PrivateTeacher

Maîtriser les Sciences Exactes

STATISTIQUES

Cours Ciblé: Statistique en Psychologie
L'analyse de variance ANOVA
Statistique de Fisher, Facteurs et p-valeur

Julien RUPPEN

02 April, 2026

Abstract

Le test t de Student mesure la différence entre deux groupes compte tenu de la variance des observations. La différence entre deux moyennes en effet peut être non-nulle et ne pas être significative pour autant à cause d'un recouvrement important entre les observations. La variance joue donc un rôle central dans la quantification de cette différence. L'ANOVA généralise ce raisonnement à un nombre quelconque de groupes. Dès que l'on compare trois conditions ou plus, la statistique t n'est plus adaptée : on lui substitue alors la statistique F de Fisher, qui est un rapport entre la variabilité inter-groupes à la variabilité intra-groupe. Ce document présente cette mesure et montre quelle sont les question auxquelle elle permet de répondre en psychologie clinique et expérimentale.

Contents

1 Introduction **3**
1.1 Contexte et motivation . . . 3
1.2 La logique de la statistique F . . . 4

2 Trois formes de soutien psychologique **5**
2.1 Un exemple concret . . . 5
2.2 Question de recherche . . . 5
2.3 Méthode de résolution . . . 6
2.4 Commande R . . . 6
2.5 Interprétation des Résultats . . . 6
2.6 Vérification . . . 6

3 Un même patient mesuré plusieurs fois **7**
3.1 Un exemple concret . . . 7
3.2 Question de recherche . . . 8
3.3 Méthode de résolution . . . 8
3.4 Commande R . . . 9
3.5 Interprétation des Résultats . . . 9
3.6 Vérification . . . 10

4 Au sujet des facteurs **10**
4.1 Qu'est-ce qu'un facteur . . . 10
4.2 Les niveaux de la variance . . . 10
4.3 ANOVA à deux facteurs . . . 10
4.4 En résumé . . . 11
4.5 Mais pourquoi toute cette variance ? . . . 11

5 L'effet dépend-il du genre ? **12**
5.1 Un exemple concret . . . 12
5.2 Question de recherche . . . 13
5.3 Méthode de résolution . . . 13
5.4 Commande R . . . 13
5.5 Interprétation des Résultats . . . 13
5.6 Vérification . . . 14

6 Conclusion **14**
6.1 Ce que nous avons appris . . . 14
6.2 Travailler avec méthode . . . 15
6.3 Prendre des points de repère . . . 16

1. Introduction

1.1. Contexte et motivation

En psychologie clinique, on se retrouve souvent en situation de comparer des individus soumis à des conditions expérimentales ou thérapeutiques différentes. Tant qu'il n'y a que deux conditions qui distingue les individu, formant ainsi deux groupes, il est facile de les distinguer à l'aide du test t de Student. Dès que le nombre de conditions expérimentales augmentes, et avec lui le nombre de groupes, le test de Student devient insuffisant.

1.1.1 Trop de tests tue les tests

Pour comparer plusieurs groupes entre eux, on peut toujours, bien entendu effectuer des comparaisons deux à deux à l'aide de l'instrument test de Student. Avec trois groupes par exemple, cela représente trois tests t distincts. Or chaque test est conduit au seuil α = 0.05, ce qui implique une probabilité de 5 % de commettre une erreur de type I par test. En effectuant trois tests, la probabilité de commettre au moins une erreur augmente, ce qui constitue une sérieuse limitation de l'utilisation des test t lorsque le nombre de groupe est important. Les conclusions sont alors peu fiables car on multiplie les tests, augmentant de fait le nombre de faux positifs.

1.1.2 Une Solution élégante

L'**analyse de variance** (ANOVA) résout ce problème en effectuant une seule comparaison globale, quel que soit le nombre de groupes, tout en maintenant le risque alpha à son niveau nominal.

1.1.3 Représentation Graphique

Le graphique ci-dessous illustre la situation suivante: trois groupes ayant été soumis à trois conditions expérimentales différentes sont comparée sur la base d'une valeur commune: Le score PHQ. Chaque groupe est représenté par une distribution, et chaque distribution possède sa propre moyenne. La question est de savoir si ces différences observées est réelle c'est à dire, différentes de ce que le hasard pourrait produire sans qu'aucun traitement ne soit intervenu.

1.1.4 Question de recherche

Un chercheur qui dispose de plusieurs groupes mesurés sur une même variable quantitative se pose typiquement la question suivante : le facteur qui distingue les groupes a-t-il un effet statistiquement significatif ? mesurée ?

