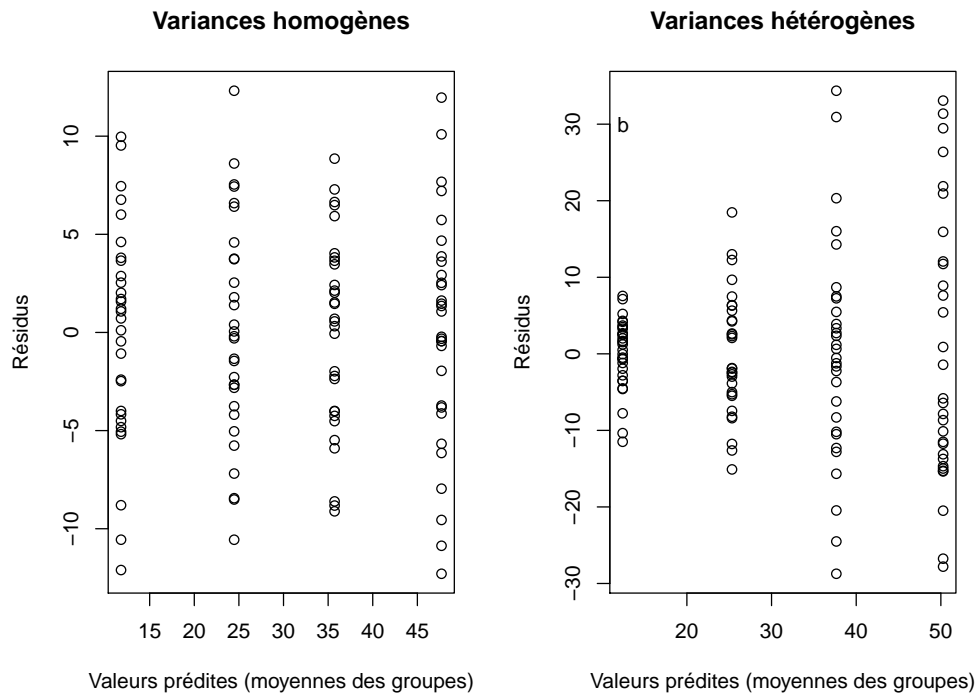


FIGURE 3 – Graphique de résidus en fonction des valeurs prédites illustrant l'homogénéité des variances (a) et l'hétérogénéité des variances (b). Notez le patron en forme d'entonnoir en b qui est indiqué par une variance qui augmente avec les valeurs prédites.

Exemple 6.5 Nous poursuivons l'exemple 6.4 en vérifiant si la condition de normalité est respectée. Pour vérifier la normalité, on extrait les résidus de l'ANOVA et on examine le graphique quantile-quantile. On constate que la condition de normalité est assez bien respectée, quoiqu'il y ait quelques valeurs aux extrémités des queues qui dévient de la normalité (fig. 4a). Le graphique des résidus en fonction des valeurs prédites suggère que les variances sont plutôt homogènes, bien que le groupe avec la plus grande moyenne ait une variabilité légèrement plus faible que les autres (fig. 4b).

```
> ##graphique quantile-quantile
> qqnorm(res, ylab = "Quantiles observés",
          xlab = "Quantiles théoriques",
          main = "Graphique quantile-quantile de normalité",
          cex.lab = 1.2)
> ##cex.lab indique que les étiquettes seront 1.2 fois plus grande
> ##que la normale
> ##ajout de droite théorique
> qqline(res)
> ##résidus vs valeurs prédites
> plot(res ~ pred, ylab = "Résidus", xlab = "Valeurs prédites",
       cex.lab = 1.2)
> ##cex.lab indique que les étiquettes seront 1.2 fois plus grande
> ##que la normale
```

Nous avons maintenant toutes les notions en main afin de réaliser l'ANOVA et de vérifier le respect des conditions. C'est ce que nous ferons dans le prochain exemple.

