

200	LE MODELE CROISE FIXE.....	2
200.1	UTILITE	2
200.2	NOTION D'INTERACTION DE DEUX FACTEURS	3
200.2.1	<i>Additivité</i>	3
200.2.2	<i>Non additivité</i>	6
200.3	HOMOGENEITE DES VARIANCES	8
200.4	EGALITE DES MOYENNES	8
200.4.1	<i>Principe</i>	8
200.4.2	<i>Détermination du modèle</i>	9
200.4.3	<i>Test de Hartley</i>	10
200.4.4	<i>Hypothèses nulles</i>	10
200.4.5	<i>Détermination des tests</i>	10
200.4.6	<i>Interprétation</i>	14
200.5	COMPARAISONS MULTIPLES	16
200.5.1	<i>Principe</i>	16
200.5.2	<i>Exemple (suite)</i>	18
200.6	MODELES A PLUS DE DEUX CRITERES CROISES FIXES	20

200 Le modèle croisé fixe

200.1 Utilité

De nombreux facteurs influencent généralement une variable. Par exemple, la taille ou le poids dépendent de l'âge, mais aussi du sexe. L'évolution du nombre de globules blancs au cours d'une infection dépend du temps, mais aussi du traitement.

L'intérêt d'étudier simultanément deux facteurs se comprend à travers la question de l'interaction entre les deux facteurs. Cette question revient à dire : l'effet d'un facteur est-il influencé par le niveau de l'autre ?

Par exemple : la croissance du chômage au cours du temps est-elle la même pour les villes de Liège et Charleroi ? L'évolution du nombre de naissance au cours du temps est-elle la même dans un pays qui octroie un congé parental pour les deux parents et dans un pays où seule la mère a un congé de maternité ? La croissance du volume de déchets recyclés est-elle la même dans un milieu urbain et dans un milieu rural ?

Des critères fixes sont toujours croisés, et le modèle de l'analyse prendra en compte l'effet global de chaque facteur (en faisant abstraction de l'autre) et l'effet de chaque facteur pour chaque niveau de l'autre.

Si l'étude simultanée de deux facteurs permet de répondre à des questions essentielles, multiplier les critères fixes dans une même expérience en en faisant varier trois ou quatre facteurs simultanément génère des modèles dans lesquels les interactions sont nombreuses et complexes et les conclusions peuvent s'avérer très difficiles à établir.

200.2 Notion d'interaction de deux facteurs

200.2.1 Additivité

Un géographe compare l'évolution du taux de chômage entre 2000 et 2005 à Liège et Charleroi. Cette expérience implique deux critères de classification de 2 niveaux, fixes et croisés: temps (année 2000 et 2005) et ville (Liège et Charleroi)

Etablissons un premier modèle sur base de celui de l'ANOVA1 : l'observation (x : nombre de chômeur/1000 habitants) dépend d'une moyenne générale, d'un facteur temps (a_i), d'un facteur ville (b_j), et d'une variabilité d'un échantillon à l'autre ($E_{(ij)k}$).

$$x_{(ij)k} = \mu + a_i + b_j + E_{(ij)k}$$

Simplifions en négligeant momentanément le bruit $E_{(ij)k}$ et étudions les possibilités de ce modèle :

$$\mu_{(ij)} = \mu + a_i + b_j$$

Soit le temps a un effet, et non la ville: $\mu_{(ij)} = \mu + a_i$

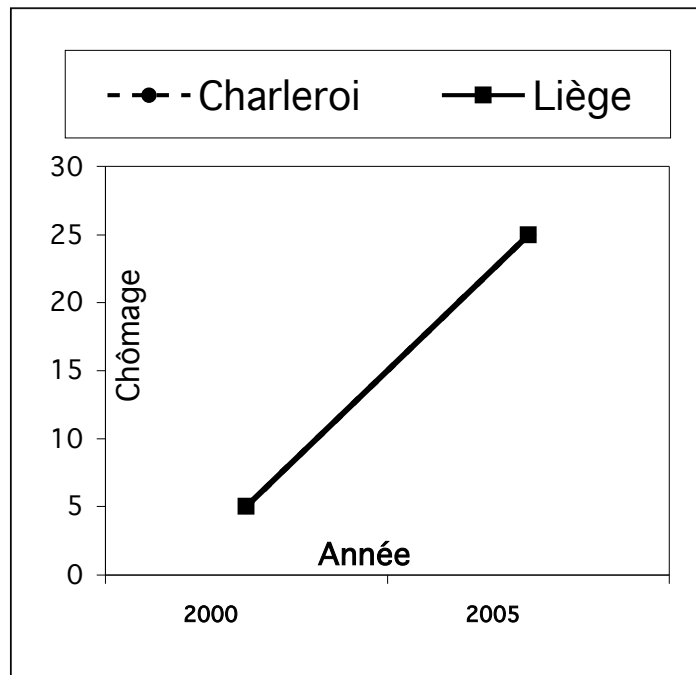


Figure 200 -200-1 Le taux de chômage est supérieur en 2005. Pas de différence entre villes.