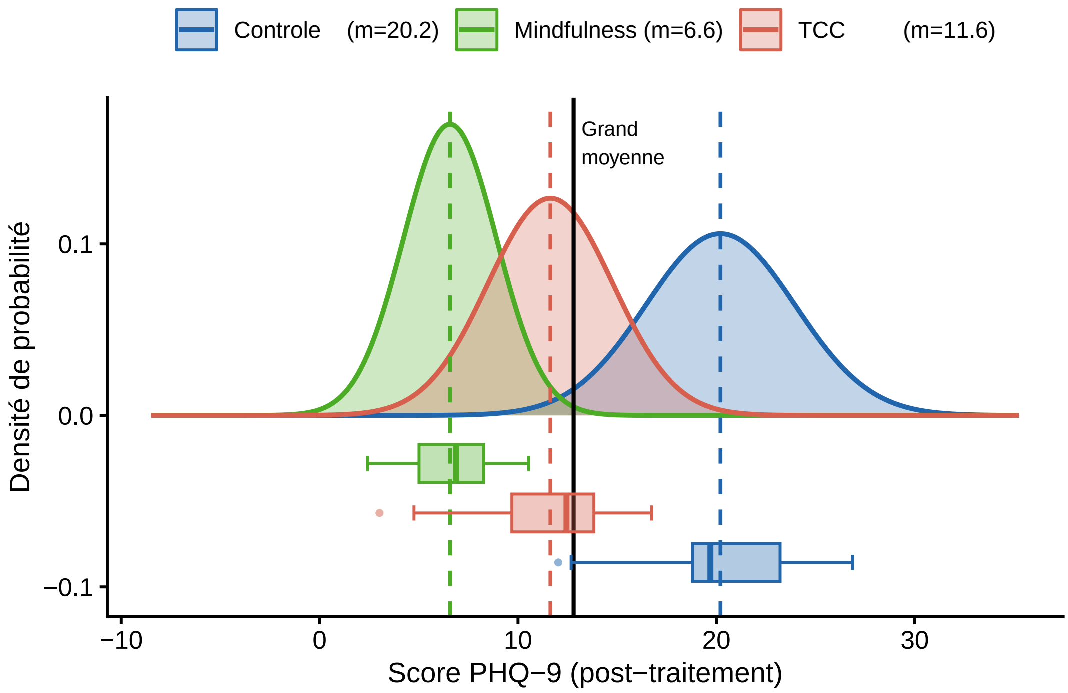

Figure 1: Scores PHQ-9 post-traitement dans les trois groupes. Les pointillés indiquent les moyennes de chaque groupe ; la ligne noire solide indique la grande moyenne.

1.2. La logique de la statistique F

L'ANOVA repose sur le raisonnement suivant: Il existe deux type de variance entre les groupes: La première est la variance inter-groupes. Cette variance est une mesure de la distance de chaque groupe par rapport à la moyennes des moyennes, appelée aussi **Grand moyenne** La deuxième est la variance intra-groupe. Il s'agit de la variance moyenne de l'ensemble des groupes. Chaque groupe vient avec sa propre variance. La moyenne de ces variance est la variance intra-groupe. Ces variances sont utile car elle permettent de construire une mesure de l'espacement entre les groupe. Pour mesure cet espacement, on construit la statistique F de la manière suivante:

$$F = \frac{\text{variance inter-groupes}}{\text{variance intra-groupe}}$$

Autrement dit, la statistique F est le rapport entre la variance inter-groupe et la variance intra-groupe. Cette mesure est utile car elle permet de distinguer entre deux cas de figure suivant: 1) Tous les groupes sont superposé les uns sur les autre, autrement dit la variance inter-groupe est faible par rapport à la variance moyenne des groupe: variance inter-groupe $<$ variance intra-groupe. Dans cette situation, la valeur de F est donc petite, proche de zéro 2) Les groupes sont espacé par rapport à leur variance propre. Autrement dit on a: variance inter-groupe $>$ variance intra-groupe. F prendra donc une valeur importante, différente et plus grande que zéro. En d'autre terme nous avons la une mesure de l'espacement entre les groupes. F=0 signifie que les groupe sont superposé et $F \neq 0$ signifie que les groupe sont espacé. C'est cela l'utilité de cette mesure. Ces valeurs extrême de F serviront de point de repère à partir de maintenant.

2. Trois formes de soutien psychologique

2.1. Un exemple concret

Certains étudiants universitaires souffrent d'anxiété élevée avant les examens. Afin d'évaluer l'efficacité d'un traitement, plusieurs étudiants ont été recrutés et répartis aléatoirement dans trois groupes :

- **Contrôle** : liste d'attente, aucune intervention sur ce groupe d'étudiants.
- **Relaxation** : programme de relaxation progressive (4 séances sur 2 semaines).
- **TCC brève** : thérapie cognitivo-comportementale individuelle.

Deux semaines après la fin des interventions, chaque participant complète un questionnaire qui évalue STAI-Y1 (*State-Trait Anxiety Inventory*). Un score élevé indique une anxiété importante. Les effectifs sont de 260, 250 et 240 participants respectivement.

