
Tables des matières

Dédicace	i
Remerciements	ii
Résumé	iii
Avant propos	v
Liste des abréviations	1
Liste des Figures	2
Liste des Tableaux	3
Introduction	4
Chapitre I : PRESENTATION GENERALE DE LA DGI	5
SECTION I : ENVIRONNEMENT INTERNE DE LA DGI	
1.1. Historique.....	
5	
1.2. Evolution de la DGI.....	5
1.3. Organisation et Structure.....	8
1.3.1. Organisation.....	8
SECTION II : EVOLUTION FISCALE CAMEROUNAISE.....	11
2.1 Recettes des impôts sur le revenu et la consommation.....	11
2.1.1 Recettes budgétaires.....	11
2.1.2 Recettes non budgétaires.....	11
2.2 Recettes d'enregistrement, du timbre, de la curatelle et des taxes spécifiques.....	11
2.2.1 Recettes budgétaires.....	11
2.2.2 Recettes non budgétaires.....	11
Chapitre II : ANALYSE DESCRIPTIVE DES RECETTES FISCALES	16
I. Présentation des Données.....	16
II. Evolution des recettes fiscales camerounaises	
II.1. Recettes totales.....	19
II.2. Recettes d'enregistrement, du timbre, de la curatelle et des taxes spécifiques	20
II.3. Recettes des impôts sur le revenu et la consommation.....	21
II.4. Recettes budgétaires.....	22
II.5. Recettes non budgétaires.....	25
III. Tentative d'interprétation des graphiques.....	28
Chapitre III : MODELE DE REGRESSION LINEAIRE	30
I. Régression Linéaire Simple.....	30
II. Régression Linéaire Multiple	42

III. Méthode de sélection des modèles	50
Chapitre IV : APPLICATIONS	48
I. Régression sur toutes les variables.....	48
II. Sélection de modèle.....	50
Chapitre IV: MODELE IS-LM	52
I- Connaissances générales.....	52
II- Construction de la courbe IS.....	55
III- Calcul du multiplicateur fiscal et déplacement de la courbe IS.....	57
 CONCLUSION.....	 60
 ANNEXES.....	 61
 BIBLIOGRAPHIE.....	 63

DEDICACE

Je dédie ce travail

Au Seigneur

Jésus Christ mon Seul Maître pour toujours

REMERCIEMENTS

Le travail que nous présentons n'est sans doute que le fruit d'une somme d'efforts qui, pris individuellement, ont concouru à sa réalisation, le soin m'étant donné par cette occasion d'adresser ma gratitude :

Au **TOUT-UISSANT** pour tous ses bienfaits ;

A tous les enseignants du Master de Statistique appliquée qui n'ont ménagé aucun effort pour nous donner une formation de qualité, en particulier :

- ❑ Pr. Henri **GWET**, pour tout ce qu'il fait pour la réussite du MASTER ;
- ❑ Pr. O. **HILI**
- ❑ Pr. J.C. **THALABARD** ;
- ❑ Pr. Jean **BRETAGNOLLE**
- ❑ Dr. **NDONG NGUEMA**
- ❑ Dr. Michel **NDOUMBE NKENG**;
- ❑ Dr. Maxime **KIKI**.
- ❑ Dr. J-F. **DUPUY**
- ❑ Dr. J-M. **MARIN**

A Monsieur **Joseph Clément MBONG** Chef de division des statistiques, de la curatelle et de l'immatriculation pour son encadrement professionnel

A Madame **AGBOR Dorothy NGOGKO** pour sa disponibilité permanente

A Messieurs **LOWE , Ambroise MBAH, Pierre ONANA**, tous chargés d'études Assistants à la cellule de la statistique et des simulations fiscales.

A tout le personnel de la **DIRECTION GENERALE DES IMPOTS**, Au personnel de **L'INSTITUT NATIONAL DE LA STATISTIQUE**

A la famille **NGUINA**, notamment ma grand-mère **MENGUE** Valentine, pour son soutien psychologique

A la famille **ONGUI** pour son assistance permanente

A la Famille **FOKOU** pour toute leur sympathie

A la Famille **FOGUEM** pour toute leur sympathie

A la Famille **EPANDA** pour tous les efforts consentis

A la Famille **TIENTIEU** pour son oreille attentive

A ma Famille paternelle par **Elessogue**

A tous mes camarades car ils n'ont ménagé aucun effort pour me venir en aide

RESUME

Les développements récents de la macroéconomie ont donné lieu à deux grands débats. Le premier oppose l'Ecole des politiques économiques actives à Celle des politiques économiques passives. Le second quant à lui est relatif à une politique économique qui doit être conduite comme règle ou discrétion. Les tenants des politiques économiques actives voient en la théorie économique – des classiques aux keynésiens en passant par la théorie de la croissance exogène et endogène – un ensemble de règles qu'il faut appliquer lorsqu'une économie est en détresse ou lorsqu'on veut atteindre un certain niveau de croissance en négligeant l'équivalence ricardienne. Ces différentes règles tiennent lieu d'outils puissants lorsqu'il s'agit de conduire une politique économique. Ainsi, on distingue deux grands types d'instruments de politique économique : la politique monétaire et la politique budgétaire. Cette dernière consiste entre autre à faire varier les dépenses publiques, à modifier l'endettement public, et à une augmentation ou une réduction des impôts. Une modification de la politique fiscale – et/ou les autres instruments – a des impacts différés sur l'économie. Il devient donc intéressant pour le décideur de maîtriser l'ampleur sur la courbe de demande agrégée – principalement sur la courbe IS – du multiplicateur fiscal suivant un horizon temporel bien défini.

INTRODUCTION

L'économiste britannique Nicholas Kaldor assigne quatre principaux objectifs à la politique économique : l'emploi, la croissance, la stabilité des prix et l'équilibre extérieur. Pour atteindre ces objectifs, les deux instruments majeurs qui s'offrent aux décideurs sont la politique monétaire et la politique budgétaire. S'agissant de la politique monétaire, il est question pour l'Etat de pratiquer une politique expansive ou restrictive de la masse monétaire en augmentant ou en diminuant l'offre de monnaie ; les instruments y relatifs sont l'encadrement du crédit, le refinancement et les réserves obligatoires, et l'intervention sur le marché monétaire. La politique budgétaire quant à elle, stimule la demande agrégée par le biais d'instruments tels que la variation des dépenses publiques - aides aux entreprises, investissements en infrastructures, aides à la recherche -, la modification de l'endettement public - déficit pour doper l'activité économique, réduction du déficit ou excédent pour réduire la demande agrégée-, et à une augmentation ou une réduction des recettes fiscales - réduction d'impôts, baisse des taux d'imposition, aides fiscales(exonérations, etc.).

L'impôt n'a pas de définition juridique propre : il se définit le plus souvent dans le cadre d'une approche pragmatique et par l'analyse de ses caractéristiques. Selon Jeze (1930), l'impôt est une prestation pécuniaire, requise des particuliers, par voie d'autorité, à titre définitif et sans contrepartie en vue de la couverture des charges publiques. Bien que critiquable sur certains points, elle peut servir de fondement à la définition suivante : l'impôt est un prélèvement pécuniaire obligatoire auprès des agents économiques (ménages et entreprises), établi par la loi, à titre définitif et sans contrepartie, en vue d'assurer la couverture des charges publiques en général. L'impôt se distingue ainsi des cotisations sociales (où il y a une contrepartie : les prestations sociales), des redevances pour l'utilisation d'un service public, des taxes parafiscales destinées à un objet économique et social particulier. L'expression « fiscalité » peut prendre plusieurs significations. Au sens étroit, la fiscalité est l'ensemble des impôts qui pèsent sur les agents économiques. Au sens étendu, elle englobe toute la politique budgétaire des administrations publiques. La fiscalité est ainsi un instrument au service d'une politique, d'une vision de la vie en commun, organisé par les Pouvoirs Publics gestionnaires des services publics, qui nécessite la perception en monnaie des ressources. La politique fiscale qui consiste « une fois arrêté le principe de prélèvement fiscal, à déterminer les caractéristiques générales de l'impôt en fonction des données économiques et psychologiques ».

Une fois levées toutes les équivoques sur les champs lexical et sémantique de la fiscalité, il convient dès à présent de voir l'ampleur des recettes fiscales sur l'équilibre des du marché des biens et services synthétisé ici par la courbe IS.

L'économie de l'offre prévalait avant les années 1930. Selon elle, ce sont la technologie et les facteurs de production qui déterminent le niveau de production. Lorsque survint la grande dépression des années 1930, ni la technologie et les facteurs de production n'avaient considérablement changé, et ne pouvaient donc expliquer la crise. C'est alors que John Maynard Keynes s'est penché de l'autre composante de l'économie qui est la demande. Il parvint à expliquer dans son ouvrage publié en 1936 et intitulé : « *Théorie générale de l'emploi de l'intérêt et de la monnaie* » que la crise économique d'alors était due à une insuffisance d'absorption des produits fabriqués. C'est plu tard que Hicks et Hansen se sont attelés à modéliser cette pensée keynésienne qui donnera naissance au modèle IS - LM. Malgré qu'elle soit sujette à des critiques depuis l'avènement des Libéraux, il est clair que ce modèle reste valable parce qu'aucune alternative sérieuse n'a encore pu être proposée.

Au lendemain du nouveau gouvernement camerounais du 07 Septembre 2007, et du conseil ministériel qui s'en est suivi, le Chef de l'Etat prescrivait aux membres du gouvernement de mettre en place une fiscalité incitative. Le défi majeur pour le Ministère de l'Economie, de la Planification et de l'Aménagement du Territoire et celui des Finances est alors de trouver un juste milieu du compromis « trop de recettes – moindres recettes » afin que l'Etat puisse au moins assurer ses missions régaliennes.

En supposant toutes choses égales par ailleurs, au niveau de la recouvrement des recettes fiscales – entendu ici au sens strict – les questions principales que l'on pourrait se poser évidemment sont celles de savoir : Quel sera le niveau des recettes en argent dans l'avenir ? Et dans quelles proportions il déplacera la courbe IS ?

Une tentative de réponse à ces différentes interrogations donnera cinq articulations à notre étude.

La première consistera à présenter la Direction Générale des Impôts (DGI) car c'est dans ses structures que notre travail est réalisé.

La seconde quant à elle sera consacrée à une analyse des données qui serviront à la construction d'un modèle valide pour prédire le montant des recettes fiscales.

Dans la troisième partie, nous présenterons un exposé sur le modèle de régression linéaire.

A la quatrième partie, nous appliquerons le modèle de régression linéaire multiple à nos données pour obtenir ainsi la prédiction de nos recettes totales pour la fin du premier semestre de l'année budgétaire 2007.

Enfin, un bref exposé sur le modèle IS – LM sera donné et on calculera enfin le multiplicateur fiscal pour déterminer l'ampleur du déplacement de la courbe IS.

MODELE DE REGRESSION LINEAIRE

Lorsque nous avons une série d'observations, il est coutume de chercher à préciser une liaison éventuelle qui existe entre deux grandeurs et/ou entre une grandeur et plusieurs autres.

Les données peuvent être en coupe transversale, ou peuvent être en des séries temporelles ou chronologiques. On peut aussi examiner des données dites de panel, présentant deux dimensions, tels le salaire annuel net, la consommation alimentaire, l'épargne et l'impôt sur le revenu extraits pour un groupe de 300 salariés des personnels du secteur des télécommunication suivie pendant 20 ans.

