

Quatrième partie

Quatrième partie : Efficacité des
contrats : Analyse

La quatrième partie de notre thèse fait une étude statistique destinée à apporter une réponse à la question suivante : comment les relations annexes (ces relations ont été définies dans les parties précédentes de la thèse) déterminent les comportements des clients vis-à-vis de leurs échéances de crédit ? Le but est donc d'analyser statistiquement le rapport probable entre les relations annexes et les comportements de respect ou de non respect des clients vis-à-vis des échéances.

Présentation des données de l'étude et des traitements à effectuer

Cette mise en place est destinée à montrer les différentes variables de l'étude, les traitements qui doivent être réalisés ainsi que les exercices préliminaires à savoir les corrélations simples de Pearson et les tests de comparaisons de moyennes. Ces exercices sont nécessaires pour comprendre les traitements futurs et pour valider certains résultats.

Précisons que pour traiter nos données, nous avons retenu le logiciel SPSS : "Statistical Package for the Social Sciences", en particulier la version 10.1. C'est à notre avis, l'un des meilleurs outils de traitement des données pour des analyses statistiques en sciences sociales. Ce logiciel n'est pas le seul permettant de réaliser les types de traitement dans notre thèse. D'autres logiciels tels que SAS, Stata, SPAD, Statview, Eviews, ... sont aussi performants et auraient été utiles pour réaliser les mêmes types d'analyse. Mais, nous avons utilisé le logiciel SPSS avant la thèse, durant notre stage au PADME et à la FECECAM. Ce logiciel a servi pour la saisie de l'enquête, l'enregistrement des données collectées et le traitement effectué. Nous avons continué avec ce logiciel.

Echantillons d'enquête et base de données

Les données de notre étude sont présentées par la figure 9.3 ci-dessous.

FIG. 9.3 – Données de l'étude

Echantillons d'enquête				Population	
Total clients					
Individuels et					
caut. Solid.					
Fececam		64			
Padme		832		34804	
Padme	Dont contrats en caution solidaire	272	32,70%	13658	39,20%
	Dont contrats individuels	560	67,30%	21146	60,80%
	Clients individuels avec TRECLT > = 0,95	344		9166	
	Dont Femmes	231	67,20%	6150	67,10%
	Dont Hommes	113	32,80%	3016	32,90%
	Dont Recommandés	330	95,90%	9101	99,30%
	Dont Avalisés	252	73,30%	7267	79,30%

annees	nbrenc	nbrcredi	valenc	fondep	penalit	encm	fondepdm	penalim
1994	251	239	122164527	1948539	629616	486711,26	8152,88	2508,43
1995	816	707	310699079	5431132	4739027	380758,68	7681,94	5807,63
1996	1355	1163	441633512	14347985	8769422	325928,79	12337,05	6471,90
1997	1910	1821	753515045	48276272	10056496	394510,49	26510,86	5265,18
1998	3044	2642	1133224466	99246200	6589541	372281,36	37564,80	2164,76
1999	4618	3874	1848335813	136117144	8040469	400245,95	35136,07	1741,11
2000	5948	5001	2716316429	196454629	10633556	456677,27	39283,07	1787,75
2001	7002	6079	3583263023	234356774	13703329	511748,50	38551,86	1957,06
2002	7714	6915	4445201682	263922064	16836339	576251,19	38166,60	2182,57
2003	9211	8204	6340050968	356592692	24053882	688312,99	43465,71	2611,43
2004	10161	9374	8421580522	508211320	39778422	828814,14	54214,99	3914,81
2005	9143	9605	8311234031	479138951	48364701	909027,02	49884,33	5289,81

Pour réaliser les traitements, nous avons utilisé les 9166 clients comme un groupe témoin. Le but est d'analyser l'évolution des taux de respect des échéances pour l'ensemble des clients en crédits individuels qui sont au nombre de 21146.

Choix des variables

Variable dépendante : taux de respect des échéances

Cette étude utilise une variable dépendante qui est le taux de respect des échéances de la clientèle.

Dans toute l'étude, cette variable est notée *treclt*. Elle est obtenue grâce à la formule suivante :

$$treclt = \frac{nbesr}{nbe}$$

où :

nbesr est le nombre d'échéances honorées sans retard sur la période et *nbe* est le nombre total d'échéances de la période.

Pour obtenir ce taux, nous calculons d'abord le nombre total d'échéance pour un client (*nbe*) en multipliant le nombre de crédit obtenu (variable ancienneté) par 12. Par hypothèse, on suppose qu'un cycle de crédit est reparti en douze échéances. Ensuite, on calcule le nombre d'échéances honorées sans retard sur la période correspondante à l'ancienneté d'un client (*nbesr*) en soustrayant du nombre total d'échéances (*nbe*) le nombre de retards enregistré (variable retards). Enfin, on calcule le taux de respect des échéances de la clientèle (*treclt*) en rapportant le nombre d'échéances honorées sans retard (*nbesr*) sur le nombre total d'échéances (*nbe*) comme indiqué dans la formule précédente. Le résultat est proposé par la figure 9.4 pour certains clients en crédits individuels sur lesquels porteront les analyses ultérieures.

FIG. 9.4 – Taux de respect des échéances de la clientèle

échantillon				population			
taux de respect des éch. de la clientèle				taux de respect des éch. de la clientèle			
N	Valide		344	N	Valide		9166
	Manquante		0		Manquante		0
Moyenne			,9926	Moyenne			,9810
Centiles	25		,9896	Centiles	25		,9667
	50		1,0000		50		,9833
	75		1,0000		75		1,0000

	Fréquence	Pour cent	Pourcentage cumulé		Fréquence	Pour cent	Pourcentage cumulé
,95	1	,3	,3	,95	288	3,1	3,1
,96	1	,3	,6	,96	13	,1	3,3
,97	1	,3	8,7	,97	95	1,0	4,3
,98	2	,6	10,5
...
,99	3	,9	30,5	,99	3	,0	64,1
1,00	239	69,5	100,0	,99	5	,1	64,2
Total	344	100,0		1,00	2	,0	64,2
				Total	9166	100,0	

Le respect des échéances de la clientèle (*treclt*) est une variable continue.

Avant tout, nous devons savoir la distribution de cette variable dépendante. Pour cela, on fait le test de normalité.

Test de normalité de la variable dépendante : *treclt*

Le test de normalité qui est effectué a pris en compte uniquement les clients qui ont les "bons" taux. Le résultat est proposé par la figure 9.5 ci-dessous. Ce test montre que la variable taux de respect des échéances pour les clients qui ont les "bons" taux suit une distribution normale. Le coefficient d'asymétrie (ou skewness) est inférieur à $|1|$ et le coefficient d'aplatissement (ou de concentration, "Kurtosis") est inférieur à $|1,5|$.

FIG. 9.5 – Test de normalité de la variable : *treclt*

Statistiques		
taux de respect des éch. de la clientèle		
N	Valide	9168
	Manquante	0
Asymétrie		-.222
Erreur std. d'asymétrie		.028
Aplatissement		-1,402
Erreur std. d'aplatissement		.051

Catégorisation des clients par tranche de taux de respect des échéances

Dans cette étude, la variable dépendante taux de respect des échéances de la clientèle est une variable métrique de proportion. Elle varie de manière conséquente entre des valeurs faibles qui sont inférieures à la valeur critique (médiane) de 50% et des valeurs fortes qui sont très proches de 100%. Dans l'analyse, il est important de prendre en compte et de distinguer les clients selon qu'ils ont des taux de respect des échéances (*treclt*) inférieurs ou supérieurs à la valeur critique de 50%. Dans cette perspective, l'analyse des comportements de respect ou de non respect des échéances par les clients doit prendre en compte et distinguer des catégories abstraites de clients selon les taux auxquels ils sont identifiés au moment de la collecte des données de cette étude.

Nous avons ainsi distingué quatre groupes ou catégories de clients selon leur taux de respect des échéances et sur chaque catégorie est portée une observation particulière. Nous distinguons les clients aux "bons" taux, les clients aux taux moyens "supérieurs", les clients aux taux moyens "inférieurs" et les clients aux "mauvais" taux.

1. pour les clients aux "bons" taux : $treclt \geq 95\%$
2. pour les clients aux taux moyens "supérieurs" : $85\% \leq treclt < 95\%$
3. pour les clients aux taux moyens "inférieurs" : $50\% \leq treclt < 85\%$.
4. pour les clients aux "mauvais" taux : $treclt < 50\%$

L'essentiel des traitements effectués concerne la première catégorie de clients, c'est-à-dire ceux qui ont les $trecht \geq 95\%$.

variables indépendantes

Dans notre étude, les variables indépendantes sont potentiellement nombreuses eu égard à l'importance des informations récoltées dans le questionnaire (annexe 5). Celui-ci est établi avec 75 questions c'est-à-dire 75 possibilités de variables indépendantes. Par conséquent, nous avons procédé à des regroupements et à faire des tris.

La figure 9.6 montre les variables retenues pour cette étude.

FIG. 9.6 – Variables de l'étude

	variable dépendante		variables explicatives	
	type	modalités	type	modalités
taux de respect des échéances	continu	infini		
ancienneté			discrète	1 à 22
genre			discrète	1 = féminin 2 = masculin
type de crédit			discrète	1 = individuel 2 = caut. Solid
recommandation			discrète	1 = recommandé 2 = non recom
aval			discrète	1 = avalisé 2 = non avalisé
Variables complémentaires				
Encours moyen de crédits	continu			
Fonds et dépôts de garanties moyens	continu			
Pénalités (de retards) moyennes	continu			

La variable ancienneté est une variable importante pour cette étude. Mais son utilisation comme variable explicative pose un problème de raisonnement important. Doit-on penser que l'ancienneté a un effet probable sur le taux de respect des échéances ou l'inverse ? Car, il est plus logique de considérer que c'est le taux de respect des échéances qui conditionne l'ancienneté et non pas le contraire. Or, si nous voulons comprendre l'évolution du taux de respect des échéances pour un client, son ancienneté n'apporterait aucune information importante³. L'idée qui paraît plus logique c'est de considérer que moins un client fera de retard sur ses échéances et plus il augmentera ses chances de durer dans l'institution en obtenant des renouvellements de crédits. Donc, sur cette base, l'ancienneté serait expliquée par le taux de respect

³Même si certaines informations recueillies durant les enquêtes laissent penser que plus certains clients durent dans le portefeuille et moins ils respectent leurs échéances.

des échéances qui devient une variable explicative. Mais est-ce la seule? On peut penser que non. Or, l'ancienneté étant connue, avec des valeurs limitées de 1 à 22⁴, nous ne pouvons pas raisonnablement la considérer comme une variable continue. Ce faisant, elle ne peut être prise comme variable dépendante dans cette étude. Nous l'utiliserons comme variable indépendante dans une analyse précise : l'analyse de variance factorielle et pour le reste, elle sera prise comme variable de pondération.

La recommandation et l'aval ont été définis précédemment dans les relations du contrat de dette (voir la section 5.1.1).

D'autres variables ont été choisies. Ce sont :

- **le montant moyen d'encours de crédit par client : (encm)**. Il est déterminé de la manière suivante. D'abord, on cherche la valeur de l'encours moyen de crédit (encours de base) en divisant la valeur de l'encours annuel par le nombre de crédits en cours pour la même année. Le nombre de crédits en cours correspond au nombre de crédits déboursés pour l'année en cours plus les crédits non soldés de l'année précédente. Le chiffre trouvé correspond au montant moyen d'encours de crédit pour chaque client. Ensuite, connaissant l'ancienneté de chaque client, le montant correspondant est affecté à chacun en fonction de son ancienneté. La base des données fournit des données allant de 1994 à 2005 ; ce qui entraîne que l'ancienneté maximal est de 12 ans. Or la même base de données fournit des anciennetés supérieures à 12 pour certains clients. Cela est possible car certains clients sollicitent et obtiennent des renouvellements de crédits avant la clôture de leur précédent contrat. Pour la cohérence de l'analyse, nous affecterons une ancienneté = 12 dans ces cas particuliers.

- **montant moyen des fonds et dépôts de garanties pour les crédits en cours par client : (fondepm)**. Il est déterminé de la manière suivante. D'abord, on cherche le montant moyen des fonds et dépôts de garanties en divisant la valeur des fonds et dépôts de garanties reçus annuellement par l'IMF PADME par le nombre de crédits déboursés durant l'année. Ensuite, le chiffre trouvé est affecté à chaque client en fonction de son ancienneté.

- **montant moyen des frais de pénalités de retard pour crédit en cours par client : (penalim)**. Il est déterminé de la manière suivante. D'abord, on cherche le montant moyen des pénalités de retards en divisant le montant total reçu pour pénalités par le nombre de crédits en cours durant l'année. Ensuite, le chiffre trouvé est affecté à chaque client en fonction du nombre de retards effectués.

Traitements

Présentation des analyses effectuées dans l'étude

Pour notre étude, les étapes suivantes sont suivies : une analyse de variance à 1 facteur, une analyse de variance multiple, une régression linéaire multiple, et une analyse de covariance. Ces analyses se présentent en deux niveaux. D'abord, nous avons examiné l'évolution de la variable

⁴A ce propos une explication s'impose. Elle a été faite plus bas

dépendante taux de respect des échéances en fonction d'une série de variables discrètes : genre, recommandation et aval. Pour cela, les analyses effectuées sont :

1. analyse de variance à 1 facteur (ou ANOVA à 1 facteur) qui utilise comme variable indépendante (VI) l'ancienneté. C'est le facteur
2. analyse de variance factorielle (ou ANOVA à plusieurs facteurs) qui utilise comme variables indépendantes : genre, aval et recommandation avec ancienneté comme variable de pondération.

Ensuite, nous avons examiné l'évolution de la variable dépendante taux de respect des échéances en fonction des variables discrètes : genre, recommandation et aval et des variables continues : encours moyen (encm), fonds et dépôts moyens (fondep) et pénalités moyennes (penalim). Pour cela, les analyses effectuées sont :

1. régression linéaire simple qui utilise comme variable indépendante (VI) l'encours moyen. C'est le regresser. Cette étude est réalisée également avec chacun des deux autres VI continues : fonds et dépôts moyens et pénalités moyennes.
2. régression linéaire multiple qui utilise comme variable indépendante (VI) les trois variables continues citées ci-dessus.
3. analyse de covariance qui utilise comme variables indépendantes : genre, recommandation, aval (qui sont les variables discrètes) et les trois variables continues : encours moyen, fonds et dépôts moyens et pénalités moyennes.

Préliminaires : corrélation de Pearson et test de student

Avant les analyses mentionnées au point précédent, nous avons effectué les traitements préliminaires suivants : analyse de corrélation simple de Pearson et test du t de Student.

Corrélation simple de Pearson

Avant une étude de données, il est important de calculer un coefficient de corrélation afin d'établir l'existence d'un lien entre la ou les variables explicative-s et la ou les variables dépendante-s. Mais ce calcul n'est possible qu'à condition d'avoir une variable indépendante quantitative et une variable dépendante quantitative. Les données de notre étude permettent de réaliser ces calculs. La procédure de calcul d'une corrélation avec SPSS est relativement simple.

Le coefficient de corrélation (r) est basé sur la covariance qui s'obtient par la formule :

$$COV(X, Y) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \quad (9.1)$$

et :

$$r = \frac{COV(X, Y)}{S_X S_Y} \quad (9.2)$$

Le calcul effectué est présenté dans la figure 9.7.

FIG. 9.7 – Coefficients de corrélation

		tredt	encom	fondep m	penalim m	ancienn eté trié	genre	recom manda tion	aval
tredt	Cor. Pearson	1	,143**	-,063**	-,552**	-,032**	,017	-,067**	-,119**
	Sig. (bilatérale)		,000	,000	,000	,003	,105	,000	,000
	N	9166	9166	9166	9166	9166	9166	9166	9166
encom	Cor. Pearson	,143**	1	-,749**	-,379**	-,857**	,220**	-,048**	-,003
	Sig. (bilatérale)	,000		,000	,000	,000	,000	,000	,752
	N	9166	9166	9166	9166	9166	9166	9166	9166
fondep m	Cor. Pearson	-,063**	-,749**	1	,015	,537**	-,163**	-,007	-,018
	Sig. (bilatérale)	,000	,000		,149	,000	,000	,527	,080
	N	9166	9166	9166	9166	9166	9166	9166	9166
penali m	Cor. Pearson	-,552**	-,379**	,015	1	,430**	-,062**	,171**	,096**
	Sig. (bilatérale)	,000	,000	,149		,000	,000	,000	,000
	N	9166	9166	9166	9166	9166	9166	9166	9166
ancien neté trié	Cor. Pearson	-,032**	-,857**	,537**	,430**	1	-,182**	,062**	-,034**
	Sig. (bilatérale)	,003	,000	,000	,000		,000	,000	,001
	N	9166	9166	9166	9166	9166	9166	9166	9166
genre	Cor. Pearson	,017	,220**	-,163**	-,062**	-,182**	1	,076**	,651**
	Sig. (bilatérale)	,105	,000	,000	,000	,000		,000	,000
	N	9166	9166	9166	9166	9166	9166	9166	9166
recom manda tion	Cor. Pearson	-,067**	-,048**	-,007	,171**	,062**	,076**	1	-,014
	Sig. (bilatérale)	,000	,000	,527	,000	,000	,000		,170
	N	9166	9166	9166	9166	9166	9166	9166	9166
aval	Cor. Pearson	-,119**	-,003	-,018	,096**	-,034**	,651**	-,014	1
	Sig. (bilatérale)	,000	,752	,080	,000	,001	,000	,170	
	N	9166	9166	9166	9166	9166	9166	9166	9166

** La corrélation est significative au niveau 0.01 (bilatéral).

Test de comparaison de moyennes

Dans les chapitres précédents, voir les chapitres 7, 8 et 9, l'examen des données disponibles à savoir les données d'échantillons d'enquêtes au PADME et la base de données de la clientèle dans cette institution a permis de déduire que :

1. le comportement de respect ou de non respect vis-à-vis des échéances par les clients ne pouvait pas être déterminé par leur genre masculin ou féminin. C'est-à-dire qu'on ne pouvait pas soutenir par exemple l'idée selon laquelle les femmes seraient de facto plus respectueuses des échéances par rapport aux hommes ;
2. le comportement de respect ou de non respect vis-à-vis des échéances par les clients ne

pouvait pas être déterminé par le type de crédit que ces derniers sollicitent. C'est-à-dire qu'on ne pouvait pas soutenir par exemple l'idée selon laquelle les clients en groupes de caution solidaire seraient de facto plus respectueux des échéances par rapport aux clients en crédits individuels ;

3. le comportement de respect des échéances par les clients pourrait être déterminé par leur relation de recommandation (ou de parrainage). C'est-à-dire qu'il serait possible de soutenir l'idée que les clients recommandés par leurs proches parents amis et connaissances seraient plus respectueux des échéances par rapport à ceux qui ne seraient pas recommandés ;
4. le comportement de respect des échéances par les clients pourrait être déterminé par leur relation d'aval. C'est-à-dire qu'il serait possible de soutenir l'idée que les clients avalisés par leurs proches parents amis et connaissances seraient plus respectueux des échéances par rapport à ceux qui ne seraient pas avalisés ;
5. Parmi les clients qui sont recommandés mais non avalisés et les clients qui ne sont pas recommandés mais qui sont avalisés, ce sont les premiers qui semblent être plus respectueux des échéances par rapport aux seconds. C'est-à-dire que le comportement de respect vis-à-vis des échéances serait plus prononcé chez les clients recommandés par rapport aux clients avalisés ;
6. le comportement de non respect des échéances est plus prononcé parmi les clients qui ne sont ni recommandés, ni avalisés alors que le comportement de respect des échéances est plus prononcé parmi les clients qui sont à la fois recommandés et avalisés

Le but des tests de comparaison de moyennes est de tester statistiquement ces déductions. Ces tests ne sont possibles que par exemple, lorsqu'on a une recherche comportant une variable dépendante quantitative et une variable explicative qualitative ayant deux modalités. C'est notre cas si nous considérons la variable dépendante taux de respect des échéances et chacune des variables explicatives discrètes.

Dans la logique d'un test d'hypothèses comme le test du t de Student ou le khi deux de Pearson, etc., il y a toujours deux hypothèses statistiques. La première - l'hypothèse nulle ou H_0 - postule qu'il n'y a pas de différence entre les moyennes de deux groupes observés : Groupe 1 = Groupe 2 ou Mesure1 = Mesure2. La seconde - l'hypothèse alternative ou H_1 - correspond habituellement à l'hypothèse de la recherche. Contrairement à l'hypothèse nulle, on postule qu'il existe une différence entre les moyennes des deux groupes (ou des deux mesures) : Groupe 1 \neq Groupe 2 ou Mesure1 \neq Mesure2. Donc \neq signifie n'égale pas.

Généralement, il y a deux types d'hypothèses alternatives ; l'hypothèse unilatérale (ou unicaudale) et l'hypothèse bilatérale (ou bicaudale). L'hypothèse unilatérale est formulée lorsqu'on désire vérifier une hypothèse de recherche. Par exemple est-il vrai que les femmes sont plus respectueuses des échéances par rapport aux hommes ? Si nous répondons oui à cette question, alors nous formulons une hypothèse statistique unilatérale : $F > H$ (on pourrait

également formuler l'hypothèse inverse $F < H$). C'est l'un ou l'autre, ou bien on affirme $F > H$, ou bien on affirme $F < H$.

L'hypothèse bilatérale est choisie lorsqu'on se contente de formuler un objectif de recherche. Par exemple : est-il vrai de dire que les femmes ont un comportement différent vis-à-vis du respect des échéances par rapport aux hommes? Dans ce cas, différent signifie logiquement deux choses : $F > H$ et $F < H$, d'où le terme "bilatérale". C'est l'un et l'autre. Il est possible que $F > H$ et il se peut également que $F < H$.

Les calculs effectués avec SPSS sont proposés ci-dessous.

Vérification avec la variable genre. Les données porte sur la catégorie des clients aux "bons" taux : $treclt \geq 95\%$

L'hypothèse à vérifier (H1) est la suivante : Le taux de respect des échéances est en moyenne supérieur chez les clients de sexe féminin par rapport aux clients de sexe masculin. H1 : $F > H$. Les tests effectués sont présentés dans la figure 9.8.

FIG. 9.8 – Test de Student par la variable : genre

Statistiques de groupe					
genre (sexe du client)		N	Moyenne	Ecart-type	Erreur standard moyenne
treclt	féminin	6160	,9808	,01676	,00021
	masculin	3016	,9814	,01768	,00032

Test d'échantillons indépendants								
		Test de Levene sur l'égalité des variances		Test-t pour égalité des moyennes				
		F	Sig.	t	ddl	Sig. (bilatérale)	Différence moyenne	Différence écart-type
treclt	Hypothèse de variances égales	63,719	,000	-1,621	9164	,105	-,0006	,00038
	Hypothèse de variances inégales			-1,592	5714,5	,112	-,0006	,00039

Le premier réflexe est de regarder le tableau " statistiques de groupe " pour vérifier si la moyenne du taux de respect des échéances pour les clients de sexe féminin est supérieure à celle des clients de sexe masculin. Dans nos résultats, les clients de sexe féminin obtiennent une moyenne de 0,98081, alors que les clients de sexe masculin ont une moyenne de 0,98143. La différence entre les deux moyennes est -.00062. Ensuite, on regarde le test d'homogénéité

des variances dans le tableau du "test d'échantillons indépendants" pour sélectionner la ligne à choisir pour la lecture des résultats du test.