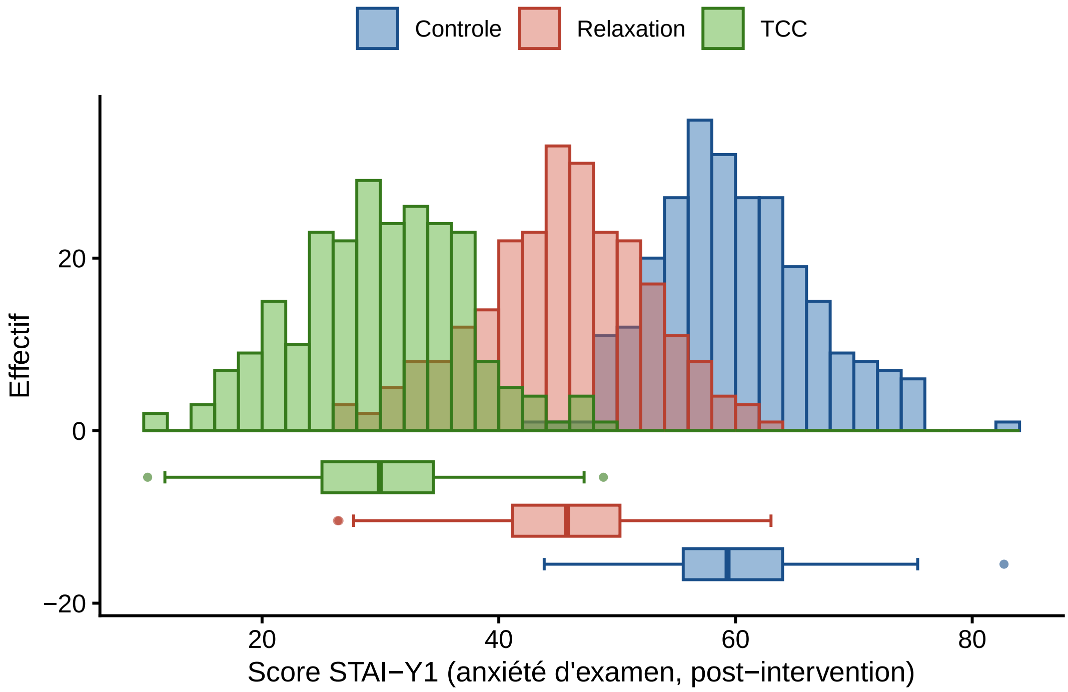

Figure 2: Distribution des scores STAI-Y1 post-intervention dans les trois groupes. Les boxplots en dessous résument la tendance centrale et la dispersion de chaque groupe.

Cette situation est typique d'un cas d'application de ANOVA : on dispose de trois groupes indépendants, d'une seule **variable dépendante quantitative et continue**, et on veut savoir si les moyennes diffèrent au-delà du hasard.

2.2. Question de recherche

Un praticien ou un chercheur confronté à ces données se pose naturellement la question suivante : le type de soutien psychologique reçu a-t-il un effet significatif sur le niveau d'anxiété des étudiants avant l'examen ?

2.3. Méthode de résolution

On applique la séquence d'étapes suivantes:

Étape 1 — Choisir une mesure adaptée Trois groupes indépendants, variable quantitative continue: ANOVA à un facteur

Étape 2 — Déterminer si la mesure est statistiquement significative. On calcule F puis on lit la p-valeur associée.

Étape 3 — Prendre une décision Si p $<$ 0.05 la valeur de F est significativement différente de 0. si p $>$ 0.05 la valeur de F est proche de 0.