Soit la ville a un effet, et non le temps: $\mu_{(ij)} = \mu + b_i$

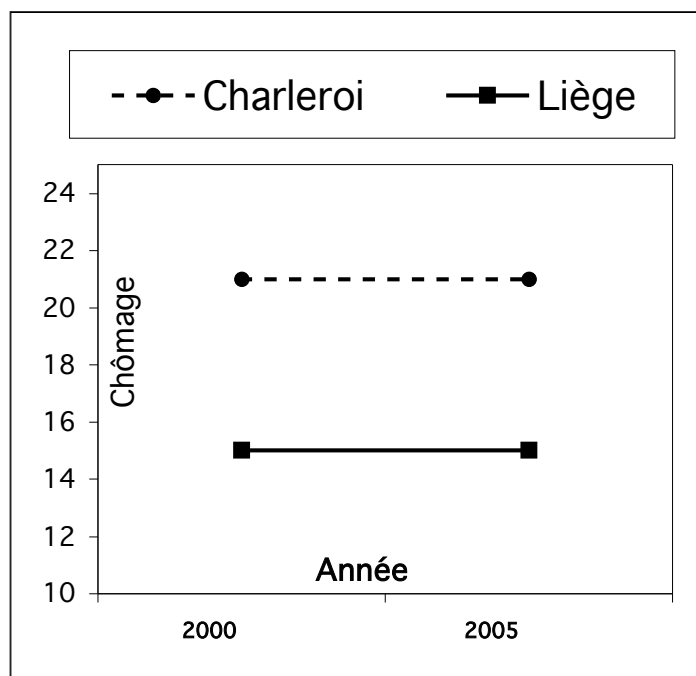


Figure 200 -200-2 Le taux de chômage est supérieur à Charleroi Pas de différence entre les années.

Soit le temps et la ville ont un effet : $\mu_{(ij)} = \mu + a_i + b_j$

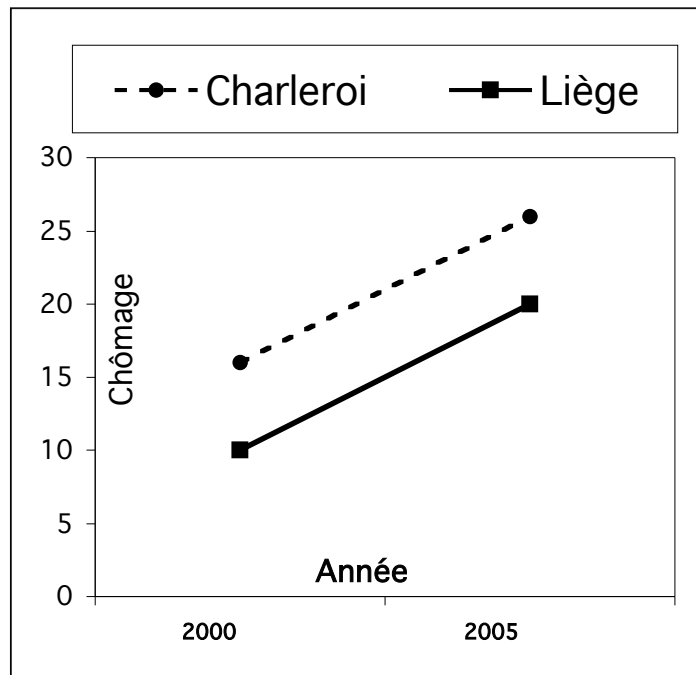


Figure 200 -200-3 Le taux de chômage est supérieur en 2005 et il est supérieur à Charleroi

Ce modèle est limité car il peut uniquement décrire que l'effet du temps est le même pour chaque ville ou réciproquement que la différence entre les villes est constante au cours du temps.

200.2.2 Non additivité

Imaginons maintenant une situation tout à fait réaliste :

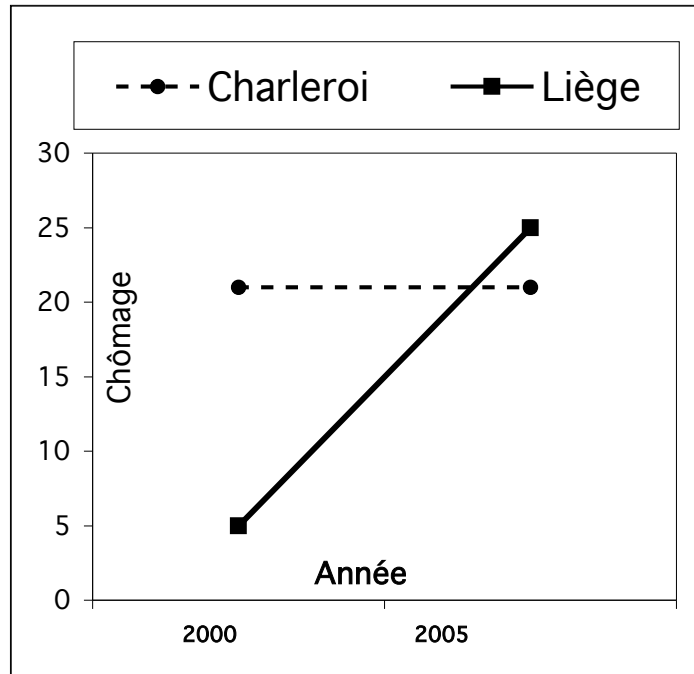


Figure 200 -200-4 le taux de chômage augmente à Liège entre 2000 et 2005, mais à Charleroi il reste constant.