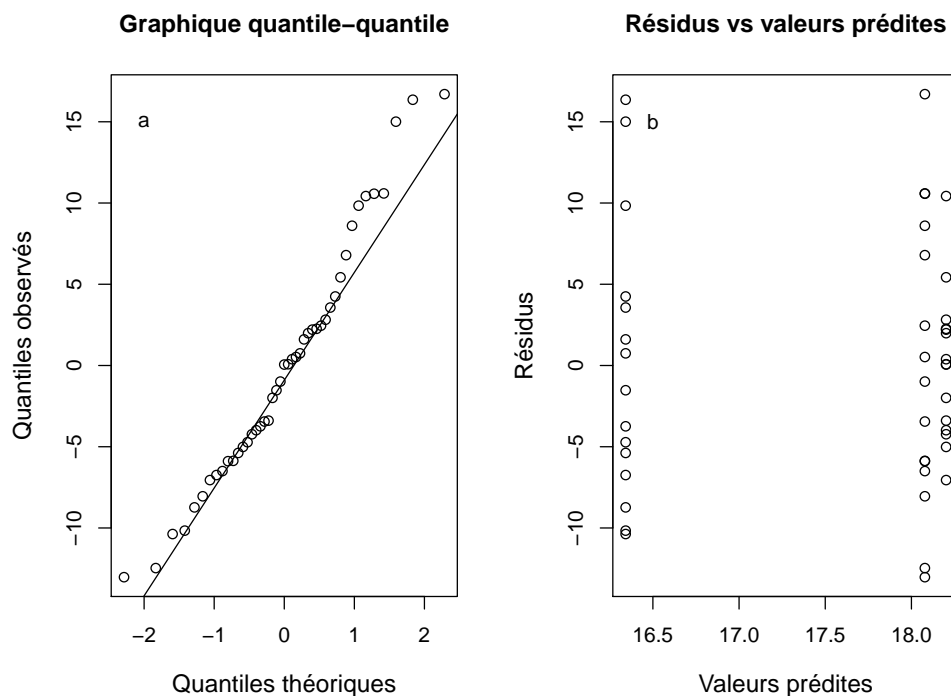


FIGURE 4 – Graphique quantile-quantile des résidus de l’ANOVA sur les concentrations d’ozone (a) et résidus en fonction des valeurs prédites (b).

Exemple 6.6 On étudie l’effet de la compétition sur la biomasse produite par une espèce de plante. Pour ce faire, on effectue une expérience en procédant à cinq régimes d’élagage aussitôt le débourrement⁴ complet des feuilles au printemps :

- témoin (contrôle, aucun élagage des plantes voisines) ;
- élagage de 25 % des feuilles des voisins (n25) ;
- élagage de 50 % des feuilles des voisins (n50) ;
- élagage de 10 % des racines des voisins (r10) ;
- élagage de 5 % des racines des voisins (r5).

Les données sont incluses dans le jeu de données `competition.txt`.

```
> comp <- read.table("competition.txt", header = TRUE)
> head(comp)
```

4. La période pendant laquelle les bourgeons s’épanouissent.

	Biomasse	Elagage
1	551	n25
2	457	n25
3	450	n25
4	731	n25
5	499	n25
6	632	n25

Nous avons les hypothèses statistiques suivantes :

$H_0 : \mu_{\text{contrôle}} = \mu_{\text{n25}} = \mu_{\text{n50}} = \mu_{\text{r10}} = \mu_{\text{r5}}$ (non-différence)

H_a : au moins une moyenne diffère parmi toutes les moyennes

$\alpha = 0.05$

Avant de faire l'analyse, on peut visualiser rapidement les données et la variabilité de chaque groupe avec `boxplot()` (fig. 5). On exécute l'ANOVA à un critère :

```
> ##ANOVA
> m1 <- aov(Biomasse ~ Elagage, data = comp)
```

Avant de se lancer dans l'interprétation, il faut vérifier les suppositions de l'ANOVA, notamment celles concernant l'homoscédasticité et la normalité des résidus. On constate que la supposition d'homogénéité des variances est assez bien respectée et une seule observation se démarque des autres avec une valeur > 150 (fig. 6a). Le graphique quantile-quantile avec les résidus indique que la supposition de normalité est respectée pour ce jeu de données (fig. 6b).

```
> ##on organise la fenêtre graphique
> ##pour avoir 1 rangée et deux colonnes
> ##de graphiques
> par(mfrow = c(1, 2))
```

