

## Chapitre 14

# STATISTICA

## Analyse de variance

### Présentation générale

### *Analyses a priori et a posteriori*

## Analyse de variance

### A quoi cela sert?

Objectif de l'Analyse de Variance. D'une manière générale, l'objectif d'une analyse de variance (ANOVA) vise à tester les différences significatives entre les moyennes. Si nous ne comparons que deux moyennes, l'ANOVA nous donnera les mêmes résultats qu'un test t pour des échantillons indépendants (pour comparer deux groupes différents d'observations), ou qu'un test t pour des échantillons appariés (pour comparer deux variables relatives au même ensemble d'observations).

Pourquoi le nom analyse de variance ? Il peut sembler étrange qu'une procédure destinée à comparer des moyennes soit appelée analyse de variance. Ce nom provient du fait que pour tester la significativité statistique entre des moyennes, nous devons en fait comparer (c'est-à-dire, analyser) les variances.

LIRE: Principes Fondamentaux - ANOVA à Plusieurs Facteurs

### Effet d'interaction

L'ANOVA présente un autre avantage par rapport aux simples tests t : elle nous permet en effet de détecter les interactions entre les variables, et donc, de tester des hypothèses plus complexes.

LIRE: Principes Fondamentaux - Interactions

## Analyse de contrastes

### A quoi cela sert?

Pourquoi Comparer des Moyennes Individuelles ? Souvent, nous formulons nos hypothèses expérimentales de façon plus spécifique qu'en termes simples d'effets principaux ou interactions. Nous pouvons formuler l'hypothèse spécifique qu'un ouvrage pédagogique particulier va améliorer le niveau en mathématiques des hommes, mais pas celui des femmes, tandis qu'un autre manuel sera efficace quel que soit le sexe, mais globalement moins efficace pour les hommes. Nous pouvons donc conjecturer l'existence d'une interaction : l'efficacité du manuel dépend du (est modifiée par le) sexe de l'individu. Nous avons une certaine hypothèse sur la nature de l'interaction : nous attendons une différence significative entre les sexes pour un ouvrage mais pas pour l'autre. Ce type de prévision spécifique est souvent testé par une analyse de contraste.

Analyse de Contrastes. En bref, l'analyse de contrastes nous permet de tester la significativité statistique des différences spécifiques prévues sur certaines parties particulières de notre modèle complexe. Il s'agit d'une composante majeure et indispensable de l'analyse de tout modèle ANOVA complexe. L'ANOVA/MANOVA possède une fonction souple d'analyse de contrastes permettant à l'utilisateur de spécifier et d'analyser pratiquement tout type de comparaison (voir les Notes pour une description de la manière de spécifier des contrastes).

## Analyse *post hoc*

### A quoi cela sert?

Nous trouvons parfois dans nos expériences des effets inattendus. Même si, dans la plupart des cas, un expérimentateur créatif saura expliquer la plupart de ces structures de moyennes, il serait inopportun d'analyser et d'évaluer cette structure comme si nous l'avions prévue dès le départ. Le problème ici est que nous tirons parti de la chance lorsque nous effectuons plusieurs tests post-hoc, c'est-à-dire, sans hypothèse *a priori*. Pour illustrer ce point, considérons "l'expérience" suivante. Écrivons un nombre compris entre 1 et 10 sur 100 morceaux de papier. Plaçons ensuite ces papiers dans un chapeau et tirons 20 échantillons (morceaux de papier) constitués de 5 observations chacun, puis calculons les moyennes (des nombres écrits sur les morceaux de papier) de chaque groupe. Quelle chance pensez-vous avoir de trouver deux moyennes d'échantillons significativement différentes ? C'est très probable ! Si vous sélectionnez les moyennes extrêmes obtenues à partir des 20 échantillons, il est très différent de ne prendre que 2 échantillons dans le chapeau (ce qu'implique le test de l'analyse de contrastes). Sans entrer dans les détails, il existe des tests appelés Tests post-hoc basés sur le premier scénario (prendre les extrêmes des 20 échantillons), c'est-à-dire basés sur l'hypothèse que nous avons choisi pour notre comparaison les moyennes les plus extrêmes (différentes) sur un total de  $k$  moyennes dans le plan. Ces tests appliquent des "corrections" destinées à compenser l'avantage de la sélection post-hoc des comparaisons les plus extrêmes. L'ANOVA/MANOVA offre toute une gamme de tests. Dès que nous trouvons des résultats incorrects ou inattendus dans une expérience, il faut utiliser ces procédures post-hoc afin de tester leur significativité statistique.

