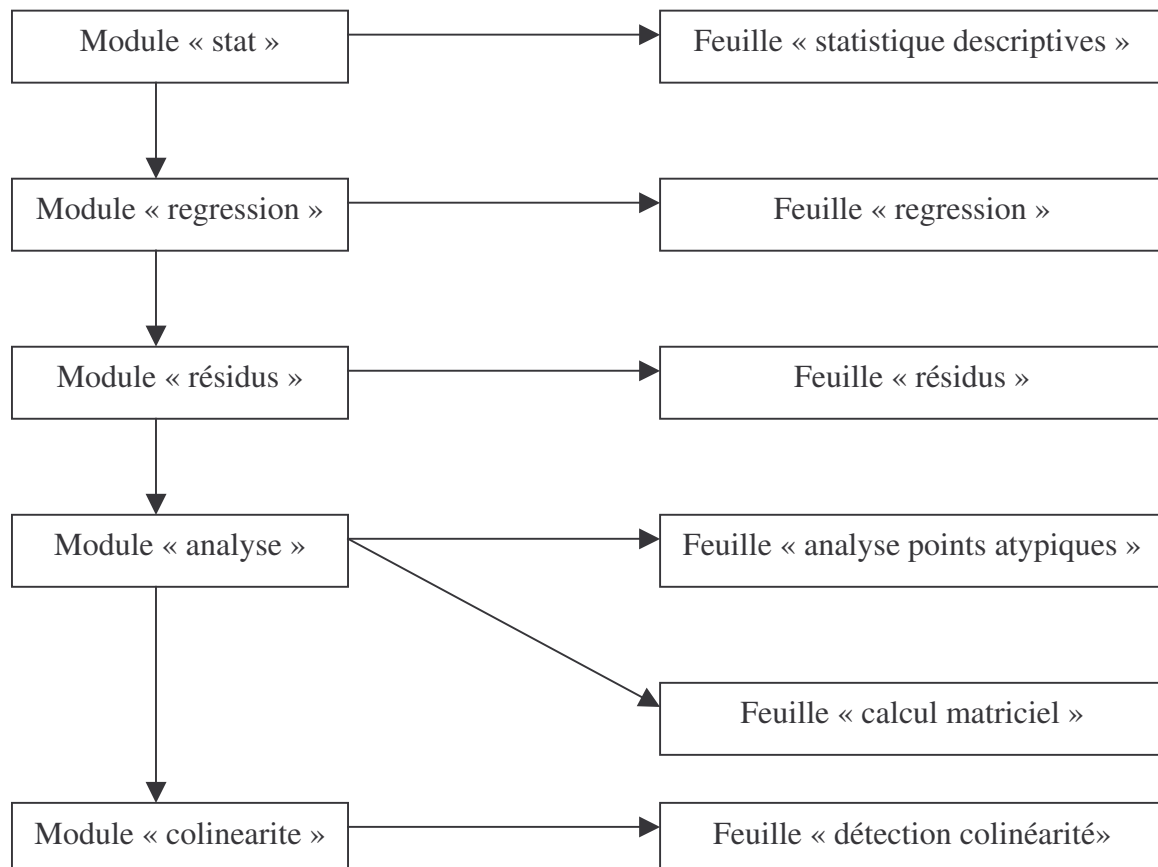


1. Introduction

Ce manuel de l'utilisateur a été extrait du rapport de stage rédigé par Mlle **NGUYEN LAO Bao Truc**, en stage au sein du laboratoire ERIC de Juillet à Septembre 2005.

Son travail consistait à élaborer des macros destinées à automatiser les principales étapes de la régression multiple dans le tableur EXCEL. Les opérations décrites sont en relation directe avec le cours d'économétrie dispensé en Licence IDS.