La variable qu'on cherche à prédire est appelée variable expliquée ou variable dépendante et les variables qui servent à prédire sont des variables explicatives, prédicteurs ou variables indépendantes.

La régression linéaire permet :

- d'établir le niveau de liaison entre des variables numériques
- déterminer l'intensité relative de différentes variables explicatives d'une variable dépendante.
- de savoir s'il faut augmenter ou non une variable explicative dans un modèle de prédiction d'une variable expliquée.

I - LA REGRESSION LINEAIRE SIMPLE (RLS)

C'est la plus simple et la plus répandue des techniques de régression. La RLS est un modèle singulier de régression dans lequel :

- Il n'y a qu'un seul prédicteur (numérique),
- Le modèle est linéaire dans la variable dans les paramètres

Comme toutes les techniques de prédiction elle a deux finalités:

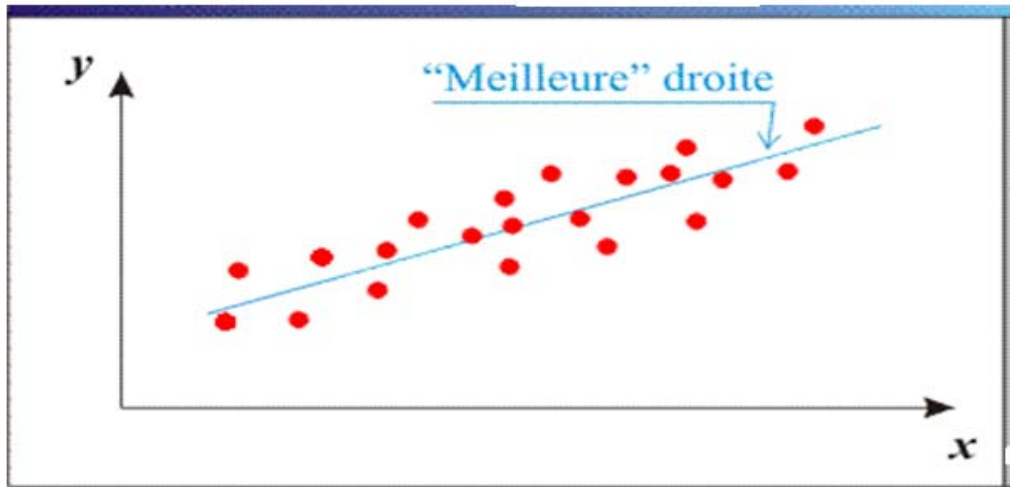
- Construire un modèle dont les paramètres soient interprétables par le praticien en terme de propriétés de la population dont est issu l'échantillon. On espère bien entendu que les paramètres du modèle seront de bons estimateurs des paramètres de la population sous-jacente.
- Se servir du modèle pour faire des prédictions.

I.1- Problème

La Régression Linéaire Simple traite de la question suivante :

- Une grandeur y est mesurée
- Pour un certain nombre de valeurs d'une autre grandeur x

Quel est le meilleur ajustement linéaire que l'on puisse trouver pour ces deux grandeurs ?



En termes plus concret, admettons qu'il existe un lien quelconque entre deux variables X et Y tel que $Y = \varphi(X)$. X représente l'Impôt sur les sociétés et Y les Recettes fiscales totales. Le problème qui se pose est de déterminer quelle est la nature de la fonction φ , ou mathématiquement parlant, quelle est son équation.

I-2 Solution

Pour étudier φ de façon expérimentale, nous allons prendre les données mensuelles de Juillet 1999 à Décembre 2006 de l'Impôt sur les sociétés et les Recettes fiscales totales. Ensuite nous appliquons des méthodes statistiques, graphiques diagrammes etc. pour essayer de comprendre quelle est la nature de la fonction φ . Dans le cas de la RLS, ces méthodes consistent à postuler que la relation φ peut être approchée par une fonction affine F dont les propriétés sont connues puis évaluer à quel point cette fonction F est capable de rendre compte des données observées. La RLS est un cas particulier d'approche statistique dans lequel on suppose que F est une droite. En d'autres termes, on part du principe qu'il existe une droite F dont l'équation est une bonne approximation de φ . Partant de cette hypothèse on applique des calculs mathématiques sur les données qui nous permettent d'évaluer les paramètres de F . Dans un second temps, on va essayer d'évaluer à quel point F est ajusté les données, c'est-à-dire à quel point, F représente un modèle de données observées. Si cet ajustement est jugé satisfaisant, on pourra considérer, jusqu'à nouvel ordre, que la droite F est une bonne représentation de la fonction « vraie », φ . Bien entendu, si φ est trop différente d'une droite, on ne parviendra pas à trouver une droite F qui ressemble suffisamment à φ pour que l'on s'en satisfasse.

I.3- Analyse descriptive des variables considérées

L'intérêt de ces analyses est surtout de repérer si les postulats de la régression linéaire sont remplis (comme la linéarité de la relation entre la variable prédite et le prédicteur) et si certaines valeurs ne distinguent pas vraiment du reste. Cela peut arriver en cas d'erreur de

saisie, ou encore un sujet peut n'avoir pas mal compris (ou pas voulu suivre) la consigne et produit des résultats tout à fait hors norme, même parfois incohérents.

I.3.1- Notion de corrélation linéaire

Ajuster les courbes de régression revient à modéliser une relation (causale ou non) entre une variable indépendante et une variable dépendante. Toutefois, on peut s'intéresser à chercher, si deux variables sont liées mais sans supposer que l'une est la cause de l'autre. En ce cas, X est liée à Y exactement comme Y est liée à X .

Cela n'a alors plus de sens de s'intéresser à deux droites de régression distinctes, mais à une seule, qui correspond à l'axe de symétrie principal du nuage de points. On dit que la droite de régression « synthétise » le nuage de points. Elle passe bien sûr par le centre de gravité G du nuage de points, de coordonnées $G(\mu_x, \mu_y)$. Sa pente est $\frac{\sigma_y}{\sigma_x}$ et son signe est celui de $\text{cov}(X,Y)$.

Et l'on obtient alors le coefficient de corrélation linéaire, ou coefficient de Bravais-Pearson (les anglo-saxons retiennent simplement le nom de Pearson), noté r et tel que

$$r = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y} \quad \text{avec}$$

$$\sigma_x = \sqrt{\text{Var}(X)} \quad \text{et} \quad \sigma_y = \sqrt{\text{Var}(Y)}$$

Autrement dit face à un nuage de points, on peut s'intéresser à trois droites : (1) la droite de régression de Y en X , (2) la droite de régression de X en Y , (3) l'axe de symétrie du nuage de points.

Le coefficient de corrélation linéaire varie entre -1 et $+1$:

- Une valeur nulle indique l'absence de relation linéaire (soit parce qu'il y a pas de relation du tout entre les variables considérées, soit parce que la relation n'est pas linéaire) ;
- Une valeur proche de 1, les variables sont corrélées positivement ;
- Une valeur proche de 0, les variables sont corrélées négativement ;

Le graphique qui suit montre la corrélation entre les Recettes totales et l'Impôt sur les sociétés (corrélation positive). On a ici $r=0.84$ qui est proche de 1

Le coefficient de corrélation est un nombre sans dimension (il n'a pas d'unité, contrairement à une longueur ou un poids).

On notera qu'il existe une relation entre les pentes des droites de régression et le coefficient de corrélation. Ainsi la pente de la droite de régression en X peut être donnée par :

$$b=r\frac{\sigma_Y}{\sigma_x}$$

tandis que la pente de la droite de régression en Y est :

$$b'=r\frac{\sigma_x}{\sigma_y}$$

La part de la variance totale du nuage de points qui est expliquée par l'équation de régression est directement donnée par r^2

Pour tester si une corrélation est significativement différente de 0, on peut calculer la statistique t :

$$t=r\sqrt{\frac{n-2}{1-r^2}}$$

I.3.2- Statistique descriptive

Confère chap.2

I.3.3- Analyse graphique

(Confère chap2)

I.4- Ajustement à l'aide d'une droite : méthode des moindres carrés

Lorsque qu'on veut modéliser les données, la question qui se pose est de trouver l'équation de la droite de régression.

I.4.1- Présentation du modèle

$$Y_t = a_0 + a_1 X_t \quad t = \text{juil}99, \dots, \text{dec}2006$$

I.4.2- Le terme aléatoire

Le modèle tel qu'il vient d'être conçu n'est qu'une caricature de la réalité. Les Recettes totales dépendent aussi de bien d'autres impôts et taxes diverses. C'est pourquoi nous allons compléter notre modèle en ajoutant le terme aléatoire (ε_t) qui synthétise l'ensemble des informations non pris en compte dans notre modèle : $Y_t = a_0 + a_1 X_t + \varepsilon_t$. Le terme de ε_t

mesure l'écart entre les valeurs réellement observées de Y_t et les valeurs qui auraient été observées si la relation spécifiée avait été rigoureusement exacte. Le terme ε_t regroupe donc trois facteurs :

- Une erreur de spécification, c'est à dire le fait que la seule variable explicative n'est pas suffisante pour restituer toute l'information du phénomène expliqué ;
- Une erreur de mesure, les données ne représentent pas exactement le phénomène ;
- Une erreur de fluctuation d'échantillonnage, d'un échantillon à un autre, les observations et donc les estimations sont plus ou moins différentes.

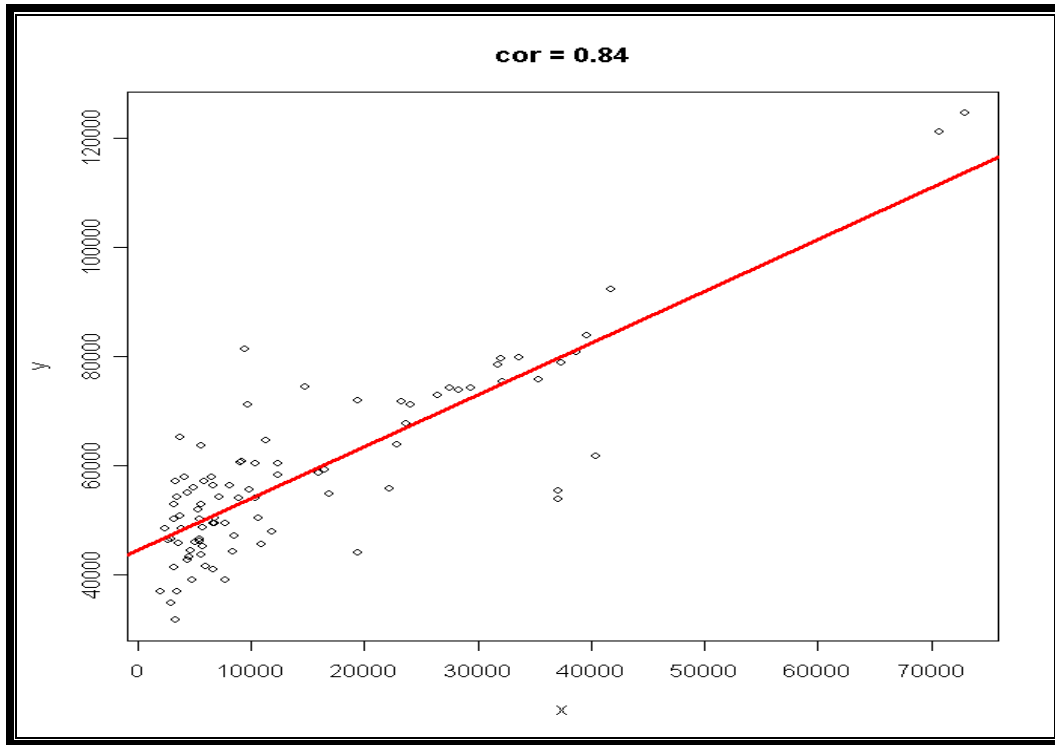
I.4.3- Estimation des paramètres

I.4.3.1- Hypothèses

- $H1$: le modèle est linéaire en X_t
- $H2$: les valeurs X_t sont observées sans erreur
- $H3$: $E(\varepsilon_t) = 0$,
- $H4$: $E(\varepsilon_t^2) = \sigma_\varepsilon^2$
- $H5$: $E(\varepsilon_t, \varepsilon_{t'}) = 0$ si $t \neq t'$
- $H6$: $Cov(x_t, \varepsilon_t) = 0$

I.4.3.2- Formulation des estimateurs

En traçant un graphique, des couples de données liant les Recettes fiscales totales et l'Impôt sur les sociétés on obtient :



L'estimateur des coefficients a_0 et a_1 est obtenu en minimisant la distance au carrée entre chaque observation et la droite, d'où le nom d'Estimateur des Moindres Carrés Ordinaires (MCO).