Dans un test t, il y a trois données importantes : le résultat du test t (= -1,621 dans notre cas), le degré de liberté ddl (= 9164 dans notre cas) et la valeur de p (ou Signification (bilatérale) = 0,105 dans notre cas lorsque les variances sont égales). Le t et le ddl permettent à SPSS de calculer la valeur de p (Sig.(bilatérale)). Cette valeur est l'erreur alpha, soit la probabilité ou le risque de commettre une erreur en déclarant qu'il existe une différence entre les deux groupes observés. C'est ce résultat qui permet de confirmer ou d'infirmer notre hypothèse statistique (H1). Le seuil de signification pris en compte est de 0,05 (ou 5%). Les règles de lecture des résultats du test t pour échantillons indépendants sont les suivantes⁵.

Dans notre exemple, le test de Levene sur l'égalité des variances ("Levene's Test for Equality of Variances") a une signification de 0,000 (inférieure à 5%). La lecture du résultat se fera sur la deuxième ligne. La valeur du t de Student est de -1,592. (Le signe (-) ne nous sert à rien). Le degré de liberté est de 5714,5. La probabilité p est de 0.112 (ou 11,2%).

Vu que nous avons posé une hypothèse dite "orientée", c'est-à-dire que nous avons spécifié le sens de la différence en posant que les clients de sexe féminin ont une moyenne du taux de respect des échéances qui est supérieure, nous pouvons diviser la probabilité (11,2%) par deux. On trouve alors une probabilité de 5,6% qui est supérieure à 5%. La lecture du résultat final se fera sur la première ligne.

Le test invite à retenir l'hypothèse des variances égales. C'est-à-dire qu'il n'y pas de différence significative entre les moyennes de taux de respect des échéances entre les deux groupes. Donc l'hypothèse de départ est rejetée. On ne peut pas soutenir que le taux de respect des échéances est en moyenne supérieur chez les clients de sexe féminin par rapport aux clients de sexe masculin dans la catégorie des clients qui ont les "bons" taux. Ce n'est pas vrai. Mais le contraire se serait pas vrai non plus ; c'est-à-dire qu'on ne peut pas soutenir que $F < H$. Statistiquement, les groupes ont des variances égales même si on constate une légère différence dans les moyennes de ces deux groupes.

Le même procédé est utilisé pour analyser les autres variables indépendantes : recommandation, aval et type de crédit. Pour cette dernière variable, les données concernent les clients individuels et les clients en caution solidaire, tous ayant un $treclt \geq 95\%$.

Vérification avec la variable recommandation

Hypothèse testée H1 : les clients recommandés respectent en moyenne plus les échéances par rapport aux clients non recommandés de la même catégorie. Ce supposé est approuvé

⁵Si l'hypothèse statistique est unilatérale, on divise la valeur de Sig. (bilatérale) par 2. Si Sig. (bilatérale) est supérieure à 0,05, alors, on accepte l'hypothèse nulle et on conclue qu'il n'y a pas de différence significative entre les deux groupes testés. Si Sig. (bilatérale) est inférieure à 0,05, on doit rejeter l'hypothèse nulle et conclure qu'il y a une différence significative entre les deux groupes dans le sens proposé par le tableau des statistiques

par les " statistiques de groupe ". La moyenne obtenue par les clients recommandés (0,98111) est supérieure à celle obtenue par les clients non recommandés soit 0,96743. L'écart est de 0,01368. Le test de Levene a une signification de 0,000. Le t de Student qui est lu est de 8,461. Le degré de liberté est de 65,596. La probabilité est de 0,000. Notre hypothèse de départ étant " orientée ", on divisera la probabilité lue par deux, on obtient 0,000 ; inférieur à 0,05. Ce qui nous conduit à retenir l'hypothèse des variances inégales. Le test confirme donc l'inégalité des moyennes observée entre les clients recommandés et les clients non recommandés. Le sens du test effectué montre que le premier groupe c'est-à-dire les clients recommandés ont en moyenne le score supérieur. Ainsi l'hypothèse " orientée " de départ est confirmée par le test. Il y a bien un écart de comportement vis-à-vis du respect des échéances chez les clients aux " bons " taux entre ceux qui sont recommandés et ceux qui ne sont pas recommandés. Et se sont les clients recommandés qui sont en moyenne plus respectueux des échéances par rapport aux clients non recommandés.

On approfondit ce test en couplant les variables recommandation et genre. On a ainsi : les femmes recommandées et ayant leur $treclt \geq 95\%$; les femmes non recommandés et ayant leur $treclt \geq 95\%$; les hommes recommandés et ayant leur $treclt \geq 95\%$ et enfin les hommes non recommandés et ayant leur $treclt \geq 95\%$. A chaque fois, la variable dépendante reste le taux de respect des échéances de la clientèle et la variable indépendante devient la variable genre. Les résultats de ses tests sont présentés dans le tableau de synthèse.

Vérification avec la variable aval

Hypothèse testée : les clients qui sont avalisés respectent en moyenne plus les échéances par rapport aux clients non avalisés de la même catégorie. Ce supposé est approuvé par les " statistiques de groupe ". La moyenne obtenue par les clients avalisés (0,98205) est supérieure à celle obtenue par les clients non avalisés soit 0,97704. L'écart est de 0,00501. Le test de Levene a une signification de 0,207. Le t de Student qui est lu est de 11,464. Le degré de liberté est de 9164. La probabilité est de 0,000. Notre hypothèse de départ étant " orientée ", on divise la probabilité lue par deux, on obtient 0,000 ; inférieur à 0,05. Ce qui conduit à retenir l'hypothèse des variances inégales. Le test confirme l'inégalité des moyennes observées entre les clients avalisés et les clients non avalisés. Ce sont les clients avalisés qui sont en moyenne plus respectueux des échéances par rapport aux clients non avalisés. D'autres tests ont été effectués en couplant les variables aval*genre, recommandation*aval, et recommandation*aval*genre. Les résultats de ses tests sont présentés dans le tableau de synthèse.

Vérification avec la variable type de crédit. Les données portent sur l'ensemble des clients individuels et en caution solidaire ayant les "bons" taux : $treclt \geq 95\%$

Hypothèse testée : H1 : le taux de respect des échéances est en moyenne supérieur chez les clients en groupes de caution solidaire par rapport aux clients en crédits individuels dans la

catégorie des clients qui ont les "bons" taux.

Le résultat du test confirme cette l'hypothèse. Parmi les clients qui ont les " bons " taux, les clients en caution solidaire ont effectivement un niveau moyen de respect des échéances supérieur à celui des clients en crédits individuels.

Synthèse des tests de comparaison de moyenne

Nous proposons dans une figure récapitulative les tests du t de Student qui ont été calculés, les hypothèses testées et les conclusions de chaque test. Puis nous proposons les interprétations que nous pouvons en faire. Figure 9.9.

FIG. 9.9 – Récapitulatif des tests du t de Student

Variables		hypothèses	Test	
			$treclt \geq 0,95$	$treclt < 0,5$
genre	Féminin (F) Masculin (H)	$F > H$	Indécision ; F = H	Rejet : $F < H$
typcrt	Individuel (I) Caut. Solid (CS)	$F > H$	Confirmée : $F > H$	t non significatif
typcrt * genre	Féminin (F)	$F.I. > H.I$	Rejet : $F.I < H.I$	Rejet : $F.I < H.I$
	Masculin (H)	$F.CS > H.CS$	Rejet : $F.CS < H.CS$	Rejet : $F.CS < H.CS$
recommandation	recommandé (R) non recommandé (NR)	$R > NR$	Confirmée : $R > NR$	t non significatif
recommand * genre	Féminin (F)	$F.R. > H.R$	Rejet : $F.R < H.R$	t non calculé
	Masculin (H)	$F.NR > H.NR$	Rejet : $F.NR < H.NR$	Rejet : $F.NR < H.NR$
aval	Avalisé (A) non avalisé (NA)	$A > NA$	Confirmée : $A > NA$	t non calculé
aval * genre	Féminin (F)	$F.A > H.A$	Rejet : $F.A < H.A$	t non calculé
	Masculin (H)	$F.NA > H.NA$	Confirmée : $F.NA > H.NA$	Rejet : $F.NA < H.NA$
genre*recomm and * aval	R.A > R.NA NRA > NR.NA F.R.A > H.R.A F.R.NA > H.R.NA F.NR.A > H.NR.A F.NR.NA > H.NR.NA		Confirmée : $R.A > R.NA$ Rejet : $NR.A < NR.NA$ Rejet : $F.RA < H.RA$ Confirmée : $F.R.NA > H.R.NA$ Rejet : $F.NR.A < H.NR.A$ t non calculé	} t non calculé Rejet : $F.NR.NA < H.NR.NA$

Nous avons soumis seize (16) hypothèses "orientées" au test du t de Student. L'idée générale des hypothèses dans les tests réalisés suppose la supériorité des taux de respect des échéances pour les clients de sexe féminin par rapport aux clients de sexe masculin. A la fin de ces tests et au regard des résultats obtenus, les conclusions suivantes peuvent être tirées concernant les clients qui ont les "bons" taux.

Sur les seize hypothèses testées, un seul test n'a pu être calculé pour absence d'observation. Il concerne les clients non recommandés et non avalisés.

Tous les tests avec la variable *genre* conduisent à rejeter les hypothèses testées. Notamment, c'est le cas avec les variables : genre, type de crédit*genre, recommandation*genre et aval*genre. C'est dans un cas uniquement que le test impliquant la variable genre a été confirmé. Il s'agit du cas du test de la variable genre sur les clients non avalisés.

Plusieurs tests (6) ont confirmé leurs hypothèses. Ils concernent notamment la variable type de crédit et les variables recommandation et aval prises de façon isolée c'est-à-dire uniquement sans les " coupler " avec d'autres variables.

Plusieurs autres tests (8) ont rejeté leurs hypothèses. Ils concernent notamment les variables couplées : type de crédit*genre, recommandation*genre et aval*genre lorsque dans ce dernier cas, les clients sont avalisés. Il y a aussi la variable couplée : recommandation*aval lorsque les clients sont non recommandés, ou si les clients sont recommandés, l'hypothèse est rejetée lorsque les clients sont avalisés et que le test "couple" les variables recommandation, aval et genre.

Les tests qui ont été effectués ont conduit à des résultats qui confirment certaines hypothèses. Parmi ces hypothèses, il y a celle qui porte sur la variable type de crédit. Le fait que le test avec cette variable confirme l'hypothèse de base soulève une interrogation. Ce n'est pas la conclusion du test qui est surprenante. C'est plutôt le fait que la proportion de contrats de dette pour les clients en groupe de caution solidaire soit relativement faible en volume et surtout en valeur comparativement aux contrats de dette pour les clients en crédits individuels. Et pourtant, le test confirme que les crédits en caution solidaire connaissent significativement des taux de respect des échéances qui sont supérieurs aux crédits individuels. Ce comportement d'offre de microcrédits ne paraît pas à ce titre être un comportement économiquement rationnel. Y a-t-il une raison cohérente et rationnelle pour justifier la préférence des crédits individuels dans ce contexte ? Surement ! Une piste pour expliquer pourquoi les contrats de dette pour les clients en crédits individuels sont en volume et en valeur plus nombreux comparativement aux contrats pour les clients en caution solidaire pourrait être les revenus d'intérêts tirés par les institutions prêteuses sur les crédits individuels⁶. Mais cela n'explique pas pourquoi les montants des crédits en caution solidaire sont faibles. Il y aurait peut-être une présomption de relation (négative) entre le montant accordé pour un microcrédit en caution solidaire et les comportements des membres du groupe de caution vis-à-vis du paiement de la dette ? C'est une piste.

Les tests qui utilisent uniquement les variables recommandation et aval confirment les hypothèses de "supériorité" des clients recommandés et des clients avalisés vis-à-vis du respect

⁶Le taux d'intérêt étant par définition une proportion du montant du crédit, plus le montant est élevé, plus élevé sera également le revenu d'intérêt. Or, traditionnellement, les montants de crédits sont plus élevés pour les clients individuels. Du coup, plus nombreux ils seront, et plus conséquents seront les revenus d'intérêt pour les institutions prêteuses.

des échéances par rapport aux clients non recommandés et des clients non avalisés. Cela veut dire qu'en prenant en compte uniquement les relations annexes des clients en crédits individuels et toute autre considération mise à part, les clients qui "possèdent" ses relations annexes ont les taux de respect des échéances les plus élevés. Cela pourrait justifier l'emploi de ses relations annexes en tant que critères d'établissement des contrats de dette et de sélection des clients et cela pourrait justifier parallèlement l'offre de microcrédits. On comprendrait alors pourquoi les clients qui sont à la fois recommandés et avalisés sont en proportion très importante par rapport à ceux qui ne sont ni recommandés ni avalisés. D'un point de vue économique, on comprend ce comportement d'offre de microcrédits qui est logique, cohérent et rationnel. Or, là se pose un problème. Puisque les clients qui sont recommandés et les clients qui sont avalisés ont forcément un genre (ou sexe). Nous devons donc traiter obligatoirement la question de l'homogénéité dans le test sur la variable recommandation et sur la variable aval. Or, lorsqu'on applique le test d'homogénéité des variances (le test de Levene sur les variances égales), on a des conclusions mitigées qui globalement rejettent les hypothèses de "supériorité" des clients de sexe féminin vis-à-vis du respect des échéances par rapport aux clients de sexe masculin. Par exemple, pour la variable couplée : recommandation et genre (recommandation*genre), les tests rejettent les hypothèses de "supériorité" des clients de sexe féminin vis-à-vis du respect des échéances par rapport aux clients de sexe masculin. De même, pour la variable couplée : aval et genre (aval*genre), les tests rejettent les hypothèses de "supériorité" des clients de sexe féminin vis-à-vis du respect des échéances par rapport aux clients de sexe masculin lorsque les clients sont avalisés.

Si les conclusions des tests changent selon que chaque variable indépendante est testée uniquement ou selon que les variables sont couplées avec d'autres variables, en particulier la variable genre (avec le traitement de l'homogénéité des variances), cela veut dire qu'il existe une corrélativité entre les variables indépendantes de l'étude.

Bref

Dans cette mise au point, nous avons montré les données de l'étude et présenté les variables sur lesquelles porte le traitement. Les traitements préliminaires ont permis de montrer que des relations possibles existent entre les variables indépendantes choisies et la variable dépendante taux de respect des échéances.

Le chapitre suivant entame l'analyse des données proprement dite.

Chapitre 10

Analyse de variance

10.1 Analyse de variance à 1 facteur (ANOVA)

10.1.1 Principes, données et calculs

Principes

L'analyse de variance à un facteur (ou ANOVA à 1 facteur) est utilisée dans le cas des tests de comparaison de moyennes avec plusieurs (plus de deux) échantillons indépendants ou groupes pour une variable discrète. Cela ne donne pas la possibilité de recourir aux tests de comparaison de moyennes classiques comme ceux qui ont été réalisés dans le chapitre précédent (voir l'exemple dans le paragraphe IV). Nous avons montré que des différences existent parmi les moyennes des groupes de clients. Nous avons testé et vérifié que ces différences sont significatives. Mais cela ne suffit pas. Nous devons savoir précisément quelles sont les moyennes qui diffèrent. Pour cela, il existe deux types de tests destinés à faire cette étape supplémentaire de comparaison des moyennes. Il y a les contrastes à priori et les tests *post hoc*. Les contrastes sont des tests définis avant de conduire l'expérience, et les tests *post hoc* sont effectués après l'expérience. Il est parfois possible de tester également les tendances à travers les catégories.

Dans une analyse de variance à un facteur, la variable indépendante (le facteur) doit avoir plus de deux modalités. La variable indépendante prise ici est la variable ancienneté-triée. Elle a huit modalités. 1= nouveaux clients, 2= anciens clients avec [1 à 2] renouvellements, 3= anciens clients avec [3 à 4] renouvellements, 4= anciens clients avec [5 à 6] renouvellements, 5 = anciens clients avec [7 à 9] renouvellements, 6= anciens clients avec [10 à 14] renouvellements, 7= anciens clients avec [15 à 20] renouvellements et 8= anciens clients avec plus de 20 renouvellements. L'ancienneté détermine le nombre cumulé de contrats de dette obtenus par un client. On considère par principe qu'un contrat de dette est signé pour une durée d'une année - c'est le délai du crédit - et qu'il s'établit sur des échéances mensuelles. L'analyse qu'il nous revient de faire consiste à tester si le taux de respect moyen des échéances est différent selon l'ancienneté des clients. Avant de répondre à cette question, mettons en place les notations et

les données du problème.

Etapas

Avant de procéder à une analyse de variance, nous devons formuler nos hypothèses statistiques (H_0 et H_1). H_0 = hypothèse nulle; elle postule qu'il n'y a pas de différence entre les moyennes des groupes. Groupe 1 = Groupe 2 = Groupe 3, etc. H_1 = hypothèse alternative; elle postule qu'il existe une différence entre les moyennes des groupes. On a : Groupe 1 \neq Groupe 2 \neq Groupe 3 etc. (le symbole \neq signifiant n'égal pas). L'existence de cette différence permet d'inférer que la variable explicative (VI) est bien la cause de la variable dépendante (VD). On notera que l'hypothèse alternative d'une analyse de variance est toujours bidirectionnelle.

Le résultat du test est présenté dans la figure 10.1.

FIG. 10.1 – ANOVA à 1 facteur : ancienneté

ANOVA^a

taux de respect des ech. de la clientèle

	Somme des carrés	ddl	Moyenne des carrés	F	Signification
Inter-groupes	,110	7	,016	56,212	,000
Intra-groupes	2,561	9153	,000		
Total	2,671	9165			

a. type de crédit = individuel

Descriptives^a

taux de respect des ech. de la clientèle

	N	Moy.	Ecart-type	Erreur std	Intervalle de confiance à 95% pour la moyenne		Min.	Max.
					Borne inf.	Borne sup.		
nouveaux clients	108	1,0000	,0000	,00000	1,0000	1,0000	1,00	1,00
ancien client, [1 à 2] renouvellements	2960	,9834	,0181	,00033	,9828	,9841	,96	1,00
ancien client, [3 à 4] renouvellements	2510	,9769	,0172	,00034	,9763	,9776	,95	1,00
ancien client, [5 à 6] renouvellements	1898	,9805	,0155	,00035	,9798	,9812	,95	1,00
ancien client, [7 à 9] renouvellements	1198	,9813	,0155	,00045	,9804	,9821	,95	1,00
ancien client, [10 à 14] renouvellements	455	,9833	,0152	,00071	,9819	,9847	,95	1,00
ancien client, [15 à 20] renouvellements	35	,9969	,0039	,00066	,9955	,9982	,98	1,00
ancien client, [21 et +] renouvellements	2	,9981	,0027	,00189	,9740	1,0222	1,00	1,00
Total	9166	,9810	,0171	,00018	,9807	,9814	,95	1,00

a. type de crédit = individuel

La signification du test : Sig. = 0,000 < 0,05 montre qu'il y a une différence significative entre les huit groupes. Cependant, il y a un risque de déclarer que les huit groupes sont différents alors qu'il est possible que seulement deux ou trois groupes le soient. Pour résoudre ce problème, nous devons faire un test *post-hoc* qui indiquera lesquels des huit groupes sont

différents. Ce test compare les huit groupes deux à deux. SPSS calcule ce test.

Comparaisons multiples : post-hoc

Dans cette étude, nous n'appliquons que le test *post hoc*. Les *post hoc* permettent des comparaisons deux à deux si (et uniquement si) l'analyse de variance est significative.

Ce test indiquera (dans la feuille des résultats) les paires significativement différentes. C'est le but du test *post hoc*. Dans le tableau croisé proposé dans la feuille des résultats, une étoile est placée à l'intersection des lignes et colonnes dont les variables sont différentes des autres. Si l'analyse de variance n'est pas significative, ce tableau croisé n'apparaît pas dans la feuille des résultats. Les résultats sont présentés dans la figure 10.2.

FIG. 10.2 – Test Post-Hoc

Comparaisons multiples ^a

Variable dépendante: taux de respect des éch. de la clientèle
LSD

(I) ancienneté trié	(J) ancienneté trié	Diff. de moy. (I-J)	Erreur standard	Sig.	Intervalle de confiance à 95%	
					Borne inf.	Borne sup.
nouveaux clients	ancien client, [1 à 2] renouvellements	,0166*	,00164	,000	,0134	,0198
	ancien client, [3 à 4] renouvellements	,0231*	,00164	,000	,0198	,0263
	ancien client, [5 à 6] renouvellements	,0195*	,00165	,000	,0162	,0227
	ancien client, [7 à 9] renouvellements	,0187*	,00168	,000	,0154	,0220
	ancien client, [10 à 14] renouvellements	,0167*	,00179	,000	,0132	,0202
	ancien client, [15 à 20] renouvellements	,0031	,00325	,339	-,0033	,0095
	ancien client, [21 et +] renouvellements	,0019	,01193	,874	-,0215	,0253
ancien client, [1 à 2] renouvellements	nouveaux clients	-,0166*	,00164	,000	-,0198	-,0134
	ancien client, [3 à 4] renouvellements	,0065*	,00045	,000	,0056	,0074
	ancien client, [5 à 6] renouvellements	,0029*	,00049	,000	,0019	,0039
	ancien client, [7 à 9] renouvellements	,0022*	,00057	,000	,0010	,0033
	ancien client, [10 à 14] renouvellements	,0002	,00084	,848	-,0015	,0018
	ancien client, [15 à 20] renouvellements	-,0135*	,00284	,000	-,0190	-,0079
	ancien client, [21 et +] renouvellements	-,0147	,01183	,215	-,0379	,0085
ancien client, [3 à 4] renouvellements	nouveaux clients	-,0231*	,00164	,000	-,0263	-,0198
	ancien client, [1 à 2] renouvellements	-,0065*	,00045	,000	-,0074	-,0056
	ancien client, [5 à 6] renouvellements	-,0036*	,00051	,000	-,0046	-,0026
	ancien client, [7 à 9] renouvellements	-,0043*	,00059	,000	-,0055	-,0032
	ancien client, [10 à 14] renouvellements	-,0063*	,00065	,000	-,0080	-,0047
	ancien client, [15 à 20] renouvellements	-,0199*	,00285	,000	-,0255	-,0144
	ancien client, [21 et +] renouvellements	-,0212	,01183	,074	-,0443	,0020
ancien client, [5 à 6] renouvellements	nouveaux clients	-,0195*	,00165	,000	-,0227	-,0162
	ancien client, [1 à 2] renouvellements	-,0029*	,00049	,000	-,0039	-,0019
	ancien client, [3 à 4] renouvellements	,0036*	,00051	,000	,0026	,0046
	ancien client, [7 à 9] renouvellements	-,0007	,00062	,235	-,0019	,0005
	ancien client, [10 à 14] renouvellements	-,0027*	,00087	,002	-,0045	-,0010
	ancien client, [15 à 20] renouvellements	-,0164*	,00285	,000	-,0220	-,0108
	ancien client, [21 et +] renouvellements	-,0176	,01183	,137	-,0408	,0056
ancien client, [7 à 9] renouvellements	nouveaux clients	-,0187*	,00168	,000	-,0220	-,0154
	ancien client, [1 à 2] renouvellements	-,0022*	,00057	,000	-,0033	-,0010
	ancien client, [3 à 4] renouvellements	,0043*	,00059	,000	,0032	,0055
	ancien client, [5 à 6] renouvellements	,0007	,00062	,235	-,0005	,0019
	ancien client, [10 à 14] renouvellements	-,0020*	,00092	,029	-,0038	-,0002
	ancien client, [15 à 20] renouvellements	-,0156*	,00287	,000	-,0212	-,0100
	ancien client, [21 et +] renouvellements	-,0168	,01183	,155	-,0400	,0064
ancien client, [10 à 14] renouvellements	nouveaux clients	-,0167*	,00179	,000	-,0202	-,0132
	ancien client, [1 à 2] renouvellements	-,0002	,00084	,848	-,0018	,0015
	ancien client, [3 à 4] renouvellements	,0063*	,00065	,000	,0047	,0080
	ancien client, [5 à 6] renouvellements	,0027*	,00087	,002	,0010	,0045
	ancien client, [7 à 9] renouvellements	,0020*	,00092	,029	,0002	,0038
	ancien client, [15 à 20] renouvellements	-,0136*	,00293	,000	-,0194	-,0079
	ancien client, [21 et +] renouvellements	-,0148	,01185	,211	-,0381	,0084
ancien client, [15 à 20] renouvellements	nouveaux clients	-,0031	,00325	,339	-,0095	,0033
	ancien client, [1 à 2] renouvellements	,0135*	,00284	,000	,0079	,0190
	ancien client, [3 à 4] renouvellements	,0199*	,00285	,000	,0144	,0255
	ancien client, [5 à 6] renouvellements	,0164*	,00285	,000	,0108	,0220
	ancien client, [7 à 9] renouvellements	,0156*	,00287	,000	,0100	,0212
	ancien client, [10 à 14] renouvellements	,0136*	,00293	,000	,0079	,0194
	ancien client, [21 et +] renouvellements	-,0012	,01216	,920	-,0250	,0226
ancien client, [21 et +] renouvellements	nouveaux clients	-,0019	,01193	,874	-,0253	,0215
	ancien client, [1 à 2] renouvellements	,0147	,01183	,215	-,0085	,0379
	ancien client, [3 à 4] renouvellements	,0212	,01183	,074	-,0020	,0443
	ancien client, [5 à 6] renouvellements	,0176	,01183	,137	-,0056	,0408
	ancien client, [7 à 9] renouvellements	,0168	,01183	,155	-,0064	,0400
	ancien client, [10 à 14] renouvellements	,0148	,01185	,211	-,0084	,0381
	ancien client, [15 à 20] renouvellements	,0012	,01216	,920	-,0226	,0250

* La différence de moyennes est significative au niveau ,05.