## Analyse de variance

### Hypothèse de travail:

#### Écarts à la loi normale.

On considère que la variable dépendante est mesurée sur au moins un niveau d'échelle d'intervalle (voir les Concepts Élémentaires). En outre, la variable dépendante doit être normalement distribuée au sein des groupes. Le module ANOVA/MANOVA comporte divers graphiques et statistiques permettant de tester la validité de cette hypothèse.

→ Conséquences de leurs violations. Globalement, le test F est remarquablement robuste aux écarts à la normalité (voir Lindman, 1974, pour une synthèse).

#### Homogénéité des Variances.

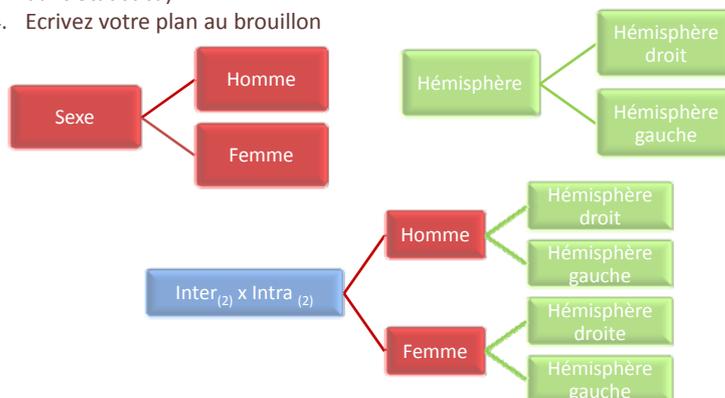
Les variances des différents groupes du modèle sont supposées identiques ; cette hypothèse est appelée hypothèse d'homogénéité des variances.