La résolution analytique est la suivante :

La régression sur toutes les données fiscales camerounaises est

$$E[Y_t / X_t] = a_0 + a_1 X_t \quad (1)$$

Une estimation de $E[Y_t / X_t]$ qui donne l'équation suivante

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X_t \quad (2)$$

Explicitement, les coefficients inconnus de notre relation stochastique

$$Y_t = a_0 + a_1 X_t + \varepsilon_t = E[Y_t / X_t] + \varepsilon_t \quad (3)$$

Sont nos paramètres à estimer.

L'erreur associée avec la *tième* donnée est :

$$\varepsilon_t = Y_t - a_0 - a_1 X_t \quad (4)$$

Pour toutes valeurs de \hat{a}_0 et \hat{a}_1 on peut estimer ε_t avec le résidu,

$$e_t = Y_t - \hat{a}_0 - \hat{a}_1 X_t \quad (5)$$

Des définitions

$$Y_t = a_0 + a_1 X_t + \varepsilon_t \quad (6)$$

$$= \hat{a}_0 + \hat{a}_1 X_t + e_t \quad (7)$$

I.4.3.3- Méthode des moindres carrés

En minimisant la somme des résidus au carré on obtient

$$\text{Min} \sum_t e_t^2 = \sum_t (Y_t - \hat{a}_0 - \hat{a}_1 X_t)^2 \quad (8)$$

En opérant par dérivation par rapport à a_0 et a_1 afin de trouver le minimum de cette fonction, on obtient les résultats suivants

$$\frac{\theta(\sum_t e_t^2)}{\theta \hat{a}_0} = \sum_t 2(Y_t - \hat{a}_0 - \hat{a}_1 X_t)(-1) \quad (9)$$

$$= -2 \sum_t (Y_t - \hat{a}_0 - \hat{a}_1 X_t) = 0 \quad (10)$$

Ou de façon équivalente :

$$\sum_t e_t = 0 \quad (11)$$

Et

$$\frac{\theta(\sum_t e_t^2)}{\theta \hat{a}_1} = \sum_t 2(Y_t - \hat{a}_0 - \hat{a}_1 X_t)(-X_t) = 0 \quad (12)$$

$$= -2 \sum_t X_t (Y_t - \hat{a}_0 - \hat{a}_1 X_t) = 0 \quad (13)$$

Ce qui implique

$$\sum_t X_t e_t = 0 \quad (14)$$

Nous avons par la suite

$$\sum_t Y_t = n \hat{a}_0 + \left(\sum_t X_t \right) \hat{a}_1, \quad (15)$$

$$\sum_t X_t Y_t = \left(\sum_t X_t \right) \hat{a}_0 + \left(\sum_t X_t^2 \right) \hat{a}_1 \quad (16)$$

On obtient la solution en divisant d'abord par n . Le résultat est

$$\bar{Y} = \hat{a}_0 + \hat{a}_1 \bar{X}. \quad (17)$$

En tirant \hat{a}_0 on a

$$\hat{a}_0 = \bar{Y} - \hat{a}_1 \bar{X} \quad (18)$$

Nous avons aussi

$$\sum_t X_t = n \bar{X} \quad (19)$$

En arrangeant

$$\sum_t X_t Y_t - n \bar{X} \bar{Y} = b \left(\sum_t X_t^2 - n \bar{X}^2 \right) \quad (20)$$

$$b = \frac{\sum_t X_t Y_t - n \bar{X} \bar{Y}}{\sum_t X_t^2 - n \bar{X}^2} \quad (21)$$

$$\frac{\sum_t (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_t (X_t - \bar{X})^2} \quad (21)$$

I.4.3.4- Propriété statistique des estimateurs

Les valeurs \hat{a}_1 et \hat{a}_0 calculées comme précédemment par la méthode des MCO, ne sont plus simplement les coefficients d'une droite géométrique satisfaisante, mais des estimations statistiques des coefficients a_0 et a_1 .

Les propriétés des estimateurs des moindres carrés, dépendent des caractéristiques de l'aléa ε_t .

On suppose que les aléas : ε_t , suivent la même la normale : $\mathcal{N}(0, \sigma)$, centrée, et l'écart type : σ (inconnu), et sont indépendants. L'hypothèse de normalité n'est pas déraisonnable si l'on admet que ces aléas résultent de l'addition de nombreux effets secondaires mineurs indépendants. Sous ces conditions :

- Les résidus calculés : e_t , approchent les aléas ε_t , et la quantité $\hat{\sigma}_\varepsilon^2 = \frac{1}{n-2} \sum_t e_t^2$ liée à la somme des carrés des résidus, est une bonne estimation de l'écart type σ , de l'aléa. Elle est appelée écart type résiduel.
- Les estimateurs \hat{a}_0 et \hat{a}_1 sont les « meilleurs possibles »
- Les estimateurs \hat{a}_0 et \hat{a}_1 suivent des lois normales : $\mathcal{N}(a_0, \hat{\sigma}_{\hat{a}_0})$ et $\mathcal{N}(a_1, \hat{\sigma}_{\hat{a}_1})$, donc les espérances a_0 et a_1 sont des quantités estimées : ces estimateurs sont sans biais

- On peut également estimer les variances des estimateurs \hat{a}_1 et \hat{a}_0 donc celui du

premier est
$$\hat{\sigma}_{\hat{a}_1}^2 = \frac{\hat{\sigma}_\varepsilon^2}{\sum_t (X_t - \bar{X})^2}$$

I.4.3.5- Test de significativité d'un coefficient

Bien qu'il soit possible de construire les intervalles de confiance pour les coefficients du modèle, on préfère en régression linéaire donner les estimations ponctuelles en écrivant l'équation estimée, et tester simplement la significativité des coefficients.

Il s'agit de tester si, pour un niveau de confiance donné (en général 95%), l'intervalle de confiance peut ou non contenir la valeur 0. En effet si la valeur véritable du coefficient peut être 0, il n'est même pas certain que la variable explicative (ou le terme constant) intervienne réellement dans le modèle.

Sachant que pour un risque α , l'intervalle de confiance pour a_1 est :

$$\left[\hat{a}_1 - t_\alpha \cdot \hat{\sigma}_{\hat{a}_1}; \hat{a}_1 + t_\alpha \cdot \hat{\sigma}_{\hat{a}_1} \right]$$

Le test intervient à examiner si le rapport :

$$\left| \frac{\text{coefficient...estimé}}{\text{écart - type...estimé}} \right| = \left| \frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}} \right|$$

I.4.4 -L'équation d'analyse de la variance

L'équation fondamentale de l'analyse de la variance :

$$\sum_t (Y_t - \bar{Y})^2 = \sum_t (\hat{Y}_t - Y)^2 + \sum_t e_t^2$$

$$SCT = SCE + SCR$$

Cette équation va nous permettre de juger de la qualité de l'ajustement d'un modèle. En effet, plus la variance expliquée est proche de la variance totale, meilleur est l'ajustement du nuage de points par la droite des moindres carrés. Il est d'usage de calculer le rapport :

$$R^2 = \frac{\sum_t (\hat{Y}_t - \bar{Y})^2}{\sum_t (Y_t - \bar{Y})^2} = 1 - \frac{\sum_t e_t^2}{\sum_t (Y_t - \bar{Y})^2}$$

R^2 est appelé le coefficient de détermination, et R le coefficient de corrélation multiple (dans le cas particulier du modèle de régression à une seule variable indépendante, il est égal au coefficient de corrélation linéaire entre X et Y)

I.4.5- Prédiction dans le modèle de régression linéaire simple

L'équation estimée permet de faire des prévisions ou plus généralement de calculer une valeur de la variable explicative X_t , si cette dernière est connue en $n+1$ soit X_{n+1} . La prévision est donnée par :

$$\hat{Y}_{n+1} = \hat{a}_1 X_{n+1} + \hat{a}_0$$

En supposant le modèle toujours valable, la valeur véritable serait en fait :

$$Y_{n+1} = a_1 X_{n+1} + a_0$$

Nous allons montrer que cette prévision est sans biais. L'erreur de prévision est égale à :

$$\begin{aligned} e_{n+1} &= Y_{n+1} - \hat{Y}_{n+1} \\ &= (a_0 - \hat{a}_0) + (a_1 - \hat{a}_1) X_{n+1} + \varepsilon_{n+1} \end{aligned}$$

en se référant aux hypothèses du modèle, on a :

$$E(e_{n+1}) = 0$$

La prévision sans biais est donc obtenue par l'application directe du modèle de régression estimé. Cependant, dans la pratique, il n'est que de peu utile de connaître la prévision si nous ne savons pas quel degré de confiance nous pouvons lui accorder. Nous allons donc calculer la variance de l'erreur de prévision qui nous permet de déterminer un intervalle de confiance (de prédiction) bornant la prévision. La variance de l'erreur de prévision est donnée par :

$$V(e_{n+1}) = V(Y_{n+1} - \hat{Y}_{n+1}) = \hat{\sigma}_\varepsilon^2 \left[\frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum_t (X_t - \bar{X})^2} + 1 \right]$$

L'hypothèse de normalité permet alors de déterminer un intervalle de confiance à $(1 - \alpha)\%$ pour la prévision. Soit :

$$Y_{n+1} = \hat{Y}_{n+1} + t_{n-2}^{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum_t (X_t - \bar{X})^2} + 1}$$

II- MODELE DE REGRESSION LINEAIRE MULTIPLE

La régression linéaire simple cherchait à « expliquer » les valeurs prises par une variable Y_t par les valeurs prises par une autre variable X_t . Ces deux variables étant supposée à priori avoir une relation linéaire :