L'application est distribuée sous la forme d'un classeur XLS, elle comporte un exemple de données et une série de macros regroupées dans des modules organisés de la manière suivante :

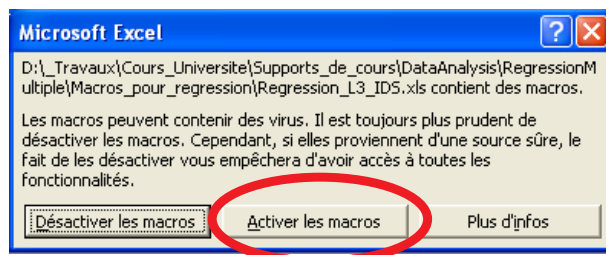


Dans ce qui suit, nous reprenons les grandes lignes du rapport de Mlle **NGUYEN LAO Bao Truc** en apportant quelques éclaircissements lorsque cela est nécessaire.

2. Manuel de l'utilisateur

2.1. Ouverture du classeur

Ouvrez le classeur « Regression_L3_IDS.XLS » après l'avoir récupéré le classeur sur le site web. Attention, lors de l'ouverture, une boîte de dialogue apparaît, demandant si l'on souhaite activer les macros ou pas sur ce classeur : il est absolument nécessaire d'activer les macros sur l'ensemble du classeur, sinon le programme ne fonctionnera pas.



Activer les macros pour que le programme de régression puisse s'exécuter.

2.2. Préparation des données

Par défaut, le classeur contient un jeu de données type, correspondant au formatage que vous devrez adopter pour les calculs sur vos propres données.

La configuration à respecter scrupuleusement est la suivante :

- Le classeur ne doit comporter qu'une seule feuille.
- Les données doivent être dans la feuille « Feuil1 », il n'est pas possible de renommer cette feuille.
- La première colonne (colonne A) des données correspond aux étiquettes des observations. Si vous disposez d'un jeu de données sans labels, vous devez en créer une, fictive, en mettant par exemple des numéros d'observation.
- La première ligne (ligne 1) correspond au nom des variables.
- La dernière variable à droite correspond à la variable endogène, les autres colonnes correspondent aux variables exogènes. Le nombre de variables exogènes n'est pas limité, le programme s'adapte automatiquement à votre sélection.
- Attention, mis à part la colonne des étiquettes, toutes les données doivent être numériques.

1	A	B	C	D	E	F	G
1	Cigarette	TAR (mg)	NICOTINE (mg)	WEIGHT (g)	CO (mg)		
2	Alpine	14.1	0.86	0.9853	13.6		
3	Benson&Hedges	16	1.06	1.0938	16.6		
4	CamelLights	8	0.67	0.928	10.2		
5	Carlton	4.1	0.4	0.9462	5.4		
6	Chesterfield	15	1.04	0.8885	15		
7	GoldenLights	8.8	0.76	1.0267	9		
8	Kent	12.4	0.95	0.9225	12.3		
9	Kool	16.6	1.12	0.9372	16.3		
10	L&M	14.9	1.02	0.8858	15.4		
11	LarkLights	13.7	1.01	0.9643	13		
12	Marlboro	15.1	0.9	0.9316	14.4		
13	Merit	7.8	0.57	0.9705	10		
14	MultiFilter	11.4	0.78	1.124	10.2		
15	NewportLights	9	0.74	0.8517	9.5		
16	Now	1	0.13	0.7851	1.5		
17	OldGold	17	1.26	0.9186	18.5		
18	PallMallLight	12.8	1.08	1.0395	12.6		
19	Raleigh	15.8	0.96	0.9573	17.5		
20	SalemUltra	4.5	0.42	0.9106	4.9		
21	Tareyton	14.5	1.01	1.007	15.9		
22	VV	7.3	0.61	0.9806	8.5		
23	ViceroyRichLight	8.6	0.69	0.9693	10.6		
24	VirginiaSlims	15.2	1.02	0.9496	13.9		
25	WinstonLights	12	0.82	1.1184	14.9		

Nom des variables
(Ligne 1)

Étiquette des observations
(Colonne A)

Variables exogènes
(colonne B à ...)