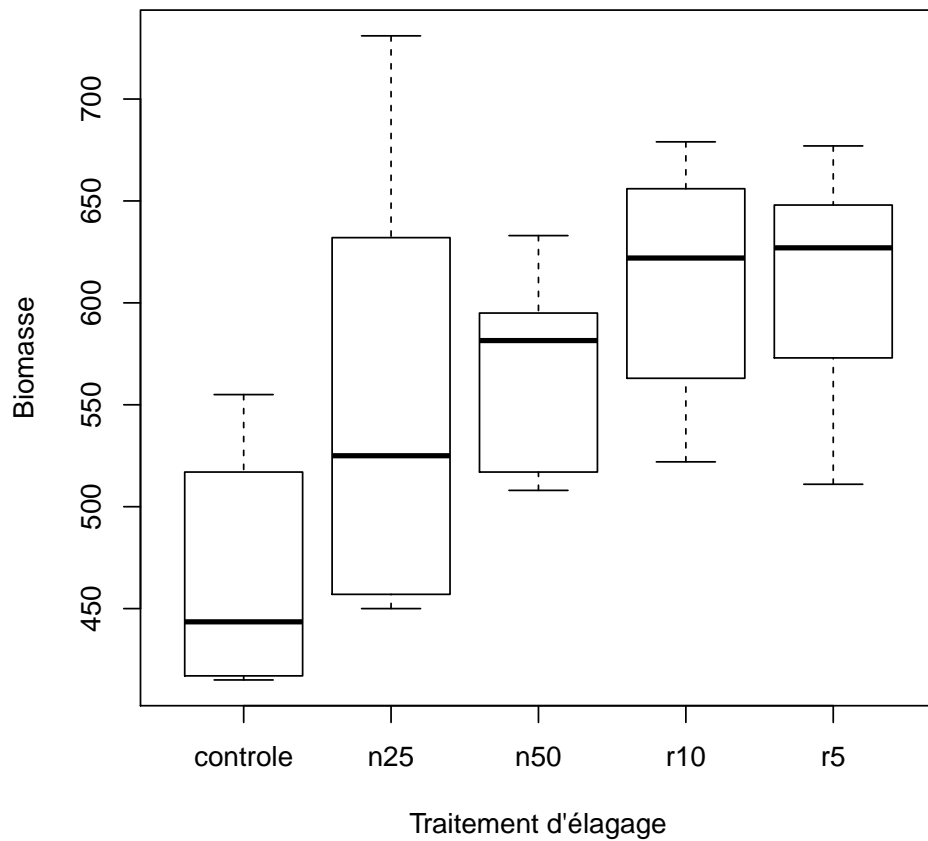


FIGURE 5 – Diagramme de boîtes et moustaches de la biomasse pour chaque traitement d'élague.

```
> ##homoscédasticité
> plot(residuals(m1) ~ fitted(m1),
       ylab = "Résidus", xlab = "Valeurs prédites",
       main = "Résidus vs valeurs prédites",
       cex.lab = 1.2)
> ##normalité
> qqnorm(residuals(m1), ylab = "Quantiles observés",
        xlab = "Quantiles théoriques",
        main = "Graphique quantile-quantile",
```



```
cex.lab = 1.2)
> qqline(residuals(m1))
```

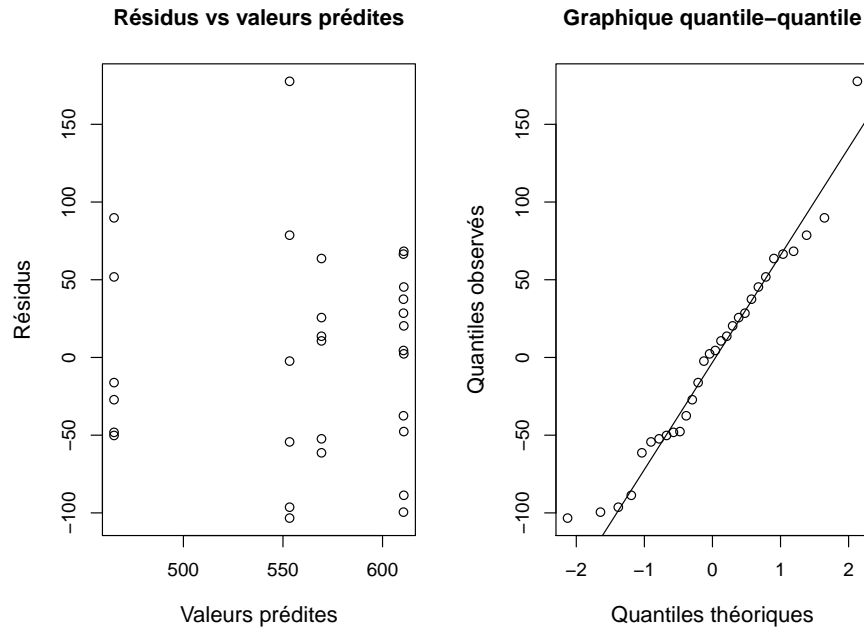


FIGURE 6 – Diagnostics de l’ANOVA.