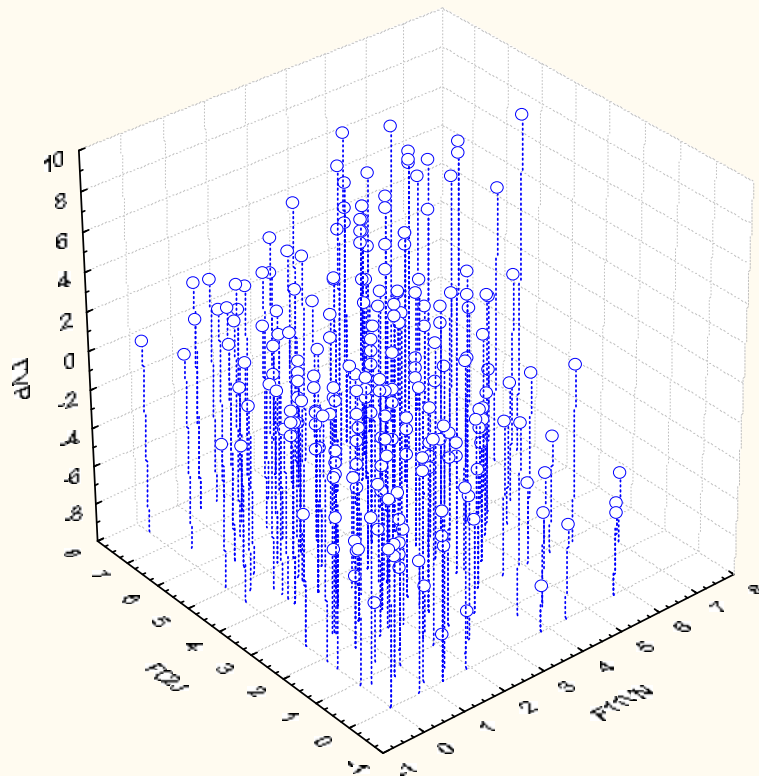
$$Y_t = a_0 + a_1 X_t + \varepsilon_t \quad \text{où}$$

ε_t résume l'ensemble des informations non prises en compte par le modèle.

Or une grandeur dépend rarement d'une seule variable, et les statisticiens considèrent généralement un modèle à plusieurs prédicteurs.

Ici nous ne cherchons pas seulement à modéliser un nuage de points dans un plan (2 dimensions) mais plus généralement dans un espace à k dimensions

Nuage de Points en 3D (Données pour cours de régression.sta 43v*242c)

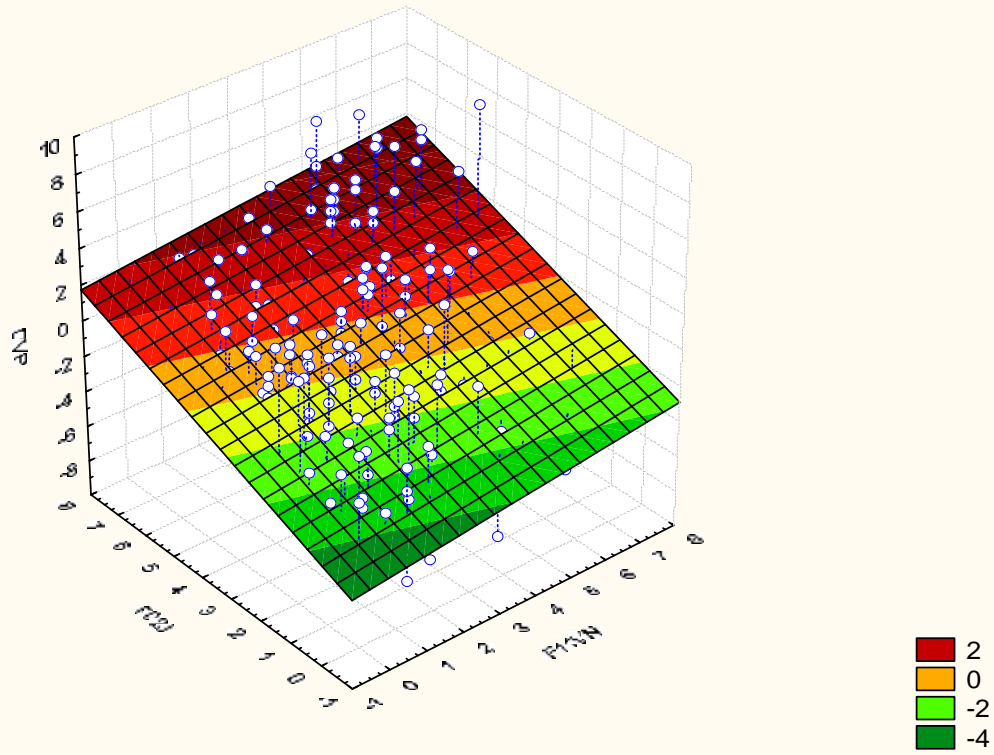


Le fait d'être dans un espace à k dimensions ne change pratiquement rien au fond des vues plus hauts. Ainsi, imaginons un espace à trois dimensions, X_0, X_1 et Y , l'équation de l'ajustement dans cet espace s'écrira :

$$Y = a + b_1 X_1 + b_2 X_2$$

Nous n' avons plus une « droite de régression » mais un « plan de régression »

Nuage de Points en 3D (Données pour cours de régression.sta 43v*242c)
 $DVP = -4.0474 + 0.2359 * x + 0.751 * y$



Ou encore plus généralement dans un espace à k dimensions, nous avons un « hyperplan de régression » :

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_k X_k$$

Cela devient difficile à faire une figure.

Dans l'illustration ci-dessus, le plan des moindres carrés est celui qui minimise la somme des carrés des résidus.

En dimension supérieure, il faut se contenter de dire que la régression linéaire multiple va déterminer l'hyperplan de dimension k minimisant la somme des écarts entre les points représentant les données de l'hyperplan

I- Estimation des paramètres

Hypothèses

- $H1$: le modèle est linéaire
- $H2$: les valeurs $X_1, X_2, X_3 \dots X_i$ sont observées sans erreur
- $H3$: $E(\varepsilon_i) = 0$,

- $H4: E(\varepsilon_t^2) = \sigma_\varepsilon^2$
- $H5: E(\varepsilon_t, \varepsilon_{t'}) = 0$ si $t \neq t'$
- $H6: Cov(x_{i,t}, \varepsilon_t) = 0$
- $H7$: absence de colinéarité entre les variables explicatives, cela impose que la matrice $(X'X)$ est régulière et que la matrice inverse $(X'X)^{-1}$ existe
- $H8: \frac{(X'X)}{n}$ tend vers une matrice finie non singulière.
- $H9: n > k + 1$, le nombre d'observations est supérieur au nombre de séries explicatives.

Analytiquement on obtient l'estimateur des moindres carrés en suivant les étapes suivantes :

Les coefficients des moindres carrés minimisent la somme des résidus au carré :

$$\sum_t \bar{e}_t^{-2} = \sum_t (Y_t - \bar{B}'X_t)^2$$

Où \bar{B} est un choix arbitraire du vecteur des coefficients, le problème de minimisation consiste alors à choisir \bar{B} en

$$\text{Minimisant } \bar{S}(\bar{B}) = \bar{\varepsilon}'\bar{\varepsilon} = (Y - X\bar{B})'(Y - X\bar{B})$$

En développant, on a :

$$\bar{\varepsilon}'\bar{\varepsilon} = Y'Y - \bar{B}'X'Y - Y'X\bar{B} + \bar{B}'X'X\bar{B}$$

Ou

$$S(\bar{B}) = Y'Y - 2\bar{B}'X'Y + \bar{B}'X'X\bar{B}$$

La condition nécessaire pour un minimum est :

APPLICATIONS

Nous voulons construire un modèle de prédiction des recettes fiscales totales sur les différentes nomenclatures qui la constituent. Explicitement, nous cherchons à estimer les coefficients du modèle ci-contre :

$$Y_t = a_0 + a_1 X_{1,t} + a_2 X_{2,t} + a_3 X_{3,t} + \dots + a_{15} X_{15,t} + a_{16} X_{16,t} + \varepsilon_t \text{ où}$$

Y_t représente les recettes fiscales totales mensuelles

$X_{i,t}$ les différents motifs sources de recettes $i=1\dots 16$; $t=juil99\dots dec2006$

I- ESTIMATION DES COEFFICIENTS DE LA REGRESSION

```
#####RESULTAT DE LA REGRESSION SUR TOUTES LES VARIABLES#####
summary(reg)

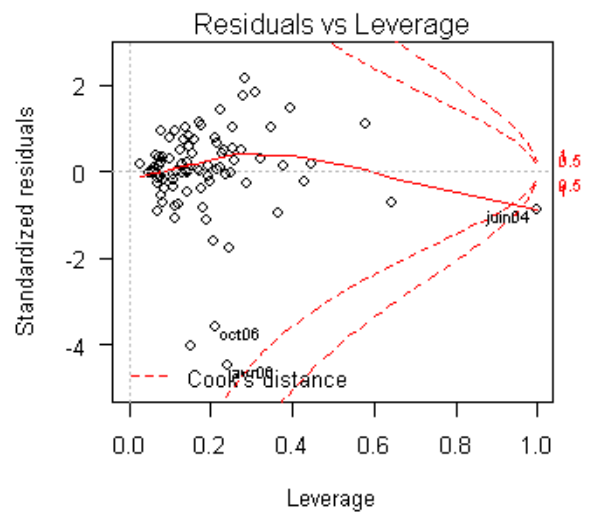
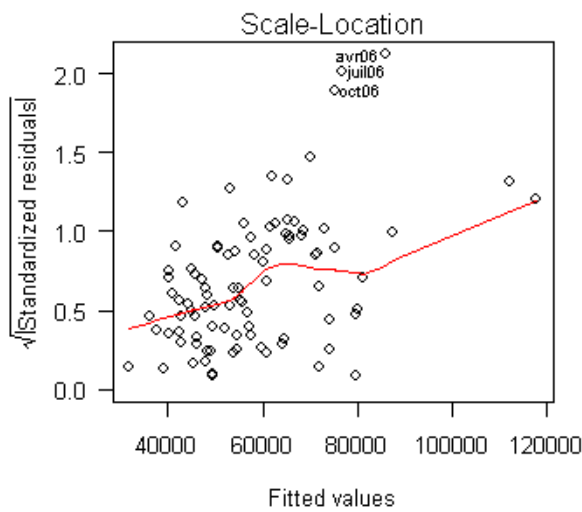
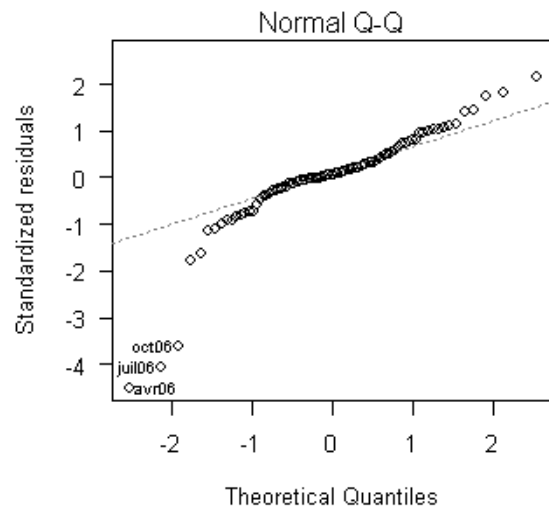
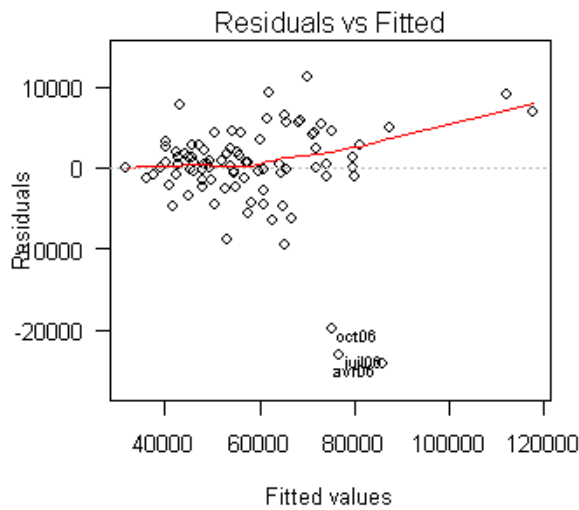
Call:
lm(formula = RECETTESTOTALES ~ ., data = data1)

Residuals:
    Min       1Q   Median       3Q      Max
-24221.2  -1225.9   340.9   2699.1  11277.9

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -31.66361  5600.84953  -0.006  0.99550
IMPSURREVPHYS  0.95646   0.53416   1.791  0.07751 .
IMPOTSALACONSOMMATION 0.93303   0.18223   5.120  2.39e-06 ***
IMPOTSOCIETES  0.85083   0.05646  15.070 < 2e-16 ***
TAXESASSIMAUVID 0.84964   0.41327   2.056  0.04337 *
PRELEVEMENTSDIV 0.16561   0.81778   0.203  0.84008
DROITSDENREGISTREM 2.01368   2.25399   0.893  0.37459
TAXESDENREGISTREM 2.70343   4.62889   0.584  0.56100
DROITSDETIMBRE  3.31331   1.24017   2.672  0.00930 **
TAXEFORJEUX    6.81388   4.61743   1.476  0.14433
RECETTESMINJUSTICE -8.06037   8.16354  -0.987  0.32673
TAXESFORESTIERES 0.90688   0.58907   1.540  0.12800
TAXESDELEVAGE -13.39462  18.05831  -0.742  0.46062
RECETTESDOMANIALES -3.86627  14.58146  -0.265  0.79164
RECETTESNONBUDGETAIRES 1.44399   0.49713   2.905  0.00486 **
DROITSDETRANSIT -0.62733   2.39878  -0.262  0.79443
TAXESMINIERES  -2.23794   8.04079  -0.278  0.78155
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6173 on 73 degrees of freedom
Multiple R-Squared:  0.8807,    Adjusted R-squared:  0.8546
F-statistic: 33.69 on 16 and 73 DF,  p-value: < 2.2e-16
```