Nous observons ici un phénomène extrêmement courant : l'effet d'un facteur (ici : le temps) n'est pas le même suivant les niveaux d'un autre facteur (ici : la ville). Dans ce cas de figure, la différence entre les villes évolue au cours du temps : en 2000 Charleroi est plus haut que Liège, en 2005, c'est l'inverse. Le taux de chômage évolue à Liège et pas à Charleroi.

Dans d'autres cas de figure, il pourrait évoluer dans les 2 villes, mais plus vite dans l'une que dans l'autre, diminuer dans l'une tandis qu'il augmente dans l'autre etc...

Cet effet différencié d'un facteur suivant le niveau d'un autre facteur s'appelle **l'interaction** entre deux facteurs.

Il y a interaction entre deux facteurs lorsque l'effet de l'un n'est pas le même suivant les niveaux de l'autre.

Cet effet se visualise par un non parallélisme de la représentation graphique des moyennes.

L'interaction ne peut pas être exprimée par le modèle additif :

$$\mu_{(ij)} = \mu + a_i + b_j$$

Il faut donc ajouter un terme au modèle qui exprime la non additivité des facteurs. Ce terme est spécifique de chaque combinaison des niveaux des facteurs (ab_{ij}) :

$$\mu_{(ij)} = \mu + a_i + b_j + ab_{ij}$$

L'interaction peut prendre différentes formes, on parlera de synergie¹ si l'effet d'un facteur est renforcé par l'autre ($ab_{ij} > 0$) ou d'antagonisme² si l'effet d'un facteur est inhibé par l'autre, voire changé de sens ($ab_{ij} < 0$)

Dans sa forme extrême, l'interaction peut carrément masquer les différences globales. Dans ce cas le modèle se résume à :

$$\mu_{(ij)} = \mu + ab_{ij}$$

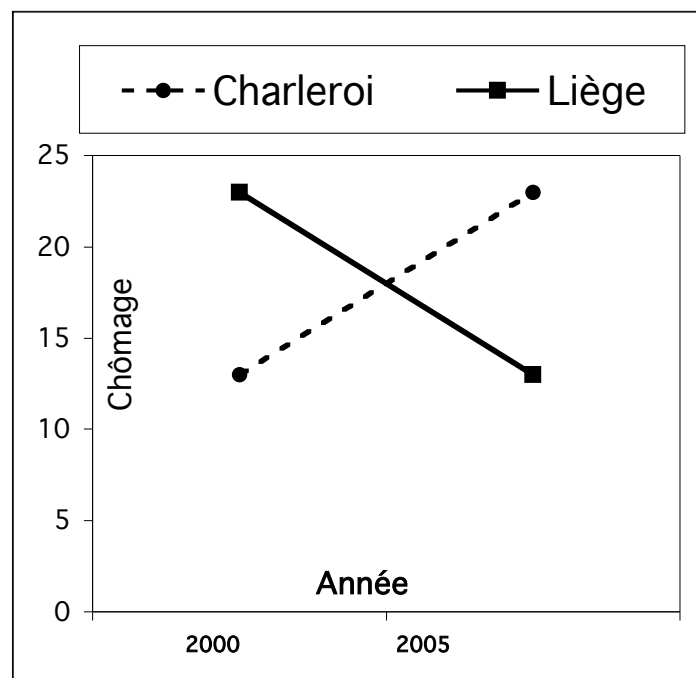


Figure 200 -200-5 En moyenne le taux de chômage est le même dans les deux villes et aux deux périodes. Il y a cependant une effet ville et un effet temps mais l'un compense exactement l'autre

On peut constater que si l'on mélange les données relatives aux deux villes la différence entre les deux années ne sera pas mise en évidence.

¹ Synergie : mise en commun de plusieurs actions concourant à un effet renforcé avec une économie de moyens.

² Antagonisme : opposition fonctionnelle entre deux systèmes, deux organes, deux substances biochimiques.

Le modèle complet de l'Anova 2 fixe est donc :

$$X_{(ij)k} = \mu + a_i + b_j + ab_{ij} + E_{(ij)k} \quad \text{Équation 200-1}$$

200.3 Homogénéité des variances

Le fait que l'on puisse considérer que σ^2 estimée par CMR, est égale dans toutes les populations définies par la combinaison des niveaux des différents critères de classification est l'une des conditions les plus importantes pour la validité de l'inférence.