On peut procéder à l’interprétation des résultats.

```
> ##on extrait les résultats
> summary(m1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Elagage	4	85356	21339	4.3	0.0088	**
Residuals	25	124020	4961			

Signif. codes:

0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

On note dans `summary(m1)` que la dernière ligne du tableau `Signif. codes` donne les symboles qui décrivent le seuil de signification statistique du facteur `Elagage`.

Ces caractères ne sont affichés qu'en présence de termes significatifs dans l'analyse. Ici, **Elagage** a les caractères ** sur sa ligne, indiquant que $P \leq 0.01$. Ces symboles n'ont d'autre fonction que d'attirer l'attention de l'analyste sur les termes significatifs dans l'analyse. Il est peu probable d'observer une valeur de F de 4.3 dans une expérience avec des échantillons (groupes) qui ont la même taille que le jeu de données original tirés d'une population où H_0 est vraie. La valeur exacte du P est de 0.0088 ($P(F_{4,25} \geq 4.3015) = 0.0088$). On rejette l'hypothèse nulle et on conclut qu'il y a un effet du traitement d'élagage : au moins une moyenne des groupes d'élagage diffère des autres. Toutefois, l'ANOVA ne nous permet pas d'identifier quels traitements diffèrent entre eux. Pour ce faire, on doit utiliser un autre type de test qui sera présenté dans la prochaine section.

4 Comparaisons multiples : trouver les différences

L'ANOVA permet de déterminer si la variance expliquée par le facteur à l'étude est supérieure à la variance résiduelle (c.-à-d., la variance inexpliquée). Ainsi, le tableau d'ANOVA nous permet de rejeter ou non H_0 . Lorsqu'on rejette l'hypothèse nulle après avoir effectué un test t pour comparer deux groupes indépendants, on sait qu'il existe une différence entre ces deux groupes. Dans une ANOVA qui a plus de deux groupes, le rejet de H_0 ne nous révèle pas où se trouvent les différences⁵. Les comparaisons multiples permettent d'identifier les différences entre les moyennes des groupes.

Comme nous l'avons mentionné au début de ce texte, effectuer des comparaisons multiples augmente la probabilité de commettre une erreur de type I. L'ANOVA est conçue

5. Le tableau d'ANOVA classique ne permet pas de décortiquer où se trouvent les différences entre les groupes. Néanmoins, il est possible d'utiliser des contrastes orthogonaux pour décomposer la somme des carrés du facteur à l'intérieur du tableau d'ANOVA. Ce concept est plus avancé et ne sera pas couvert dans le cours.

pour comparer tous les groupes simultanément afin d'évaluer l'effet global du traitement. Lorsque l'ANOVA entraîne le rejet de l'hypothèse nulle, il s'avère nécessaire d'exécuter des comparaisons multiples entre les groupes afin de trouver où se trouvent les différences. Ces comparaisons effectuées après l'ANOVA augmentent aussi la probabilité de commettre une erreur de type I – la probabilité de commettre ce type d'erreur est supérieure au seuil α que nous avons fixé. Puisque les comparaisons multiples sont effectuées après l'ANOVA, on parle souvent de tests *a posteriori* ou *post hoc* (*post hoc tests*, *a posteriori tests*).

4.1 Erreurs associées aux comparaisons multiples

On peut commettre des erreurs à deux niveaux lorsqu'on fait des comparaisons multiples : une **erreur au niveau de la comparaison** (*comparison-wise error*) et une **erreur au niveau de l'expérience** (*experiment-wise error*). L'erreur au niveau de la comparaison se traduit par la déclaration d'un faux positif lors d'une comparaison spécifique entre deux groupes. La probabilité de commettre cette erreur pour une comparaison donnée est déterminée par α . L'erreur au niveau de l'expérience, quant à elle, correspond à déclarer au moins un faux positif parmi toutes les comparaisons effectuées. Plus on a de groupes (et de comparaisons), plus notre erreur au niveau de l'expérience augmente. Le prochain exemple illustre les différences entre les deux niveaux d'erreur.

Exemple 6.7 Pour faciliter la compréhension de l'erreur liée à la comparaison entre les groupes et l'erreur liée à l'expérimentation, nous vous proposons l'exercice de simulation suivant. Nous simulons des données où l'hypothèse nulle est vraie, c'est-à-dire qu'il n'y a aucune différence entre les moyennes des groupes. On crée trois groupes de $n = 50$ tirés d'une population suivant une distribution normale avec $\mu = 5$ et $\sigma = 3$ (c.-à-d., $N(5, 3)$). En d'autres mots, nous simulons les données de trois groupes et ces données sont toutes tirées de la même popu-

lation pour satisfaire H_0 . Il y a trois comparaisons possibles : groupe 1 *vs* groupe 3, groupe 2 *vs* groupe 3 et groupe 1 *vs* groupe 2. Nous voulons savoir quelle est la probabilité de commettre une erreur liée à la comparaison et une erreur liée à l'expérience. On peut décrire l'algorithme de la simulation comme suit :

1. on génère des données aléatoires pour chacun des groupes conformément à H_0 (`rnorm(n = 50, mean = 5, sd = 3)`);
2. on réalise les trois comparaisons (1 *vs* 3, 2 *vs* 3, 1 *vs* 2) à l'aide de trois tests t (`t.test()`);
3. on détermine pour chaque test t effectué, si $P \leq \alpha$ (p. ex., $P \leq 0.05$);
4. on répète les étapes 1 à 3 un grand nombre de fois (p. ex., 1000 fois).

On peut réaliser ce genre d'exercice avec des boucles dans R. Le document *Les boucles avec R* couvre les concepts de base et donne des exemples pour construire des boucles qui permettront de réaliser des tâches répétitives.

À la fin de la simulation, nous pouvons constater le nombre de fois où nous avons rejeté H_0 par erreur. Ici, nous savons que H_0 est vraie puisque nous avons simulé des données qui respectent l'hypothèse nulle (aucune différence entre les moyennes des groupes). À chaque fois que nous avons rejeté H_0 dans cette simulation, nous avons commis une erreur.