^a type de crédit = individuel

10.1.2 Analyse des résultats

Dans une analyse de variance significative, il y a trois données importantes : le df ou degré de liberté (ddl) = 7 dans notre cas, le résultat du F (=56,212 dans notre cas), et la valeur de p (Sig.) = 0,000 dans notre cas.

$F = \text{somme des carrés des écarts inter-échantillons} / k-1$ divisée par la somme des carrés des écarts de la régression /n-k, [Tribout, 2007], [Saporta, 2006], [Pupion, 2008]

$F = \frac{SCI/k-1}{SCE/n-k} \leftrightarrow F(k-1; n-k)$. Pour le test : k=8 ; n=9166 ; SCI=0,110 ; SCE=2,561.
On calcule $F_{Obs} = \frac{0,110/7}{2,561/9158} = 56,212 \leftrightarrow F_{0,95}(7; 9158)$

Les hypothèses du test sont :

$H_0 : \mu_1 = \mu_2 \dots = \mu_k$

H_1 : au moins 2 des espérances sont différentes

F a comme règle de décision au seuil de signification $\alpha (= 5\%)$:

c'est rejet de H_0 si $F > F_{(1-\alpha)}$. La région critique est de la forme $C_r = [f_{0,95}; +\infty[$

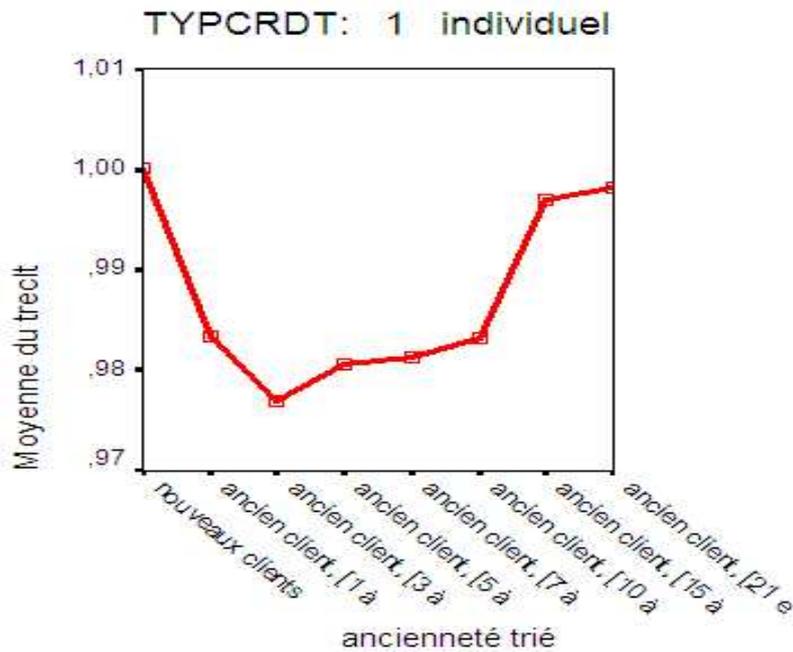
F_{Obs} appartient à la zone de rejet de l'hypothèse nulle.

Le F et le ddl permettent d'établir la valeur de p (Sig.). Cette valeur de p est l'erreur alpha, soit la probabilité ou le risque de commettre une erreur en déclarant qu'il existe une différence entre les huit groupes testés. C'est ce résultat qui permet de confirmer ou d'infirmer notre hypothèse alternative (H_1). Le seuil de signification est de 0,05. Si la valeur de Sig. ou valeur de p est supérieure à 0,05, alors on accepte l'hypothèse nulle et on conclue qu'il n'y a pas de différence significative entre les groupes. Si la valeur de Sig. est inférieure à 0,05, on doit faire deux choses :- d'abord rejeter l'hypothèse nulle et conclure qu'il y a une différence significative entre les huit groupes puis procéder à un test post-hoc afin de savoir laquelle des comparaisons de groupes, pris deux à deux, est significativement différente.

10.1.3 Diagramme de moyenne

Figure 10.3

FIG. 10.3 – Diagramme de moyenne



Ce diagramme montre que le taux de respect des échéances pour les clients aux "bons" taux n'a pas une évolution linéaire simple mais au contraire, la progression est contrastée selon l'ancienneté. En voyant cette image, on pourrait dire qu'il serait injuste de croire que les clients deviennent plus respectueux des échéances avec l'ancienneté. Car, le taux de respect des échéances n'augmente pas avec l'ancienneté. Globalement, c'est plutôt le contraire. Les clients deviennent moins respectueux des échéances avec l'ancienneté si on compare la situation des nouveaux clients (ceux qui sont à leur premier crédit) et celle des plus anciens clients actifs de la base des données. Ces derniers ont globalement un taux moyen de respect des échéances qui est inférieur à celui des premiers. Mais ce constat général n'est pas linéaire non plus. Les "bons" clients n'ont pas des comportements de non respect croissants vis-à-vis des échéances à mesure qu'ils prennent de l'ancienneté dans l'institution prêteuse. La situation est plus complexe. En fait, on peut constater trois phases dans l'évolution du taux de respect des échéances pour un "bon" client. Au premier crédit, le taux est quasiment à 100%. Puis à partir du deuxième crédit c'est-à-dire du premier renouvellement, le niveau de respect des échéances baisse mais pas suffisamment pour empêcher l'obtention de nouveaux renouvellements. Ainsi, pendant une période moyenne de dix renouvellements, le seuil de respect des échéances plafonne à un niveau relativement moyen pour la catégorie c'est-à-dire autour de 98%. Puis, à partir du dixième renouvellement, le seuil de respect des échéances remonte mais avec une progressivité faible.

En bref dans cette section

Elle a montré que les variables indépendantes de cette étude, tout en étant en rapport avec la variable dépendante, sont en relation les unes avec les autres. Cette section a servi à comparer les moyennes obtenues par des groupes de clients indépendants qui se différencient sur un facteur à savoir l'ancienneté. Les résultats conduisent à penser qu'il existe des effets entre les différents groupes de clients qui sont significativement différents les uns par rapport aux autres dans leur comportement vis-à-vis du respect des échéances.

Dans la section suivante, nous allons comparer les moyennes obtenues par des groupes de clients indépendants qui se différencient sur deux ou plusieurs facteurs. C'est le principe de l'analyse de variance multiples ou variances factorielles.

10.2 Analyse de variances multiples (MANOVA)

Cette section pouvait s'intituler "ANOVA à plusieurs facteurs". Dans la section précédente, l'analyse de la variance concernait un seul facteur. Mais, si plusieurs facteurs indépendants peuvent agir, nous devons utiliser une ANOVA à plusieurs facteurs qu'on appelle "MANOVA". Contrairement à l'ANOVA à 1 facteur, il faut proposer plusieurs H_0 . Une ANOVA à plusieurs facteurs évite de recourir à plusieurs ANOVA à 1 facteur pour tester la même chose. En plus, une ANOVA à plusieurs facteurs permet de tester les interactions entre les facteurs, [Tribout, 2007], [Saporta, 2006], [Pupion, 2008].

Cette section compare donc les moyennes obtenues par des groupes de clients indépendants qui se différencient sur deux ou plusieurs facteurs. En effet, en utilisant l'ancienneté comme variable de pondération, nous pouvons chercher à comprendre les facteurs qui influencent l'évolution du taux de respect des échéances pour les "bons" clients. Les facteurs (ou variables indépendantes) pris en compte sont : le genre, la recommandation et l'aval. C'est pour cela que nous faisons cette analyse des variances factorielles à effet simple (ou ANOVA factorielle simple). Concrètement, on utilise le modèle linéaire général "General Linear Model" (Univarié) avec SPSS.

Choix des données

Variable dépendante (VD) = taux de respect des échéances

Variables indépendantes ou facteurs (VI) = genre ; recommandation ; aval

Variable de pondération = ancienneté-trié

10.2.1 Principes et calculs

GLM - Univarié (précédemment appelée "Factoriel complet") fournit un modèle de régression et une analyse de la variance pour une variable dépendante par un ou plusieurs facteurs

ou variables indépendantes dites variables de critère. Les variables de critère divisent la population en groupes. Cette procédure de régression linéaire généralisée permet de tester les hypothèses nulles à propos des effets des autres variables indépendantes sur la moyenne de différents regroupements de la variable dépendante. Par ce modèle, on peut rechercher les interactions entre les facteurs ainsi que les effets des différents facteurs.

Généralement, on utilise l'analyse factorielle lorsqu'on souhaite examiner l'effet d'un facteur sur la variable dépendante ou l'effet d'un facteur à un niveau précis d'un autre facteur sur la variable dépendante. Ces préoccupations sont les deux motifs de cette section.

Pour ce test, les calculs ont été effectués de manière suivante. Variable dépendante : taux de respect des échéances de la clientèle ; facteurs : genre, aval, recommandation. On utilise ces variables car on désire connaître le niveau de l'effet d'une de ses variables sur les autres, par exemple entre le genre et l'aval, le genre et la recommandation ou entre la recommandation et l'aval ; Poids WLS = ancienneté-trié. L'ancienneté est la variable de pondération. Le "Modèle" du test est : l'interaction. On compare les effets simples (LSD : simples).

Dans les résultats nous montrons les facteurs inter-sujets, le test d'égalité des variances des erreurs de Levene, le test des effets inter-sujets et l'estimation des paramètres, les moyennes marginales estimées et enfin les diagrammes des profils.

10.2.2 Diagnostic des résultats

Les résultats des tests effectués sont proposés dans la figure 10.4. Les représentations des diagrammes des profils sont dans une figure à part.

FIG. 10.4 – GLM test Unuvarié

Analyse de variance univariée

Facteurs inter-sujets

		Etiquette de valeur	N
genre (sexe du client)	1	féminin	6076
	2	masculin	2982
recommandation	1	recommandé	8993
	2	non recommandé	65
avalisation	1	avalisés	7224
	2	non avalisés	1834

Test d'égalité des variances des erreurs de Levene

Variable dépendante: taux de respect des ech. de la clientèle

F	ddl1	ddl2	Signification
56,839	6	9051	,000

Teste l'hypothèse nulle que la variance des erreurs de la variable dépendante est égale sur les différents groupes.

- a. Plan : Intercept+GENRE * RECOMMAN+GENRE * AVAL+RECOMMAN * AVAL
- b. Régression des moindres carrés pondérés - Pondérés par ancienneté trié

Tests des effets inter-sujets

Variable dépendante: taux de respect des ech. de la clientèle

Source	Somme des carrés de type III	ddl	Moyenne des carrés	F	Signification	Eta au carré partiel
Modèle corrigé	,355 ^a	6	5,908E-02	100,741	,000	,063
Constante	198,872	1	198,872	339060,3	,000	,974
GENRE * RECOMMAN	4,366E-04	1	4,366E-04	,744	,388	,000
GENRE * AVAL	1,810E-03	1	1,810E-03	3,086	,079	,000
RECOMMAN * AVAL	1,750E-02	1	1,750E-02	29,844	,000	,003
GENRE * RECOMMAN * AVAL	,000	0	,	,	,	,000
Erreur	5,308	9051	5,865E-04			
Total	20170,108	9058				
Total corrigé	5,663	9057				

- a. R deux = ,063 (R deux ajusté = ,062)
- b. Régression des moindres carrés pondérés - Pondérés par ancienneté trié

Estimations des paramètres

Variable dépendante: taux de respect des ech. de la clientèle

Paramètre	B	Erreur standard	t	Sig	Intervalle de confiance à 95%		Eta au carré partiel
					Borne inférieure	Limite supérieure	
Constante	,997	,008	123,493	,000	,981	1,013	,628
[GENRE=1] * [RECOMMAN=1]	-3,24E-02	,008	-3,939	,000	-4,847E-02	-1,626E-02	,002
[GENRE=1] * [RECOMMAN=2]	-9,29E-03	,005	-1,870	,062	-1,903E-02	4,494E-04	,000
[GENRE=2] * [RECOMMAN=1]	-1,90E-02	,008	-2,352	,019	-3,485E-02	-3,168E-03	,001
[GENRE=2] * [RECOMMAN=2]	0 ^a	,	,	,	,	,	,
[GENRE=1] * [AVAL=1]	-2,93E-02	,008	-3,464	,001	-4,587E-02	-1,271E-02	,001
[GENRE=1] * [AVAL=2]	0 ^a	,	,	,	,	,	,
[GENRE=2] * [AVAL=1]	-3,23E-02	,008	-3,892	,000	-4,850E-02	-1,601E-02	,002
[GENRE=2] * [AVAL=2]	0 ^a	,	,	,	,	,	,
[RECOMMAN=1] * [AVAL=1]	4,543E-02	,008	5,463	,000	2,913E-02	6,173E-02	,003
[RECOMMAN=1] * [AVAL=2]	0 ^a	,	,	,	,	,	,
[RECOMMAN=2] * [AVAL=1]	0 ^a	,	,	,	,	,	,
[RECOMMAN=2] * [AVAL=2]	0 ^a	,	,	,	,	,	,
[GENRE=1] * [RECOMMAN=1] * [AVAL=1]	0 ^a	,	,	,	,	,	,
[GENRE=1] * [RECOMMAN=1] * [AVAL=2]	0 ^a	,	,	,	,	,	,
[GENRE=1] * [RECOMMAN=2] * [AVAL=1]	0 ^a	,	,	,	,	,	,
[GENRE=2] * [RECOMMAN=1] * [AVAL=1]	0 ^a	,	,	,	,	,	,
[GENRE=2] * [RECOMMAN=1] * [AVAL=2]	0 ^a	,	,	,	,	,	,
[GENRE=2] * [RECOMMAN=2] * [AVAL=1]	0 ^a	,	,	,	,	,	,
[GENRE=2] * [RECOMMAN=2] * [AVAL=2]	0 ^a	,	,	,	,	,	,

- a. Ce paramètre est mis à zéro car il est redondant.
- b. Régression des moindres carrés pondérés - Pondérés par ancienneté trié

Moyennes marginales estimées

genre (sexe du client)*recommandation

Estimations^a

Variable dépendante: taux de respect des ech. de la clientèle

genre (sexe du client)	recommandation	Moyenne	Erreur standard	Intervalle de confiance à 95%	
				Borne inférieure	Limite supérieure
féminin	recommandé	,973	,001	,971	,974
	non recommandé	,973	,006	,961	,985
masculin	recommandé	,984	,000	,984	,985
	non recommandé	,981	,004	,973	,989

a. Régression des moindres carrés pondérés - Pondérés par ancienneté trié

aval*genre (sexe du client)

Estimations^a

Variable dépendante: taux de respect des ech. de la clientèle

avalisation	genre (sexe du client)	Moyenne	Erreur standard	Intervalle de confiance à 95%	
				Borne inférieure	Limite supérieure
avalisés	féminin	,970	,002	,965	,974
	masculin	,978	,001	,976	,980
non avalisés	féminin	,976	,005	,966	,986
	masculin	,987	,004	,979	,995

a. Régression des moindres carrés pondérés - Pondérés par ancienneté trié

aval*recommandation

Estimations^a

Variable dépendante: taux de respect des ech. de la clientèle

avalisation	recommandation	Moyenne	Erreur standard	Intervalle de confiance à 95%	
				Borne inférieure	Limite supérieure
avalisés	recommandé	,986	,000	,985	,986
	non recommandé	,961	,002	,957	,966
non avalisés	recommandé	,971	,001	,970	,973
	non recommandé	,992	,008	,976	1,009

a. Régression des moindres carrés pondérés - Pondérés par ancienneté trié

aval*genre (sexe du client)*recommandation

Estimations^a

Variable dépendante: taux de respect des ech. de la clientèle

avalisation	genre (sexe du client)	recommandation	Moyenne	Erreur standard	Intervalle de confiance à 95%	
					Borne inférieure	Limite supérieure
avalisés	féminin	recommandé	,981	,000	,980	,981
		non recommandé	,958	,004	,950	,967
	masculin	recommandé	,991	,001	,990	,992
		non recommandé	,965	,002	,961	,968
non avalisés	féminin	recommandé	,965	,002	,962	,968
		non recommandé	,988	,009	,969	1,006
	masculin	recommandé	,978	,000	,977	,979
		non recommandé	,997	,008	,981	1,013

a. Régression des moindres carrés pondérés - Pondérés par ancienneté trié

Test des effets inter-sujets

Les trois variables sont trois facteurs inter-sujets.

Le test des effets inter-sujets est significatif pour le couple de variable : recommandation*aval. Il n'est pas significatif pour les couples de variables : genre*recommandation et genre*aval. Alors, on doit voir le détail dans le tableau d'estimation des paramètres. Pour faciliter la lecture de ce tableau, nous précisons les règles de codage des données¹.

Estimation des paramètres inter-sujets

Dans le tableau des estimations des paramètres, le test est significatif pour le [genre = 1]*[recomman = 1] c'est-à-dire pour les femmes qui sont recommandées par des tiers. C'est-à-dire que le fait d'être recommandé a un effet qui influence le comportement d'une femme vis-à-vis du respect des échéances. Ou inversement, le fait d'être une femme influence la connaissance générale qu'on peut avoir sur le comportement des clients qui sont recommandés par des tiers vis-à-vis des échéances. De même, le test est significatif pour [genre = 2]*[recomman = 1] c'est-à-dire pour les hommes qui sont recommandés par des tiers. Donc, il y a un effet du fait d'être recommandé qui influence le comportement des clients de sexe masculin vis-à-vis du respect des échéances. Ou inversement, on peut dire qu'il y a un effet du genre masculin sur le comportement de respect des échéances pour les clients qui sont recommandés par des tiers. De la même manière, le test est significatif pour les femmes qui sont avalisées, pour les hommes qui sont avalisés et pour les clients qui sont à la fois recommandés et avalisés. C'est-à-dire que pour un client, qu'il soit une femme ou un homme, le fait d'être avalisé a un effet sur son comportement de respect des échéances. De même, le fait d'être avalisé produit un effet en matière de respect des échéances sur les clients qui sont recommandés.

Par contre, pour les clients de sexe féminin et qui ne sont pas recommandés, le test est calculé mais il n'est pas significatif. La signification est de 0,062 (soit 6,2%, ce qui est supérieur au seuil de 5% autorisé par le modèle). Pour tous les autres cas, c'est-à-dire pour les hommes non recommandés, les clients non avalisés (féminins comme masculins), pour les clients qui ne sont pas recommandés et qui sont avalisés, pour tous ces clients, le test n'est pas calculé. De même, lorsqu'on "injecte" la variable genre dans l'analyse conjointe des variables recommandation*aval, c'est-à-dire lorsqu'on croise les trois variables (genre*recommandation*aval) simultanément, aucun paramètre n'est testé.

¹Chacune des variables indépendantes a deux modalités. La variable genre a été codée : 1 = féminin, 2 = masculin. La variable recommandation a été codée : 1 = recommandé, 2 = non recommandé. Enfin, la variable avalisation a été codée : 1 = avalisé, 2 = non avalisé ; en précisant que les clients non avalisés proposent leur propre garantie. La garantie étant exigée en tant que condition d'accès au crédit individuel. Tous les clients pris en compte dans cette analyse sont en crédits individuels et ont un taux moyen de respect des échéances au moins égal à 95% quelque soit leur ancienneté.

10.2.3 Analyse des effets

Dans cette section, nous présentons les diagrammes de profils qui présentent les différents effets montrés par le modèle. En effet, après avoir montré l'aspect statistique de l'ANOVA factorielle à savoir le test des effets inter-sujets et l'estimation des paramètres, il est nécessaire de proposer la représentation graphique de ses effets. Comme nous l'avons observé dans les aspects statistiques, les trois variables indépendantes (genre, recommandation et aval) ne peuvent pas agir tous les trois simultanément. Le test n'étant pas calculé pour cause de redondance. En revanche, plusieurs résultats significatifs ont été obtenus lorsque les facteurs sont pris deux à deux. Or, les variables sont indépendantes les unes des autres. Donc, les effets d'au moins deux facteurs sont possibles. Certains effets des trois facteurs ont même été représentés. Les effets qui ont été représentés sont en particulier les effets d'un facteur sur la variable dépendante ou les effets d'un facteur sur un autre facteur et sur la variable dépendante. Pour représenter ses effets, on utilise deux axes ; un axe inférieur pour représenter (d'un bout à l'autre) les deux modalités du premier facteur et ; deux lignes différentes (nous les proposons en couleur rouge et verte) pour représenter les valeurs des variables indépendantes en fonction de chacun des niveaux de l'autre facteur. Les différentes représentations possibles sont présentées dans la figure 10.5.