La régression sur toutes les variables nous donnent plusieurs estimateurs des coefficients du modèle qui ne sont pas significatifs avec un $\bar{R}^2=0.8546$ proche de 1. La courbe des résidus se présente comme suit :



Les résidus semblent être alignés sur une même droite d'où un soupçon de normalité de ceux-ci.

Malgré la bonne qualité du $\overline{R^2}$ et des résidus, nous ne pouvons pas retenir ce modèle car la majorité des coefficients ne sont pas significatifs et on note aussi un grand nombre de prédicteurs (16) ce qui donnera des mauvaises estimations de prévision sur une période donnée. D'où le recours aux méthodes de sélection de modèle.

II- SELECTION DU MODELE

II.1- Sélection de modèle par élimination pas à pas

Il est question de retirer du modèle la variable ayant la plus grande P-value (dont le test de student est le moins significatif) . Et arrêter le processus lorsque tous les coefficients sont considérés significativement (5%) différents de 0

```
#####RESULTAT DE LA REGRESSION AVEC RETRAIT DE "DROITSDENREGISTREMENT"#####  
summary(reg1)  
  
Call:  
lm(formula = RECETTESTOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +  
  IMPOTSOCIETES + TAXESASSIMAUXXID + PRELEVEMENTSDIV + TAXESDENREGISTREM +  
  DROITSDETIMBRE + TAXEFORJEUX + RECETTESMINJUSTICE + TAXESFORESTIERES +  
  TAXESDELEVAGE + RECETTESDOMANIALES + RECETTESNONBUDGETAIRES + DROITSDETRANSIT + TAXESMINIERES  
Residuals:  
    Min       1Q   Median       3Q      Max  
-24004.7 -1124.5   409.2   2677.3  10477.6  
  
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)    1779.61073  5213.89237    0.341  0.73383  
IMPSUREVPERSPHYS    0.94160    0.53318    1.766  0.08152 .  
IMPOTSALACONSOMMATION  0.96140    0.17920    5.365  8.9e-07 ***  
IMPOTSOCIETES      0.85504    0.05618   15.219 < 2e-16 ***  
TAXESASSIMAUXXID   0.91747    0.40568    2.262  0.02666 *  
PRELEVEMENTSDIV    0.23139    0.81335    0.284  0.77683  
TAXESDENREGISTREM  2.90787    4.61692    0.630  0.53075  
DROITSDETIMBRE     3.41616    1.23313    2.770  0.00708 **  
  
TAXEFORJEUX        7.34178    4.57321    1.605  0.11267  
RECETTESMINJUSTICE -9.37505    8.01885   -1.169  0.24610  
TAXESFORESTIERES   0.94351    0.58684    1.608  0.11214  
TAXESDELEVAGE     -10.33975   17.70741   -0.584  0.56105  
RECETTESDOMANIALES -4.54137   14.54199   -0.312  0.75570  
RECETTESNONBUDGETAIRES  1.24961    0.44638    2.799  0.00653 **  
DROITSDETRANSIT    -1.57835    2.14670   -0.735  0.46451  
TAXESMINIERES      -1.09396    7.92734   -0.138  0.89062  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 6165 on 74 degrees of freedom  
Multiple R-Squared:  0.8794,    Adjusted R-squared:  0.855  
F-statistic: 35.98 on 15 and 74 DF,  p-value: < 2.2e-16
```

```
#####RESULTAT DE LA REGRESSION AVEC RETRAIT DE "TAXESMINIERES"#####  
summary(reg2)  
  
Call:  
lm(formula = RECETTESTOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +  
  IMPOTSOCIETES + TAXESASSIMAUXXID + PRELEVEMENTSDIV + TAXESDENREGISTREM +  
  DROITSDETIMBRE + TAXEFORJEUX + RECETTESMINJUSTICE + TAXESFORESTIERES +  
  TAXESDELEVAGE + RECETTESDOMANIALES + RECETTESNONBUDGETAIRES +  
  DROITSDETRANSIT)  
Residuals:  
    Min       1Q   Median       3Q      Max  
-24152.4 -1122.0   410.1   2740.2  10437.8  
  
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)    1761.3192  5178.0087    0.340  0.73469  
IMPSUREVPERSPHYS    0.9556    0.5200    1.838  0.07004 .  
  
IMPOTSALACONSOMMATION  0.9625    0.1779    5.412  7.18e-07 ***  
IMPOTSOCIETES      0.8532    0.0542   15.741 < 2e-16 ***  
TAXESASSIMAUXXID   0.9054    0.3935    2.301  0.02418 *  
PRELEVEMENTSDIV    0.2545    0.7906    0.322  0.74837  
TAXESDENREGISTREM  2.8299    4.5522    0.622  0.53605  
DROITSDETIMBRE     3.3999    1.2195    2.788  0.00672 **  
TAXEFORJEUX        7.2988    4.5326    1.610  0.11154  
RECETTESMINJUSTICE -9.2805    7.9371   -1.169  0.24600  
TAXESFORESTIERES   0.9195    0.5569    1.651  0.10287  
TAXESDELEVAGE     -11.2074   16.4449   -0.682  0.49764  
RECETTESDOMANIALES -5.1477   13.7714   -0.374  0.70961
```

```
RECETTESNONBUDGETAIRES 1.2610 0.4359 2.893 0.00499 **
DROITSDETRANSIT -1.5721 2.1321 -0.737 0.46322
```

```
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 6124 on 75 degrees of freedom
Multiple R-Squared: 0.8794, Adjusted R-squared: 0.8569
F-statistic: 39.06 on 14 and 75 DF, p-value: < 2.2e-16
```

```
##### REGRESSION AVEC RETRAIT DE « PRELEVEMENTSDIV »
```

```
> summary(reg3)
```

```
Call:
```

```
lm(formula = RECETTESTOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
    IMPOTSOCIETES + TAXESASSIMAUXXID + TAXESDENREGISTREM + DROITSDETIMBRE +
    TAXEFORJEUX + RECETTESMINJUSTICE + TAXESFORESTIERES + TAXESDELEVAGE +
    RECETTESDOMANIALES + RECETTESNONBUDGETAIRES + DROITSDETRANSIT)
```

```
Residuals:
```

```
Min 1Q Median 3Q Max
-24060.8 -1121.4 416.1 2734.0 10320.8
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1847.96508	5140.42726	0.359	0.72022
IMPSUREVPERSPHYS	0.94841	0.51640	1.837	0.07018 .
IMPOTSALACONSOMMATION	0.94830	0.17128	5.537	4.22e-07 ***
IMPOTSOCIETES	0.85524	0.05351	15.982	< 2e-16 ***
TAXESASSIMAUXXID	0.93590	0.37966	2.465	0.01595 *
TAXESDENREGISTREM	2.86354	4.52404	0.633	0.52866
DROITSDETIMBRE	3.39449	1.21213	2.800	0.00647 **
TAXEFORJEUX	7.11674	4.47064	1.592	0.11556
RECETTESMINJUSTICE	-8.93592	7.81809	-1.143	0.25664
TAXESFORESTIERES	0.88004	0.53999	1.630	0.10729
TAXESDELEVAGE	-8.33744	13.73772	-0.607	0.54573
RECETTESDOMANIALES	-4.48935	13.53822	-0.332	0.74110
RECETTESNONBUDGETAIRES	1.27548	0.43096	2.960	0.00411 **
DROITSDETRANSIT	-1.59633	2.11821	-0.754	0.45340

```
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 6088 on 76 degrees of freedom
Multiple R-Squared: 0.8792, Adjusted R-squared: 0.8586
F-statistic: 42.56 on 13 and 76 DF, p-value: < 2.2e-16
```

```
##### REGRESSION AVEC RETRAIT DE «RECETTESDOMANIALES»
```

```
summary(reg4)
```

```
Call:
```

```
lm(formula = RECETTESTOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
    IMPOTSOCIETES + TAXESASSIMAUXXID + TAXESDENREGISTREM + DROITSDETIMBRE +
    TAXEFORJEUX + RECETTESMINJUSTICE + TAXESFORESTIERES + TAXESDELEVAGE +
    RECETTESNONBUDGETAIRES + DROITSDETRANSIT)
```