Variable endogène
(Dernière colonne dans la plage de données)

Pour réaliser les calculs sur vos propres données, il est donc nécessaire de supprimer les valeurs pré existantes dans la feuille, et d'y placer votre tableau, en respectant la configuration initiale.

NOTA : Bien entendu, votre tableau peut comporter un nombre d'observations différent (65535 observations maximum). De même, vous pouvez travailler sur un fichier comportant un nombre largement plus élevé de variables exogènes (253 variables exogènes maximum).

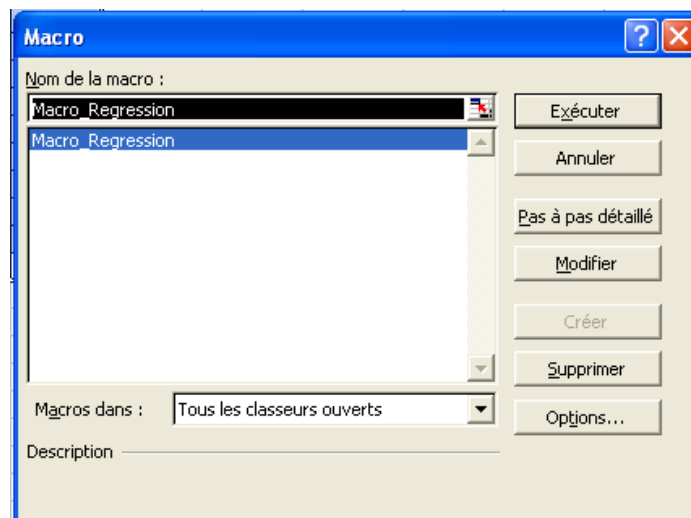
2.3. Exécution du programme

Avant d'exécuter le programme, assurez-vous que le classeur ne comporte que la seule feuille « Feuil1 ». **En cas de ré-exécution de la macro, il est nécessaire de supprimer les feuilles qui ont été générées automatiquement par l'analyse.**

Pour lancer l'analyse complète, sélectionnez la plage de données entière, y compris les étiquettes et les noms de variables.

1	Cigarette	TAR (mg)	NICOTINE (mg)	WGEICHT (g)	CO (mg)
2	Alpine	14.1	0.86	0.9853	13.6
3	Benson&Hedges	16	1.06	1.0938	16.6
4	CamelLights	8	0.67	0.928	10.2
5	Carlton	4.1	0.4	0.9462	5.4
6	Chesterfield	15	1.04	0.8885	15
7	GoldenLights	8.8	0.76	1.0267	9
8	Kent	12.4	0.95	0.9225	12.3
9	Kool	16.6	1.12	0.9372	16.3
10	L&M	14.9	1.02	0.8858	15.4
11	LarkLights	13.7	1.01	0.9643	13
12	Marlboro	15.1	0.9	0.9316	14.4
13	Merit	7.8	0.57	0.9705	10
14	MultiFilter	11.4	0.78	1.124	10.2
15	NewportLights	9	0.74	0.8517	9.5
16	Now	1	0.13	0.7051	1.5
17	OldGold	17	1.26	0.9186	18.5
18	PallMallLight	12.8	1.08	1.0395	12.6
19	Raleigh	15.8	0.96	0.9573	17.5
20	SalemUltra	4.5	0.42	0.9106	4.9
21	Tareyton	14.5	1.01	1.007	15.9
22	VV	7.3	0.61	0.9806	8.5
23	ViceroyRichLight	8.6	0.69	0.9593	10.6
24	VirginiaSlims	15.2	1.02	0.9496	13.9
25	WinstonLights	12	0.82	1.1184	14.9

Activez le menu OUTILS / MACRO / MACROS :



Il est dès lors possible d'exécuter l'analyse de régression en sélection la procédure MACRO_REGRESSION et cliquer le bouton « Exécuter ».

La rapidité du calcul dépend fortement du nombre de variables et du nombre d'observations, l'exécution est un peu saccadée car plusieurs feuilles sont générées et le programme les manipule simultanément.