On considère que l'analyse de la variance, lorsque n est constant, est relativement robuste par rapport à cette condition. Cela signifie que le risque d'erreur α théorique n'est pas fortement modifié par une éventuelle hétérogénéité de variance. Toutefois, lors d'un rejet de H_0 pour un critère fixe à plus de deux niveaux, l'analyse sera suivie de contrastes pour déterminer quels sont les niveaux qui diffèrent. Ces contrastes sont moins robustes. Il faut donc accorder une grande importance à ce test préliminaire, bien que la plupart des logiciels ne l'effectuent pas.

La condition la plus importante est d'assurer l'indépendance des observations sur lesquelles on estime ces variances, ce qui implique d'avoir clairement identifié tous les critères de classification.

Lorsque les effectifs sont égaux, le test le plus simple à réaliser est le test de Hartley.

Il faut toutefois bien remarquer que ce test s'effectue sur les variances calculées sur les n valeurs qui constituent les répétitions de la mesure dans des conditions jugées identiques.

Dans certains modèles, il y a un grand nombre de variances à prendre en considération, et dans d'autres, il n'y a pas de duplicats, et le test est impossible.

200.4 Egalité des moyennes

200.4.1 Principe

Lorsque deux variances sont estimées avec respectivement k et l degrés de liberté dans des populations de même variance σ^2 , leur rapport suit une distribution de Fisher avec k et l degrés de liberté.

Il paraîtra donc peu vraisemblable qu'elles proviennent de populations de même variance si le rapport est significativement supérieur à 1. Nous

considérons donc, au seuil α , que deux carrés moyens (variances) n'ont pas la même valeur attendue si

$$\frac{CM1_k}{CM2_l} > F_{k,l;1-\alpha}$$

200.4.2 Détermination du modèle

Dans un département où coexistent trois variétés d'une même plante, un laboratoire de botanique a étudié si la concentration en chlorophylle des feuilles dépendait de la variété et de l'avancement de la saison. Par variété et par saison, on a mesuré la concentration en chlorophylle sur quatre plantes. Que peut-on en tirer comme conclusions?

		VARI 1	VARI 2	VARI 3
Saison	1	16	19	15
		15	17	18
		21	21	19
		17	19	21
	2	20	15	15
		23	17	18
		17	16	14
		20	20	12
	3	30	21	19
		26	24	25
		23	25	23
		22	19	21
	4	27	16	16
		24	19	17
		20	13	15
		21	24	18

Tableau 200 -200-1 Résultats expérimentaux de l'étude de la concentration en chlorophylle des feuilles en fonction de la variété et de l'avancement de la saison.

Modèle : $x_{(ij)k} = \mu + a_i + b_j + ab_{ij} + E_{(ij)k}$

Critère fixe saison (a), $n_a = 4$, $i = 1,4$

Critère fixe variété (b), $n_b = 4$, $j = 1,3$

critères croisés : interaction (ab)

concentration en chlorophylle $x_{(ij)k}$, $n = 4$, $k = 1,4$

200.4.3 Test de Hartley

	VARI 1	VARI 2	VARI 3
Mois 1	6.9	2.7	6.3
Mois 2	6.0	4.7	6.3
Mois 3	12.9	7.6	6.7
Mois 4	10.0	22.0	1.7

Tableau 200 -200-2 Variances calculées sur les duplicats de l'analyse du taux de chlorophylle.

Question : peut-on considérer que la variance de la concentration en chlorophylle des feuilles est constante dans les 12 populations correspondant à la combinaison variété/saison ?

$$H_0 : \sigma^2_{ij} = \sigma^2$$

H_1 : au moins une des variances est différente

Test de Hartley :

$$H_{3;12} = 22/1.7 = 13.2$$

$$H_{3;12;0.95} = 124$$

H observé est plus petit que la limite au seuil α 5%, il y a donc acceptation de H_0 avec un risque d'erreur β inconnu.

Jusqu'à preuve du contraire nous considérerons que les variances sont homogènes. Si nous nous trompons, considérons que les variances ne sont pas dramatiquement différentes (ce qui nous aurait donné de grandes chances de Rho, et que le test ne pourrait pas en être dramatiquement affecté).

200.4.4 Hypothèses nulles

$\delta^2_a = 0$: il n'y a pas d'effet de la saison, toutes variétés confondues

$\delta^2_b = 0$: il n'y a pas d'effet de la variété, toutes saisons confondues

$\delta^2_{ab} = 0$: l'effet de la saison est le même pour toutes les variétés (ou réciproquement)

200.4.5 Détermination des tests

La démarche suivie dans ce paragraphe suppose la maîtrise des règles de détermination des espérances de carrés moyen vues au module 210.

Le lecteur non averti devra admettre que comme dans un modèle à un critère fixe le test d'hypothèse est réalisé par le rapport CMF/CMR, dans le modèle croisé fixe, les 3 tests du modèle croisés fixes se feront également en rapport au CMR et peut se rendre au point suivant.