2.4. Commande R

```
anova_sect2 <- aov(score ~ groupe, data = data_sect2)
summary(anova_sect2)
```

2.5. Interprétation des Résultats

```
            Df Sum Sq Mean Sq F value Pr(>F)
 groupe       2 112898   56449    1187 <2e-16 ***
Residuals   747  35510      48
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ce résultat nous dit que le facteur "Groupe" produit une valeur de F 1187.5. La p-valeur associée est p = 1.04e-232 Cette p-valeur est essentiellement 0 est très largement inférieure au seuil $\alpha = 0.05$, on rejette donc H_0. On dira donc: la valeur de F est statistiquement significative (significativement différente de 0). Les trois groupes sont donc probablement différent.

2.5.1 Réponse à la questions

Oui, les traitements considérés ont un impacte sur la valeur STAI-Y1.

2.6. Vérification

Sur le graphique, les trois histogrammes sont visuellement bien séparés : les pics se trouvent respectivement autour de 60, 45 et 30. Un F de 1187 reflète précisément cela : le numérateur (variance entre groupes) est très grand par rapport au dénominateur (variance intra-groupe).

3. Un même patient mesuré plusieurs fois

Dans la section précédente, les trois groupes étaient **indépendants** : chaque participant appartenait à un seul groupe. Il s'agit d'un design très courant, mais qui représente à désavantage important: la variabilité entre individus (le fait que certaines personnes sont naturellement plus anxieuses que d'autre) fausse les observations. On dit que l'erreur intra-groupe vient *bruiter* la détection de l'effet du traitement.

3.0.1 Solution

L'ANOVA à **mesures répétées** adopte une logique différente. Plutôt que de recruter des individus et de les placer dans des groupes distincts, on mesure les **mêmes individus** plusieurs fois. Dans ce design, **chaque participant devient son propre contrôle**. La variabilité intra-groupe est donc diminuée et la valeur de F s'en trouve augmentée. Le test gagne ainsi en **puissance statistique** : il détecte plus facilement de vraies différences avec le même nombre de sujets.

3.0.2 Explication

La variabilité intra-groupe est une mesure de la diversité des observation au sein d'un groupe. Cette variabilité à **deux causes distinctes:** La première est la variabilité normale entre les individus. Il s'agit de la diversité naturelle des individus. la seconde est une **variabilité résiduelle** Elle est due à des erreur de mesures ou encore à des fluctuations aléatoires de cette mesure (un même patient peut répondre différemment selon qu'il le remplit le matin, ou le soir par exemple) Il est important de réaliser que cette variabilité résiduelle **est incompressible** (on ne peut pas l'éliminer) A cela viens s'ajouter la variabilité inter-individuelle pour former le terme total :

$$\text{variance intra-groupe} = \text{variance inter-individuelle} + \text{variance résiduelle}$$

Le design à mesures répétées soustrait la variance inter-individuelle et réduit donc la variance intra-groupe :

$$\text{variance intra-groupe}_{\text{répété}} = \text{variance intra-groupe} - \text{variance inter-individuelle}$$

C'est ainsi que la valeur de F s'en trouve augmentée et la puissance améliorée.

3.1. Un exemple concret

Un chercheur évalue l'effet d'un programme de relaxation guidée (8 semaines) sur l'intensité de la douleur chronique chez 40 patients fibromyalgiques. Il n'y a pas de groupes distincts : ce sont les **mêmes patients** mesurés à des intervalle de temps différentes. La mesure choisie est l'Échelle Visuelle Analogique (EVA, 0–10). Un score plus élevé signifie une douleur plus intense.

- **T0** — avant le programme (baseline)
- **T1** — après 4 semaines (mi-parcours)

- **T2** — après 8 semaines (fin du programme)

Le graphique ci-dessous représente les trajectoires individuelles (lignes grises) de chaque patient entre les trois temps de mesure, avec les boxplots par temps et la ligne de moyenne en noir. La ligne de moyenne descend régulièrement de 7.7 à 6 puis à 4.2.

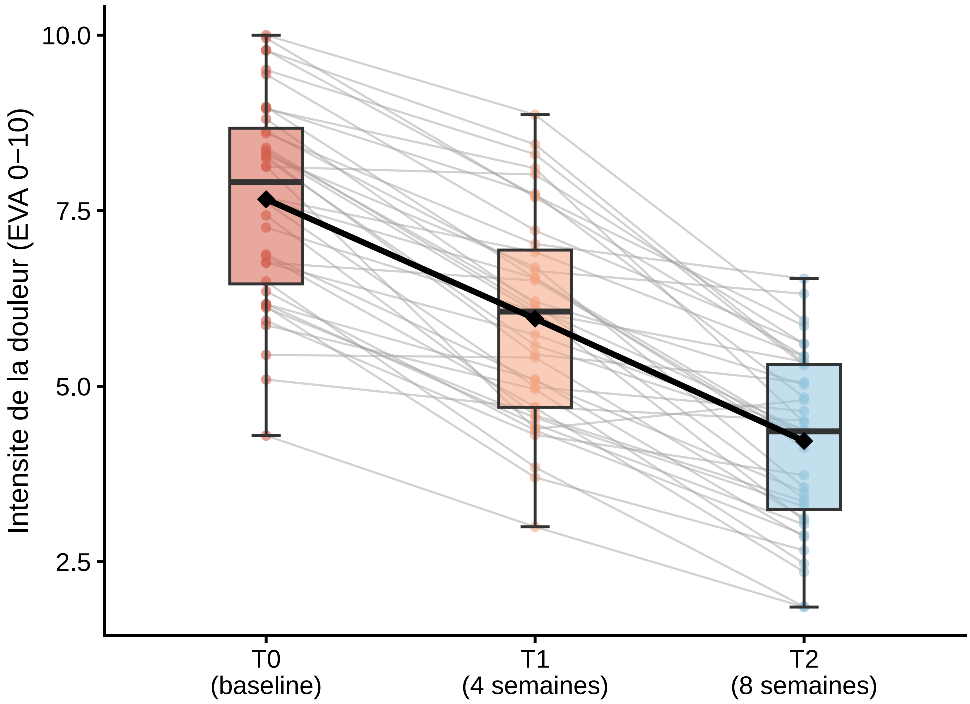

Figure 3: Trajectoires individuelles (spaghetti) de l'intensité de la douleur EVA sur les trois temps de mesure. La ligne noire relie les moyennes observées.

3.2. Question de recherche

La question que se pose le chercheur est alors naturellement : l'intensité de la douleur évolue-t-elle significativement au fil du temps ?

3.3. Méthode de résolution

On applique les mêmes trois étapes identique pour toute les ANOVA

Étape 1 — Choisir une mesure adaptée. Un facteur intra-sujet (trois niveaux), une variable quantitative continue, des mesures corrélées au sein de chaque sujet : l'ANOVA à mesures répétées

Étape 2 — Déterminer si la mesure est statistiquement significative. On calcule F et on lit la p-valeur associée.

Étape 3 — Prendre une décision. Si $p < 0.05$ on rejette H_0 (la valeur F est significativement différente de zéro) et les groupes peuvent etre considéré différent. Si la valeur $p > 0.05$ alors F est proche de zéro et les groupe sont comparable.

3.4. Commande R