	A	B	C	D	E	F	G	H	I	J
1		TAR (mg)	NICOTINE (mg)	WEIGHT (g)						
2	TAR (mg)	1	0.95989525	0.28351665						
3	NICOTINE (mg)		1	0.28610572						
4	WEIGHT (g)			1						
5		carré de la matrice de corrélation								
6		TAR (mg)	NICOTINE (mg)	WEIGHT (g)						
7	TAR (mg)	1	0.9213969	0.08038169						
8	NICOTINE (mg)		1	0.08185648						
9	WEIGHT (g)			1						
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										

Feuille initiale contenant les données

Feuilles de résultats, générées automatiquement

2.4. Les feuilles de résultats

Plusieurs feuilles sont générées automatiquement, observons-les tour à tour en détaillant leur contenu.

Statistiques descriptives

Cette feuille présente les statistiques descriptives sur les données, plusieurs indicateurs sont disponibles, elles permettent de se donner une idée sur les caractéristiques des données. Par exemple, si l'écart type est égal à zéro, cela veut dire que la variable est composée d'une seule valeur, il s'agit d'une constante, elle est inutilisable pour la régression.

	A	B	C	D	E
1	Cigarette	TAR (mq)	NICOTINE (mg)	WEIGHT (g)	CO (mq)
2	Moyenne	11.48333333	0.828333333	0.96217083	12.0708333
3	Ecart-type	4.41515834	0.26558658	0.07945121	4.24135737
4	Min	1	0.13	0.7851	1.5
5	Max	17	1.26	1.124	18.5
6	Médiane	12.6	0.88	0.95345	12.8

Régression

Cette feuille retrace les résultats de la régression, elle complète les statistiques standard d'EXCEL en ajoutant les tests classiques de signification des coefficients.

	A	B	C	D	E
2		WEIGHT (g)	NICOTINE (mg)	TAR (mg)	constante
3	a	2.07934422	0.51846956	0.80758035	-0.55169763
4	sigma a	3.17841712	3.25233113	0.19548169	2.97128094
5		0.93497531	1.15982622	#N/A	#N/A
6		95.8584963	20	#N/A	#N/A
7		386.845646	26.9039373	#N/A	#N/A
9		Test de significativité individuelle de coefficients			
10	t de Student	0.65420747	0.15941475	4.54047821	
11	t absolu	0.65420747	0.15941475	4.54047821	
12	p-value	0.52043066	0.87494102	0.00019909	
14		Tableau d'analyse de variance			
15		SC Expliquée	SC	Degrés de lib	Carrés moyens
16			386.845646	3	128.948549
17		SC Résiduelle	26.9039373	20	1.34519686
18		SC Totale	413.749583	23	
20		Evaluation globale de la régression			
21	F	95.8584963			
22	DDL1	3			
23	DDL2	20			
24	p-value	4.8503E-12			

Résidus

La feuille « Résidus » reprend la projection des valeurs à partir des coefficients de la droite de régression et produit les résidus. Plusieurs indicateurs statistiques sont calculés, les coefficients d'asymétrie et d'aplatissement permettent d'évaluer, de manière très approximative (très asymptotiquement), la normalité des résidus.

La moyenne des résidus est forcément égale à zéro pour une régression avec constante

P-value largement supérieure à zéro : la conformité à la loi normale de l'erreur n'est pas contredite par les résidus observés