```
Residuals:
```

```
Min 1Q Median 3Q Max
-24299 -1040 320 2729 10187
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2004.62598	5089.00265	0.394	0.69473
IMPSUREVPERSPHYS	0.95781	0.51263	1.868	0.06551 .
IMPOTSALACONSOMMATION	0.94510	0.17001	5.559	3.75e-07 ***
IMPOTSOCIETES	0.85586	0.05317	16.096	< 2e-16 ***
TAXESASSIMAUXXID	0.92031	0.37456	2.457	0.01626 *
TAXESDENREGISTREM	2.71082	4.47445	0.606	0.54640
DROITSDETIMBRE	3.24727	1.12137	2.896	0.00492 **
TAXEFORJEUX	6.93910	4.41271	1.573	0.11993
RECETTESMINJUSTICE	-8.60924	7.71082	-1.117	0.26767
TAXESFORESTIERES	0.87804	0.53682	1.636	0.10600
TAXESDELEVAGE	-8.95053	13.53384	-0.661	0.51037
RECETTESNONBUDGETAIRES	1.29528	0.42433	3.053	0.00311 **
DROITSDETRANSIT	-1.55950	2.10303	-0.742	0.46062

```
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 6053 on 77 degrees of freedom
Multiple R-Squared: 0.879, Adjusted R-squared: 0.8602
F-statistic: 46.64 on 12 and 77 DF, p-value: < 2.2e-16
```

```
##### REGRESSION AVEC RETRAIT DE «TAXESDENREGISTREM»
```

```
summary(reg5)
```

```
Call:
lm(formula = RECETTESTOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
    IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDETIMBRE + TAXEFORJEUJ +
    RECETTESMINJUSTICE + TAXESFORESTIERES + TAXESDELEVAGE + RECETTESNONBUDGETAIRES +
    DROITSDETRANSIT)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-24172.4 -1098.9   227.7   2817.9  10083.7
```

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      2061.33860  5067.45499   0.407  0.68528
IMPSURREVPERSPHYS  0.97728    0.50954   1.918  0.05878 .
IMPOTSALACONSOMMATION  0.95425    0.16865   5.658 2.43e-07 ***
IMPOTSOCIETES     0.87026    0.04737  18.372 < 2e-16 ***
TAXESASSIMAUXXID  0.91080    0.37271   2.444  0.01679 *
DROITSDETIMBRE    3.40440    1.08653   3.133  0.00244 **
TAXEFORJEUJ      7.26568    4.36185   1.666  0.09978 .
RECETTESMINJUSTICE -9.09649    7.63759  -1.191  0.23726
TAXESFORESTIERES  0.84364    0.53164   1.587  0.11659
TAXESDELEVAGE    -9.02340   13.47828  -0.669  0.50517
RECETTESNONBUDGETAIRES  1.29293    0.42259   3.060  0.00304 **
DROITSDETRANSIT  -1.70388    2.08099  -0.819  0.41541
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 6028 on 78 degrees of freedom
Multiple R-Squared:  0.8785,    Adjusted R-squared:  0.8613
F-statistic: 51.26 on 11 and 78 DF,  p-value: < 2.2e-16
```

```
##### REGRESSION AVEC RETRAIT DE «TAXESDELEVAGE»
summary(reg6)
```

```
Call:
lm(formula = RECETTESTOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
    IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDETIMBRE + TAXEFORJEUJ +
    RECETTESMINJUSTICE + TAXESFORESTIERES + RECETTESNONBUDGETAIRES +
    DROITSDETRANSIT)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-24603.3 -1091.7   404.3   2748.4  10153.4
```

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      2793.11520  4930.85267   0.566  0.57269
IMPSURREVPERSPHYS  0.90913    0.49753   1.827  0.07143 .
IMPOTSALACONSOMMATION  0.91367    0.15683   5.826 1.18e-07 ***
IMPOTSOCIETES     0.86674    0.04691  18.476 < 2e-16 ***
TAXESASSIMAUXXID  0.93671    0.36940   2.536  0.01319 *
DROITSDETIMBRE    3.16049    1.02005   3.098  0.00269 **
TAXEFORJEUJ      6.62168    4.23957   1.562  0.12231
RECETTESMINJUSTICE -7.69916    7.32116  -1.052  0.29617
TAXESFORESTIERES  0.83677    0.52968   1.580  0.11816
RECETTESNONBUDGETAIRES  1.32194    0.41889   3.156  0.00227 **
DROITSDETRANSIT  -2.23478    1.91724  -1.166  0.24728
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 6007 on 79 degrees of freedom
Multiple R-Squared:  0.8778,    Adjusted R-squared:  0.8623
F-statistic: 56.73 on 10 and 79 DF,  p-value: < 2.2e-16
```

```
##### REGRESSION AVEC RETRAIT DE «RECETTESMINJUSTICE»
summary(reg7)
```

```
Call:
lm(formula = RECETTESTOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
    IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDETIMBRE + TAXEFORJEUJ +
    TAXESFORESTIERES + RECETTESNONBUDGETAIRES + DROITSDETRANSIT)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-25910.4 -1247.8   285.8   2785.8  10216.9
```

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1887.52548  4858.28964   0.389  0.69867
IMPSURREVPERSPHYS  0.78956    0.48468   1.629  0.10724
IMPOTSALACONSOMMATION  0.95046    0.15298   6.213 2.2e-08 ***
IMPOTSOCIETES     0.86525    0.04692  18.440 < 2e-16 ***
TAXESASSIMAUXXID  0.95593    0.36919   2.589  0.01142 *
DROITSDETIMBRE    3.10957    1.01957   3.050  0.00310 **
TAXEFORJEUJ      2.58745    1.80613   1.433  0.15587
TAXESFORESTIERES  0.92760    0.52294   1.774  0.07990 .
```

```

RECETTESNONBUDGETAIRES    1.27816    0.41709    3.064    0.00297 **
DROITSDETRANSIT          -2.53038    1.89778   -1.333    0.18621
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

Residual standard error: 6011 on 80 degrees of freedom
Multiple R-Squared:  0.8761,    Adjusted R-squared:  0.8621
F-statistic: 62.83 on 9 and 80 DF,  p-value: < 2.2e-16
```

```

##### REGRESSION AVEC RETRAIT DE «DROITSDETRANSIT»
summary(reg8)
```

```

Call:
lm(formula = RECETTESTOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
    IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDETIMBRE + TAXEFORJEUX +
    TAXESFORESTIERES + RECETTESNONBUDGETAIRES)
```

```

Residuals:
    Min       1Q   Median       3Q      Max
-25843.6 -1108.6   536.7   2334.0  10158.6
```

```

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.055e+03  4.841e+03   0.218  0.82798
IMPSUREVPERSPHYS  9.035e-01  4.794e-01   1.885  0.06306 .
IMPOTSALACONSOMMATION 9.541e-01  1.537e-01   6.208 2.17e-08 ***
IMPOTSOCIETES    8.729e-01  4.679e-02  18.654 < 2e-16 ***
TAXESASSIMAUXXID  9.134e-01  3.696e-01   2.472  0.01555 *
DROITSDETIMBRE   3.319e+00  1.012e+00   3.279  0.00154 **
TAXEFORJEUX     1.918e+00  1.743e+00   1.100  0.27455
TAXESFORESTIERES  8.894e-01  5.246e-01   1.695  0.09389 .
RECETTESNONBUDGETAIRES 1.124e+00  4.026e-01   2.791  0.00655 **
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

Residual standard error: 6040 on 81 degrees of freedom
Multiple R-Squared:  0.8733,    Adjusted R-squared:  0.8608
F-statistic: 69.79 on 8 and 81 DF,  p-value: < 2.2e-16
```

```

##### REGRESSION AVEC RETRAIT DE «TAXEFORJEUX»
summary(reg9)
```

```

Call:
lm(formula = RECETTESTOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
    IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDETIMBRE + TAXESFORESTIERES +
    RECETTESNONBUDGETAIRES)
```

```

Residuals:
    Min       1Q   Median       3Q      Max
-25845.1 -1154.5   522.8   2639.2  10085.2
```

```

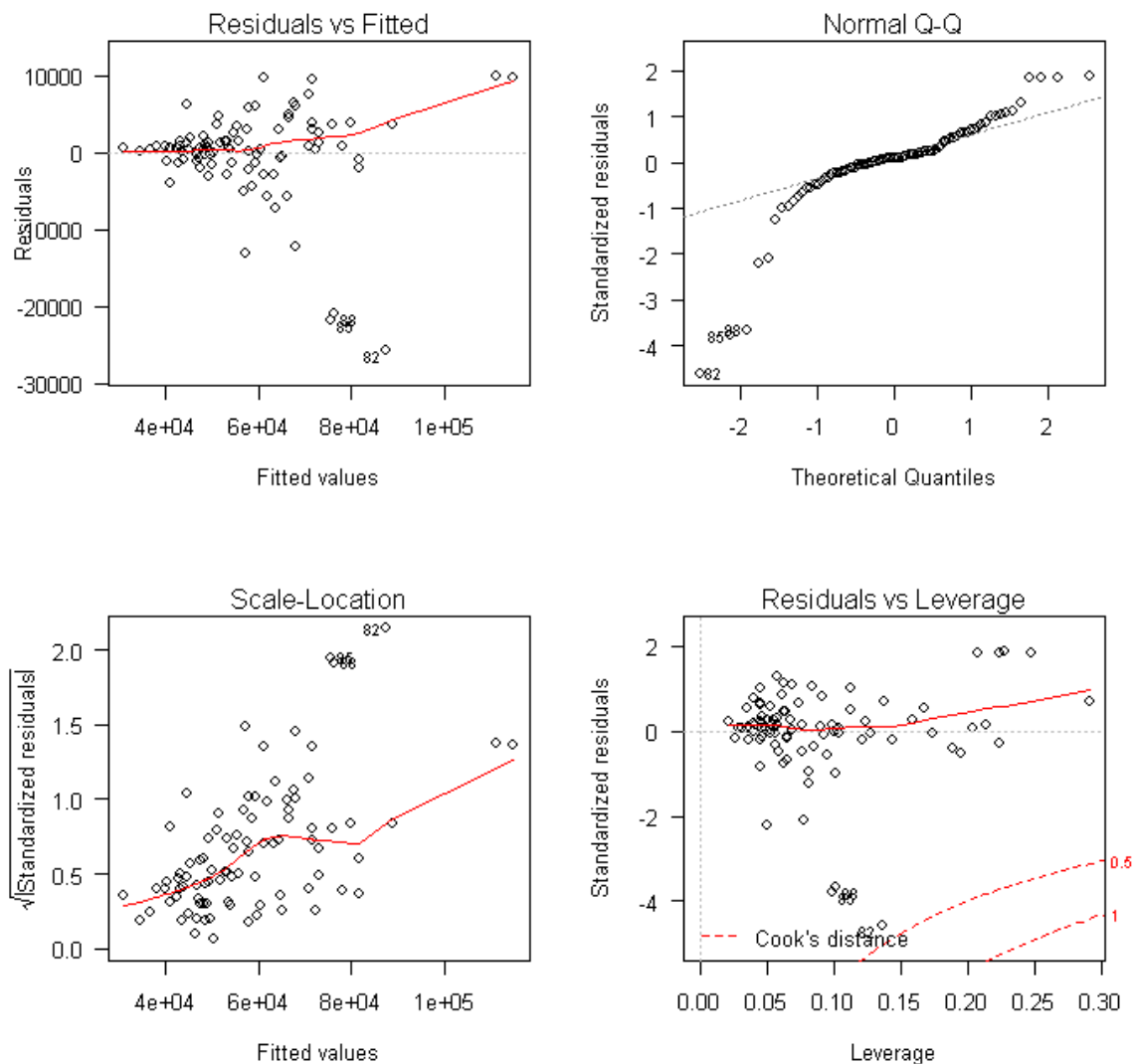
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.007e+02  4.821e+03   0.104  0.91753
IMPSUREVPERSPHYS  8.686e-01  4.789e-01   1.814  0.07339 .
IMPOTSALACONSOMMATION 9.774e-01  1.524e-01   6.412 8.61e-09 ***
IMPOTSOCIETES    8.681e-01  4.665e-02  18.608 < 2e-16 ***
TAXESASSIMAUXXID  9.562e-01  3.680e-01   2.598  0.01110 *
DROITSDETIMBRE   3.257e+00  1.012e+00   3.218  0.00185 **
TAXESFORESTIERES  9.261e-01  5.243e-01   1.766  0.08105 .
RECETTESNONBUDGETAIRES 1.201e+00  3.968e-01   3.027  0.00330 **
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

Residual standard error: 6048 on 82 degrees of freedom
Multiple R-Squared:  0.8714,    Adjusted R-squared:  0.8604
F-statistic: 79.39 on 7 and 82 DF,  p-value: < 2.2e-16
```

Nous constatons que le modèle 9 est celui dont toutes les variables sont significatives ; en plus de cela, on a $\bar{R}^2=0.8604$ qui est proche de 1. Pour vraiment juger de la validité du modèle, nous allons tracer son graphique pour vérifier si les résidus suivent à peu près une loi normale.



Il est donc clair que les résidus semblent être à peu près sur une même droite ; donc ce modèle semble avoir un caractère prédictif.

II.2 Procédure de sélection automatique ascendante ou forward

Ici on part de la régression sur toutes les variables et automatiquement, on obtient le meilleur modèle.