```
anova_sect3 <- aov(score ~ temps + Error(sujet/temps), data = data_sect3)
summary(anova_sect3)
```

3.5. Interprétation des Résultats

```
Error: sujet
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 39  193.9   4.973

Error: sujet:temps
          Df Sum Sq Mean Sq F value Pr(>F)
 temps      2  237.2   118.6   299.4 <2e-16 ***
Residuals 78   30.9     0.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3.5.1 Réponse

Oui : l'intensité de la douleur évolue significativement au fil des trois temps

3.5.2 Syntaxe Précision

La syntaxe `Error(sujet/temps)` est spécifique à R et mérite attention. Elle indique à `aov()` comment distinguer les deux terme de la variance intra groupe.

`Error: sujet` désigne la variabilité entre sujets `Error: sujet:temps` désigne la variabilité résiduelle

C'est ce qui permet à R de savoir qu'il s'agit d'un disigne mesure-répétée.

La même analyse peut être conduite avec `anova_test()` du package `rstatix`. Les deux syntaxes en parallèle :

```
aov(score ~ temps + Error(sujet/temps), data = data_sect3)
```

```
anova_test(data = data_sect3, dv = score, wid = sujet, within = temps)
```

Les deux fonctions calculent le même F — elles implémentent le même modèle. L'avantage de `anova_test()` est qu'elle ajoute automatiquement le test de sphéricité de Mauchly et la correction de Greenhouse-Geisser si nécessaire, ce que `aov()` ne fait pas.

> **Remarque.** Les deux termes de la variance intra-groupe sont identiques dans les deux syntaxes, bien que nommés différemment : `Error(sujet/temps)`

dans `aov()` et `wid = sujet, within = temps` dans `anova_test()` désignent la même partition — variance inter-individuelle (`sujet`) et variance résiduelle (`sujet:temps`).

3.6. Vérification

La confirmation visuelle est immédiate sur le graphique des trajectoires individuelles. La ligne de moyenne descend de 7.7 à 6 à 4.2, sans jamais remonter. La différence est donc bien visible au cours du temps.

4. Au sujet des facteurs

4.1. Qu'est-ce qu'un facteur

Un facteur nous l'avons vu est un moyen de distinguer des groupes. On peut très bien le voir comme une variable catégorielle qui permet de classer des individu dans des catégories. Dans ce context, il y a donc autant de modalité de cette variable qu'il y a de groupe différent. Les facteur cependant sont d'avantage qu'un moyen de classement. Les facteur jouent un rôle différent: il représente une source de variance que chercheur cherche à identifier. Pour cette raison la terminologie est différente: on parlera de modalité pour une variable catégorielle alors qu' on parlera de niveau pour un facteur. Ce sont deux choses différentes. Un point important à noter est qu'un facteur peut avoir plusieurs niveau. Pour cette raison on peut très bien avoir fait une expérience sur 10 groupe différent, si ces 10 groupes sont comme 10 niveaux d'un seul facteur, on parlera d'une ANOVA à un facteur (One-Way ANOVA)

4.2. Les niveaux de la variance

Dans le cadre de l'ANOVA, un niveau n'est donc pas seulement une condition expérimentale, il est une source de variance que l'on cherche à expliquer. L'idée de départ est que la variance totale des observatoin n'est pas homogène: une partie de la variance provient de différences entre les niveaux (les groupes ont des moyennes différentes parce que la condition expérimentale produit un effet), et une autre partie provient de différences à l'intérieur de chaque niveau (les individus d'un même groupe ne sont pas identiques, pour des raisons qui n'ont rien à voir avec le traitement). En distingant les niveau de la variance, on distinque quelle est la part de variance qui provient du traitement et quelle est la part de variance est due à d'autre raison. Vous comprenez maintenant pourquoi on appel cela une ANALYSE DE VARIANCE: C'est une analyse qui découpe les différentes contributions de la variance totale.