FIG. 10.5 – Effets des facteurs : les diagrammes de profils

Diagramme des effets d'interaction

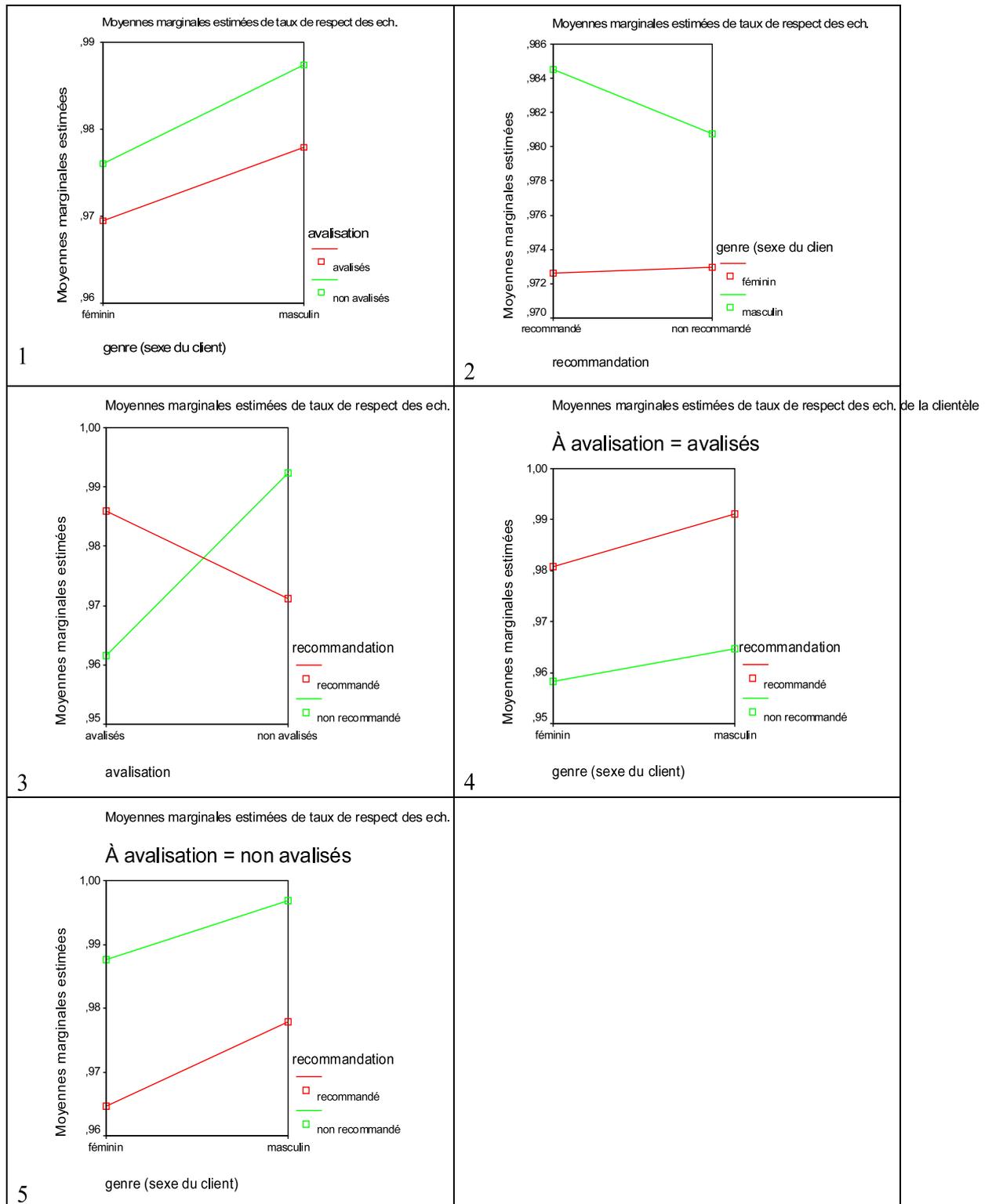
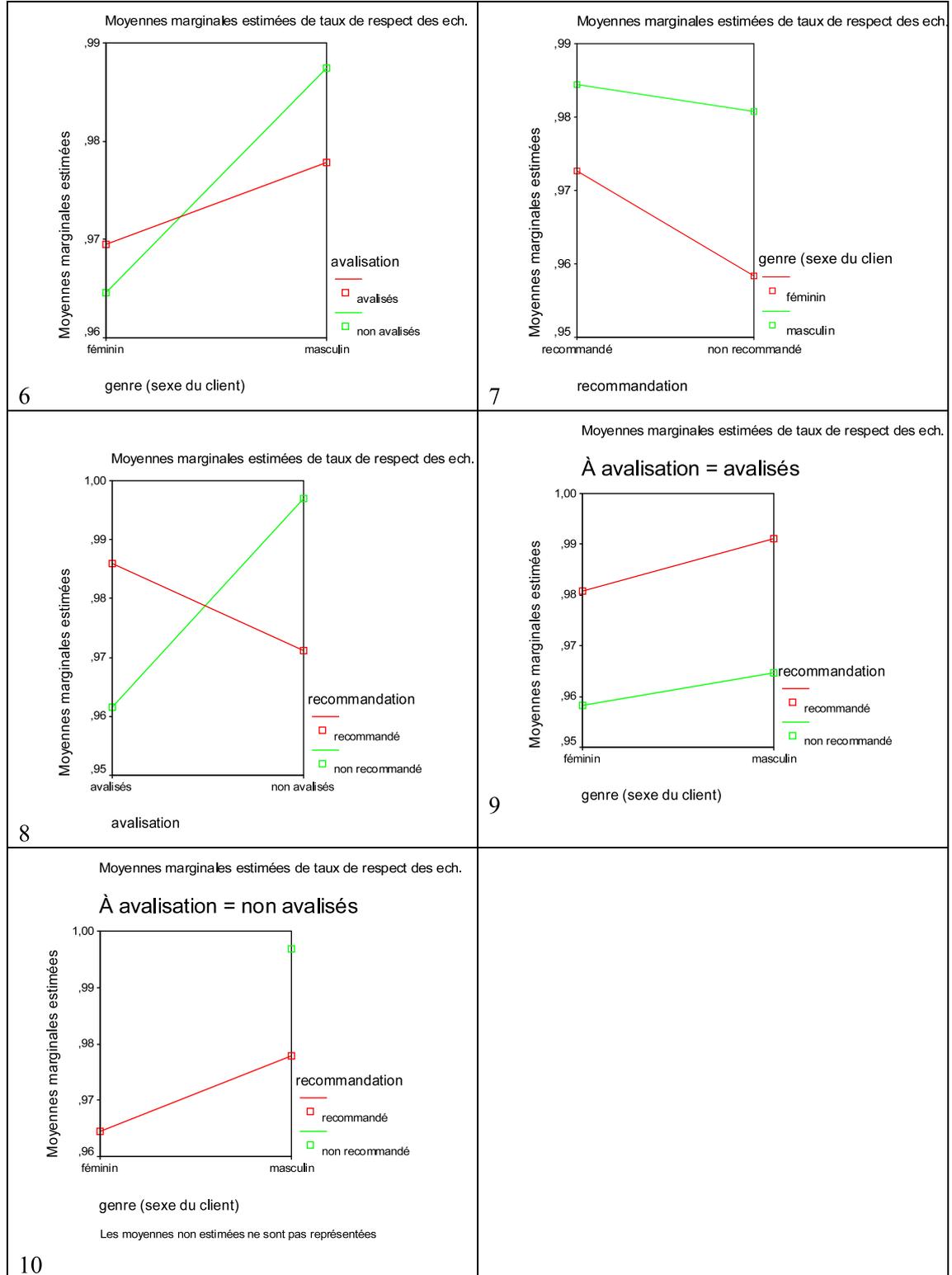


Diagramme des effets principaux



Les diagrammes qui sont numérotés de 1 à 5 sont obtenus lors de l'exécution du modèle avec effet d'interaction. Les diagrammes 6 à 10 sont générés par le modèle avec effets principaux. La différence c'est que pour les effets d'interaction, on peut avoir les effets simples. Les effets principaux et les effets d'interaction sont générés automatiquement par le modèle GLM-univarié. Pour les effets simples, le modèle GML-univarié ne les génère pas automatiquement. On doit les produire "à la main". Pour cela, on utilise la fonction "syntaxe" de SPSS, et on écrit les commandes nécessaires pour exécuter les requêtes prévues.

Avant de faire la lecture de ses graphiques, il est bon de définir les effets rencontrés sur les graphiques, c'est-à-dire les effets principaux, les effets d'interaction et les effets simples. Globalement, on dit qu'il y a un effet lorsque les lignes (en rouge et en vert) dévient de leur axe horizontal.

Effets principaux

Un effet est dit principal lorsque c'est l'effet d'une variable indépendante sur une variable dépendante. En d'autres termes, on compare les données associées à chacune des modalités d'une variable indépendante. Mais s'il y a plusieurs variables indépendantes pour une variable dépendante, comme c'est notre cas (avec les variables indépendantes : genre, recommandation et aval) ; dans ce cas, on examine l'effet principal de l'une et l'effet principal de l'autre des variables indépendantes indépendamment l'une de l'autre. Puis on examine l'effet dit d'interaction c'est-à-dire l'effet d'action de deux variables indépendantes l'une sur l'autre sur la variable dépendante étudiée.

Les effets principaux concernent en particulier les diagrammes 6 à 10. L'effet est observé lorsque les lignes (en rouge et en vert) dévient de leur axe horizontal. Globalement, on dit qu'il y a un effet principal lorsque l'effet d'un facteur est indépendant d'un autre. Chaque diagramme présente le résultat du test de la variance factorielle pour une hypothèse impliquant directement les variables qui figurent dans le diagramme en question. Par exemple, pour le diagramme 6, les hypothèses testées sont :

- premièrement : si genre = 1 (c'est-à-dire : féminin), la moyenne marginale estimée du treclt (qui est la variable dépendante) sera plus élevée lorsque avalisation = 1 (c'est-à-dire : avalisé) que lorsque avalisation = 2 (c'est-à-dire : non avalisé)

- deuxièmement : si genre = 2 (c'est-à-dire masculin), la moyenne marginale estimée du treclt (qui est la variable dépendante) sera plus élevée lorsque avalisation = 1 (c'est-à-dire : avalisé) que lorsque avalisation = 2 (c'est-à-dire : non avalisés).

Pour les diagrammes 6 à 10, nous faisons ci-dessous les lectures suivantes pour décrire les effets :

un effet principal du facteur *genre* plus une interaction (qualitative) : diagramme 6.

Il y a un effet principal. C'est l'effet du facteur genre. Car, les comportements vis-à-vis

du respect des échéances sont positifs pour les clients de sexe masculin et négatif pour les clients de sexe féminin, quelque soit l'évolution du facteur avalisation. Donc, le facteur genre a un effet sur la variable dépendante indépendamment du facteur avalisation. C'est un effet principal. En plus, il y a une interaction (qualitative). L'interaction est matérialisée par le croisement formel des deux lignes².

Il y a interaction car si nous observons le comportement positif vis-à-vis du respect des échéances (marqué par l'élévation du seuil de la moyenne marginale estimée du trecht), on constate qu'un client de sexe féminin à un score plus élevé qu'il est avalisé et inversement, il a un score moins élevé qu'il est non avalisé. Ce raisonnement est complètement inversé si on considère la situation d'un client de sexe masculin. Il y a donc une interaction entre les facteurs genre et aval qui exercent des effets réciproques l'un sur l'autre. Ici, on parle d'interaction qualitative puisque la direction des effets d'un facteur sur la variable dépendante est différente selon les niveaux de l'autre facteur.

un effet principal du facteur *genre* plus un effet d'interaction (quantitative) : diagramme 7.

Dans le diagramme 7, il y a un effet principal : c'est l'effet du facteur genre. Les clients de sexe masculin obtiennent un score plus élevé en matière de respect des échéances quelque soit le statut par rapport la recommandation. Il y a là, un effet principal du facteur genre. En plus il y a une interaction entre les deux facteurs. Il s'agit d'une d'interaction quantitative car la direction des effets d'un facteur sur l'autre varie en amplitude selon les niveaux de l'autre facteur mais ne conserve pas la même direction.

pas d'effet principal mais une interaction (qualitative) : diagramme 8.

Dans ce diagramme, il n'y a pas d'effet principal. Cette absence d'effet principal masque une interaction. Effectivement, les clients qui sont recommandés obtiennent un score plus positif pour le respect des échéances que s'ils sont avalisés ; et inversement, les clients qui ne sont pas recommandés obtiennent un score plus positif pour le respect des échéances que s'ils ne sont pas avalisés. Il y a donc une interaction entre les deux effets : recommandation et aval sans effet principal. On parle ici d'interaction qualitative car la direction des effets d'un facteur sur la variable dépendante est différente selon les niveaux de l'autre facteur.

deux effets principaux des facteurs *genre* et *recommandation* : diagramme 9.

Dans ce diagramme, il y a deux effets principaux. Un effet principal du facteur recommandation et un effet principal du facteur genre.

pas d'interprétation : pour le diagramme 10.

Il y a une seule modalité du facteur recommandation qui est représentée concernant les clients qui sont recommandés. On aperçoit que ce sont les clients de sexe masculin qui

²Il n'est pas nécessaire que le croisement de produise formellement sur le diagramme pour parler d'interaction.

ont un score positif en matière de respect des échéances.

Effets d'interactions

On parle d'effet d'interaction lorsque l'effet d'un facteur sur la variable dépendante est différent en fonction des modalités d'un autre facteur.

Pour les diagrammes 1 à 5, nous faisons ci-dessous les lectures suivantes pour décrire les effets. On a :

- un effet principal : dans le diagramme 1, c'est l'effet du facteur *aval* ; et dans le diagramme 2, l'effet du facteur *genre*.
- pas d'effet, mais une interaction qualitative : dans le diagramme 3 ;
- deux effets principaux : dans le diagramme 4, il y a un effet du facteur *recommandation* et un effet du facteur *genre*. ; dans le diagramme 5, il y a les effets des facteurs *recommandation* et *genre*.

Effets simples

On décompose l'effet d'interaction en effets simples. L'effet simple est l'effet d'une variable indépendante sur la variable dépendante étudiée sous l'une des modalités d'une autre variable indépendante. L'effet simple peut être aussi la décomposition en comparaison deux à deux d'un effet principal d'une variable indépendante comprenant plus de deux modalités. Ce n'est pas notre cas.

Pour exécuter les effets simples, nous avons écrit la syntaxe SPSS suivante :

```
MANOVA treclt BY genre(1,2) recomman(1,2) aval(1,2) /DESIGN = genre WITHIN recomman (1) genre WITHIN recomman (2) genre WITHIN aval (1) genre WITHIN aval (2) recomman WITHIN aval (1) recomman WITHIN aval (2)
```

Description :

treclt = taux de respect des échéances de la clientèle (c'est la variable dépendante). genre, recomman, aval sont les trois variables indépendantes, chacune est accompagnée par ses deux modalités : 1 ou 2.

lorsqu'on exécute la syntaxe ci-dessus, on obtient les résultats que présente la figure 10.6.

FIG. 10.6 – Analyse de variance : effets simples

```

***** Analysis of Variance -- design 1 *****

Tests of Significance for TRECLI using UNIQUE sums of squares
Source of Variation      SS      DF      MS      F      Sig of F

WITHIN+RESIDUAL          2,54    9160     ,00
GENRE WITHIN RECOMMA     ,00      1     ,00     15,28     ,000
N(1)
GENRE WITHIN RECOMMA     ,00      1     ,00      ,10     ,749
N(2)
GENRE WITHIN AVAL(1)     ,02      1     ,02     82,50     ,000
GENRE WITHIN AVAL(2)     ,00      0      ,      ,      ,
RECOMMAN WITHIN AVAL     ,02      1     ,02     71,98     ,000
(1)
RECOMMAN WITHIN AVAL     ,00      1     ,00     10,69     ,001
(2)

(Model)                   ,13      5     ,03     94,37     ,000
(Total)                   2,67    9165     ,00

R-Squared =                ,049
Adjusted R-Squared =      ,048

```

Comme il y a deux modalités des facteurs recommandation et aval, le logiciel propose deux lignes pour chaque facteur. Pour la modalité codée 1 du facteur recommandation (c'est-à-dire pour les clients qui sont recommandés par des tiers), il y a un effet simple qui est significatif pour le facteur genre : ("sig of F" = 0,000). Ce qui veut dire que parmi les clients qui sont recommandés, l'effet genre (ou sexe) est significatif sur la probabilité de respect des échéances pour ces clients. Le test ne dit pas le sens de l'effet. Parallèlement, pour la modalité codée 2 du facteur recommandation (c'est-à-dire pour les clients qui ne sont pas recommandés par des tiers), l'effet simple du facteur genre n'est pas significatif : ("sig of F" = 0,749). Ce qui est largement supérieur au seuil de 5% d'erreur admise pour ce test. Cela voudrait dire que parmi les clients qui n'ont pas été recommandés par des tiers, l'effet genre (ou sexe) n'est pas significatif sur la probabilité de respect des échéances pour ces clients. Il est inutile qu'on s'intéresse au sens d'un effet qui n'est de toute façon pas significatif.

Pour la modalité codée 1 du facteur aval (c'est-à-dire pour les clients qui sont avalisés par des tiers), il y a un effet simple qui est significatif pour le facteur genre : ("sig of F" = 0,000). Ce qui veut dire que parmi les clients qui sont avalisés, l'effet genre (ou sexe) est significatif sur la probabilité de respect des échéances pour ces clients. C'est tout ce qu'on sait. Le test ne dit pas le sens de l'effet. Mais, l'effet simple du facteur genre sur les clients qui ne sont pas avalisés n'a pas été calculé par le test. En conséquence, on ne peut tirer aucune conclusion quant à l'effet genre (ou sexe) sur la probabilité de respect des échéances pour les clients non avalisés.

Les tests du facteur recommandation sur les clients repartis dans les deux modalités du facteur aval sont significatifs. Cela veut dire que parmi les clients qui sont avalisés, l'effet recommandation est significatif sur la probabilité de respect des échéances pour ces clients. Le test n'indique pas le sens de l'effet.

Si le test indique les facteurs qui ont des effets significatifs, il n'indique pas le sens (ou la direction de ses effets). Est-ce positif ou négatif? On ne saurait répondre. De plus, le test analyse un facteur sur les modalités des autres facteurs. Or, pour chaque facteur, nous avons défini deux modalités. Donc, lorsque le test indique qu'un facteur (par exemple le facteur genre) a un effet significatif sur une modalité d'un autre facteur (par exemple la modalité 1 du facteur recommandation), le test ne dit pas la proportion des clients sur chacune des modalités du facteur genre. Le test donne un résultat "général" de l'effet du facteur genre et dit si cet effet est significatif ou non significatif. Dans notre cas il est significatif. Il est donc important (pour les besoins de décisions stratégiques en matière de gestion des portefeuilles de crédits) de savoir les proportions de clients dans chaque modalité du facteur genre d'une part et du facteur recommandation d'autre part.

Pour résumer

Dans cette section, les résultats des tests effectués concluent qu'il existe des effets significatifs des facteurs genre, recommandation et aval entre différents groupes de clients dans leur comportement vis-à-vis du respect des échéances. Nous avons montré et testé notamment des effets principaux, des effets d'interaction et des effets simples de ses facteurs. Les facteurs analysés ici sont pondérés par l'ancienneté mais nous n'avons pas pris en compte d'autres facteurs et ou variables indépendantes potentiels qui pourraient influencer l'effet des facteurs analysés. Ces autres variables potentielles sont appelées co-variables.

Dans le chapitre suivant, nous allons montrer cette relation potentielle entre la variable d'étude (taux de respect des échéances) et ces autres variables dans le cadre des analyses de régression linéaire.

Chapitre 11

Analyse des régressions linéaires

Ce chapitre a pour but d'établir dans un modèle économique, l'estimation (ou ajustement) de la relation linéaire entre la variable dépendante taux de respect des échéances et une ou plusieurs variables explicatives ou prédicteurs.

11.1 Principes

L'ajustement linéaire ou régression consiste à rechercher la "droite des moindres carrés" de type $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ qui passe le plus près possible de toutes les observations dans la population.

Y = variable dépendante (variable à expliquer) de type continu. Pour nous, la variable : taux de respect des échéances de la clientèle (treclt).

X_i = variables indépendantes (ou explicatives), de type continu. Ici, ce sont : encours moyen (encm), fonds et dépôts moyens (fondep) et pénalités moyennes (penalim).

α = constante. C'est la valeur de Y quand X est = 0

β_i = coefficients de la droite de régression.

SPSS permet de calculer des estimations de ses coefficients β_i et α pour l'échantillon. On cherche à vérifier :

- que la relation linéaire testée est significative (c'est-à-dire que les coefficients β_i de la droite sont significativement différents de 0. Cela montre que la relation linéaire n'est pas due au hasard.

- que la droite de régression $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ résume bien l'ensemble des observations, c'est-à-dire que la part de variance de la variable à expliquer Y qui est résumée dans la droite est élevée.

Par rapport à la matrice des corrélations évoquée précédemment (voir le paragraphe IV), la régression teste non seulement l'existence d'une corrélation, mais en plus, elle teste la nature précise de cette relation. D'autre part, si la relation trouvée est significative et que le pourcentage de variance expliquée est élevé, il est possible et intéressant de faire des estimations

de la variable expliquée à partir des variables explicatives. En outre, contrairement au test de corrélation qui prend des couples de variables (deux à deux), il est possible de faire rentrer plusieurs variables simultanément dans une procédure de régression dans le cadre d'une régression multiple.

Dans un premier temps, voyons le cas d'une régression simple (ou régression bivariée).

11.2 Régression linéaire simple

11.2.1 Données et calculs

Les données prises en compte sont :

- variable dépendante (Y) = *treclt*
- variable indépendante (X) = *encm*

Dans le modèle linéaire, la relation bivariee entre Y et X est décrite (ou résumée) par l'équation : $Y = \alpha + \beta X$.

Si on reprend notre exemple, on pose donc : $treclt = \alpha + \beta.encm$.

Selon cette équation, le taux de respect des échéances dépend linéairement de l'encours moyen de crédit reçu. Ce taux varie d'un terme constant représenté par le coefficient β pour chaque encours de crédit supplémentaire. Autrement dit, la relation entre Y et X est modélisée par une droite dont la pente est β . Cette équation permet aussi de calculer un taux de respect des échéances prédit (ou estimé) pour chaque encours de crédit. Ce taux prédit (ou estimé) peut être plus ou moins différent du taux effectivement observé au même encours de crédit. En fait, le raisonnement sur les paramètres α et β suppose que les valeurs de ceux-ci sont directement accessibles et mesurables. Or, ces paramètres α et β sont ceux de la population des clients individuels (les 21146 clients). Leurs valeurs sont inconnues et doivent être estimées à partir des données observées dans l'échantillon. Pour les calculs, nous avons pris le groupe témoin (= 9166 clients).