		$n_a = 4$	$n_b = 3$	$n = 4$
		i	j	k
$\delta^2 a$	a_i	0	3	4
$\delta^2 B$	b_j	4	0	4
$\delta^2 aB$	ab_{ij}	0	0	4
σ^2	$E_{(ij)k}$	1	1	1

Tableau 200 -200-3 Application de la règle 1 au modèle croisé fixe.

	d.l	E(CM)
a_i	3	$12 \delta^2 a + \sigma^2$
b_j	2	$16 \delta^2 B + \sigma^2$
ab_{ij}	6	$4 \delta^2 aB + \sigma^2$
$E_{(ij)k}$	36	σ^2

Tableau 200 -200-4 Application des règles 2, 3 et 4 pour la détermination des valeurs attendues des carrés moyens et des degrés de liberté.

Vérification des degrés de liberté : $N - 1 = 4.3.4 - 1 = 47 = 3 + 2 + 6 + 36$

Ho	CM	E(CM)	F obs
$\delta^2 a = 0$	Cma	$12 \delta^2 a + \sigma^2$	Cma/CMR
$\delta^2 b = 0$	Cmb	$16 \delta^2 B + \sigma^2$	Cmb/CMR
$\delta^2 ab = 0$	Cmab	$4 \delta^2 aB + \sigma^2$	Cmab/CMR
	CMR	σ^2	

Tableau 200 -200-5 Différentes Ho et rapport de carrés moyens pour les tester.

Ho	d.l	F	$F_{0.95}$
$\delta^2 a = 0$	3	F 3,36	2.87
$\delta^2 b = 0$	2	F 2,36	3.26
$\delta^2 ab = 0$	6	F 6,36	2.36
σ^2	36		

Tableau 200 -200-6 Degrés de liberté des rapports de CM et valeur critique au seuil α 5%.

si le rapport $Cma/CMR > 2.87$, il y aura $Rho \delta^2 a = 0$ et on conclura que la moyenne du taux de chlorophylle, toutes variétés confondues, est différente pour au moins une saison ;

si le rapport $Cmb/CMR > 3.26$, il y aura $Rho \delta^2_b = 0$ et on conclura que la moyenne du taux de chlorophylle, toutes saisons confondues, est différente pour au moins une variété ;

si le rapport $Cmab/CMR > 2.36$, il y aura $Rho \delta^2_{ab} = 0$ et on conclura que l'effet de la saison sur le taux de chlorophylle n'est pas le même pour toutes les variétés.

Toutes ces conclusions sont assorties d'un risque d'erreur de 5%.

En cas d'Aho, on conclura que l'effet recherché n'est pas mis en évidence, avec un risque d'erreur inconnu.

❖ Etablir la table des moyennes

Saison	VARI 1	VARI 2	VARI 3	Mx_i
1	17.3	19.0	18.3	18.2
2	20.0	17.0	14.8	17.3
3	25.3	22.3	22.0	23.2
4	23.0	18.0	16.5	19.2
Mx_j	21.4	19.1	17.9	Mx 19.4

Tableau 200 -200-7 Tables des moyennes calculées sur les duplicats de l'analyse du taux de chlorophylle ((Mx_{ij})). Moyennes globales sur la saison (Mx_i), sur les variétés (Mx_j) et moyenne générale (Mx).

❖ Etablir les équations

Les équations algébriques restent valides que l'on tienne compte de l'indépendance des données ou non. Nous allons donc utiliser l'équation de l'anova 1 et celle de l'anova 2 :

$$SCET = SCEF + SCER$$

Dans cette expression, on considère les données individuelles $x_{(ij)k}$, les moyennes des 12 groupes Mx_{ij} et la moyenne générale Mx

SCER peut être calculé par $SCET - SCEF$

$$SCET = SCEa + SCEb + SCEab + SCER$$

$$\text{Donc : } SCEF = SCEa + SCEb + SCEab$$

SCEa peut être calculée à partir des moyennes Mx_i et de la moyenne générale

SCEa peut être calculée à partir des moyennes Mx_j et de la moyenne générale

SCEab peut être calculé par $SCEF - SCEa - SCEb$

❖ Calculer les SCE dans le tableur Excel

Utilisons la fonction Excel somme.carres.ecarts, et multiplier le résultat par le nombre d'observations sur lesquelles les moyennes en jeu sont calculées

Sur les données $x_{(ij)k}$ (tableau 13) pour SCET

$$\text{SCET} = \text{somme.carres.ecarts}(x_{ijk}) * 1$$

Sur les moyennes Mx_{ij} (tableau 19, à l'exception de la dernière ligne et de la dernière colonne)

$$\text{SCEF} = \text{somme.carres.ecarts}(Mx_{ij}) * 4$$

$$\text{SCER} = \text{SCET} - \text{SCEF}$$

Sur les moyennes Mx_i (tableau 19, dernière colonne)

$$\text{SCEa} = \text{somme.carres.ecarts}(Mx_i) * 4 * 3$$

Sur les moyennes Mx_j (tableau 19, dernière ligne)

$$\text{SCEb} = \text{somme.carres.ecarts}(Mx_j) * 4 * 4$$

$$\text{SCEab} = \text{SCEF} - \text{SCEa} - \text{SCEb}$$

SCET	705.8
SCEF	425.1
SCEb	244.6
SCEa	101.4
SCEab	79.1
SCER	280.8

Tableau 200 -200-8 Calcul des SCE calculées sur les données de l'analyse du taux de chlorophylle.