4.3. ANOVA à deux facteurs

Le premier facteur nous l'avons vu, le traitement, **produit** l' effet principal: Il répond à la questions suivante: en moyenne sur l'ensemble des participants, le traitement est il efficace globalement ?

En ajoutant un deuxième facteur, c'est à dire un deuxième moyen de distinguer des groupes, le **genre** par exemple, on se donne le moyen de savoir si les hommes et les femmes sont des groupe différent au départ, indépendamment du traitement. Si l'on mesure la dépression par exemple, le deuxième facteur permettra de dire si les hommes ont un niveau de dépression moyen différent de celui des femme. C'est ce qu'on appel le deuxième effet principal.

La présence d'un deuxième facteur enfin présente une possibilité très intéressante: elle permet de quantifier **l'effet croisé** entre les deux facteurs. Cet effet croisé est très utile car permet de répondre à la question: Le traitement a-t-il un effet différent au sein des sous-groupes, entre le homme et les femme dans notre exemple.

4.4. En résumé

Un design à deux facteurs permet de tester trois choses en un seul modèle :

1. L'**effet principal du traitement** : en moyenne sur les deux genre, le traitement est il efficace ?
2. L'**effet principal du genre** : en moyenne les hommes et les femmes ont-ils des scores différents indépendamment du traitement
3. L'**interaction traitement $\times$ genre** : l'effet du traitement est-il le même pour les hommes et pour les femmes ?

On introduit ainsi une sorte de hiérarchie entre les facteurs qui nous permettra de dinstinguer encore plus finement les source de la variance au sein de nos observations. De chacun de ces trois terme, c'est le term d'interaction, qui est le plus riche d'informations. Si l'interaction est significative, les effets principaux ne peuvent pas être interprétés correctement, car ils masquent une réalité plus complexe. Pour cette raison, une règle pratique consiste à **lire l'interaction en premier**.

4.5. Mais pourquoi toute cette variance ?

Pouquoi fait on tous ces efforts pour identifier les différents niveaux de variance au juste ? Parce que la variance résiduelle est le dénominateur du test F. Plus on identifie et isole des sources de variance légitimes — traitement, genre, interaction — moins il en reste dans le résidu. Un résidu plus petite rend le test plus sensible : ***il est plus facile de détecter un effet réel quand le bruit de fond est bien séparé du signal.***

5. L'effet dépend-il du genre ?

5.1. Un exemple concret

Un chercheur recrute 120 adultes souffrant de dépression modérée à sévère, répartis équitablement dans quatre cellules selon un design **2 × 2** :

	Contrôle (liste d'attente)	TCC (8 séances)
Homme	n = 30	n = 30
Femme	n = 30	n = 30

La dépression est mesurée avec l'Inventaire de Dépression de Beck (BDI-II, 0–63). Un score plus élevé indique une dépression plus sévère.