La méthode des moindres carrés est couramment utilisée pour estimer les paramètres de l'équation de régression linéaire, [Tribout, 2007], [Saporta, 2006], [Pupion, 2008]. L'équation des moindres carrés a dans le cas de la régression bivariée, la forme :

$$\hat{Y} = a + b.x$$

Alors, a et b sont les coefficients de l'équation de régression tels qu'ils sont calculés à partir des données observées (de l'échantillon). De manière conventionnelle, on écrit \hat{Y} afin d'indiquer qu'il s'agit de la valeur prédite de Y et non de sa valeur réellement observée. Les coefficients a et b sont obtenus par les formules suivants :

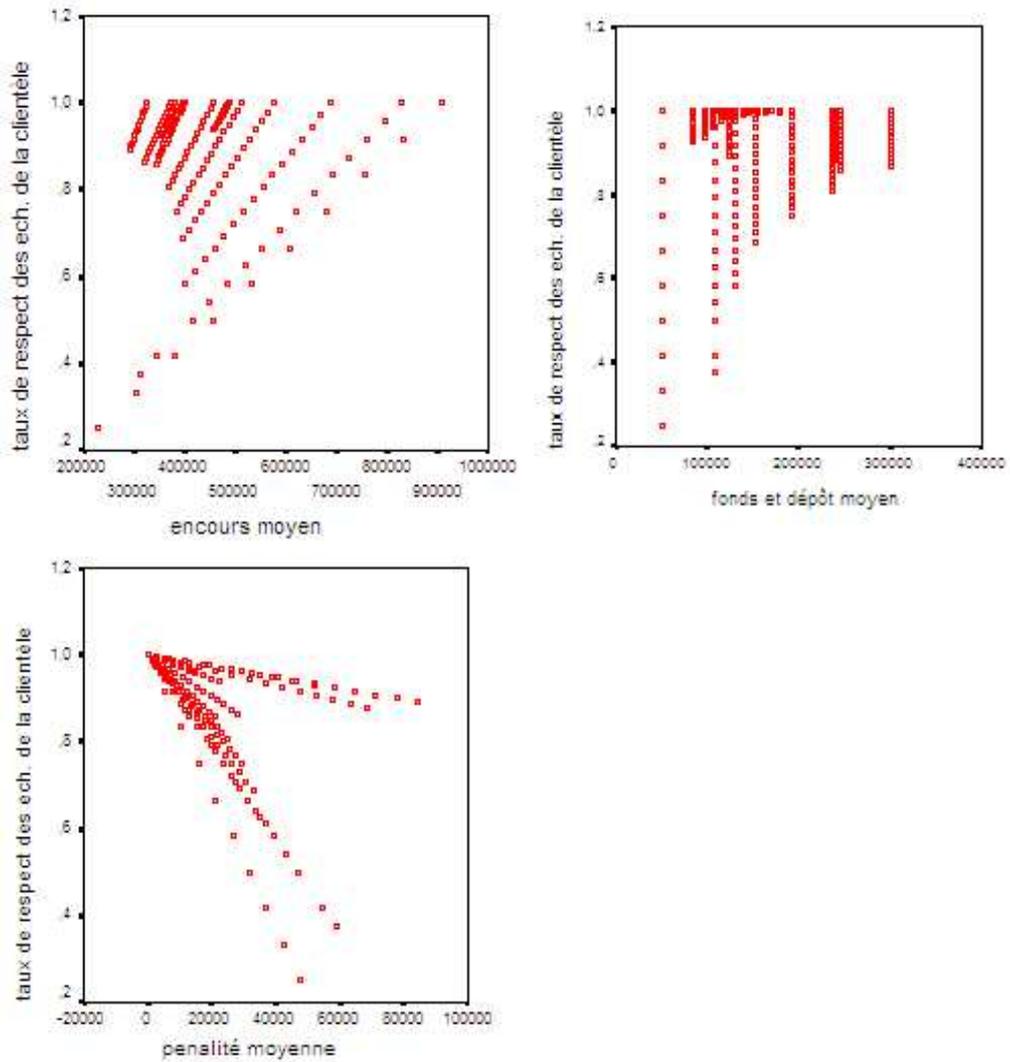
$$b = \frac{COV(X,Y)}{VAR(X)} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} \quad (11.1)$$

et

$$a = \bar{Y} - b.\bar{X} \quad (11.2)$$

Pour décrire la relation entre le taux de respect des échéances et l'encours moyen, on peut commencer par faire un diagramme de dispersion (scatterplot) avec l'encours moyen en abscisse et le taux de respect des échéances en ordonnée. Ce diagramme est utile pour déterminer si la relation entre Y et X est au moins approximativement linéaire, figure 11.1

FIG. 11.1 – Diagramme de dispersion



On observe une tendance à l'augmentation du taux de respect des échéances à mesure que l'encours moyen augmente. C'est le constat inverse pour les pénalités de retard et les fonds et dépôts de garanties. On constate sur tendance à la diminution des taux de respect des échéances à mesure que les montants des fonds et dépôts augmentent et à mesure que les pénalités moyennes augmentent.

A la demande, le logiciel SPSS calcul les paramètres a et b vus plus haut. Sans demande additionnelle, SPSS fournit de manière standard les trois tableaux de résultats suivants : Récapitulatif du modèle, ANOVA et Coefficients. Les résultats des calculs effectués sont présentés dans la figure 11.2.

FIG. 11.2 – Régression linéaire simple

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,143 ^a	,020	,020	,01690

	Moyenne	Ecart-type	N
treclt	,9810	,01707	9166
encom	667876,3252	160803,04371	9166

a. Valeurs prédites : (constantes), encours moyen
b. Variable dépendante : taux de respect des ech. de la clientèle

Modèle		Somme des carrés	ddl	Carré moyen	F	Signification
1	Régression	,054	1	,054	190,801	,000 ^a
	Résidu	2,816	9164	,000		
	Total	2,871	9165			

a. Valeurs prédites : (constantes), encours moyen
b. Variable dépendante : taux de respect des ech. de la clientèle

Modèle	Coefficients non standardisés		Coefficients standardisés	t	Sig	Intervalle de confiance à 95% de B	
	B	Erreur standard	Bêta			Borne inférieure	Borne supérieure
1 (constante)	,973	,001		1526,187	,000	,971	,974
encours moyen	1,615E-08	,000	,143	13,806	,000	,000	,000

a. Variable dépendante : taux de respect des ech. de la clientèle

11.2.2 Analyse des résultats

Tableau "Récapitulatif du modèle"

R = coefficient de corrélation. Il a été calculé précédemment (paragraphe IV)

R-deux = coefficient de détermination. C'est une mesure de la qualité du modèle. Il me-

sure la différence entre deux grandeurs. D'un côté, il utilise la somme des carrés des erreurs de prédiction produits par une équation des moindres carrés sans aucune variable indépendante, avec seulement la constante, qui est alors égale à la moyenne de y (SCE) tel que : $SCE = \Sigma(\hat{Y}_i - \bar{Y})^2$. De l'autre côté, il utilise la somme des carrés des erreurs de prédiction produits par l'équation des moindres carrés sous inspection (CSR) tel que $SCR = \Sigma(Y_i - \hat{Y}_i)^2$. La somme des écarts totaux est SCT telle que $SCT = SCE + SCR$. La différence entre SCE et SCR exprimée sous forme de proportion renvoie à l'amélioration de la prédiction due au fait qu'on tient compte de X . Exprimée en formule, cela donne : $R\text{-deux} = (E1-E2) / E1$. Ou, $R\text{-deux} = SCE/SCT$. Seulement 2% de la regression du treclt est dû à l'encm.

Tableau "ANOVA"

Somme des carrés = somme des erreurs de prédiction ; (évoquées ci-dessus)

F = test de H_0 . C'est le résultat d'un test de Fischer-Snedecor $F(n_1; n_2)$ selon lequel l'ensemble des coefficients des X sont égales à zéro. H_0 = absence de relation entre treclt (ou y) et encm (ou x), auquel cas le modèle n'est tout simplement pas " meilleur " que celui où ne figure aucun X . L'estimation de la variable de décision du test de Fischer = $F = \frac{SCE}{SCR/n-2} = \frac{Rdeux}{(1-Rdeux)/n-2} \mapsto F(1; n-2)$. Ici, F appartient à la zone de rejet de l'hypothèse nulle ; donc on rejette H_0^1 .

ddl = degré de liberté. Il est utilisé avec F pour calculer la P value (ou signification avec SPSS)

Signification = niveau de satisfaction du modèle

Tableau "Coefficients"

constante = valeur de a (= 0,973)

B = valeur du coefficient (b) dans l'équation de régression ; $b = 1,515.(10)^{-8}$. Le signe de (b) donne le sens de la relation entre treclt et encm. $b > 0$, signifie que treclt croît quand encm croît dans le même sens, positivement ou négativement. si b était négatif ($b < 0$), on dirait que treclt décroît quand encm croît ou inversement treclt croît quand encm décroît. Enfin, si b était nul ($b=0$), on dirait qu'il y a absence de relation entre treclt et encm.

Beta = valeur du coefficient b de l'équation de régression lorsqu'on standardise les valeurs de x (encm) et de y (treclt). la relation entre x et y est dès lors exprimée en termes d'écart type et non plus par rapport aux échelles originelles. On lit " de combien varie treclt en termes d'écart type lorsque encm augmente d'un écart type ". Dans le cas d'une régression bivariée (comme c'est le cas ici), Beta est égale au coefficient de Corrélation (R) indiquée dans le tableau " Récapitulatif du modèle ".

¹Le test de Fischer-Snedecor est présenté et appliqué dans l'analyse de la covariance, chapitre 12

$t = b / \text{erreur standard}$. Cette valeur renvoie à un test de t (avec un degré de liberté) selon lequel la valeur du coefficient β dans la population est égale à 0. Autrement dit, on teste l'hypothèse d'indépendance selon laquelle il y a une absence de relation entre encm et treclt dans la population. Les tests de Student et de Fischer conduisent à des résultats identiques dans le cas de la régression simple. On remarque que la valeur de f ($=190,601$) est égale à la valeur de t ($=13,806$) élevée au carré : $190,601 = (13,806)^2$

Signification = valeur p . Elle donne le résultat du test de t .

Intervalle de confiance = borne inférieure et la borne supérieure entre lesquelles se trouve, selon toute probabilité (à 99%), la valeur du coefficient β dans la population. Avec les valeurs de l'intervalle de confiance, on peut aussi calculer un éventail de valeurs prédites pour treclt dans la population.

Résumé du modèle

$a = 0,973$ et $b = 1,515.(10)^{-8}$, d'où : $\hat{Y} = a + b.x$ devient, en remplaçant \hat{Y} par treclt et x par encm :

$$\widehat{\text{treclt}} = 0,973 + 1,515.(10)^{-8}.\text{encm}$$

Exemple avec la variable : fonds de dépôt moyen

$$\widehat{\text{treclt}} = 0,984 - 1,723.(10)^{-8}.\text{fondepm}$$

Voir la figure 11.3 ci-dessous.

FIG. 11.3 – Régression linéaire simple avec la VI : fondep

Récapitulatif du modèle^a

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,063 ^a	,004	,004	,01704

- a. Valeurs prédites : (constantes), fonds et dépôt moyen
 b. Variable dépendante : taux de respect des éch. de la clientèle

ANOVA^b

Modèle		Somme des carrés	ddl	Carré moyen	F	Signification
1	Régression	,010	1	,010	36,096	,000 ^a
	Résidu	2,660	9164	,000		
	Total	2,671	9165			

- a. Valeurs prédites : (constantes), fonds et dépôt moyen
 b. Variable dépendante : taux de respect des éch. de la clientèle

Coefficients^a

Modèle	Coefficients non standardisés		Coefficients standardisés	t	Sig.	Intervalle de confiance à 95% de B	
	B	Erreur standard				Borne inf.	Borne sup.
1 (constante)	,984	,001		1862,703	,000	,983	,985
fondep	-1,723E-08	,000	-,063	-8,008	,000	,000	,000

- a. Variable dépendante : taux de respect des éch. de la clientèle

Exemple avec la variable : pénalité moyenne

$$\widehat{treclt} = 0,987 - 1,815.(10)^{-6}.penalim$$

Voir la figure 11.4 ci-dessous.

FIG. 11.4 – Régression linéaire simple avec la VI : penalim

Récapitulatif du modèle^a

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,652 ^a	,304	,304	,01424

a. Valeurs prédites : (constantes), pénalité moyenne

b. Variable dépendante : taux de respect des éch. de la clientèle

ANOVA^b

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	,812	1	,812	4006,655	,000 ^a
	Résidu	1,858	9164	,000		
	Total	2,671	9165			

a. Valeurs prédites : (constantes), pénalité moyenne

b. Variable dépendante : taux de respect des éch. de la clientèle

Coefficients^a

Modèle		Coefficients non standardisés		Coefficient standardisés	t	Sig.	Intervalle de confiance à 95% de B	
		B	Erreur standard				Bêta	Borne inf.
1	(constante)	,987	,000		5571,707	,000	,987	,987
	penalim	-1,815E-06	,000	-,652	-,83,289	,000	,000	,000

a. Variable dépendante : taux de respect des éch. de la clientèle

11.3 Régression linéaire multiple

11.3.1 Données et calculs

L'équation des moindres carrés pour la régression multivariée est :

$$\widehat{Y} = a + b_1.x_1 + b_2.x_2 + \dots + b_k.x_k$$

Les données utilisés :

- Variable dépendante = treclt

- Variables indépendantes = encm, fondemp, penalim

Les calculs sont effectués avec SPSS et les résultats sont présentés dans la figure 11.5

FIG. 11.5 – Régression linéaire multiple

Statistiques descriptives

	Moyenne	Ecart-type	N
treclt	,981	,017	9166
enclm	557876,325	160803,044	9166
fondep	173409,638	62048,457	9166
penalim	3356,201	5187,435	9166

Récapitulatif du modèle^d

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation	Durbin-Watson
1	,552 ^a	,304	,304	,01424	
2	,556 ^b	,309	,309	,01419	
3	,587 ^c	,345	,344	,01382	,014

a. Valeurs prédites : (constantes), pénalité moyenne

b. Valeurs prédites : (constantes), pénalité moyenne, encours moyen

c. Valeurs prédites : (constantes), pénalité moyenne, encours moyen, fonds et dépôt moyen

d. Variable dépendante : taux de respect des ech. de la clientèle

ANOVA^d

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	,812	1	,812	4005,555	,000 ^a
	Résidu	1,858	9164	,000		
	Total	2,671	9165			
2	Régression	,826	2	,413	2051,546	,000 ^b
	Résidu	1,845	9163	,000		
	Total	2,671	9165			
3	Régression	,920	3	,307	1605,181	,000 ^c
	Résidu	1,751	9162	,000		
	Total	2,671	9165			

a. Valeurs prédites : (constantes), pénalité moyenne

b. Valeurs prédites : (constantes), pénalité moyenne, encours moyen

c. Valeurs prédites : (constantes), pénalité moyenne, encours moyen, fonds et dépôt moyen

d. Variable dépendante : taux de respect des ech. de la clientèle

Coefficients^a

Modèle		Coefficients non standardisés		Coefficients standardisés	t	Sig.	Intervalle de confiance à 95% de B		Statistiques de colinéarité	
		B	Erreur std.				Borne inf.	Borne sup.	Tolérance	VIF
		Bêta								
1	(constante)	,987	,000		5571,707	,000	,987	,987		
	penalim	1,815E-06	,000	-,552	-63,289	,000	,000	,000	1,000	1,000
2	(constante)	,992	,001		1598,522	,000	,991	,993		
	penalim	1,911E-06	,000	-,581	-61,910	,000	,000	,000	,856	1,168
	enclm	8,224E-09	,000	-,077	-8,257	,000	,000	,000	,856	1,168
3	(constante)	1,024	,002		650,008	,000	1,021	1,027		
	penalim	2,237E-06	,000	-,680	-66,848	,000	,000	,000	,692	1,445
	enclm	3,722E-08	,000	-,351	-22,869	,000	,000	,000	,304	3,286
	fondep	8,663E-08	,000	-,315	-22,190	,000	,000	,000	,355	2,815

a. Variable dépendante : taux de respect des ech. de la clientèle

Diagnosics de colinéarité

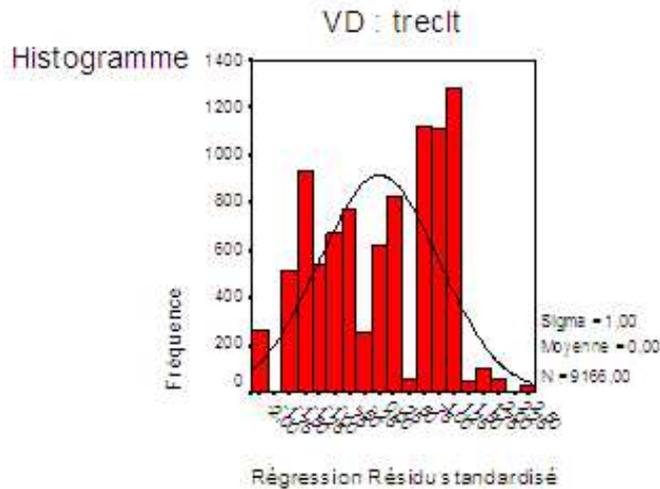
Modèle	Dimension	Valeur propre	Index de conditionnement	Proportions de la variance			
				(constante)	penalim	encom	fondep
1	1	1,543	1,000	,23	,23		
	2	,457	1,838	,77	,77		
2	1	2,323	1,000	,01	,06	,01	
	2	,647	1,895	,01	,72	,02	
	3	3,083E-02	8,680	,98	,22	,97	
3	1	3,169	1,000	,00	,02	,00	,00
	2	,665	2,183	,00	,62	,00	,00
	3	,160	4,450	,00	,03	,06	,15
	4	5,532E-03	23,934	1,00	,33	,94	,85

a. Variable dépendante : taux de respect des ech. de la clientèle

Statistiques des résidus

	Minimum	Maximum	Moyenne	Ecart-type	N
Prévision	,9152	1,0028	,9810	,01002	9166
Résidu	-,0264	,0348	,0000	,01382	9166
Prévision standardisée	-8,564	2,176	,000	1,000	9166
Résidu standardisé	-1,908	2,515	,000	1,000	9166

a. Variable dépendante : taux de respect des ech. de la clientèle



11.3.2 Analyse des résultats

Modèle général

$$\widehat{Y} = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_k \cdot x_k$$

$$x_1 = \text{penalim}; x_2 = \text{encm}; x_3 = \text{fondepmm}$$

$$a = 1,024; b_1 = -2,237 \cdot (10)^{-6}; b_2 = -3,722 \cdot (10)^{-8}; b_3 = -8,663 \cdot (10)^{-8}$$

$$\widehat{\text{treclt}} = 1,024 - 2,237 \cdot (10)^{-6} \cdot \text{penalim} - 3,722 \cdot (10)^{-8} \cdot \text{encm} - 8,663 \cdot (10)^{-8} \cdot \text{fondepmm}$$

Interprétation des résultats

Tous les éléments de l'analyse des résultats faite précédemment (11.2.2) sont valables.

Remarques sur les valeurs de R-deux : La valeur de R-deux dans la régression multiple est de 34,5. Or, les valeurs des R-deux des régressions simples sont respectivement de 2%; 0,4% et 30,4% pour les variables encm, fondepmm et penalim. C'est-à-dire que si les variables encm et fondepmm permettent d'améliorer la connaissance sur le treclt, ces variables ont un apport très limité en comparaison avec la variable penalim. Cette dernière pourrait être utilisée seule dans l'estimation du treclt.

-Dans récapitulatif du modèle, le test de Durbin-Waston permet de détecter la présence d'auto-corrélation positive entre les résidus du modèle de régression linéaire.

-Dans ANOVA, le carré moyen : C'est la somme des carrés divisée par le nombre de degrés de liberté.

-Dans le tableau "Coefficients", la colonne des statistiques de colinéarité propose dans ses sous-colonnes : Tolérance² et VIF deux notions de la colinéarité opposées l'une par rapport à l'autre. Le VIF est le facteur d'inflation de la variance. C'est la réciproque de la tolérance de colinéarité. Le principe est que la variance du coefficient de régression (B) augmente de la même façon que le facteur d'inflation de la variance, ce qui en fait un estimateur instable. Des valeurs élevées de VIF impliquent la présence de multi colinéarité. La multi colinéarité est une situation indésirable où les corrélations entre les variables indépendantes sont fortes. Or, la présence de multi colinéarités très élevées impliquent que les paramètres du modèle c'est-à-dire les coefficients de régression B seront estimés avec imprécision et les prédictions pourront être entachées d'irrégularités même si $R\text{-deux}$ (R^2) a une valeur élevée. Par conséquent, il est important de diagnostiquer ces problèmes de multi colinéarité et de proposer des remèdes à ces problèmes. Pour traiter la question de la multi colinéarité, il y a deux solutions courantes : on étudie la matrice des corrélations des estimateurs des coefficients ($X'X$) et on réalise une analyse factorielle en composante principale (A.C.P.) des VI. Cet exercice est réalisé par de nombreux

²La tolérance d'une variable est la proportion de sa variance qui n'est pas expliquée par d'autres variables explicatives de l'équation. Une variable à très faible tolérance apporte peu d'informations à un modèle, et peut entraîner des problèmes de calcul.

logiciels comme SAS et SPSS. Ces logiciels calculent des diagnostics de colinéarité basés sur les valeurs propres de la matrice normée $X'X$.

-Dans le tableau "Diagnostics de colinéarité", l'élément important est contenu dans la colonne : valeur propre. Pour chaque VI, si la valeur propre proposée est supérieure à 1, alors on conclut que la VI concernée est en multi colinéarité avec une autre VI. Or, ce n'est pas le cas à l'issue du traitement de la multi colinéarité. Aucune VI introduite dans le modèle n'affiche une valeur propre supérieure à 1. Par conséquent, on conclut que les VI sont introduites dans le modèle et chacune contribue à la variance de la VD de façon indépendante des autres. En conclusion, le VIF observé dans le tableau des coefficients a été correctement traité. On peut donc valablement utiliser les coefficients de la régression d'une part pour interpréter les résultats du test et d'autre part, on peut procéder à des prédictions sans accroître les risques d'erreur.

le tableau "Statistiques des résidus" présente les statistiques des résidus. Le résidu est la valeur réelle de la variable expliquée moins la prévision calculée par l'équation de régression, pour chaque variable. Le terme d'erreur résume les informations absentes du modèle. Il importe de vérifier ces hypothèses afin de pouvoir interpréter les résultats. Ces hypothèses liées au terme d'erreur sont : Les résidus doivent avoir une distribution symétrique (ils doivent suivre une loi normale), leur variance doit être constante et les erreurs i ($i = 1 ; \dots ; n$) doivent être indépendantes.

Résumé du chapitre

Ce chapitre a permis d'estimer l'évolution du taux de respect des échéances de la clientèle à partir d'une série de variables explicatives ou prédictives.

Dans le chapitre qui va suivre, nous allons tester la relation entre le taux de respect des échéances et les variables (facteurs) genre, recommandation et aval en supprimant statistiquement l'effet indirect des co-variables : pénalité moyenne, encours moyen et fonds et dépôt moyen. C'est l'analyse de la covariance.