	C	D	E	F	G	H	I	J	K	L	M
1	NICOTINE (mg)	WEIGHT (g)	CO (mg)	Prédiction	Résidus						
2	0.86	0.9853	13.6	14.458	-0.8578	Indicateur		Valeur	Ecart-type	Statistique	p-value
3	1.06	1.0938	16.6	16.474	0.1264	Observations		24			
4	0.67	0.928	10.2	8.826	1.374	Moyenne		4.5E-15			
5	0.4	0.9462	5.4	5.2622	0.1378	Ecart-type		1.08154			
6	1.04	0.8885	15	15.149	-0.1487	Skewness		0.17177	0.44222	0.38842	0.6977
7	0.76	1.0267	9	9.7879	-0.7879	Kurtosis		-0.7123	0.73741	0.64048	0.52186
8	0.95	0.9225	12.3	12.865	-0.565	Lamda de bera Jarque		0.53688		0.53688	0.76457
9	1.12	0.9372	16.3	16.712	-0.4116						
10	1.02	0.8858	15.4	15.044	0.356						
11	1.01	0.9643	13	14.137	-1.1369						
12	0.9	0.9316	14.4	15.255	-0.8545						
13	0.57	0.9705	10	8.685	1.315						
14	0.78	1.124	10.2	12.308	-2.1083						
15	0.74	0.8517	9.5	9.5912	-0.0912						
16	0.13	0.7851	1.5	2.0358	-0.5358						
17	1.26	0.9186	18.5	17.101	1.3995						
18	1.08	1.0395	12.6	13.531	-0.9308						
19	0.96	0.9573	17.5	15.96	1.5396						
20	0.42	0.9106	4.9	5.5536	-0.6536						
21	1.01	1.007	15.9	14.936	0.9642						
22	0.61	0.9806	8.5	8.2829	0.2171						
23	0.69	0.9693	10.6	9.4547	1.1453						
24	1.02	0.9496	13.9	15.443	-1.5429						
25	0.82	1.1184	14.9	12.85	2.05						

Analyse des points atypiques

Cette feuille calcule une série d'indicateurs destinée à mettre en évidence l'existence éventuelle d'observations atypiques dans l'échantillon. Les indicateurs les plus intéressants sont le LEVERAGE qui indique une anomalie au niveau des variables exogènes, et le RSTUDENT qui indique une anomalie dans la régression. Les observations (atypiques) pour lesquelles la valeur de l'indicateur calculé est supérieure à la valeur seuil sont mis en évidence avec un formatage approprié.

Microsoft Excel - Regression_L3_IDS.xls

	A	B	C	D	E	F	G	H	I	J	K
1											
2									0.33333		2.09302
3	Cigarette	const	TAR (mg)	NICOTINE	WEIGHT	CO (mg)	Prédiction	Résidus	Leverage	R.Standar	RSTUDENT
4	Alpine	1	14.1	0.86	0.9853	13.6	14.4578	-0.85785	0.17073	0.81221	0.80503
5	Benson&Hedges	1	16	1.06	1.0938	16.6	16.4736	0.12645	0.18491	0.12076	0.11774
6	CamellLights	1	8	0.67	0.928	10.2	8.82595	1.37405	0.08571	1.23899	1.25681
7	Carlton	1	4.1	0.4	0.9462	5.4	5.26225	0.13775	0.16686	0.13012	0.12688
8	Chesterfield	1	15	1.04	0.8885	15	15.1487	-0.14871	0.13349	0.13774	0.13432
9	GoldenLights	1	8.8	0.76	1.0267	9	9.78791	-0.78791	0.15725	0.74	0.73135
10	Kent	1	12.4	0.95	0.9225	12.3	12.865	-0.56504	0.0979	0.51293	0.50326
11	Kool	1	16.6	1.12	0.9372	16.3	16.7116	-0.41158	0.11961	0.3782	0.36995
12	L&M	1	14.9	1.02	0.8858	15.4	15.044	0.35603	0.13361	0.32979	0.32232
13	LarkLights	1	13.7	1.01	0.9643	13	14.1369	-1.13692	0.07631	1.01993	1.02102
14	Marlboro	1	15.1	0.9	0.9316	14.4	15.2545	-0.85451	0.23146	0.84041	0.83398
15	Merit	1	7.8	0.57	0.9705	10	8.68496	1.31504	0.09483	1.19174	1.20514
16	MultiFilter	1	11.4	0.78	1.124	10.2	12.3083	-2.10831	0.25992	2.11301	2.3368
17	NewportLights	1	9	0.74	0.8517	9.5	9.59117	-0.09117	0.15539	0.08553	1.08338
18	Now	1	1	0.13	0.7851	1.5	2.03578	-0.53578	0.46035	0.62883	0.61906
19	OldGold	1	17	1.26	0.9186	18.5	17.1005	1.39947	0.2528	1.39822	1.43474
20	PallMallLight	1	12.8	1.08	1.0395	12.6	13.5308	-0.93076	0.3566	0.97006	0.96858
21	Raleigh	1	15.8	0.96	0.9573	17.5	15.9604	1.53964	0.19485	1.47941	1.52797
22	SalemUltra	1	4.5	0.42	0.9106	4.9	5.55362	-0.65362	0.13248	0.61215	0.60232
23	Tareyton	1	14.5	1.01	1.007	15.9	14.9358	0.96423	0.06874	0.86749	0.8557
24	VV	1	7.3	0.61	0.9806	8.5	8.28291	0.21709	0.08508	0.19587	0.19207
25	ViceroyRichLight	1	8.6	0.69	0.9693	10.6	9.45475	1.14525	0.06353	1.02367	1.02496
26	VirginiaSlims	1	15.2	1.02	0.9496	13.9	15.4429	-1.54291	0.0335	1.38957	1.42492
27	WinstonLights	1	12	0.82	1.1184	14.9	12.85	2.05005	0.23556	2.02162	2.20902
28											
29											