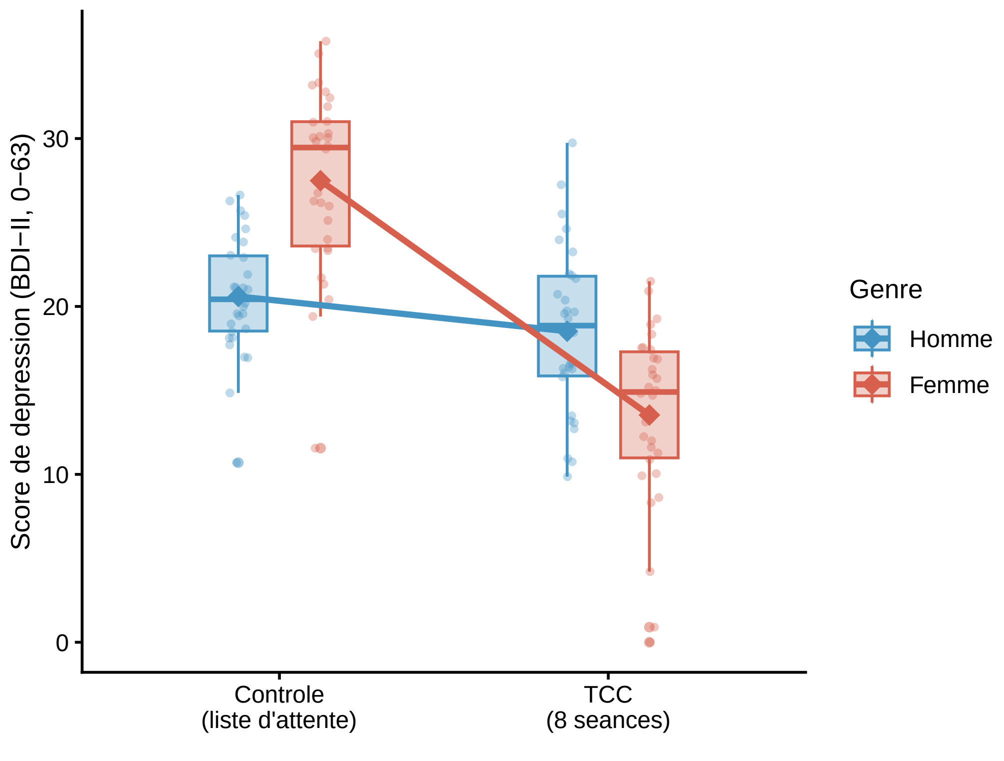

Figure 4: Scores BDI-II par condition (Controle vs TCC) et par genre. La ligne d'interaction relie les moyennes observées. Un croisement des lignes signale une interaction.

Ce graphique illustre la section précédente: chaque cellule représente le croisement d'un niveau du premier facteur (traitement) avec un niveau du second facteur (genre). Les deux lignes relient les moyennes observées pour chaque sous-groupe. Si les deux facteurs agissaient de manière indépendante, ces lignes seraient parallèles. Ici, les lignes se croisent nettement : au niveau Contrôle, les femmes présentent un score moyen de 27.5 contre 20.6 pour les hommes. Au niveau TCC, la situation s'inverse : les femmes descendent à 13.5 tandis que les hommes restent à 18.5. C'est la signature visuelle d'un effet croisé entre les deux facteurs.

5.2. Question de recherche

L'effet du traitement sur la dépression diffère-t-il selon le genre du patient ?

5.3. Méthode de résolution

Étape 1 — Choisir une mesure adaptée. Deux facteurs catégoriels croisé (between-subject) et une variable dépendante quantitative continue Ce design est typique d'une ANOVA à deux facteurs (Two-Way ANOVA)

Étape 2 — Déterminer si les effets sont statistiquement significatifs. Le modèle produit trois valeurs de F — une par source de variance. Pour chacune, la question est la même : cette valeur de F est-elle statistiquement significative ? On commence par l'interaction, car si elle est significative, les effets principaux ne peuvent pas être lus indépendamment.

Étape 3 — Prendre une décision. Si l'interaction est significative, l'effet du traitement n'est pas uniforme selon le genre. Les effets principaux décrivent alors des effet moyens qui doivent être interprétés en tenant compte de l'interaction. **Si l'interaction n'est pas significative, les effets principaux peuvent être interprétés directement.**

5.4. Commande R

```
anova_sect4 <- aov(score ~ traitement * genre, data = data_sect4)
summary(anova_sect4)
```

L'opérateur `*` développe automatiquement en `traitement + genre + traitement:genre` : les deux effets principaux plus leur terme d'interaction.