Chapitre 12

Analyse de covariance (ANCOVA)

L'analyse de la covariance combine l'ANOVA et l'analyse de la régression linéaire. Le but est de déterminer l'effet d'une ou plusieurs variable-s discontinue-s sur une variable continue qui est l'objet de l'étude en contrôlant l'effet d'une ou de plusieurs autres variable-s continue-s (les co-variables) qui ont des impacts présumés sur l'évolution de la variable étudiée. Concrètement, pour analyser l'effet des variables genre, recommandation et aval sur le taux de respect des échéances pour un client, les variables comme le montant moyen de crédit reçu ou le montant moyen des pénalités de retard ou encore le montant moyen des fonds et dépôts de garantie peut avoir une influence supposée qu'il faudrait contrôler. Nous avons montré dans le chapitre précédent (chapitre 11) que ces variables sont liées à la variable taux de respect des échéances et qu'elles déterminent significativement l'évolution de cette dernière. Ces variables de contrôle sont appelées " co-variables ". On les suspecte de co-varier avec la variable dépendante et d'affecter indirectement la relation qui existe entre les facteurs et la variable dépendante. Le but de l'analyse de covariance (ou ANCOVA) est de tester la relation initiale en supprimant statistiquement l'effet indirect de la (ou les) covariable-s. Cela revient à tester l'effet des variables indépendantes (discrètes) sur la variable dépendante (continue) une fois que l'effet de la (ou les) co-variable-s (continue-s) sur la variable dépendante est enlevé.

12.1 Principes

Pour le calcul d'une ANCOVA, on combine à la fois une régression linéaire et une ANOVA. Cela est possible car, ces deux types d'analyse (la régression et l'ANOVA) reposent sur la décomposition de la somme des carrés totaux. Voir l'ouvrage de Brigitte Tribout, [Tribout, 2007] pour plus de détails. " Ce sont les variations, donc la variance de la variable expliquée qui sont examinées. Pour la régression, l'hypothèse de normalité des erreurs n'est pas indispensable, mais si elle est vérifiée, les lois conditionnelles de Y sont également gaussiennes et, en particulier la non-nullité de la pente de la droite de régression peut être testée. La décomposition est obtenue à l'aide de l'expression suivante : $Y_i - \bar{Y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})$. Pour l'ANOVA,

les hypothèses portent sur les variables aléatoires parentes X_i qui doivent avoir la même variance σ_ϵ^2 . L'hypothèse de normalité des X_i est nécessaire. La décomposition est obtenue par : $X_{ij} - \bar{X} = (\bar{X}_i - \bar{X}) + (X_{ij} - \bar{X}_i)$. [Op. cit. pp. 638-641]

Le modèle général d'une ANCOVA avec plusieurs variables indépendantes est :

$$Y_{ij} = \alpha_i + \beta(X_{ij} - \bar{X}_i) + \epsilon_{ij} \quad (12.1)$$

- β est la pente de la régression de Y sur X_1 regroupé sur tous les niveaux de la variable discrète X_2

- α_i est la différence entre la moyenne pour chaque niveau i et la moyenne générale.

Pour une ANCOVA avec k variables indépendantes, ($k \geq 2$) on note trois hypothèses nulles :

$H_{01} : \beta_i = \beta = 0 \equiv$ absence d'effet

$H_{02} : \beta_i = \text{constante} \equiv$ droites parallèles

$H_{03} : \alpha_i = 0$ pour tous les $i \equiv$ même ordonnée à l'origine

12.2 Calculs

Dans l'ANCOVA, nous testons donc l'hypothèse nulle de l'absence de différence entre les moyennes des groupes une fois que l'effet de la co-variable est retiré. L'hypothèse alternative (H1) est donc que les moyennes des groupes se distinguent. Considérons une situation où plusieurs (g) estimations indépendantes de la pente de la relation linéaire entre la variable pénalité moyenne (penalim = x) et le taux de respect des échéances (trecht = y). Par exemple, on peut disposer de $g = 8$ droites de régression qui décrivent le taux de respect des échéances en fonction de la pénalité moyenne. Mais, les droites de régression concernées sont obtenues sur la base d'informations concernant les clients. Ces informations peuvent être résumées ainsi :

1. F-R-A = féminin et recommandée et avalisée
2. F-R-NA = féminin et recommandée et non avalisée
3. F-NR-A = féminin et non recommandée et avalisée
4. F-NR-NA = féminin et non recommandée et non avalisée
5. M-R-A = masculin et recommandé et avalisé
6. M-R-NA = masculin et recommandé et non avalisé
7. M-NR-A = masculin et non recommandé et avalisé
8. M-NR-NA = masculin et non recommandé et non avalisé

Toutes ces possibilités sont analysées par trois variables : la variable dépendante quantitative, continue : "taux de respect des échéances", la variable explicative quantitative, continue : "pénalité moyenne", et la variable explicative qualitative multiclasse "genre-recommandation-aval" qui a 8 classes (d'où $g = 8$).

On peut alors se poser une question préliminaire : les g droites seront-elles égales? Autrement dit, explique-t-on la même quantité de variation en regroupant les données des g groupes et en calculant une seule régression qu'en calculant g régressions? Cette hypothèse serait notre H_0 préliminaire qui postule que les droites de régression sont confondues. Si cette H_0 est rejetée, c'est-à-dire s'il existe une différence significative entre les g droites de régression, l'analyse du respect des échéances poserait alors trois types de questions :

1. La pénalité moyenne influence-t-elle le taux de respect des échéances indépendamment des modalités du client c'est-à-dire le fait qu'il soit d'un sexe donné et qu'il soit ou non recommandé et ou avalisé?
2. Les modalités du client influencent-elles le rapport entre la pénalité moyenne et le taux de respect des échéances? Ou encore, la pénalité moyenne influence-t-elle le rapport entre les modalités du client et le taux de respect des échéances?
3. Les modalités du client influencent-elles le taux de respect des échéances indépendamment des pénalités?

Concernant la question 1, on peut la tester à l'aide d'une série de g régressions linéaires simples. Pour répondre à la question 2 qui traite de l'interaction entre les modalités des clients et les pénalités, on doit tester si les pentes des g droites de régression ne sont pas toutes nulles, ($H_{01} : \beta_i = \beta = 0$), et si elles diffèrent entre elles ou si les droites sont parallèles (aux fluctuations aléatoires près), $H_{02} : \beta_i = \text{constante}$. Des droites parallèles signifieraient que le rapport entre le taux de respect des échéances et les pénalités ne dépend pas des modalités des clients. Dans ce cas, c'est-à-dire (si les droites sont parallèles), alors on doit tester la question 3 en vérifiant si les droites (parallèles) sont situées à des hauteurs différentes (donc, si leur ordonnée à l'origine diffère) ou si les droites sont confondues en une seule (aux fluctuations aléatoires près); ($H_{03} : \alpha_i = 0$); mêmes ordonnées à l'origine. Des droites situées à des hauteurs différentes signifieraient que les modalités des clients influencent le taux de respect des échéances indépendamment de l'effet des pénalités.

12.2.1 Données

Variable dépendante (VD) = taux de respect des échéances

Variables indépendantes (ou facteurs) = genre; recommandation; aval. Pour l'application, nous utiliserons dans un premier temps la variable genre.

Co-variables = encours moyen de crédit, fonds et dépôts de garanties moyen, pénalité moyenne. Pour l'application, nous utiliserons une seule co-variable : pénalité moyenne (qui a la plus forte corrélation) avec le taux de respect.

12.2.2 Etapes

Pour réaliser une analyse de covariance, il y a des étapes à suivre, [Tribout, 2007], [Saporta, 2006], [Pupion, 2008]. La réalisation d'une ANCOVA consiste à tester les coefficients du modèle de régression complet contenant toutes les variables de l'étude ; c'est-à-dire qu'on effectue le test de nullité de la pente (notamment le test de Fischer de nullité de la pente) et ou le test de nullité de la constante concernant l'ordonnée à l'origine.

En effet, le modèle de régression suppose que la forme de la relation entre Y (variable étudiée) et X (variable explicative continue, régresseur) et X (variable explicative discrète, facteur) est linéaire. Cependant, si on obtient une estimation de la pente qui est nulle, le modèle se réduit à l'équation $Y = \alpha_i + \epsilon_{ij}$ et les variables Y et X ne sont pas liées. Il convient donc de tester la nullité de la pente β . Généralement, on utilise pour cela le test de Fischer-Snedecor. Ce test consiste à poser les hypothèses suivantes au seuil de 5% ou 2,5% ou 1%

$$H_0 : \frac{(n-2)SCE}{SCR} = 1$$

$$H_1 : \frac{(n-2)SCE}{SCR} \succ 1$$

SCE = somme des carrés des écarts de l'estimation, SCR = somme des carrés des résidus.

Dans le cadre d'une régression simple le test de Fischer-Snedecor revient à tester l'hypothèse H_0 ci-dessous à un seuil critique par exemple de 5% :

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

C'est ce que nous réalisons lorsque nous faisons la régression du taux de respect des échéances par la variable pénalité moyenne.

En effet,

- si H_0 est vraie (c.-à-d. $\beta = 0$) alors, $\frac{SCR}{(n-2)}$ et SCE sont deux estimations non biaisées de σ_ϵ^2 .

- si H_0 est fautive, (c.-à-d. $\beta \neq 0$) $\frac{SCR}{(n-2)}$ reste sans biais mais SCE surestime σ_ϵ^2 .

Lorsque H_0 est vraie, le rapport des deux estimateurs doit être proche de 1. On a :

$$\frac{SCE}{SCR/(n-2)} = \frac{R^2}{(1-R^2)/(n-2)} \hookrightarrow F(1; n-2).$$

La variable de décision du test est donc :

$$F_R = \frac{SCE}{SCR/(n-2)}$$

Cette règle de décision au seuil α (du test de Fischer par exemple $\alpha = 5\%$) est formulée de la manière suivante :

Rejet de H_0 si F_R observé ($=F_{Robs}$) est supérieur à $F_{1-\alpha}$ c.-à-d :

$$F_R = \frac{SCE}{SCR/(n-2)} \succ F_{95\%}$$

avec

$$F_R \hookrightarrow F(1; n-2)$$

La valeur critique de $F_{95\%}$ c'est-à-dire la valeur de X pour $\alpha = 0,05$ est lue sur la table de Fischer-Snedecor. Si $F_{Obs} > F_{95\%}$, alors la variable appartient à une zone de rejet de l'hypothèse nulle ; donc, on conclut que H_0 est rejetée. La décision à prendre est que la régression ajuste correctement les données puisqu'on rejette H_0 . Si H_0 est rejetée, la suite de la procédure ANCOVA consiste à "séparer" le modèle sur chaque niveau de la variable discrète. Mais, si H_0 est acceptée, alors on doit tester l'ordonnée à l'origine de la pente de la régression commune. Des ordonnées à l'origine différentes signifieraient qu'il y a un effet "identique" de la variable discrète sur chacun des niveaux de la variable expliquée en fonction des niveaux de la co-variable explicative.

12.3 Résultats

Afin de faciliter la lecture des résultats de l'analyse de covariance, nous proposons dans un premier temps une ANCOVA avec 1 régresseur qui est la variable continue (penalim) et 1 facteur qui est la variable discrète : genre. Nous montrerons dans un second temps les résultats de l'ANCOVA avec 3 régresseurs : penalim, encm et fondep et 3 facteurs dont 1 facteur fixe (= genre) et 2 facteurs variables (= recommandation et aval).

Tous les calculs ont été réalisés avec le logiciel SPSS. Les résultats sont présentés successivement ci-dessous.

12.3.1 Ajustement du modèle à 1 régresseur et 1 facteur

Question 1 : La pente de la régression de treclt par penalim est-elle significative indépendamment du genre ?

Pour répondre à cette question, une régression multiple avec comme variables indépendantes : penalim et genre aurait été intéressante ; or la variable genre étant discontinue, on ne peut procéder ainsi. Mais, on peut réaliser un test des effets inter-sujets. Figure 12.1.

FIG. 12.1 – Test des effets inter-sujets entre les variables genre et penalim

Facteurs inter-sujets

		Etiquette de valeur	N
genre	1	féminin	6150
	2	masculin	3016

Tests des effets inter-sujets

Variable dépendante: trecht

Source	Somme des carrés de type III	ddl	Moyenne des carrés	F	Sig.	Eta au carré partiel	Paramètre non centré	Puissance observée ^a
Modèle corrigé	,819 ^b	3	,273	1349,034	,000	,306	4047,102	1,000
Constante	5647,023	1	5647,023	27929566,2	,000	1,000	2,8E+07	1,000
GENRE * PENALIM	5,192E-03	1	5,2E-03	25,680	,000	,003	25,680	,999
GENRE	2,134E-04	1	2,1E-04	1,055	,304	,000	1,055	,177
PENALIM	,702	1	,702	3472,783	,000	,275	3472,783	1,000
Erreur	1,852	9162	2,0E-04					
Total	8823,902	9166						
Total corrigé	2,671	9165						

a. Calculé à partir d'alpha = ,05

b. R deux = ,306 (R deux ajusté = ,306)

Sur cette figure, on peut faire les constatations suivantes :

- L'interaction entre les variables genre et penalim (genre*penalim) est testée (avec le test de Fischer-Snedecor) au seuil de 5%.

$$F(1; 9162) = \frac{5,192 \cdot 10^{-3}}{1,852/9162} = 25,680 > F_{(0,95)}$$
 F appartient à la zone de rejet de l'hypothèse nulle. Cela signifie que si nous réalisons deux régressions de treclt par la variable penalim sur chaque modalité du facteur genre, ces deux régressions seraient significativement différentes l'une de l'autre.
- Le facteur genre n'a pas d'effet significatif sur la variable étudiée (= treclt) indépendamment de la co-variable (= penalim). On dit que le facteur genre n'a pas d'effet principal.
- Par contre, la co-variable (= penalim) a un effet principal. Elle contribue pour 27,5% à la variabilité de la variable étudiée.
- Puisque l'interaction des variables genre et penalim (= genre*penalim) a un effet significatif sur la variable étudiée (= treclt), cela revient à dire que le facteur genre a un effet d'interaction sur la variable étudiée corrélativement à la variable de contrôle (= penalim). Puisqu'il n'y a pas d'effet principal du facteur genre, on peut supposer que ce facteur agirait comme un "aspirateur" pour éponger le niveau de variabilité de la variable étudiée par d'autres facteurs (comme le facteur penalim) avec lesquels le facteur genre est en interaction. On appellera l'effet d'interaction du facteur genre : effet de rétention.

A présent, réalisons les ajustements des modèles réduits pour voir quelle modalité du facteur genre exprime le plus cet effet de rétention.

12.3.2 Ajustement des modèles réduits par classes de facteur

Question 2 : La pente de la régression de treclt par penalim est-elle la même pour les deux sexes ?

La conclusion partielle qu'on peut tirer du paragraphe précédent c'est que le résultat du test de Fischer de nullité de la pente conduit à estimer la variable treclt par la variable penalim sur chacune des modalités de la variable genre. Cela présume de l'existence d'un effet de la variable genre sur la variable étudiée treclt non pas indépendamment de la variable penalim (car cet effet serait non significatif si on se réfère à la figure 12.1, voir la signification = 0,304), mais corrélativement à la variable contrôle (= penalim). Cela suppose nécessairement l'existence d'une relation (et d'un effet) réciproque entre les variables genre et pénalim.

Dans un premier temps, nous proposons les estimations des paramètres intra-classe qui correspondent à la régression de treclt par la co-variable penalim en fonction de chaque modalité du facteur genre, puis dans un second temps, nous proposerons la représentation graphique de cette mesure de l'effet du facteur genre selon ses modalités sur la variable taux de respect des échéances.

- Estimation des paramètres de la régression de treclt par penalim en prenant en compte les effets du facteur genre

La figure 12.2 propose l'estimation ($=\beta$) de chaque paramètre du modèle et l'erreur standard de cette estimation ainsi que le test ($=t$) de signification de chaque estimation.

FIG. 12.2 – Estimation des paramètres de l'ANCOVA

Estimations des paramètres

Variable dépendante: treclt

Paramètre	B	Erreur std	t	Sig.	Intervalle de confiance à 95%		Eta au carré partiel	Paramètre non centré	Puissance observée ^a
					Borne inf.	Limite sup.			
Constante	,987	,000	3253,0	,000	,987	,988	,999	3252,98	1,000
[GENRE=1] * PENALIM	-1,729E-06	,000	-51,277	,000	-1,8E-06	-1,7E-06	,223	51,277	1,000
[GENRE=2] * PENALIM	-2,054E-06	,000	-37,603	,000	-2,2E-06	-1,9E-06	,134	37,603	1,000
[GENRE=1]	-3,838E-04	,000	-1,027	,304	-1,1E-03	3,485E-04	,000	1,027	,177
[GENRE=2]	0 ^b								
PENALIM	0 ^b								

a. Calculé à partir d'alpha = ,05

b. Ce paramètre est mis à zéro car il est redondant.

Comparaisons par paires

Variable dépendante: treclt

(I) genre	(J) genre	Différence des moyennes (I-J)	Erreur standard	Signification ^a	Intervalle de confiance de la différence à 95% ^b	
					Borne inférieure	Limite supérieure
féminin	masculin	7,078E-04*	,000	,026	8,613E-05	1,330E-03
masculin	féminin	-7,078E-04*	,000	,026	-1,330E-03	-8,613E-05

Basée sur les moyennes marginales estimées

*. La différence de moyennes est significative au niveau ,05

a. Ajustement des comparaisons multiples : Différence la moins significative (équivalent à aucun ajustement).

Tests d'insuffisance d'ajustement

Variable dépendante: treclt

Source	Somme des carrés	ddl	Moyenne des carrés	F	Sig.	Eta au carré partiel	Paramètre non centré	Puissance observée ^a
Insuffisance d'ajustement	1,852	78	,024	900058,768	,000	1,000	7,E+07	1,000
Erreur pure	,000	9084	,000					

a. Calculé à partir d'alpha = ,05

Cette figure permet de faire les commentaires suivants :

- La valeur "Eta au carré partiel" fournit une estimation de l'effet de taille pour chaque effet et chaque estimation. Cette statistique décrit la proportion de la variabilité totale de treclt imputable à chaque élément estimé. La valeur de *Eta-carré* pour [genre=1]*penalim

est de 0,223 ; celle de [genre=2]*penalim est de 0,134 ; celle de [genre=1] est de 0,000 et celle de [genre=2] n'est pas estimée. Or, nous savons que la valeur de *Eta-carré* pour [genre] = 0,000 ; celle de penalim = 0,275 et celle de [genre]*penalim = 0,003.

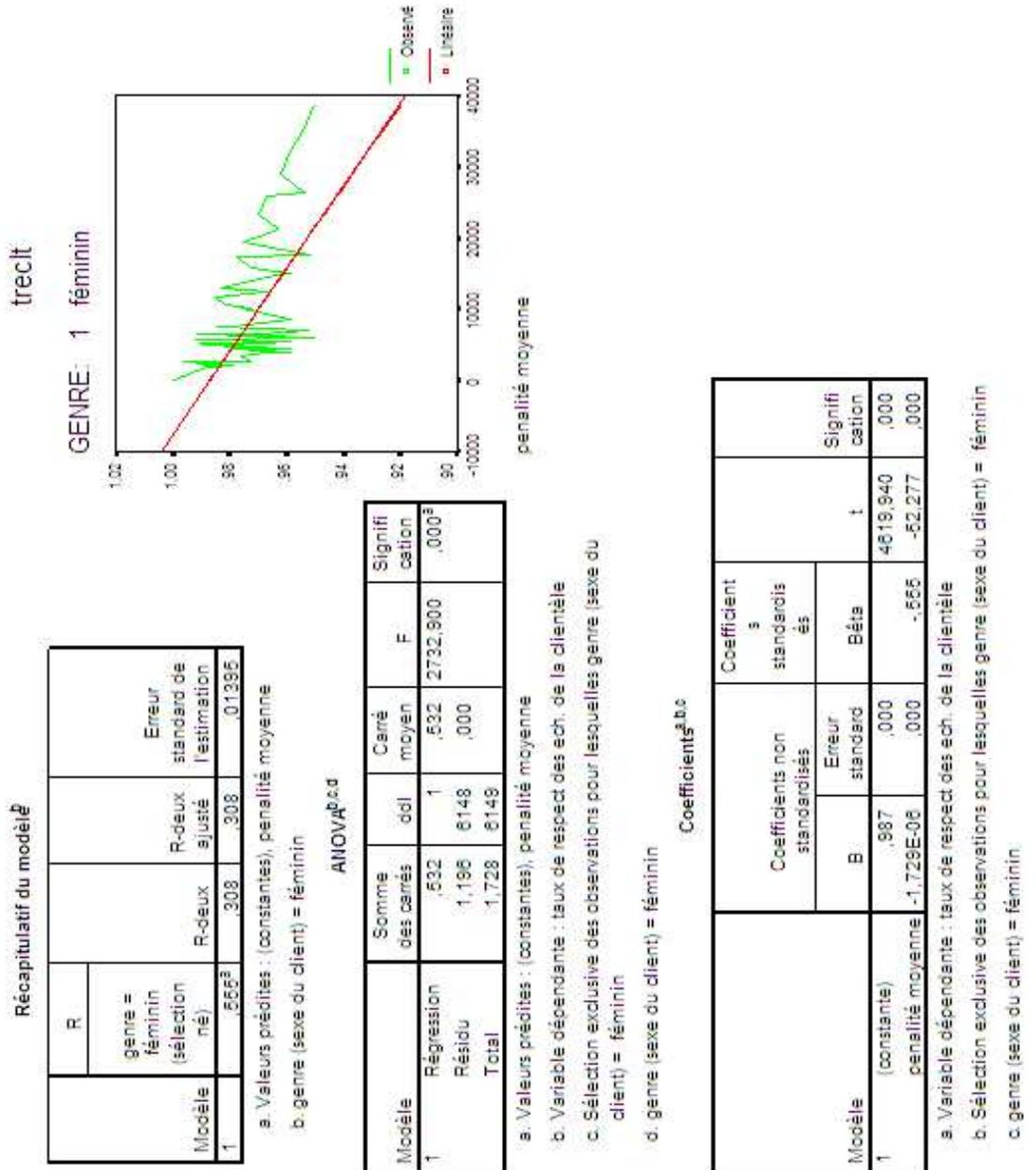
Ces données montrent deux formes d'effet pour le facteur genre. D'abord il y a l'effet de rétention qui est l'effet "global" du facteur c'est-à-dire l'effet d'interaction. On le montre en prenant : *Eta carré partiel* de [genre] = 0,000 (absence d'effet principal ; *Eta carré partiel* de penalim = 0,275 et *Eta carré partiel* de [genre]*penalim = 0,003. Le facteur [genre] a déprécié la contribution de la variable penalim à la variabilité de la variable étudiée. Ensuite, il y a la "taille" de l'effet de rétention sur les modalités 1 (= féminin) et 2 (= masculin) du facteur [genre]. En effet, on constate que *Eta carré partiel* de [genre=1]*penalim = 0,223 supérieur à *Eta carré partiel* de [genre=2]*penalim = 0,134. Cela revient à dire que la taille de l'effet de rétention est plus importante chez les clients de sexe féminin.