❖ Calculer les CM et valeur de F

Reprendre les résultats des tables 17 et 18, effectuer les rapports SCE/dl puis les rapports CMa/CMR, CMb/CMR, CMab/CMR puis récapituler le tout dans une table d'analyse de la variance :

	SCE	dl	CM	F_{obs}	$F_{0.95}$
Totale	705.8	47			
Factorielle	425.1				
Saison	244.6	3	81.5	10.4	2.87
Variété	101.4	2	50.7	6.5	3.26
Interaction	79.1	6	13.2	1.7	2.36
Résiduelle	280.8	36	7.8		

Tableau 200 -200-9 Table d'analyse de la variance sur les données de l'analyse du taux de chlorophylle.

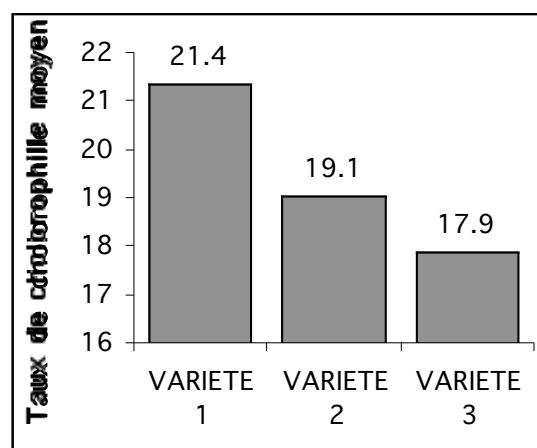
En conclusion, il y a Rho chaque fois que $F_{obs} > F_{0.95}$, c'est -à -dire pour l'effet global de la saison, et l'effet global de la variété. L'interaction des deux facteurs n'est pas significative.

200.4.6 Interprétation

Lorsque l'interaction est significative, elle doit être examinée en premier lieu, car son effet est susceptible d'invalidier, ou au moins de moduler, les conclusions tirées sur l'effet global. Dans notre exemple, l'interaction est non significative.

Le taux de chlorophylle moyen par variété, toutes saisons confondues, est significativement différent de la moyenne générale, au moins pour une variété.

Graphiquement :



aux de chlorophylle moyen par variété, toutes saisons confondues

Le graphique montre que la production globale de chlorophylle est la plus élevée pour la variété 1.

Néanmoins, à ce stade de l'analyse, nous ne pouvons pas conclure que le taux de chlorophylle est significativement plus élevé pour la variété 1 que pour la 2, ou plus pour la 2 que pour la 3. Tout au plus, on peut raisonnablement penser qu'il est plus élevé pour la 1 que pour la 3.

Le taux de chlorophylle moyen par saison, toutes variétés confondues, est significativement différent de la moyenne générale, au moins pour une saison.

Graphiquement :

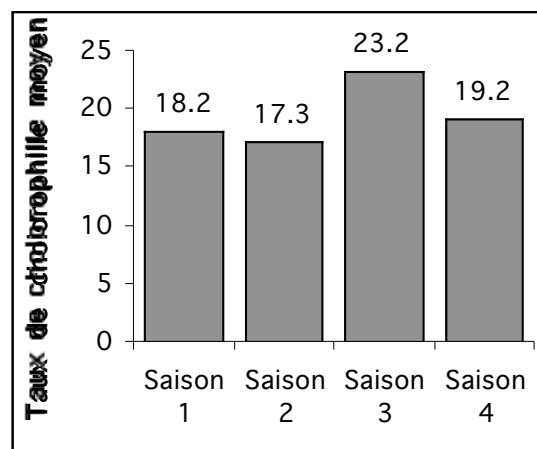


Figure 200-6 Taux de chlorophylle moyen par saison, toutes variétés confondues.

Le graphique montre que la production globale de chlorophylle est la plus élevée pour la saison 3.

Néanmoins, à ce stade de l'analyse, nous ne pouvons pas conclure que le taux de chlorophylle est significativement plus élevé pour la saison 3 que pour la 4, ou plus pour la 1 que pour la 2. Tout au plus, on peut raisonnablement penser qu'il est plus élevé pour la 3 (maximum) que pour la 2 (minimum).

200.5 Comparaisons multiples

200.5.1 Principe

Comme dans le cas de l'ANOVA I, le rejet de H_0 débouche généralement sur des H_1 équivoques. L'interprétation – en partie graphique – qui précède met en évidence que nous ne pouvons pas encore répondre à toutes les questions en maîtrisant les risques d'erreur des conclusions.