5.5. Interprétation des Résultats

```
                  Df Sum Sq Mean Sq F value   Pr(>F)
traitement         1 1927.6  1927.6  80.876 5.40e-15 ***
genre              1   27.3    27.3   1.144    0.287
traitement:genre   1 1059.1  1059.1  44.434 9.33e-10 ***
Residuals        116 2764.8    23.8
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Le tableau affiche trois lignes, une par source de variance. Pour chacune, la question est identique : la valeur de F est-elle statistiquement significative ?

Le terme d'interaction La ligne surlignée `traitement:genre` répond à notre question. On peut lire: F(1, 116) = 44.4, p = 9.33e-10. L'interaction est donc significative : il existe un effet croisé entre les deux facteurs. L'effet du traitement n'est pas le même selon le genre.

L'effet principal du traitement : F(1, 116) = 80.9, p = 5.4e-15, $\eta^2 = 0.33$. Oui, significatif. En moyenne sur les deux genres, la TCC réduit le score BDI-II (24 → 16). Cette moyenne marginale doit cependant être lue avec prudence : l'interaction significative indique qu'elle agrège des profils très différents selon le genre.

L'effet principal du genre : F(1, 116) = 1.14, p = 0.287. Non, pas significatif. En moyenne sur les deux traitements, hommes et femmes ne diffèrent pas sur le score BDI-II.

5.5.1 Réponse

Oui : l'effet du traitement sur la dépression diffère significativement selon le genre (La TCC réduit la dépression de 14 points chez les femmes contre 2.1 points chez les hommes.)

5.6. Vérification

Sur le graphique, le croisement des deux lignes confirme visuellement l'effet croisé identifié par le test : des lignes parallèles auraient signalé des effets additifs et indépendants.

La valeur F du term d'interaction 44 traduit cela précisément. Le traitement ramène les femmes de 27.5 à 13.5 (écart de 14 points), alors qu'elle ne déplace les hommes que de 20.6 à 18.5 (écart de 2.1 points). La significativité statistique de F 44 nous dit que cette asymétrie est trop grande pour être due au hasard d'échantillonnage.

6. Conclusion

6.1. Ce que nous avons appris

La statistique F de Fisher traduit une idée simple: Si les groupes se superposent alors la valeur de F vaut zéro (C'est l'hypothèse de départ, l'hypothèse H_0) A mesure que les groupe se distingue, la valeur de F devient toujours plus différente de zéro. La p-valeur mesure si cette différence de F par rapport à zéro est statistiquement significative. Lorsque F est statistiquement significatif, on rejette alors l'hypothèse de départ H_0 et on dit: Les groupes sont probablement différent !

Ce document a présenté l'analyse de variance à travers quatre situations qui dépendent toute d'une meme mesure, la statistique de Fisher F qu'il est utile de récapituler maintenant.

Voici les trois cas de figure que nous avons rencontré:

ANOVA à groupes indépendants — Le cas le plus simple de l'utilisation de la statistique de test F: une seule valeur de F, une seule source de variance et un résidu simple. Ce design permet de répondre à la question: Ces groupes sont-ils différent?

ANOVA à mesures répétées — Le principe est le même mais cette fois les groupe sont des moments au cours du temps. Dans ce design, un même individu est don mesuré plusieurs fois. L'avantage de ce design est qu'il permet de retirer la variance inter-individuelle de la variance intra-groupe (le dénominateur de F) ce qui augmente la sensibilité du test. L'interprétation repose toujours sur la valeur de F. Elle permet de répondre fois à la question: Le traitement a-t-il un effet au cours du temps ?

ANOVA à deux facteurs — Ce design repose toujours sur l'interprétation de la valeur F. Ce design cependant, offre la possibilité de distinguer plusieurs source de variance à l'aide de plusieurs facteurs. Chaque facteur permet d'identifier une variance particulière et leur croisement permet d'identifier une troisième variance issue de leur interaction. facteur A, facteur B et croisement A×B. Pour cette raison, le modèle produit trois valeurs de F: une pour chaque source de variance. On raffine ainsi la décomposition et on peut ainsi poser trois questions simultanément: **chaque facteur a-t-il un effet, et cet effet est-il le même dans tous les sous-groupes ?**

6.2. Travailler avec méthode

Comment choisir une méthode et comment interpréter les résultats

6.2.1 Choisir la méthode

1. Identifier la variable commune. Avant de choisir un modèle, identifier ce qui est mesuré sur chaque individu. Cette variable dépendante — un score, une durée, une intensité — est ce qui sera comparé entre les groupes. Elle doit être quantitative et continue.

> En psychologie, des échelles ordinales à plusieurs niveaux (Likert 5 ou 7 points) sont fréquemment traitées comme continues et soumises à une ANOVA. C'est une approximation courante et généralement acceptée dans la littérature, mais qui reste une simplification.

2. Identifier les groupes. Combien y a t il de groupes distincts ? Ces groupes sont-ils formés d'individus différents, ou s'agit-il des mêmes individus mesurés à plusieurs fois ? Y a-t-il une seule façon de grouper les individus ou y en a t il plusieurs ?

3. Choisir la méthode.

- Groupes indépendants, un facteur → ANOVA à groupes indépendants
- Mêmes individus mesurés plusieurs fois → ANOVA à mesures répétées
- Deux facteurs → ANOVA à deux facteurs

6.2.2 Inerpréter les résultat

1. Garder en tête ce que F mesure. F est toujours un rapport entre la variance entre les groupes et la variance à l'intérieur des groupes. Il mesure si les groupes sont distincts — quelle que soit la nature du design.

2. Identifier la nature du facteur. Un facteur traitement pose la question : l'intervention est-elle efficace ? Un facteur temps : y a-t-il une évolution ? Un croisement de deux facteurs : l'effet est-il uniforme dans tous les sous-groupes ? La nature du facteur précise la question à laquelle F répond.

3. F est-il statistiquement significatif ? L'interprétation de la p-valeur est toujours la meme quel que soit le design. Si p $< \alpha$, les groupes sont probablement différents. Si p $> \alpha$ — les groupe sont probablement identique.

4. Prendre une décision. Il sagit a ce stade de formuler une réponse dans le cadre de la question de recherche.

6.3. Prendre des points de repère

La valeur de F un premier point de repère. F peut prendre une valeur entre zéro et l'infini : lorsque les groupes sont identiques, F=0. Plus les groupes se distingue, plus F est différent de zéro et plus sa p-valeur est petite. F tout seul ne permet pas de donner une interprétation, c'est vraiment sa p-valeur qui permet de dire s'il est prêt de zéro ou différent de zéro.

Le graphique est le deuxième point de repère. Avant même de lire la valeur de F ou sa p-valeur, la séparation visuelle des distributions, l'alignement des boxplot et de leur quantile ou la superposition des distributions, permet déjà de se donner une intuition de la situation. Les résultats numérique doivent venir confirmer cette intuition.