- La puissance observée définit la puissance du test lorsque l'hypothèse du test est définie sur la base de la valeur observée.
- Le test d'intervalle de confiance identifie les sous-groupes homogènes de moyennes qui ne diffèrent pas les uns des autres.
- Le test de comparaison de moyenne par paire montre les différences entre les deux modalités du facteur [genre]. Ces comparaisons testent la différence entre les moyennes appariées et engendrent une matrice pour laquelle les astérisques indiquent les moyennes de groupes significativement différentes au niveau alpha 0.05. Nous avons pratiqué le test : LSD (différence de moindre signification).
- Enfin, le manque d'ajustement montre que le modèle n'ajuste pas convenablement la relation entre la variable dépendante (= treclt) et les variables indépendantes à savoir la co-variable (= penalim) et le facteur (= genre). Nous devons introduire d'autres variables et d'autres facteurs.

- Estimation de treclt par penalim avec genre = 1 (= féminin)

Somme des carrés des écarts de l'estimation : $SCE = 0,532$; Somme des carrés des écarts de l'erreur d'estimation (ou résidu) : $SCR = 1,196$. Ordonnée (a) = 0,987 ; coefficient de régression (b) = $-1,729 \cdot (10)^{-6}$. $F = \frac{0,532}{1,196/6148} = 2732,9 = t^2 = (-52,277)^2$. F appartient à la zone de rejet de H_0 sur la table de Fischer-Snedecor. Figure 12.3.

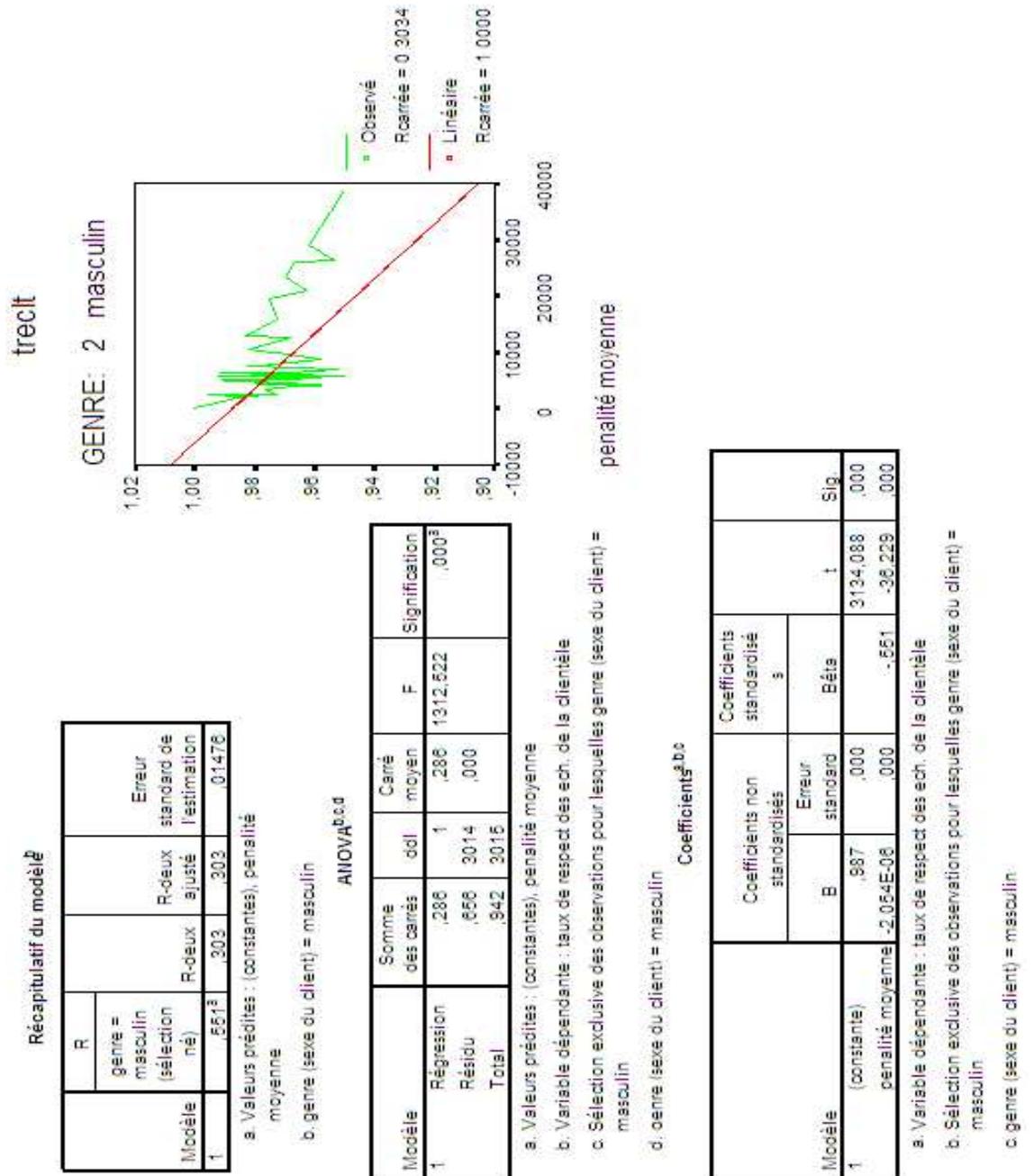
FIG. 12.3 – Effets du sexe (féminin) et de la pénalité de retard sur le respect des échéances



- Estimation de treclt par penalim avec genre = 2 (= masculin)

$SCE = 0,286$; $SCR = 0,656$; (a) = 0,987; (b) = $-2,054(.10)^{-6}$; $F = \frac{0,286}{0,656/3014} = 1312,522 = t^2 = (-36,229)^2$. F appartient à la zone de rejet de H_0 sur la table de Fischer-Snedecor. Figure 12.4

FIG. 12.4 – Effets du sexe (masculin) et de la pénalité de retard sur le respect des échéances



Conclusion partielle : Les 2 droites de régression ne sont pas confondues. Elles ont sensiblement la même ordonnée à l'origine ($a \approx 0,987$) mais les pentes sont différentes. Cela confirme le rejet de l'hypothèse de nullité de la pente de l'estimation de $treclt$ par la variable $genre*penalim$ qui a été vue dans la figure 12.1.

12.3.3 Modèle d'ANCOVA avec 3 régresseurs et 3 facteurs

a) Données

- Variable dépendante = $treclt$ (taux de respect des échéances de la clientèle)
- Variables indépendantes : régresseurs : $enclm$ (= encours de crédit moyen), $fondep$ (= fonds et dépôts de garanties annuels moyens), $penalim$ (= pénalités moyennes annuelles); facteurs : $genre$ (= facteur fixé), $recommandation$, $aval$ (= facteurs aléatoires).
- Variable de pondération = $ancienneté$

b) Statistiques descriptives

figure 12.5 ci-dessous

FIG. 12.5 – Statistiques descriptives

Facteurs inter-sujets

		Etiquette de valeur	N
genre	1	féminin	6150
	2	masculin	3016
recommandation	1	recommandé	9101
	2	non recommandé	65
évaluation	1	évalués	7267
	2	non évalués	1899

Statistiques descriptives

Variable dépendante: trect

genre	recommandation	évaluation	Moyenne	Ecart-type	N
féminin	recommandé	évalués	,9808	,01663	5998
		non évalués	,9838	,02090	138
		Total	,9809	,01674	6134
	non recommandé	évalués	,9593	,00000	16
		Total	,9593	,00000	16
		Total	,9807	,01665	6012
masculin	recommandé	évalués	,9891	,01684	1215
		non évalués	,9764	,01633	1762
		Total	,9816	,01768	2967
	non recommandé	évalués	,9644	,00374	40
		non évalués	,9969	,00926	9
		Total	,9704	,01368	49
Total	recommandé	évalués	,9883	,01714	1255
		non évalués	,9766	,01637	1761
		Total	,9814	,01768	3016
	non recommandé	évalués	,9822	,01695	7211
		non évalués	,9769	,01681	1899
		Total	,9811	,01706	9101
Total	évalués	,9627	,00420	66	
	non évalués	,9969	,00926	9	
	Total	,9674	,01295	65	
Total	évalués	,9821	,01698	7267	
	non évalués	,9770	,01684	1899	
	Total	,9810	,01707	9166	

c) Effets principaux (inter-classes)

L'analyse des effets principaux (inter-classes) et les paramètres d'estimation des facteurs selon leurs modalités sont proposés dans la figure 12.6

FIG. 12.6 – Effets principaux et paramètres d'estimation

Tests des effets inter-sujets^d

Variable dépendante: treclt

Source		Somme des carrés de type III	ddl	Moyenne des carrés	F	Sig.	Eta au carré partiel
Constante	Hypothèse	199,763	1	199,763	104400,83	,000	1,000
	Erreur	2,853E-03	1,491	1,913E-03 ^b			
GENRE	Hypothèse	4,381E-02	1	4,381E-02	117,152	,000	,013
	Erreur	3,384	9051	3,739E-04 ^c			
RECOMMAN	Hypothèse	3,278E-04	1	3,278E-04	,877	,349	,000
	Erreur	3,384	9051	3,739E-04 ^c			
AVAL	Hypothèse	4,489E-02	1	4,489E-02	120,046	,000	,013
	Erreur	3,384	9051	3,739E-04 ^c			
FONDEPM	Hypothèse	,329	1	,329	879,537	,000	,089
	Erreur	3,384	9051	3,739E-04 ^c			
ENCM	Hypothèse	,406	1	,406	1087,004	,000	,107
	Erreur	3,384	9051	3,739E-04 ^c			
PENALIM	Hypothèse	1,938	1	1,938	5183,673	,000	,364
	Erreur	3,384	9051	3,739E-04 ^c			

b. $,260 \text{ MS}(\text{RECOMMAN}) + 3,485\text{E-}02 \text{ MS}(\text{AVAL}) + ,705 \text{ MS}(\text{Erreur})$

c. $\text{MS}(\text{Erreur})$

d. Régression des moindres carrés pondérés - Pondérés par ancienneté trié

Estimations des paramètres^d

Variable dépendante: treclt

Paramètre	B	Erreur std	t	Sig.	Eta au carré partiel	Paramètre non centré	Puissance observée ^a
Constante	1,025	,002	537,874	,000	,970	537,874	1,000
[GENRE=1]	5,137E-03	,000	-10,824	,000	,013	10,824	1,000
[GENRE=2]	0 ^b	,	,	,	,	,	,
[RECOMMAN=1]	1,342E-03	,001	,936	,349	,000	,936	,155
[RECOMMAN=2]	0 ^b	,	,	,	,	,	,
[AVAL=1]	5,745E-03	,001	10,957	,000	,013	10,957	1,000
[AVAL=2]	0 ^b	,	,	,	,	,	,
FONDEPM	8,027E-08	,000	-29,657	,000	,089	29,657	1,000
ENCM	4,847E-08	,000	-32,970	,000	,107	32,970	1,000
PENALIM	1,803E-06	,000	-71,998	,000	,364	71,998	1,000

a. Calculé à partir d' $\alpha = ,05$

b. Ce paramètre est mis à zéro car il est redondant.

c. Régression des moindres carrés pondérés - Pondérés par ancienneté trié

On parle d'effet principal lorsqu'une variable indépendante (VI) a une influence directe sur la variance de la variable expliquée indépendamment d'autres variables indépendantes présentes dans le modèle d'ajustement. Les effets principaux sont présentés dans le tableau des "tests des effets inter-sujets" ou inter-classes. On effectue le test du F de Fischer-Snedecor. Dans notre cas, les effets principaux des facteurs : genre et aval sont testés. Ils sont significatifs. La taille de l'effet est calculé par l'indicateur *Eta au carré partiel* = 1,3%. Le facteur recommandation n'a pas d'effet principal significatif.

Les co-variables respectives : fondep, encm et penalim ont naturellement des effets prin-

cipaux¹. Les tailles de ses effets sont respectivement de 8,9% ; 10,7% et 36,4%.

d) Effets d'interaction

Les interactions et les paramètres d'estimation sont fournis par la figure 12.7

FIG. 12.7 – Effets d'interaction et paramètres d'estimation

Tests des effets inter-sujets

Variable dépendante: treclt

Source	Somme des carrés de type III	ddl	F	Sig.	Eta au carré partiel	Paramètre non centré	Puissance observée ^a
Modèle corrigé	2,896 ^b	13	727,915	,000	,511	9462,9	1,000
Constante	,000	0	,	,	,000	,000	,
1 GENRE * FONDEPM * ENCM	1,451E-03	2	2,371	,093	,001	4,743	,482
2 GENRE * RECOMMAN * FONDEPM * ENCM	8,286E-04	1	2,702	,100	,000	2,702	,376
3 GENRE * AVAL * FONDEPM * ENCM	5,297E-03	2	8,656	,000	,002	17,312	,969
4 GENRE * RECOMMAN * AVAL * FONDEPM * ENCM	7,149E-03	1	23,362	,000	,003	23,362	,996
5 GENRE * FONDEPM * ENCM * PENALIM	,600	2	981,024	,000	,178	1962,0	1,000
6 GENRE * RECOMMAN * FONDEPM * ENCM * PENALIM	2,130E-05	1	,070	,792	,000	,070	,058
7 GENRE * AVAL * FONDEPM * ENCM * PENALIM	2,528E-04	2	,413	,662	,000	,826	,117
8 GENRE * RECOMMAN * AVAL * FONDEPM * ENCM * PENALIM	1,289E-03	1	4,211	,040	,000	4,211	,537
Erreur	2,767	9044					
Total	20170,108	9058					
Total corrigé	5,663	9057					

a. Calculé à partir d'alpha = ,05

b. R deux = ,511 (R deux ajusté = ,511)

c. Régression des moindres carrés pondérés - Pondérés par ancienneté trié

- Estimation des modèles inter-sujets (ou inter-classes)

Le modèle inter-sujets (ou inter-classes) est constitué par l'interaction des facteurs : genre, recommandation et aval avec l'un et ou l'autre co-variable du modèle. La constitution est faite comme suit. Pour tout client, le genre est fixé, soit féminin, soit masculin. De même, pour tout client, il y a un encours de crédit (autrement le client en question ne serait pas un client actif) et il y a un fonds ou dépôt de garantie associé au crédit (autrement le client en question n'aurait pas le crédit puisque le fonds et dépôt de garantie est une condition d'accès au crédit. Le fonds et dépôt de garantie est généralement une valeur qui est amputée du montant du crédit réellement alloué au client emprunteur). Donc, nous avons, le genre, l'encours et le fonds

¹C'est une condition d'application de l'analyse de covariance.

et dépôt de garanties qui constituent le "package" de base. Les autres variables indépendantes c'est-à-dire les facteurs : recommandation et aval et la co-variable : pénalité moyenne viennent s'ajouter au "package" de base pour constituer les interactions que nous analysons. Ainsi, on a :

- un premier niveau d'interaction qui concerne l'identification du client. Les variables indépendantes (VI) prises en compte sont : genre, fondep m et enc m . On constitue l'interaction : $genre * fondep\mathit{m} * enc\mathit{m}$
- un deuxième niveau d'interaction qui concerne les clients recommandés ou non recommandés. On constitue l'interaction : $genre * recomman * fondep\mathit{m} * enc\mathit{m}$
- un troisième niveau d'interaction qui concerne les clients avalisés ou non avalisés. On constitue l'interaction : $genre * aval * fondep\mathit{m} * enc\mathit{m}$
- un quatrième niveau d'interaction qui concerne les clients recommandés ou non recommandés et avalisés ou non avalisés. On constitue l'interaction : $genre * recomman * aval * fondep\mathit{m} * enc\mathit{m}$
- un cinquième niveau d'interaction qui concerne les clients et la pénalité. On constitue l'interaction : $genre * fondep\mathit{m} * enc\mathit{m} * penalim$
- un sixième niveau d'interaction qui concerne les clients recommandés ou non recommandés et la pénalité. On constitue l'interaction : $genre * recomman * fondep\mathit{m} * enc\mathit{m} * penalim$
- un septième niveau d'interaction qui concerne les clients avalisés ou non avalisés et la pénalité. On constitue l'interaction : $genre * aval * fondep\mathit{m} * enc\mathit{m} * penalim$
- enfin, un huitième niveau d'interaction qui concerne les clients recommandés ou non recommandés et avalisés ou non avalisés et la pénalité. On constitue l'interaction : $genre * recomman * aval * fondep\mathit{m} * enc\mathit{m} * penalim$

Les renseignements fournis par ce modèle sont les suivants :

- L'ensemble du modèle est estimé par R-deux et R-deux ajusté = 51,1%.
- Les interactions de niveau 1 ; 2 ; 6 et 7 ne sont pas significatifs (au test du F de Fischer-Snedecor) au seuil de 5%.
- On peut regrouper les clients en deux groupes selon qu'on considère ou pas les pénalités. On peut avoir un groupe 1 qui est représenté principalement par le niveau 1 (= clients sans pénalités) et touchant les niveaux 2 ; 3 et 4. Et on aurait un groupe 2 représenté par le niveau 5 (= clients avec pénalités) et touchant les niveaux 6, 7 et 8
- Les effets qui peuvent être analysés concernent : les clients avalisés ou non avalisés (niveau 3) ; les clients recommandés ou non et avalisés ou non (ni-

veau 4) ; les clients et la pénalité (niveau 5) et enfin les clients recommandés ou non et avalisés ou non et la pénalité (niveau 8).

- *Analyse des effets des modèles inter-sujets (ou inter-classes)*

les clients avalisés ou non avalisés : (niveau 3) dans la figure 12.7

L'estimation de l'interaction inter-sujets : $\text{genre*aval*fondep}*encm$ (niveau 3 de l'analyse) est significative. Elle peut être analysée. Elle concerne les clients avalisés ou non avalisés. La taille de l'effet de $\text{genre*aval*fondep}*encm$ est donnée par l'indicateur *Eta au carré partiel* = 0,002 (soit 0,2%)². C'est-à-dire que 0,2% de la variance du taux de respect des échéances est dû aux clients avalisés ou non avalisés ($\text{genre*aval*fondep}*encm$). Dans ce cas, on doit voir les effets des facteurs aval et genre et dire si ces effets sont importants ou non dans l'analyse du comportement de respect des échéances pour n'importe quel client. Dans cet objectif, nous devons savoir trois choses : (1) l'effet principal de chacune des VI prises ; (2) l'effet d'interaction d'ordre 2 de toutes les VI prises et (3) l'effet d'interaction d'ordre 3 de toutes les VI prises. Toutes les interactions inter et intra-classes d'ordre 2, d'ordre 3, d'ordre 4, d'ordre 5 et d'ordre 6 sont proposées dans l'annexe 7.

Nous savons déjà que les effets principaux de toutes les VI prises en compte sont significativement positifs ; voir la figure 12.6. Les effets principaux des facteurs genre et aval sont significatifs et égaux à 0,013 chacun. L'estimation de l'interaction $\text{fondep}*encm$ produit une taille d'effet = 0,037 (voir interaction d'ordre 2 dans l'annexe 7). Les estimations des interactions d'ordre 3 produisent les effets suivants : $\text{genre*fondep}*encm$: 0,005 ; $\text{aval*fondep}*encm$: 0,006 (voir interaction d'ordre 3 dans l'annexe 7). Sur cette observation, on pourra dire que le facteur aval a un effet de taille plus élevée que celui du facteur genre. Par conséquent, on retient que le facteur aval contribue plus que le facteur genre à l'évolution du taux de respect des échéances. On aurait pu voir d'une autre manière l'effet du facteur aval en considérant l'interaction suivante : $\text{genre*fondep}*encm$. (C'est le niveau 1 de la figure 12.7). On peut dire que c'est une condition de base. La taille de l'effet correspond est de 0,001 au test de Fischer au seuil de 10%. De manière rigoureuse, nous devons accepter l'hypothèse de nullité de la pente de l'estimation de cette interaction (principe du test de Fischer). Mais pour la méthodologie, si nous supposons crédible la valeur de *Eta au carré partiel* = 0,001 et qu'on introduit le facteur aval dans le modèle de base, on obtient une amélioration de la puissance de l'observation (passant de 0,482 à 0,969) et une amélioration de la contribution à la variance du taux de respect des échéances : *Eta au carré partiel* passe de 0,001 à 0,002). On pourrait alors dire que le facteur aval a un effet multiplicateur sur la variance du taux de respect des échéances.

²On retient cette valeur de *Eta au carré partiel* = 0,2% car le tableau des interactions d'ordre 4 concernant les VI prises en compte ne produit aucune valeur. Voir l'annexe 7.

les clients recommandés ou non et avalisés ou non : (niveau 4) dans la figure 12.7

Pour l'analyse des effets de l'interaction inter-sujets : $\text{genre*recomm*aval*fondep*encm}$, on adopte le même raisonnement que précédemment, mais on considère à la fois les facteurs recommandation et aval. Nous concluons que les facteurs recommandation et aval ont ensemble un effet multiplicateur de la variance du taux de respect des échéances. La démonstration est la suivante :

Effet de l'interaction : $\text{genre*fondep*encm} = 0,005$ (lu sur le tableau des interactions d'ordre 3 dans l'annexe 7).

Lorsqu'on introduit les facteurs recommandation et aval dans ce modèle, on calcule l'effet de l'interaction produite : $\text{genre*recomm*aval*fondep*encm}$. Il est de 0,069. Résultat lu sur le tableau des interactions d'ordre 5 dans l'annexe 7. On observe donc une amélioration de l'estimation de la variance du taux de respect des échéances après introduction des facteurs recommandation et aval. L'indicateur Eta au carré partiel est passé de 0,5% à 6,9%. Il y a donc un effet multiplicateur.

Mais une distinction doit être faite à propos de la contribution des facteurs recommandation et aval dans la variance du taux de respect des échéances. D'abord, en considérant les effets principaux, on note que le facteur aval contribue plus à la variance du taux de respect (voir pour cela la figure 12.6). Le facteur recommandation n'ayant pas d'effet principal significatif.