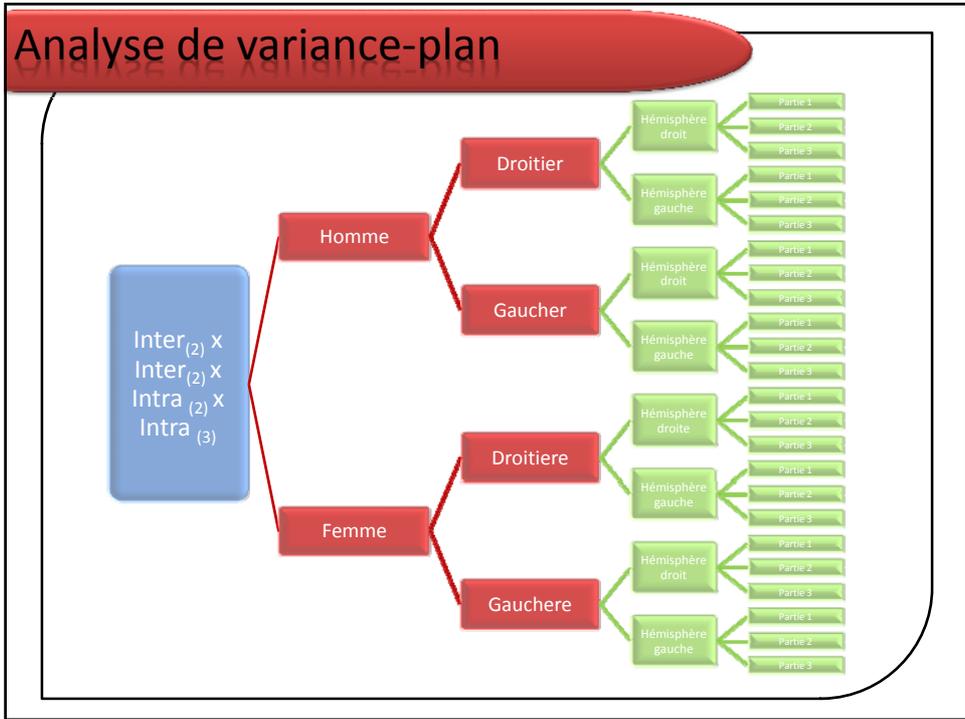
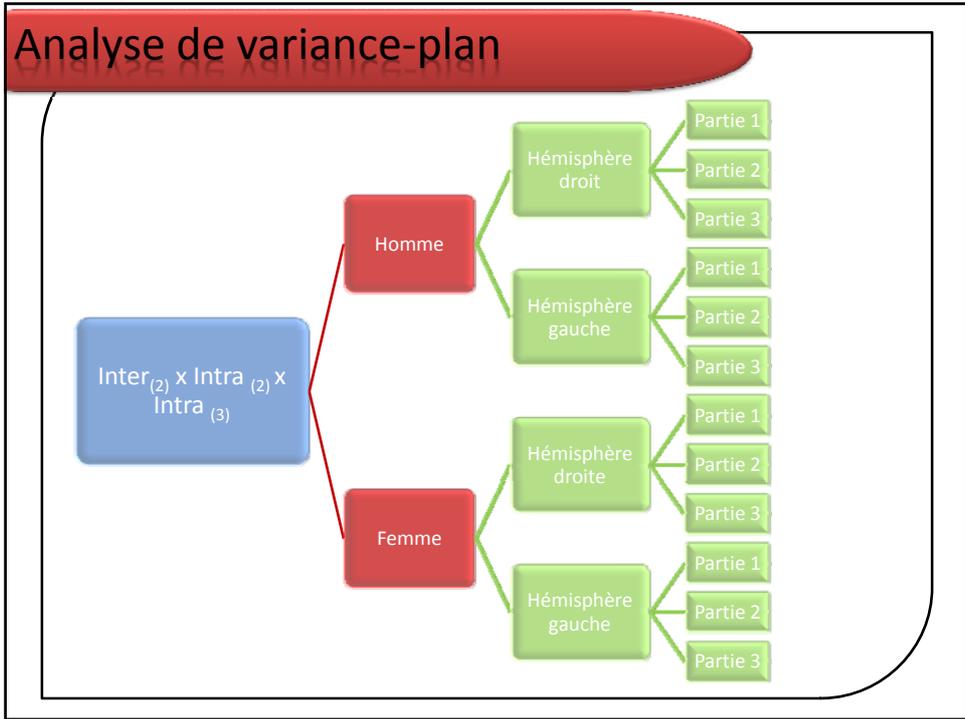
→ Effets des violations. Lindman (1974, p. 33) a montré que la statistique du F est assez robuste par rapport aux violations de cette hypothèse (hétérogénéité des variances ; voir aussi Box, 1954a, 1954b ; Hsu, 1938).

## Analyse de variance-plan

Il est IMPORTANT d'avoir bien en tête le plan de nos données. Comment faire?

1. Listez vos variables
2. Notez le nombre de modalités de chacune des variables
3. Différenciez les VI-intra-sujet (appelée *facteur catégoriel* dans Statistica) et VI inter-sujet (appelée *facteur à un ou plusieurs niveaux ou facteurs répétés* dans Statistica)
4. Écrivez votre plan au brouillon





## Analyse de variance

**Il existe plusieurs types d'ANOVA, leur choix dépend de votre plan :**

- Une variable indépendante inter-sujet
- Plusieurs variables indépendantes inter-sujets (analyse sans interaction)
- Plusieurs variables indépendantes inter-sujets (analyse avec interaction)
- Une variable indépendante intra-sujet avec ou sans variables indépendantes inter-sujets.