Ces comparaisons multiples s'effectuent par des techniques de contrastes. Un contraste L est une combinaison linéaire des moyennes Mx_i (ou Mx_j , ou Mx_{ij} ...) qui estime λ combinaison linéaire des moyennes μx_i (ou μx_j , ou μx_{ij} ...).

$$L = \sum_{i=1}^{na} c_i Mx_i$$

$$\lambda = \sum_{i=1}^{na} c_i \mu x_i$$

$$\sum_{i=1}^{na} c_i = 0$$

Lorsque n est constant, la variance d'un contraste est donnée par :

$$S_L^2 = \frac{1}{n} CMe \sum_{i=1}^{na} c_i^2$$

expression dans laquelle n représente le nombre d'observations sur lesquelles sont effectuées les moyennes qui sont comparées, et CMe représente le carré moyen qui sert à tester l'effet global (ou d'interaction) qui est décomposé par le contraste.

Différentes techniques ont été proposées pour tester l'hypothèse nulle :

$$H_0 : \lambda = 0 \text{ contre } H_1 : \lambda \neq 0$$

Dans le cadre de ce cours, nous envisageons la méthode de Scheffé, qui présente l'avantage de garantir la confiance $1 - \alpha$ quel que soit le nombre de contrastes effectués, et ce même dans le cas où l'idée de la comparaison vient de l'observation (contrastes dit a posteriori) alors que d'autres techniques soit ne garantissent pas la confiance lorsque le nombre de contrastes est élevé, soit sont plus exigeantes quant à leur définition (contrastes définis avant l'expérience, dit a priori, comparaisons indépendantes les unes des autres)...

L'inconvénient des contrastes de Scheffe est d'être peu puissants ($1 - \beta$), c'est à dire que le risque de ne pas détecter une différence est plus grand que dans d'autres approches, ce qui peut conduire à une situation paradoxale : l'ANOVA (puissante) détecte une différence qu'aucun contraste (peu puissants) n'arrive à localiser.

Toutefois ce risque ne doit pas être exacerbé : la différence de puissance n'est pas dramatique (ce qui signifie que le surplus d'observations nécessaire à obtenir une puissance comparable à d'autres méthodes n'est pas très élevé). Il faut aussi se rappeler qu'en cas de RHo, le problème de la puissance ne se pose pas : c'est bénéficier d'une confiance garantie est dans ce cas très important.

La formule de Scheffé donnant la limite de différence significative (DS) au seuil α est la suivante :

$$DS = \pm \sqrt{(na - 1) S_L^2 F_{na-1; N-na; 1-\alpha}}$$

elle se simplifie en donnant la plus petite différence significative (PPDS) pour un ensemble de comparaisons de moyennes 2 à 2 de la façon suivante :

$$PPDS = \pm \sqrt{(na - 1) \frac{2CMe}{n} F_{na-1; N-na; 1-\alpha}}$$

Cette simplification s'explique par le fait que dans ce cas, les coefficients sont tous nuls sauf un couple 1, -1, et donc dans tous les contrastes.

$$\sum c_i^2 = 2 \quad \text{et} \quad S_L^2 = 2CMe/n$$

Ces formules doivent être adaptées au contexte : n_a représente le nombre de niveaux du critères qui est décomposé en contrastes, n . le nombre d'observations de chacun des niveaux de ce critère, CMe le carré moyen qui sert à tester ce critère, $N - n_a$ le nombre de degrés de liberté de ce carré moyen.

200.5.2 Exemple (suite)

Dans un département où coexistent trois variétés d'une même espèce de plante, un laboratoire de botanique a étudié si la concentration en chlorophylle des feuilles dépendait de la variété et de l'avancement de la saison. Par variété et par saison, on a mesuré la concentration en chlorophylle sur quatre plantes.

Saison	VARI 1	VARI 2	VARI 3	Mx_i
1	17.3	19.0	18.3	18.2
2	20.0	17.0	14.8	17.3
3	25.3	22.3	22.0	23.2
4	23.0	18.0	16.5	19.2
Mx_j	21.4	19.1	17.9	Mx 19.4

Tableau 200 -200-10 Tables des moyennes calculées sur les duplicats de l'analyse du taux de chlorophylle ((Mx_{ij})). Moyennes globales sur la saison (Mx_i), sur les variétés (Mx_j) et moyenne générale (Mx).

La table d'analyse de la variance est la suivante :

	SCE	dl	CM	F_{obs}	$F_{0.95}$
Totale	705.8	47			
Factorielle	425.1				
Saison	244.6	3	81.5	10.4	2.87
Variété	101.4	2	50.7	6.5	3.26
Interaction	79.1	6	13.2	1.7	2.36
Résiduelle	280.8	36	7.8		

Tableau 200 -200-11 Table d'analyse de la variance sur les données de l'analyse du taux de chlorophylle.