Seuils

Observations à considérer avec attention

Calcul matriciel

C'est une feuille qui sert uniquement pour les calculs intermédiaires, il ne faut pas la supprimer cependant car les résultats dans les autres feuilles seraient faussés.

Détection colinéarité

Cette feuille calcule le coefficient de corrélation et le carré du coefficient de corrélation entre les variables exogènes. Elle permet de détecter les éventuelles colinéarités qui risquent de perturber les résultats de la régression. Un indicateur simple : lorsque le carré du coefficient de corrélation entre deux variables exogènes est proche du coefficient de détermination de la régression, il y a matière à s'inquiéter.

The screenshot shows an Excel spreadsheet titled 'Microsoft Excel - Regression_I3_IDS.xls'. The spreadsheet contains a correlation matrix for three variables: TAR (mg), NICOTINE (mg), and WEIGHT (g). The matrix is as follows:

	TAR (mg)	NICOTINE (mg)	WEIGHT (g)
TAR (mg)	1	0.95989525	0.28351665
NICOTINE (mg)		1	0.28610572
WEIGHT (g)			1

Below the matrix, a callout box with an orange background and black text reads: "Danger ! Le R² de la régression est de 0.9349". An orange arrow points from this box to the correlation coefficient between TAR and NICOTINE (0.95989525). The spreadsheet also shows a 'EuroView' window and a status bar at the bottom indicating 'Prêt' and 'NUM'.

3. Conclusion

Cette feuille de calcul EXCEL est un outil simple pour la réalisation de la régression, il est en adéquation avec le cours d'Econométrie dispensé en L3 – IDS (Université Lyon 2).

C'est un outil ouvert dans la mesure où il vous est possible d'accéder au code source des macros pour les modifier à votre guise (OUTIL / MACRO / VISUAL BASIC EDITOR), par exemple lorsque vous souhaitez intégrer des calculs supplémentaires ou modifier la présentation des résultats. Il est important de noter que les droits d'utilisation ou de commercialisation dans un cadre autre qu'universitaire appartiennent à Mlle **NGUYEN LAO Bao Truc**.

Enfin, cet outil est très perfectible. Il est très vraisemblable que l'on puisse améliorer grandement les temps de calculs et la présentation des résultats. De même, malgré la multiplication des tests réalisés sur un nombre important des jeux de données différents, il reste possible que, dans certains cas, les résultats ne soient pas cohérents. Vos suggestions et/ou commentaires destinés à améliorer ce classeur seront toujours le bienvenu.