Ensuite, on peut constater l'importance de l'effet du facteur aval par rapport à l'effet du facteur recommandation sur l'explication de la variance du taux de respect. Pour cela, on compare la taille de l'effet de ses deux facteurs dans l'interaction genre*fondep*encm . En prenant les données du tableau d'interaction d'ordre 4 dans l'annexe 7 on n'obtient pas de résultat probant car les effets lus sont égaux à zéro. Les lectures faites sur la figure 12.7 montrent que : si on introduit le facteur recommandation dans ce modèle de base, l'effet ne s'améliore pas. Au contraire, il baisse, passant de 0,005 à 0,000. Or, l'introduction du facteur aval améliore la taille de l'effet qui passe de 0,001 à 0,002. Le facteur qui contribue le plus à la variance du taux de respect est le facteur aval comparativement au facteur recommandation.

les clients et la pénalité : (niveau 5) dans la figure 12.7

L'analyse de l'effet de l'interaction : $\text{genre*fondep*encm*penalim}$ doit être faite en comparant avec l'effet de l'interaction : genre*fondep*encm . Cela revient à comparer le comportement des clients par rapport aux pénalités, c'est-à-dire d'un côté on observe la variance du

taux de respect des échéances sans les pénalités de retard et de l'autre on observe la variance du taux de respect une fois introduit les pénalités dans le modèle. La taille de l'effet initial c'est-à-dire le cas sans pénalité est de 0,005. Lecture faite sur le tableau des interactions d'ordre 3 dans l'annexe 7). L'introduction de la co-variable penalim améliore sensiblement la taille de l'effet qui passe à 0,178. La pénalité a de ce point de vue un effet multiplicateur évident.

les clients recommandés ou non et avalisés ou non et la pénalité : (niveau 8) dans la figure 12.7

L'analyse de l'interaction : genre*recomman*aval*fondepm*encm*penalim permet de montrer un effet "paradoxal" des facteurs recommandation et aval sur la variance du taux de respect des échéances en incluant les pénalités de retard. C'est le niveau 8 de notre analyse; ce niveau est parfaitement complémentaire au niveau 4 où les clients recommandés et ou avalisés ne sont pas concernés par les pénalités. Si on introduit la variable pénalité dans l'estimation de l'effet de respect des échéances chez les clients recommandés et ou avalisés, on constate que l'effet décroît passant de 0,003 à 0,000. De ce point de vue, la variable pénalité contribuerait à dégrader la qualité du phénomène de respect des échéances chez les clients recommandés et ou avalisés. Or, nous montrons au niveau 5 de l'analyse que la variable penalim contribue à l'amélioration de l'effet du facteur genre indépendamment des facteurs recommandation et aval. Alors nous devons vérifier si la variable penalim comprime l'effet des facteurs recommandation et aval. Le point de départ c'est qu'on considère les effets d'interaction d'ordre 2 de chacun des trois facteurs avec la co-variable penalim (voir le tableau des interactions d'ordre 2 dans l'annexe 7). On constate que la variable penalim améliore l'effet de chacun des trois facteurs par interaction avec l'un ou l'autre. En effet, l'effet principal du facteur genre = 0,013. L'effet de l'interaction genre*penalim = 0,020. L'effet principal du facteur aval = 0,013. L'effet de l'interaction aval*penalim = 0,047. L'effet principal du facteur recommandation est non significatif. Mais l'effet de l'interaction recommandation*penalim est significatif et = 0,019.

La dépréciation de l'effet qui est observée en passant du niveau d'analyse 4 au niveau d'analyse 8 (l'effet passe de 0,003 à 0,000) peut être imputable soit à l'interaction des facteurs genre*recomman ou genre*aval sur la variable penalim soit aux variables fondepem et encm.

- *Analyse des effets des modèles intra sujets (ou intra-classes)*

Pour l'analyse des effets intra-classes des facteurs genre, recommandation et aval, nous comparons deux situations, l'une où il n'y a pas de pénalité et l'autre où il y a la pénalité. Sans les pénalités de retard, l'ensemble du modèle explique seulement 5,4% de la variance du taux de respect des échéances. Avec les pénalités, le modèle explique 48,7% de la variance du

taux de respect des échéances. Cela montre que les pénalités de retard constituent la principale variable de contrôle dans notre modèle. D'ailleurs, nous avons mentionné précédemment que la variable penalim a l'effet principal le plus élevé du modèle avec une valeur de Eta au carré partiel = 0,364.

Nous prenons en compte les résultats du tableau des interactions d'ordre 5 présentés dans l'annexe 7. Les pénalités sont exclues. Nous prenons en compte également les résultats du tableau des interactions d'ordre 6 présentés dans l'annexe 7 où figurent les pénalités.

première situation : on n'applique pas la pénalité

Voir la figure 12.8 ci-dessous.

FIG. 12.8 – Résumé des effets intra-classes : pénalités exclues

Cas sans pénalités	Sig. (t) Student	Etat au carré partiel	Puissance observée	Comparaison Eta carré partiel
genre*recomman*aval*encm*fondep	0,000	0,069	1,000	
[genre=1]*[recomman=1]*[aval=1]*encm*fondep	0,000	0,039	1,000	> H-R-A
[genre=1]*[recomman=1]*[aval=2]*encm*fondep	0,018	0,001	0,658	< F-R-A < H-R-NA
[genre=1]*[recomman=2]*[aval=1]*encm*fondep	0,000	0,046	1,000	< F-R-A > H-NR-A
[genre=1]*[recomman=2]*[aval=2]*encm*fondep	nc	nc	nc	nc
[genre=2]*[recomman=1]*[aval=1]*encm*fondep	0,000	0,007	1,000	< F-R-A
[genre=2]*[recomman=1]*[aval=2]*encm*fondep	0,000	0,040	1,000	> H-R-A > F-R-NA
[genre=2]*[recomman=2]*[aval=1]*encm*fondep	0,000	0,012	1,000	> H-NR-NA < F-NR-A
[genre=2]*[recomman=2]*[aval=2]*encm*fondep	0,276	0,000	0,193	

Chez les femmes recommandées et avalisées : ([genre=1]*[recomman=1]*[aval=1]*encm*fondep), Eta au carré partiel = 0,039 (soit 3,9%). C'est-à-dire que 3,9% de la variance du taux de respect des échéances sont dus aux femmes recommandées et avalisées indépendamment

de la pénalité. Or, nous savons que la valeur de Eta au carré partiel pour l'interaction : $\text{genre}^*\text{recomman}^*\text{aval}^*\text{fondep}^*\text{encm} = 6,9\%$ (lue sur le tableau des interactions d'ordre 5 dans l'annexe 7). C'est-à-dire que la contribution des clients recommandés et avalisés à la variance du taux de respect est de 6,9% quelque soit le sexe. Si les femmes recommandées et avalisées ont une contribution de 3,9%, on doit dire que ces dernières participent plus à la variance du taux de respect des échéances par rapport aux hommes recommandés et avalisés ; d'où $F.R.R > H.R.A$

Chez les femmes recommandées et non avalisées : $([\text{genre}=1]^*[\text{recomman}=1]^*[\text{aval}=2]^*\text{encm}^*\text{fondep})$, Eta au carré partiel = 0,001 (soit 0,1%). Or Eta au carré partiel pour les femmes recommandées et avalisées = 3,9%. Sur cette base, on peut conclure sur les femmes recommandées et non avalisées ont une contribution moindre à la variance du taux de respect des échéances par rapport aux femmes recommandées et avalisées d'où $F.R.NA < F.R.A$. Par ailleurs, la même catégorie chez les hommes c'est-à-dire les hommes recommandés et non avalisés ont une valeur de Eta au carré partiel plus élevée, d'où : $F.R.NA < H.R.NA$

Chez les femmes non recommandées et avalisées : $([\text{genre}=1]^*[\text{recomman}=2]^*[\text{aval}=1]^*\text{encm}^*\text{fondep})$, Eta au carré partiel = 0,046 (soit 4,6%). Or, Eta au carré partiel pour les femmes recommandées et avalisées = 3,9%. Donc, $F.NR.A < F.R.A$. Par ailleurs, les hommes de la même catégorie c'est-à-dire non recommandés et avalisés ont une valeur de Eta au carré partiel = 1,2%. Par conséquent, $F.NR.A > H.NR.A$

Les hommes recommandés et avalisés : $([\text{genre}=2]^*[\text{recomman}=1]^*[\text{aval}=1]^*\text{encm}^*\text{fondep})$, ont un apport inférieur à la variance du taux de respect des échéances par rapport aux femmes recommandées et avalisés. $H.R.A < F.R.A$

Les hommes recommandés et non avalisés : $([\text{genre}=2]^*[\text{recomman}=1]^*[\text{aval}=2]^*\text{encm}^*\text{fondep})$, ont une contribution à la variance du taux de respect des échéances qui est supérieure par rapport aux hommes recommandés et avalisés, ainsi qu'aux femmes recommandées et non avalisées.

Le test sur les hommes non recommandés et non avalisés n'est pas significatif. De même, il n'y a pas de données sur les femmes non recommandées et non avalisées.

La dernière remarque que nous pouvons faire sur la figure 12.8, c'est que les effets des facteurs aléatoires : recommandation et aval ne sont pas facilement réparables chez les clients selon leur sexe. Par exemple, les données ne permettent pas clairement de dire que le facteur recommandation est plus important que le facteur aval chez les femmes.

deuxième situation : on applique la pénalité

Voir la figure 12.9 ci-dessous.

FIG. 12.9 – Résumé des effets intra-classes, pénalités incluses

Cas avec pénalités	Sig. (t) Student	Etat au carré partiel	Puissance observée	Comparaison Eta carré partiel
genre*recomman*aval*encm*fondepm	0,000	0,579	1,000	
[genre=1]*[recomman=1]*[aval=1]*encm*fondepm*penalim	0,000	0,422	1,000	> H-R-A
[genre=1]*[recomman=1]*[aval=2]*encm*fondepm*penalim	0,000	0,074	1,000	< F-R-A < H-R-NA
[genre=1]*[recomman=2]*[aval=1]*encm*fondepm*penalim	0,000	0,011	1,000	< F-R-A < H-NR-A
[genre=1]*[recomman=2]*[aval=2]*encm*fondepm*penalim	nc	nc	nc	nc
[genre=2]*[recomman=1]*[aval=1]*encm*fondepm*penalim	0,000	0,060	1,000	< F-R-A
[genre=2]*[recomman=1]*[aval=2]*encm*fondepm*penalim	0,000	0,235	1,000	> H-R-A > F-R-NA
[genre=2]*[recomman=2]*[aval=1]*encm*fondepm*penalim	0,000	0,034	1,000	> H-NR-NA < F-NR-A
[genre=2]*[recomman=2]*[aval=2]*encm*fondepm*penalim	0,300	0,000	0,179	

La variable de contrôle *penalim* contribue fortement à améliorer l'effet des facteurs genre, recommandation et aval.

Le cas des femmes recommandées et avalisées : ($[\text{genre}=1][\text{recomman}=1][\text{aval}=1]\text{encm}^*\text{fondep}^*\text{penalim}$) montre que $\text{Eta au carré partiel} = 0,422$ (soit 42,2%). Mais, sachant que $\text{Eta au carré partiel}$ pour les clients recommandés et avalisés = 0,579 quelque soit le sexe, on peut alors dire que les femmes recommandées et avalisées ont une influence déterminante sur la variance du taux de respect des échéances. D'où : F.R.A > H.R.A. L'introduction de la variable *penalim* a fait grossir l'effet des facteurs mais aussi on doit remarquer qu'il s'est produit un glissement de l'effet des facteurs aléatoires recommandation et aval au profit du facteur fixe : genre. Et c'est principalement sur les femmes qu'on remarque cela. La proportion de Eta carré partiel des femmes recommandées et avalisées passe de 56,5% (c'est-à-dire 0,039 sur 0,069) à 72,8% (c'est-à-dire 0,422 sur 0,579).

Le cas des femmes recommandées et non avalisées : ($[\text{genre}=1][\text{recomman}=1][\text{aval}=2]\text{encm}^*\text{fondep}^*\text{penalim}$) montre une valeur de $\text{Eta au carré partiel} = 0,074$ (soit 7,4%). C'est très inférieur à la valeur correspondante pour les femmes recommandées et avalisées. Les hommes de la même catégorie, c'est-à-dire recommandés et non avalisés ont une valeur de $\text{Eta au carré partiel}$ qui est supérieur.

Les femmes non recommandées et avalisées : ($[\text{genre}=1][\text{recomman}=2][\text{aval}=1]\text{encm}^*\text{fondep}^*\text{penalim}$) ont une influence inférieure sur la variance du taux de respect des échéances par rapport aux femmes recommandées et avalisées ainsi qu'aux hommes non recommandés et avalisés.

Du côté des hommes, l'influence de ceux qui sont recommandés et avalisés est moindre par rapport aux femmes recommandées et avalisées. D'ailleurs, parmi les hommes recommandés, ceux qui ne sont pas avalisés ont une influence plus importante que ceux qui sont avalisés. L'influence des hommes non recommandés est faible et devient non significative lorsque ces clients sont non avalisés.

12.4 Conclusion et recommandations

Pour aider la décision en matière de gestion de portefeuille de crédit, l'analyse suivante peut être faite. Lorsque les analyses montrent qu'une catégorie de clients a une influence significative sur la variance du taux de respect des échéances, c'est-à-dire quand la valeur de $\text{Eta au carré partiel}$ est significativement différente de zéro ; dans ce cas, pour la qualité du portefeuille de crédit, on doit vérifier que :

1. la catégorie identifiée a une moyenne de taux de respect des échéances qui est supérieure (respectivement inférieure) à celle d'une autre catégorie ; exemple : les femmes contre les hommes
2. l'effectif de la catégorie identifiée est suffisamment élevé (respectivement faible) par rap-

port à celui de l'autre catégorie

3. la différence d'influence sur la variance du taux de respect des échéances (valeur de Eta au carré partiel) entre les deux catégories est de même sens que le test du t de student pour la comparaison des moyennes des échantillons indépendants sur les deux catégories
4. si la différence des Eta au carré partiel est de même sens que le test de student, alors on conclura qu'il n'y a pas de politique particulière à mener. En revanche, si la différence des Eta au carré partiel est de sens contraire que le test de student, alors une action doit être menée pour corriger les écarts. Nous donnons ci-dessous quelques recommandations à partir des résultats de nos analyses.

La figure 12.10 résume ces résultats et les recommandations possibles qui peuvent être faites.

FIG. 12.10 – Résumé des analyses de données

		Désignation	Eta au carré partiel	Moyenne	comparaison Eta au carré partiel	test student	Cas	
Féminin	Recom mandé	avalisés	F.R.A	0,422	0,9808	F.R.A > H.R.A	F.R.A < H.R.A	1
		non avalisés	F.R.NA	0,074	0,9838	F.R.NA < H.R.NA	F.R.NA > H.R.NA	2
	non recom mandé	avalisés	F.NR.A	0,011	0,9583	F.NR.A < H.NR.A	F.NR.A < H.NR.A	3
		non avalisés	F.NR.NA	nc	nc	nc	nc	4
Masculin	Recom mandé	avalisés	H.R.A	0,06	0,9891	H.R.A < F.R.A	H.R.A > F.R.A	5
		non avalisés	H.R.NA	0,235	0,9764	H.R.NA > F.R.NA	H.R.NA < F.R.NA	6
	non recom mandé	avalisés	H.NR.A	0,034	0,9644	H.NR.A > F.NR.A	H.NR.A > F.NR.A	7
		non avalisés	H.NR.NA	sig > 0,05	0,9969			8

- **Conclusion 1** : on compare les deux catégories suivantes : les femmes recommandées et avalisées (F.R.A) et les hommes recommandés et avalisés (H.R.A)

- comparaison de Eta au carré partiel : F.R.A > H.R.A (0,422 > 0,060)

- t de student : F.R.A < H.R.A (0,9808 < 0,9891)

- effectif : F.R.A > H.R.A (5996 > 1215)

Interprétation : les femmes recommandées et avalisées ont une influence prépondérante sur la variance du taux de respect des échéances. Or, elles présentent une moyenne significativement inférieure à celle des hommes recommandés et avalisés. De plus, elles sont en effectif supérieur aux hommes.

Recommandations :

1. réduction de l'effectif des femmes recommandées et avalisées et ou augmentation de l'effectif des hommes recommandés et avalisés. On peut utiliser les renouvellements de crédits comme leviers. C'est le jeu de la bascule.
2. Test de l'effet du facteur recommandation et du facteur aval puis prise de décision pour le jeu de la bascule. Par exemple, on compare les catégories suivantes : les femmes recommandées (F.R) contre les hommes recommandés (H.R) et on compare les femmes avalisées (F.A) contre les hommes avalisés (H.A). Les tests de student ont été précédemment réalisés, voir la figure 9.9. Ce qui veut dire qu'on connaît les moyennes où on a : F.R < H.R et F.A < H.A et on connaît les effectifs. Il reste à comparer les valeurs de Eta au carré partiel. Ces valeurs sont contenues dans le tableau des interactions d'ordre 5 dans l'annexe 7. Pour F.R (Eta carré = 0,378), pour H.R (Eta carré = 0,149), donc F.R > H.R. Il serait logique de recommander une diminution de l'effectif des femmes recommandées et d'augmenter l'effectif des hommes recommandés. De même, pour F.A (Eta carré = 0,000), pour H.A (Eta carré = 0,003), donc F.A < H.A. Une recommandation ne s'impose pas.

- **Conclusion 2** : Les femmes recommandées et non avalisées ont une moindre influence sur la variance du taux de respect des échéances. Or, elles ont une moyenne significativement supérieure aux hommes recommandés et non avalisés. Il peut être conseillé d'augmenter l'effectif des femmes recommandées et non avalisées sans qu'il soit nécessaire de modifier l'effectif des hommes recommandés et non avalisés.

- **Conclusion 3** : Les femmes non recommandées et avalisées ont un effet moins important sur la variance du taux de respect des échéances par rapport aux hommes non recommandés et avalisés. De plus, les femmes non recommandées et avalisées ont une moyenne significativement inférieure à celle des hommes non recommandés et avalisés. Une recommandation n'est pas utile.

- **Conclusion 5** : Les hommes recommandés et avalisés ont un effet moins important sur la variance du taux de respect des échéances par rapport aux femmes recommandées et avalisées. Or, ils ont une moyenne significativement supérieure à celle des femmes recommandées et avalisées. Par conséquent, il peut être conseillé d'augmenter l'effectif des hommes recommandés et avalisés en baissant parallèlement celui des femmes recommandées et avalisés.

- **Conclusion 6** : Idem que conclusion 1.

- **Conclusion 7** : Idem que conclusion 3.

12.5 Constatation avec la population totale

La population totale des clients en crédits individuels est de 21146 clients.

Les valeurs de Eta au carré partiel par chacune des huit catégories de clients sont proposées dans le tableau des interactions d'ordre 6 dans l'annexe 7. Le modèle fourni dans cette annexe explique 57,9% de la variance du taux de respect des échéances pour l'ensemble de la population. On constate que globalement les femmes ont les scores les plus élevés en ce qui concerne l'influence de la variance par rapport aux hommes. Par exemple pour les femmes recommandées et avalisées : F.R.A (Eta au carré = 0,389). Les femmes non recommandées et non avalisées : F.NR.NA (Eta au carré = 0,337). Entre ces deux bords, les femmes recommandées et non avalisées et respectivement les femmes non recommandées et avalisées ont des scores de Eta presque similaires (Eta = 0,131 et respectivement 0,135). Cela invite à concentrer l'analyse du comportement de respect des échéances pour les femmes sur les facteurs recommandation et aval conjointement. D'autre part, on constate que le score le plus élevé chez les hommes est produit par la catégorie des hommes non recommandés et non avalisés.

Les effectifs ainsi que les moyennes de taux de respect des échéances pour les différentes catégories de clients sont aussi fournis par l'annexe 7³.

Les résultats sont présentés par la figure 12.11

³Les chiffres proposés dans l'annexe 7 sont calculés avec une pondération et sont légèrement différents de ceux obtenus par le test de student.

FIG. 12.11 – Résultat d'analyse avec les données de la population totale

t student	F.R.A	F.R.NA	F.NR.A	F.NR.NA
	< H.R.A	< H.R.NA	< H.NR.A	< H.NR.NA
Moyenne	0,9217 < 0,9571	0,8660 < 0,9597	0,8301 < 0,9062	0,7638 < 0,8520
Effectif	13120 > 1623	756 < 2531	342 < 418	845 < 1511
Eta carré	F.R.A	F.R.NA	F.NR.A	F.NR.NA
	> H.R.A	> H.R.NA	> H.NR.A	> H.NR.NA
Valeur Eta	0,389 > 0,075	0,131 > 0,034	0,135 > 0,037	0,337 > 0,261
Conclusions	1	2	3	4
Décisions	(-) F.R.A (+) H.R.A	(-) F.R.NA (+) F.R.A (+) H.R.NA	(-) F.NR.A (+) F.R.A (-) H.NR.A (+) H.R.A	(-) F.NR.NA (+) F.R.A (-) H.NR.NA

- **Conclusion 1** : Ce cas présente tous les éléments d'analyse d'un effet néfaste sur le taux de respect des échéances. En effet, les femmes recommandées et avalisées ont une grande influence sur la variance du taux de respect des échéances par rapport aux hommes. Pourtant, elles ont une moyenne significativement inférieure à celle des hommes recommandés et avalisés. En plus, elles sont en effectif très supérieur à celui des hommes. Donc, il y a à craindre que le mauvais effet soit amplifié par le grand nombre de clientes recommandées et avalisées.

Recommandation : il faudrait réduire le nombre de femmes recommandées et ou avalisées pour réduire leur effet. Ou bien il faudrait augmenter le nombre d'hommes recommandés et avalisés pour contrebalancer leur effet positif avec celui des femmes recommandées et avalisées. C'est le jeu de la bascule.

- **Conclusion 2** : Les hommes recommandés et non avalisés ont une moyenne supérieure aux femmes de la même catégorie mais ils ont une influence dérisoire sur la variance du taux de respect des échéances. L'instauration de la bascule pourrait améliorer la contribution des hommes à la variance du taux de respect. Or, les hommes sont déjà en effectif supérieur. Donc la bascule serait une politique à efficacité limitée. Dans ce cas, on compare les moyennes et

les valeurs de Eta carré pour les hommes et les femmes recommandées et non avalisés avec d'autres catégories.

Décision : On diminuera l'effectif des femmes recommandées et non avalisés au profit des femmes recommandées et avalisées. L'effectif des hommes recommandés et non avalisés peut être augmenté accessoirement.

- **Conclusion 3** : Décision : on préconise une diminution de l'effectif des femmes non recommandées et avalisées au profit des femmes recommandées et avalisés. De même, on préconise une diminution de l'effectif des hommes non recommandés et avalisés au profit des hommes recommandés et avalisés.

- **Conclusion 4** : Les hommes et les femmes ne sont ni recommandés ni avalisés. Les hommes qui sont en effectif supérieur ont aussi une moyenne significativement supérieure aux femmes. Mais ces dernières ont un plus grand impact sur la variance du taux de respect des échéances.

Décision : on préconise une diminution de l'effectif des femmes non recommandées et non avalisées au bénéfice des femmes recommandées et avalisés. Parallèlement, on préconise une diminution de l'effectif des hommes non recommandés et non avalisés.

Bilan de la quatrième partie de la thèse

La quatrième partie de notre thèse avait pour but d'analyser statistiquement le rapport probable entre les relations annexes et les comportements de respect des clients vis-à-vis des échéances de crédits. Les différents chapitres de cette partie ont utilisé les tests statistiques pour montrer que les relations annexes des clients telles que les relations de recommandation (ou parrainage) et les relations d'aval sont en mesure d'influencer de manière significative les comportements de respect des clients vis-à-vis des échéances.

Plusieurs résultats intermédiaires ont été obtenus dans les différents chapitres de cette partie. Nous en reprendrons deux. Ce sont :

- d'une part, l'importance des comportements de respect des échéances chez les clients de sexe masculin ; un fait important qui est pourtant largement sous-estimé,
- et d'autre part, l'importance des relations annexes de recommandation et d'aval dans l'analyse du comportement de respect des échéances chez les clients qui ont les "bons" taux de respect des échéances et plus largement chez l'ensemble de la clientèle en crédits individuels.

Pour clore cette quatrième partie de la thèse, une question s'impose ; c'est de savoir pourquoi les relations annexes de recommandation (ou parrainage) et les relations d'aval interviennent de manière significative dans la compréhension du comportement de respect des échéances chez les clients individuels. Les éléments de réponse à cette question constituent l'essentiel de la cinquième et dernière partie de la thèse. Cette partie cherche à comprendre pourquoi le comportement de respect des échéances (tel qu'il est montré dans la quatrième

partie) évolue ainsi ? Qu'est-ce qui justifie ce comportement ? Qu'est-ce qui l'explique ?