Le tableau 3 montre le résultat des 10 premières itérations de la simulation (les étapes 1 à 3 ci-dessus). Les trois premières colonnes correspondent aux trois comparaisons possibles. La valeur NS correspond à un résultat non significatif (H_0 non rejetée). Le symbole * indique le rejet de H_0 et, par conséquent, une erreur liée à la comparaison. La quatrième colonne correspond à l'erreur liée à l'expérience. Cette colonne prend la valeur de 1 lorsqu'au moins une comparaison a rejeté H_0 (une erreur liée à l'expérience) et la valeur de 0 lorsqu'aucune comparaison ne rejette H_0 . La première rangée du tableau 3 correspond à la première

Tableau 3 – Premières itérations de la simulation illustrant les erreurs liée à la comparaison et à l'expérience.

1 vs 2	1 vs 3	2 vs 3	Erreur liée à l'expérience
NS	NS	NS	0
NS	NS	NS	0
NS	NS	NS	0
NS	NS	NS	0
NS	NS	NS	0
NS	NS	NS	0
NS	NS	NS	0
NS	*	*	1
NS	NS	NS	0
NS	*	NS	1

itération, et on constate qu'aucune erreur de comparaison n'a été commise, et, par conséquent, aucune erreur liée à l'expérience. La huitième rangée du même tableau montre que les comparaisons 1 vs 3 et 2 vs 3 ont rejeté par erreur H_0 (erreurs liées à la comparaison) et qu'il y a donc une erreur au niveau de l'expérience. On remarquera qu'à l'intérieur de chaque colonne, la proportion du nombre d'erreurs de comparaison par rapport au nombre total d'itérations oscille autour de $\alpha = 0.05$, puisque c'est le seuil que nous avons utilisé dans les comparaisons multiples. Toutefois, en compilant le nombre total d'erreurs au niveau de l'expérience (la quatrième colonne), on obtient la probabilité de commettre une erreur au niveau de l'expérience. Ici, nous avons :

$$\frac{124 \text{ erreurs liées à l'expérience}}{1000 \text{ itérations}} = 0.124$$

Bien que nous ayons fixé le seuil α à 0.05, la probabilité de commettre une erreur au niveau de l'expérience est supérieure à ce seuil (0.124). Ce problème a motivé le développement de plusieurs types de comparaisons multiples afin de contrôler l'erreur au niveau de l'expérience.

4.2 Suppositions

Les tests de comparaisons multiples ont les mêmes suppositions de normalité et d'homoscédasticité que l'ANOVA. Les comparaisons multiples sont moins robustes face aux déviations de ces conditions, surtout à celles d'homoscédasticité, qui augmentent les erreurs de type I et II.

4.3 Structure générale

Un grand nombre de tests de comparaisons multiples ont été développés pour différentes applications. En général, les tests de comparaisons multiples suivent le même principe :

1. on effectue l'ANOVA ;
2. si on rejette H_0 , on peut faire des comparaisons multiples. Si on ne rejette pas H_0 , l'analyse est terminée. Ce choix de ne pas poursuivre avec des comparaisons multiples s'explique du fait que l'ANOVA est plus puissante que les comparaisons multiples ;
3. on classe les moyennes des groupes par ordre croissant ;
4. on calcule la différence entre la plus grande moyenne *vs* la plus petite ;
5. comme pour un test t , on divise les différences par une erreur-type pour une statistique q . Le calcul de q implique le carré moyen des erreurs, MSE , mais dépend du test ;
6. on compare le q_{obs} à un $q_{théorique}$, qui dépend du seuil α , des df du MSE et du nombre de groupes comparés. Dans la plupart des cas, α représente l'erreur au niveau de l'expérience ;
7. un $q_{obs} \geq q_{théorique}$ indique qu'il y a une différence entre la paire de moyennes comparées.

Habituellement, on commence en comparant la plus grande moyenne aux autres (*vs* la plus petite moyenne, *vs* la deuxième plus petite, ...). Si, par exemple, on ne rejette plus H_0 à partir de la comparaison de la plus grande moyenne *vs* la deuxième plus petite moyenne, les plus petites moyennes subséquentes n'ont pas besoin d'être comparées à la plus grande moyenne parce que l'on sait déjà qu'on ne rejettera plus H_0 . On passe ensuite à la comparaison de la

deuxième plus grande moyenne par rapport à la plus petite moyenne, suivie de la deuxième plus petite moyenne, de la troisième petite moyenne . . . Encore une fois, nous pourrions arrêter la comparaison entre la deuxième plus grande moyenne et les plus petites moyennes dès que l'on ne rejettera plus H_0 pour l'une des petites moyennes. On procède ainsi pour toutes les moyennes pour, enfin, terminer en comparant la deuxième plus petite moyenne à la plus petite moyenne. Par exemple, considérons les moyennes de cinq groupes organisées par ordre croissant :

<i>A</i>	<i>C</i>	<i>B</i>	<i>E</i>	<i>D</i>
3.1	9.3	10.5	10.9	11.8

Si on compare la plus grande moyenne (groupe D) au groupe C et que l'on ne rejette pas H_0 , on ne testera ni le groupe D vs le groupe B, ni le groupe D vs le groupe E. De plus, on ne testera pas le groupe E vs le groupe C, car les groupes D et C ne diffèrent pas l'un de l'autre (la valeur de E est comprise entre celles de C et D).

4.4 Test de Tukey

Parmi les tests de comparaisons multiples, mentionnons le test de Dunnett (lorsqu'on veut comparer tous les groupes à un témoin), le test de Student-Newman-Keuls (SNK, Newman-Keuls) qui dépend du nombre de moyennes séparant les moyennes comparées, le test de Scheffé, et le test de Tukey (*Tukey test*, *Tukey's Honestly Significant Difference (HSD) test*). Le test de Tukey est d'ailleurs l'un des tests de comparaisons multiples les plus recommandés. La statistique q s'obtient comme suit :

$$q = \frac{\bar{x}_{\text{groupe 1}} - \bar{x}_{\text{groupe 2}}}{SE}$$

où $\bar{x}_{\text{groupe 1}}$ correspond à la moyenne du groupe, $\bar{x}_{\text{groupe 2}}$ à la moyenne du groupe 2 et SE est l'erreur-type du test de Tukey. Cette dernière est donnée par $SE = \sqrt{\frac{MSE}{n}}$, où MSE est le carré moyen des erreurs et n est le nombre d'observations dans chaque groupe. Nous

appliquons le test de Tukey dans le prochain exemple.

Exemple 6.8 Dans l'exemple 6.6, nous avons rejeté l'hypothèse nulle à l'aide de l'ANOVA réalisée sur les données de biomasse de plantes après l'élagage de compétiteurs. On peut appliquer le test de Tukey pour trouver où se trouvent les différences entre les moyennes des groupes. On peut commencer par calculer les moyennes des groupes et les ordonner :