*Statistica vous rappelle ici quel test utiliser*

## Analyse de variance

**Il existe plusieurs types d'ANOVA, leur choix dépend de votre plan :**

- Plusieurs variables indépendantes intra-sujets avec ou sans variables indépendantes inter-sujets.

Faire un modèle linéaire général et conduire une ANOVA à mesures répétées

## Analyse de variance-coder vos VI-intra

Si un plan comporte plus d'un facteur de mesures répétées, l'ordre dans lequel ils sont spécifiés va déterminer l'affectation des variables dépendantes sur les niveaux des facteurs de mesures répétées. STATISTICA va "répartir" la liste des variables dépendantes, en fonction des niveaux consécutifs des facteurs de mesures répétées. Plus précisément, la modification la plus rapide des niveaux se fait pour le facteur de mesures répétées spécifié en dernier ; la seconde modification la plus rapide des niveaux se fait pour le facteur spécifié en avant-dernier, et ainsi de suite.

Exemple. Si vous sélectionnez 12 variables dépendantes, et que vous souhaitez spécifier un modèle de mesures répétées 2 x 2 x 3, l'affectation des variables aux niveaux doit être la suivante :

Facteur mesures répétées	niveaux des facteurs de mesures répétées											
	1	1	1	1	1	1	2	2	2	2	2	2
Premier	1	1	1	1	1	1	2	2	2	2	2	2
Deuxième	1	1	1	2	2	2	1	1	1	2	2	2
Troisième	1	2	3	1	2	3	1	2	3	1	2	3
Variable Dépendante	1	2	3	4	5	6	7	8	9	10	11	12

## Analyse de variance-coder vos VI-intra

**Conseil, suivez les étapes suivantes :**

- Coder vos données brutes dans un classeur Excel
- Ordonner vos données à l'aide d'un tableau croisé dynamique sous Excel
- Recoder sous Excel vos noms de VI-intra pour quelles soient lisibles sous Statistica

+ PLAN COMPLET (ou ANOVA Imbriquée): Dans la plupart des cas, les combinaisons de traitement manquantes sont spécifiquement arrangées, de façon à ce que les effets principaux du plan ne soient pas confondus (corrélés) les uns avec les autres. C'est le cas, par exemple, pour les Carrés Latins, qui sont souvent utilisés dans la recherche appliquée, lorsqu'il est peu pratique ou trop coûteux d'observer toutes les combinaisons de traitement possibles (consultez aussi la rubrique Carrés Latins du module Plans d'Expériences pour plus de détails et d'exemples). Ces types de modèles également présentés dans le cadre des modules ANOVA/MANOVA et Décomposition de la Variance et Modèle Mixte ANOVA/ANCOVA.

+ EFFECTIF MINIMUM DANS LES CELLULES...

+ DIFFERENCE ENTRE VARIABLE ALEATOIRE et VARIABLE FIXEE

## ANOVA-Acquis

Ce que vous devez savoir faire à l'issu de ce chapitre pour l'analyse de variance:

- Savoir ce qu'est un ANOVA et connaître le contexte d'utilisation\*\*\*
- Différencier les Viintra, Viinter, VD\*\*\*
- Construire un plan d'expérience\*\*\*
- Connaître ce qu'est une interaction et un effet simple\*\*\*
- Discriminer: l'ANOVA à un facteur, ANOVA factorielle et ANOVA à mesure répétée\*\*\*
- Savoir ce que sont des contrastes (i.e., *comparaisons planifiées, comparaisons a priori*)\*\*\*
- Savoir ce que sont des test post-hocs (i.e., *comparaisons a posteriori*)\*\*\*
- Organiser les données sur Statistica
  - Variables catégorielles (i.e., *VI-intersujet, groupe*)
  - Mesure (i.e., *VD*)
  - Mesures répétées (i.e., *VI-intrasujet, facteur à niveau*)