L'interaction étant non significative, nous décomposons l'effet variété, puis l'effet saison.

Effet variété (3 niveaux) :

$$PPDS = \pm \sqrt{2 \times \frac{2 \times 7.8}{16} F_{2;36;0.95}}$$

$$F_{2,36;0.95} = 3.27$$

$$PPDS = 2.52$$

Variété		1	2	3
		21.4	19.1	17.9
1	21.4	0	2.3	3.5
2	19.1	-2.3	0	1.2
3	17.9	-3.5	-1.2	0

Tableau 200 -200-12 Table des différences de moyennes 2 à 2 (moyennes globales par variétés). Les valeurs en gras sont –en valeur absolue – plus grande que PPDS, ces différences sont significatives au seuil 5%.

En conclusion, la teneur en chlorophylle est significativement plus basse pour la variété 3 (17.9), par rapport aux variétés 1 et 2 (risque d'erreur α 5%). Des différences éventuelles entre les autres variétés n'ont pas été mises en évidence (risque d'erreur β inconnu).

Effet saison (4 niveaux) :

$$PPDS = \pm \sqrt{(4-1) \frac{2 \times 7.8}{12} F_{3,36;0.95}}$$

$$F_{3,36;0.95} = 2.87$$

$$PPDS = 3.34$$

Saison		1	2	3	4
	Mx_i	18.2	17.3	23.2	19.2
1	18.2	0	0.9	-5	-1
2	17.3	-0.9	0	-5.9	-1.9
3	23.2	5	5.9	0	4
4	19.2	1	1.9	-4	0

Tableau 200 -200-13 Table des différences de moyennes 2 à 2 (moyennes globales par saisons). Les valeurs en gras sont –en valeur absolue – plus grande que PPDS, ces différences sont significatives au seuil 5%.

En conclusion, la teneur en chlorophylle est significativement plus élevée en saison 3 (23.2), par rapport aux saisons 1,2 et 4 (risque d'erreur α 5%). Des différences éventuelles entre les autres saisons n'ont pas été mises en évidence (risque d'erreur β inconnu).

200.6 Modèles à plus de deux critères croisés fixes

Chaque paire de critères croisés génère une interaction. De ce fait, trois critères croisés génèrent trois interactions dites de premier ordre entre les critères pris 2 à 2 et une interaction dite de deuxième ordre qui fait intervenir les trois critères.

Le modèle s'écrit :

$$\mu_{(ij)} = \mu + a_i + b_j + c_k + ab_{ij} + ac_{ik} + bc_{jk} + abc_{ijk}$$

Exemple : un biochimiste étudie l'effet protecteur sur une culture de cellules de substances anti -oxydantes provenant du raisin et solubles dans l'alcool du vin.

X : mesure de la présence de radicaux libres

Premier critère (a_i), âge des cellules : les cellules sont jeunes, ou âgées.

Deuxième critère (b_j), temps de culture : il varie de 1 à 4 jours

Troisième critère (c_k), couleur du vin : le vin est rouge ou blanc.

Les trois critères sont fixes.

L'interaction double ab_{ij} mesure la différence d'effet du temps entre cellules jeunes et âgées (ou réciproquement la différence de l'effet de l'âge des cellules aux différents temps de culture).

L'interaction double ac_{ij} mesure la différence d'effet du type de vin entre cellules jeunes et âgées (ou réciproquement la différence de l'effet de l'âge des cellules d'un type de vin à l'autre).

L'interaction double bc_{ij} mesure la différence d'effet du type de vin en fonction du temps (ou réciproquement la différence de l'effet du temps d'un type de vin à l'autre).

L'interaction triple abc_{ijk} mesure si l'interaction entre le temps et l'âge des cellules est la même d'un type de vin à l'autre, ou réciproquement si l'interaction entre le temps et le type de vin est la même suivant l'âge des cellules, ou réciproquement si l'interaction entre le type de vin et l'âge des cellules est la même au cours du temps.

La complexification du modèle et la difficulté de l'interprétation augmentent très rapidement avec le nombre de critères croisés.

Une expérience qui fait intervenir 4 critères de classifications croisés générera un tel nombre d'interactions, (dont une de troisième ordre) que les résultats risquent fort de ne pas être interprétables.

La combinaison de deux critères hiérarchisés ne génère pas d'interaction. En effet, puisque les niveaux du critère hiérarchisé changent pour chacun des niveaux du critère fixe il n'est pas possible d'isoler une différence due à l'interaction.

Modèle croisé : un individu subit le traitement A et le traitement B. Un second individu subit le traitement A et le traitement B. Le modèle expérimental permet d'identifier une différence d'effet du traitement d'un patient à l'autre.

Modèle hiérarchisé : un individu subit le traitement A. Un second individu subit le traitement B. Le modèle expérimental ne permet pas d'identifier une différence d'effet du traitement d'un patient à l'autre.

Dans un modèle expérimental, le nombre de critères hiérarchisés n'augmente pas dramatiquement la complexité de l'interprétation des résultats.