```
> ##moyennes des groupes
> moy <- tapply(X = comp$Biomasse, INDEX = comp$Elagage, FUN = mean)
> ##en ordre croissant
> sort(moy)
```

controle	n25	n50	r5	r10
465.17	553.33	569.33	610.50	610.67

Nous pouvons calculer l'erreur-type du dénominateur nécessaire au calcul de la

statistique, $SE = \sqrt{\frac{MSE}{n}}$:

```
> ##sous-jeu de données
> controle <- comp[comp$Elagage == "controle", ]
> n25 <- comp[comp$Elagage == "n25", ]
> n50 <- comp[comp$Elagage == "n50", ]
> r5 <- comp[comp$Elagage == "r5", ]
> r10 <- comp[comp$Elagage == "r10", ]
> ##SSE de controle
> SSE.controle <- sum((controle$Biomasse-mean(controle$Biomasse))^2)
> ##SSE de n25
> SSE.n25 <- sum((n25$Biomasse-mean(n25$Biomasse))^2)
```



```

> ##SSE de n50
> SSE.n50 <- sum((n50$Biomasse-mean(n50$Biomasse))^2)
> ##SSE de r5
> SSE.r5 <- sum((r5$Biomasse-mean(r5$Biomasse))^2)
> ##SSE de r10
> SSE.r10 <- sum((r10$Biomasse-mean(r10$Biomasse))^2)
> ##SSE
> SSE <- SSE.controle + SSE.n25 + SSE.n50 + SSE.r5 + SSE.r10
> SSE

[1] 124020

> ##degrés de liberté de SSE
> df.erreur <- 5 * (6 - 1)
> df.erreur

[1] 25

> ##carré moyen des erreurs
> MSE <- SSE/df.erreur
> MSE

[1] 4960.8

> ##SE pour le Tukey
> SE <- sqrt(MSE/6)
> SE

[1] 28.754

```

Par la suite, on calcule le q pour chaque comparaison multiple (tableau 4). Chaque valeur de q est comparée à ce qu'on devrait obtenir si H_0 est vraie et la fonction `ptukey()` nous fournit la probabilité cumulative en faveur de H_0 pour `nmeans`

groupes et df degrés de liberté du terme d'erreur. Comme d'habitude, on rejette H_0 lorsqu'elle est peu probable (i.e, $P \leq \alpha$).

Tableau 4 – Comparaisons multiples de Tukey entre les groupe définis par les types d'élagage.

Comparaison	Différence	<i>SE</i>	<i>q_{obs}</i>	<i>P</i>	Conclusion
r10 vs controle	145.5	28.75	5.06	0.012	rejeter H_0
r10 vs n25	57.33	28.75	1.994	0.627	ne pas rejeter H_0
r10 vs n50	41.33	28.75	on ne teste pas		
r10 vs r5	0.17	28.75	on ne teste pas		
r5 vs controle	145.33	28.75	5.054	0.012	rejeter H_0
r5 vs n25	57.17	28.75	on ne teste pas		
r5 vs n50	41.17	28.75	on ne teste pas		
n50 vs controle	104.17	28.75	3.623	0.109	ne pas rejeter H_0
n50 vs n25	16	28.75	on ne teste pas		
n25 vs controle	88.17	28.75	on ne teste pas		

On peut créer ce tableau rapidement à l'aide de la fonction `TukeyHSD()` :

```
> ##test de Tukey
> TukeyHSD(m1, which = "Elagage")

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = Biomasse ~ Elagage, data = comp)

$Elagage
              diff      lwr      upr    p adj
n25-controle  88.16667 -31.260  207.59 0.22432
n50-controle 104.16667 -15.260  223.59 0.10882
r10-controle 145.50000   26.073  264.93 0.01153
r5-controle  145.33333   25.907  264.76 0.01164
```

n50-n25	16.00000	-103.427	135.43	0.99460
r10-n25	57.33333	-62.093	176.76	0.62724
r5-n25	57.16667	-62.260	176.59	0.62975
r10-n50	41.33333	-78.093	160.76	0.84539
r5-n50	41.16667	-78.260	160.59	0.84727
r5-r10	-0.16667	-119.593	119.26	1.00000

D'après les résultats, on observe que les groupes n25, n50 et control ne diffèrent pas entre eux. Les traitements r5 et r10 ne diffèrent pas entre eux, mais les deux diffèrent du témoin (control). Toutefois, les traitement r5 et r10 ne diffèrent pas des groupes n25 et n50.

4.5 Présentation des résultats

Il est possible de présenter les résultats des comparaisons multiples de différentes manières. L'une d'elles consiste à ordonner les groupes selon la valeur des moyennes et d'unir à l'aide d'un trait les groupes qui ne diffèrent pas entre eux. On peut aussi présenter les résultats graphiquement, typiquement en présentant les moyennes \pm des barres d'erreur. Comme barre d'erreur, on peut utiliser l'erreur-type des moyennes de chaque groupe, la racine-carrée du carré moyen des erreurs (\sqrt{MSE}) ou des intervalles de confiance construits à partir de l'une ou l'autres des mesures de dispersion mentionnées. On emploie des lettres pour distinguer les groupes qui diffèrent entre eux sur le graphique. Le prochain exemple illustre ces deux méthodes avec les données de biomasse selon les différents traitements d'égale. À noter que les comparaisons multiples indiquent parfois qu'un traitement appartient à deux groupes. Cette double appartenance indique que la comparaison multiple n'a pas permis de détecter une différence entre ces deux groupes.

Les tests de comparaisons multiples ne peuvent parfois trouver de groupes qui diffèrent entre eux, alors que l'ANOVA avait rejeté H_0 . L'ANOVA est plus puissante que les tests de comparaisons multiples, ce qui explique, à l'occasion, l'absence d'une différence entre les paires de groupes. Puisque l'ANOVA est plus puissante que les comparaisons multiples, nous recommandons de ne pas procéder à des comparaisons multiples si l'ANOVA n'a pas réussi à rejeter H_0 .

Exemple 6.9 On peut représenter succinctement le résultat des comparaisons multiples sur les données de biomasse selon le degré de compétition :

controle	n25	n50	r5	r10
465.17	553.33	569.33	610.50	610.67

La présentation des résultats à l'aide de traits facilite l'interprétation. On voit rapidement que n25 et n50 ont une double appartenance : ils appartiennent au groupe constitué de controle, n25 et n50 ainsi qu'au groupe composé de n25, n50, r5 et r10. Les groupes r5 et r10 ne diffèrent pas entre eux, mais sont différents du groupe controle. En effet, le même trait relie les groupes r5 et r10 et ce même trait n'inclut pas le groupe controle. Le graphique montre la même information (présence de deux groupes), mais à l'aide de lettres (fig. 7). On remarque que les groupes n25 et n50 ont une double appartenance puisqu'ils sont identifiés avec deux lettres. On trouve le code complet pour obtenir le graphique ci-dessous :

```
> ##ordonner moyennes
> moys <- sort(moy)
> ##calculer racine carrée de MSE
```

```

> sqrt.MSE <- sqrt(MSE)

> ##calculer les limites des barres d'erreur

> lim.sup <- moys + sqrt.MSE

> lim.inf <- moys - sqrt.MSE

> ##créer graphique vide sans axe des x's

> plot(x = 0, y = 0, type = "n",
       ylim = c(min(lim.inf), max(lim.sup+10)),
       xlim = c(0, 6), xlab = "Traitement d'élagage",
       ylab = "Biomasse de plante",
       main = "Moyennes  $\pm$  racine-carrée de MSE",
       xaxt = "n", cex.lab = 1.2)

> ##ajouter axe des x's

> axis(side = 1, at = c(1, 2, 3, 4, 5),
       labels = names(moys))

> ##ajouter moyennes

> points(x = c(1, 2, 3, 4, 5),
        y = moys)

> ##ajouter barres d'erreurs

> arrows(x0 = c(1, 2, 3, 4, 5),
        y0 = lim.inf,
        x1 = c(1, 2, 3, 4, 5),
        y1 = lim.sup, length = 0.05,
        angle = 90, code = 3)

> ##ajouter les lettres, lim.sup + 10

> text(x = 1, y = 545, labels = "a")

> text(x = 2, y = 633, labels = "ab")

> text(x = 3, y = 649, labels = "ab")

```

```
> text(x = 4, y = 690, labels = "b")
> text(x = 5, y = 691, labels = "b")
```

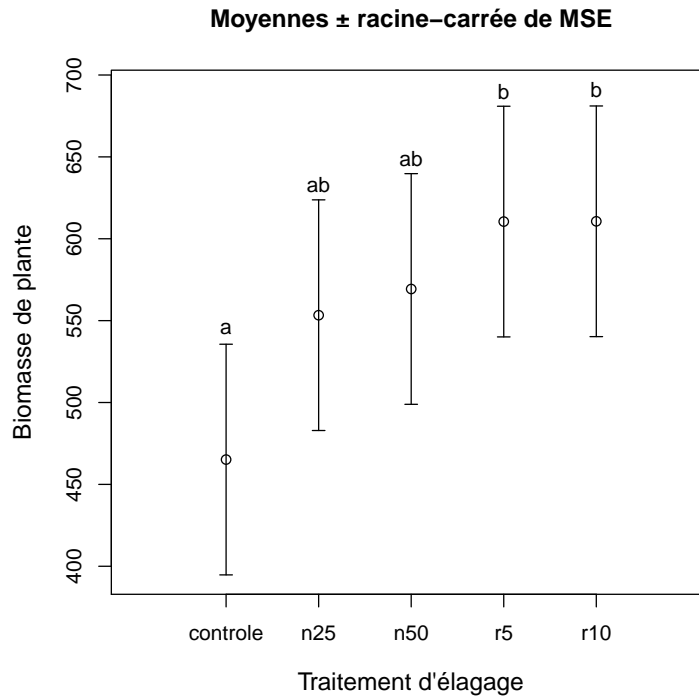


FIGURE 7 – Présentation graphique des résultats des comparaisons multiples avec le test de Tukey.

Conclusion

Nous venons d'étudier le concept de comparaisons multiples ainsi que les problèmes qui y sont associés, notamment une augmentation de la probabilité de commettre une erreur de type I (déclarer un faux positif). Au lieu d'utiliser une série de tests t , nous avons vu que l'analyse de variance (ANOVA) permet de comparer tous les groupes simultanément en estimant la variance expliquée par le facteur (variable catégorique) d'intérêt et en la comparant à la

variance résiduelle (inexpliquée). Le tableau d'ANOVA véhicule beaucoup d'information et il est important d'inclure ses détails dans la rédaction des résultats. Bien que l'ANOVA permette de déterminer si le facteur a un effet important sur la variable réponse, elle ne permet pas de déterminer où se trouvent les différences. Pour ce faire, nous avons recours à des tests de comparaisons multiples exécutés *a posteriori* qui contrôlent l'erreur au niveau de l'expérience. Pour terminer, nous avons fait quelques propositions concernant la présentation des résultats afin de faciliter l'interprétation d'une ANOVA.

Index

ajustement de Bonferroni, [3](#)

analyse de variance, [4](#)

comparaisons multiples, [25](#)

 erreur au niveau de l'expérience, [26](#)

 erreur au niveau de la comparaison, [26](#)

distribution du F , [13](#)

facteur, [4](#)

hypothèse statistique, [13](#)

présentation des résultats, [34](#)

ratio F , [12](#)

résidus, [16](#)

somme des carrés, [5](#)

somme des carrés des erreurs, [5](#)

somme des carrés du traitement, [5](#)

somme des carrés totale, [5](#)

suppositions, [16](#)

tableau d'ANOVA, [12](#)

test *a posteriori*, [26](#)

test *post hoc*, [26](#)

test t , [2](#)

variable catégorique, [4](#)

variable réponse, [4](#)

variance résiduelle, [12](#)