```
#####procédure ascendante#####
> reg <- lm(RECETTES_TOTALES ~ 1, data=data1)
> step.reg = step(reg, direction="forward", k=log(40), scope=list(lower=-1, upper=-IMPSUREVPERSPHYS +
IMPOTSALACONSOMMATION +
+ IMPOTSOCIETES + TAXESASSIMAUXXID + PRELEVEMENTSDIV + DROITSDENREGISTREM +
+ TAXESDENREGISTREM + DROITSDETIMBRE + TAXEFORJEUX + RECETTESMINJUSTICE +
+ TAXESFORESTIERES + TAXESDELEVAGE + RECETTESDOMANIALES + RECETTESNONBUDGETAIRES +
+ DROITSDETRANSIT + TAXESMINIERES), trace=TRUE)
Start: AIC= 1747.25
```

RECETTESTOTALES ~ 1

	Df	Sum of Sq	RSS	AIC
+ IMPOTSOCIETES	1	1.6465e+10	6.8579e+09	1.6410e+03
+ TAXESDENREGISTREM	1	5.0588e+09	1.8264e+10	1.7290e+03
+ IMPOTSALACONSOMMATION	1	3.9291e+09	1.9394e+10	1.7340e+03
+ TAXESMINIERES	1	2.2260e+09	2.1097e+10	1.7420e+03
+ IMPSUREVPERSPHYS	1	1.9353e+09	2.1388e+10	1.7430e+03
+ RECETTESNONBUDGETAIRES	1	1.5820e+09	2.1741e+10	1.7450e+03
+ TAXESFORESTIERES	1	1.5719e+09	2.1751e+10	1.7450e+03
+ TAXESDELEVAGE	1	1.4158e+09	2.1908e+10	1.7450e+03
+ DROITSDENREGISTREM	1	1.2818e+09	2.2042e+10	1.7460e+03
<none>			2.3323e+10	1.7470e+03
+ PRELEVEMENTSDIV	1	4.7978e+08	2.2844e+10	1.7490e+03
+ DROITSDETIMBRE	1	3.2562e+08	2.2998e+10	1.7500e+03
+ DROITSDETRANSIT	1	2.8569e+08	2.3038e+10	1.7500e+03
+ TAXESASSIMAUXXID	1	2.0948e+08	2.3114e+10	1.7500e+03
+ RECETTESDOMANIALES	1	1.9496e+08	2.3128e+10	1.7500e+03
+ TAXEFORJEUX	1	1.2770e+08	2.3196e+10	1.7500e+03
+ RECETTESMINJUSTICE	1	3.0422e+07	2.3293e+10	1.7510e+03

Step: AIC= 1640.78

RECETTESTOTALES ~ IMPOTSOCIETES

	Df	Sum of Sq	RSS	AIC
+ IMPOTSALACONSOMMATION	1	2200200793	4657704873	1610
+ IMPSUREVPERSPHYS	1	1301265237	5556640428	1626
+ TAXESASSIMAUXXID	1	554306554	6303599112	1637
+ TAXESDELEVAGE	1	489329011	6368576655	1638
+ TAXESFORESTIERES	1	475420919	6382484746	1638
+ RECETTESNONBUDGETAIRES	1	324953747	6532951918	1640
+ TAXEFORJEUX	1	299464935	6558440730	1640
+ DROITSDENREGISTREM	1	278584032	6579321634	1641
<none>			6857905666	1641
+ TAXESDENREGISTREM	1	187271168	6670634498	1642
+ DROITSDETIMBRE	1	183745927	6674159739	1642
+ TAXESMINIERES	1	152269439	6705636227	1642
+ RECETTESMINJUSTICE	1	121389103	6736516562	1643
+ RECETTESDOMANIALES	1	117899974	6740005692	1643
+ DROITSDETRANSIT	1	20719481	6837186185	1644
+ PRELEVEMENTSDIV	1	18426691	6839478974	1644

Step: AIC= 1609.64

RECETTESTOTALES ~ IMPOTSOCIETES + IMPOTSALACONSOMMATION

	Df	Sum of Sq	RSS	AIC
+ RECETTESNONBUDGETAIRES	1	611134860	4046570014	1601
+ TAXESFORESTIERES	1	515497580	4142207293	1603
+ DROITSDETIMBRE	1	455292852	4202412021	1604
+ IMPSUREVPERSPHYS	1	329864259	4327840614	1607
+ TAXESASSIMAUXXID	1	276120480	4381584393	1608
<none>			4657704873	1610
+ TAXEFORJEUX	1	137361918	4520342955	1611
+ RECETTESMINJUSTICE	1	86142771	4571562102	1612
+ TAXESDENREGISTREM	1	59028513	4598676360	1612
+ TAXESMINIERES	1	38757426	4618947447	1613
+ RECETTESDOMANIALES	1	34471620	4623233253	1613
+ PRELEVEMENTSDIV	1	14421937	4643282936	1613
+ TAXESDELEVAGE	1	9847334	4647857539	1613
+ DROITSDENREGISTREM	1	1399848	4656305025	1613
+ DROITSDETRANSIT	1	1286403	4656418470	1613

Step: AIC= 1600.67

RECETTESTOTALES ~ IMPOTSOCIETES + IMPOTSALACONSOMMATION + RECETTESNONBUDGETAIRES

	Df	Sum of Sq	RSS	AIC
+ TAXESFORESTIERES	1	378233547	3668336466	1596
+ DROITSDETIMBRE	1	298031722	3748538291	1597
+ TAXESASSIMAUXXID	1	294740321	3751829693	1598
+ IMPSUREVPERSPHYS	1	260132428	3786437585	1598
+ DROITSDENREGISTREM	1	245978072	3800591941	1599
<none>			4046570014	1601
+ TAXESMINIERES	1	110070718	3936499296	1602
+ DROITSDETRANSIT	1	70290664	3976279350	1603
+ RECETTESDOMANIALES	1	69637918	3976932096	1603
+ TAXESDENREGISTREM	1	66223551	3980346463	1603
+ TAXEFORJEUX	1	58208364	3988361650	1603
+ RECETTESMINJUSTICE	1	16466548	4030103465	1604
+ PRELEVEMENTSDIV	1	4294065	4042275949	1604
+ TAXESDELEVAGE	1	4257099	4042312915	1604

Step: AIC= 1595.53

RECETTESTOTALES ~ IMPOTSOCIETES + IMPOTSALACONSOMMATION + RECETTESNONBUDGETAIRES + TAXESFORESTIERES

	Df	Sum of Sq	RSS	AIC
+ DROITSDETIMBRE	1	316746984	3351589482	1591
+ TAXESASSIMAUXXID	1	213795421	3454541046	1594

```

+ DROITSDENREGISTREM 1 174929407 3493407059 1595
<none> 3668336466 1596
+ TAXESDENREGISTREM 1 96990278 3571346189 1597
+ DROITSDETRANSIT 1 70942607 3597393859 1597
+ RECETTESDOMANIALES 1 67895801 3600440665 1598
+ IMPSUREVSPERSPHYS 1 66662241 3601674225 1598
+ TAXEFORJEUX 1 43442419 3624894047 1598
+ TAXESMINIERES 1 25965144 3642371322 1599
+ PRELEVEMENTSDIV 1 23381434 3644955033 1599
+ RECETTESMINJUSTICE 1 13592432 3654744034 1599
+ TAXESDELEVAGE 1 2560 3668333907 1599

```

Step: AIC= 1591.09

RECETTESTOTALS ~ IMPOTSOCIETES + IMPOTSALACONSOMMATION + RECETTESNONBUDGETAIRES + TAXESFORESTIERES + DROITSDETIMBRE

```

Df Sum of Sq RSS AIC
+ TAXESASSIMAUXXID 1 232300674 3119288808 1588
<none> 3351589482 1591
+ IMPSUREVSPERSPHYS 1 105657489 3245931993 1592
+ DROITSDENREGISTREM 1 78033829 3273555654 1593
+ TAXEFORJEUX 1 56505650 3295083833 1593
+ DROITSDETRANSIT 1 35244736 3316344746 1594
+ TAXESDENREGISTREM 1 30807651 3320781831 1594
+ TAXESDELEVAGE 1 21003538 3330585945 1594
+ RECETTESMINJUSTICE 1 20410488 3331178994 1594
+ PRELEVEMENTSDIV 1 1798886 3349790597 1595
+ TAXESMINIERES 1 474370 3351115112 1595
+ RECETTESDOMANIALES 1 139727 3351449755 1595

```

Step: AIC= 1588.32

RECETTESTOTALS ~ IMPOTSOCIETES + IMPOTSALACONSOMMATION + RECETTESNONBUDGETAIRES + TAXESFORESTIERES + DROITSDETIMBRE + TAXESASSIMAUXXID

```

Df Sum of Sq RSS AIC
<none> 3119288808 1588
+ IMPSUREVSPERSPHYS 1 120298693 2998990115 1588
+ DROITSDETRANSIT 1 60877038 3058411770 1590
+ DROITSDENREGISTREM 1 48023120 3071265689 1591
+ TAXESDENREGISTREM 1 39571853 3079716955 1591
+ TAXEFORJEUX 1 34873229 3084415580 1591
+ TAXESDELEVAGE 1 14316143 3104972665 1592
+ RECETTESMINJUSTICE 1 10928117 3108360691 1592
+ TAXESMINIERES 1 10579308 3108709501 1592
+ PRELEVEMENTSDIV 1 1945227 3117343582 1592
+ RECETTESDOMANIALES 1 1787814 3117500995 1592
>

```

En régressant sur les variables sélectionnées par la méthode forward, on obtient:

```
> summary(step.reg)
```

Call:

```
lm(formula = RECETTESTOTALS ~ IMPOTSOCIETES + IMPOTSALACONSOMMATION + RECETTESNONBUDGETAIRES + TAXESFORESTIERES + DROITSDETIMBRE + TAXESASSIMAUXXID, data = data1)
```

Residuals:

```

Min      1Q  Median      3Q      Max
-27496.4 -1536.0  250.5   2708.3 13544.0

```

Coefficients:

```

Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.813e+03 4.251e+03 1.132 0.26079
IMPOTSOCIETES 8.625e-01 4.719e-02 18.278 < 2e-16 ***
IMPOTSALACONSOMMATION 1.103e+00 1.377e-01 8.012 6.15e-12 ***
RECETTESNONBUDGETAIRES 1.231e+00 4.019e-01 3.063 0.00295 **
TAXESFORESTIERES 1.360e+00 4.727e-01 2.878 0.00509 **
DROITSDETIMBRE 3.043e+00 1.019e+00 2.987 0.00371 **
TAXESASSIMAUXXID 9.265e-01 3.727e-01 2.486 0.01492 *
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

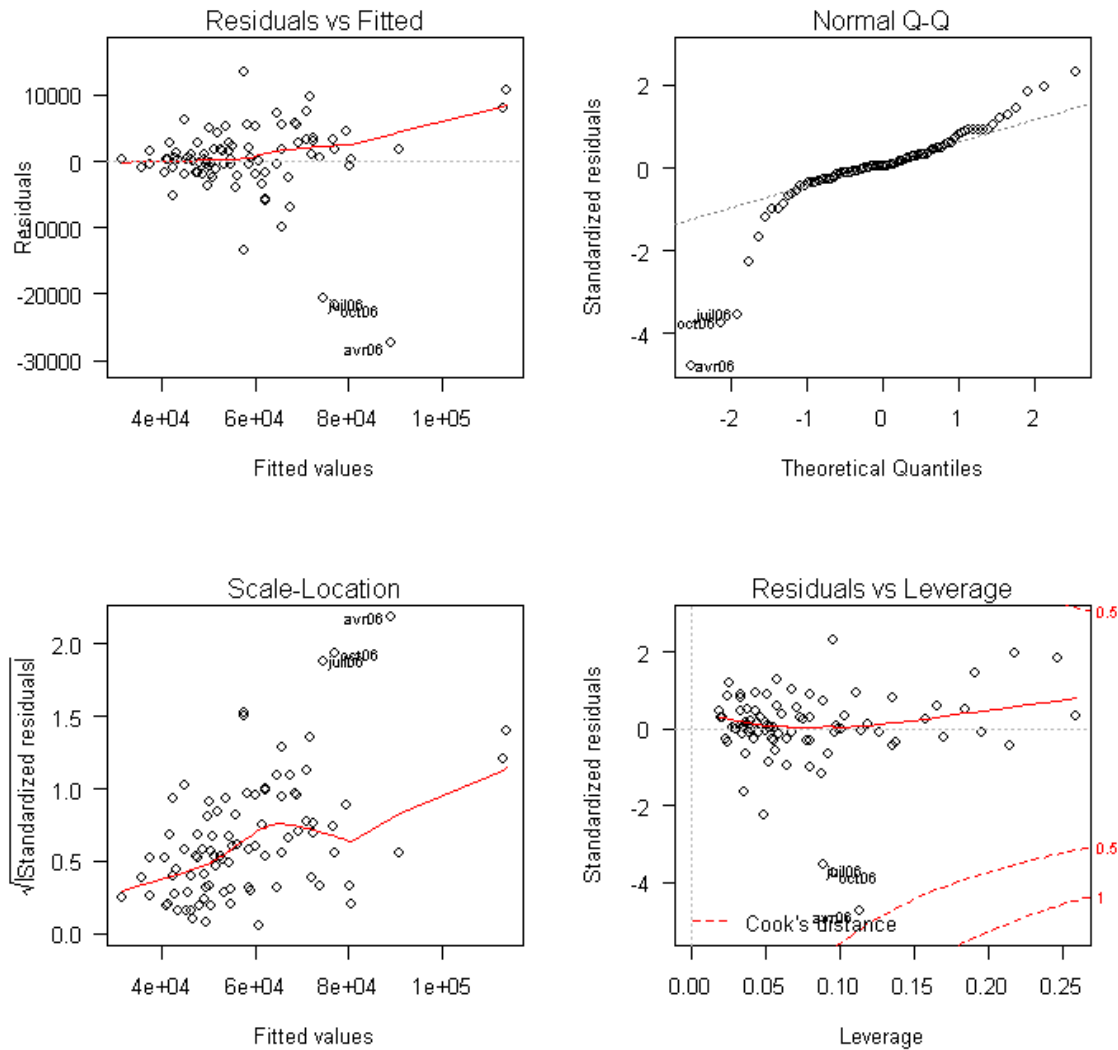
Residual standard error: 6130 on 83 degrees of freedom

Multiple R-Squared: 0.8663, Adjusted R-squared: 0.8566

F-statistic: 89.6 on 6 and 83 DF, p-value: < 2.2e-16

Dans le modèle obtenu, toutes les variables sont significatives (hautement) et son $R^2=0.8566$ qui est proche de 1. Jugeons maintenant de la qualité des résidus par le graphique ci-

après :



Les résidus semblent être à peu près alignés sur une même droite ; donc ce modèle semble avoir un caractère prédictif.

II.3- Procédure de sélection automatique descendante ou backward

On part de la régression sur toutes les variables et on obtient le meilleur modèle automatiquement.

```
##### Procédure descendante #####
> step.reg <- step(reg,direction="backward",k=log(40))
Start: AIC= 1614.9
RECETTES_TOTALES ~ IMP_SURREV_PERS_PHYS + IMPOTS_SALA_CONSOMMATION +
  IMPOTS_SOCIETES + TAXES_ASSIMAU_XID + PRELEVEMENTS_DIV + DROITS_DEN_REGISTRE +
  TAXES_DEN_REGISTRE + DROITS_DETIMBRE + TAXE_FOR_JEU + RECETTES_MINIJUSTICE +
  TAXES_FORESTIERES + TAXES_DELEVAGE + RECETTES_DOMANIALES + RECETTES_NON_BUDGETAIRES +
  DROITS_DE_TRANSIT + TAXES_MINIERES
```

	Df	Sum of Sq	RSS	AIC
- PRELEVEMENTSDIV	1	1.5627e+06	2.7834e+09	1.6110e+03
- DROITSDETRANSIT	1	2.6063e+06	2.7845e+09	1.6110e+03
- RECETTESDOMANIALES	1	2.6791e+06	2.7845e+09	1.6110e+03
- TAXESMINIERES	1	2.9520e+06	2.7848e+09	1.6110e+03
- TAXESDENREGISTREM	1	1.2998e+07	2.7949e+09	1.6120e+03
- TAXESDELEVAGE	1	2.0966e+07	2.8028e+09	1.6120e+03
- DROITSDENREGISTREM	1	3.0415e+07	2.8123e+09	1.6120e+03
- RECETTESMINJUSTICE	1	3.7151e+07	2.8190e+09	1.6120e+03
- TAXEFORJEUX	1	8.2985e+07	2.8648e+09	1.6140e+03
- TAXESFORESTIERES	1	9.0318e+07	2.8722e+09	1.6140e+03
<none>			2.7819e+09	1.6150e+03
- IMPSUREVPERSPHYS	1	1.2218e+08	2.9040e+09	1.6150e+03
- TAXESASSIMAUXXID	1	1.6107e+08	2.9429e+09	1.6160e+03
- DROITSDETIMBRE	1	2.7200e+08	3.0539e+09	1.6200e+03
- RECETTESNONBUDGETAIRES	1	3.2151e+08	3.1034e+09	1.6210e+03
- IMPOTSALACONSOMMATION	1	9.9898e+08	3.7808e+09	1.6390e+03
- IMPOTSOCIETES	1	8.6547e+09	1.1437e+10	1.7380e+03

Step: AIC= 1611.26

RECETTES TOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
 IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDENREGISTREM + TAXESDENREGISTREM +
 DROITSDETIMBRE + TAXEFORJEUX + RECETTESMINJUSTICE + TAXESFORESTIERES +
 TAXESDELEVAGE + RECETTESDOMANIALES + RECETTESNONBUDGETAIRES +
 DROITSDETRANSIT + TAXESMINIERES

	Df	Sum of Sq	RSS	AIC
- RECETTESDOMANIALES	1	1.9911e+06	2.7854e+09	1.6080e+03
- DROITSDETRANSIT	1	2.5853e+06	2.7860e+09	1.6080e+03
- TAXESMINIERES	1	4.1553e+06	2.7876e+09	1.6080e+03
- TAXESDENREGISTREM	1	1.3412e+07	2.7968e+09	1.6080e+03
- TAXESDELEVAGE	1	2.1503e+07	2.8049e+09	1.6080e+03
- DROITSDENREGISTREM	1	3.1928e+07	2.8154e+09	1.6090e+03
- RECETTESMINJUSTICE	1	3.5814e+07	2.8192e+09	1.6090e+03
- TAXEFORJEUX	1	8.1454e+07	2.8649e+09	1.6100e+03
- TAXESFORESTIERES	1	8.8788e+07	2.8722e+09	1.6100e+03
<none>			2.7834e+09	1.6110e+03
- IMPSUREVPERSPHYS	1	1.2077e+08	2.9042e+09	1.6110e+03
- TAXESASSIMAUXXID	1	1.8069e+08	2.9641e+09	1.6130e+03
- DROITSDETIMBRE	1	2.7191e+08	3.0553e+09	1.6160e+03
- RECETTESNONBUDGETAIRES	1	3.2870e+08	3.1121e+09	1.6180e+03
- IMPOTSALACONSOMMATION	1	1.0507e+09	3.8341e+09	1.6360e+03
- IMPOTSOCIETES	1	8.8945e+09	1.1678e+10	1.7370e+03

Step: AIC= 1607.64

RECETTES TOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
 IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDENREGISTREM + TAXESDENREGISTREM +
 DROITSDETIMBRE + TAXEFORJEUX + RECETTESMINJUSTICE + TAXESFORESTIERES +
 TAXESDELEVAGE + RECETTESNONBUDGETAIRES + DROITSDETRANSIT +
 TAXESMINIERES

	Df	Sum of Sq	RSS	AIC
- DROITSDETRANSIT	1	2.3401e+06	2.7878e+09	1.6040e+03
- TAXESMINIERES	1	6.3574e+06	2.7918e+09	1.6040e+03
- TAXESDENREGISTREM	1	1.2780e+07	2.7982e+09	1.6040e+03
- TAXESDELEVAGE	1	2.2239e+07	2.8077e+09	1.6050e+03
- DROITSDENREGISTREM	1	3.2515e+07	2.8179e+09	1.6050e+03
- RECETTESMINJUSTICE	1	3.4544e+07	2.8200e+09	1.6050e+03
- TAXEFORJEUX	1	7.9794e+07	2.8652e+09	1.6060e+03
- TAXESFORESTIERES	1	9.1484e+07	2.8769e+09	1.6070e+03
<none>			2.7854e+09	1.6080e+03
- IMPSUREVPERSPHYS	1	1.2092e+08	2.9063e+09	1.6080e+03
- TAXESASSIMAUXXID	1	1.7884e+08	2.9643e+09	1.6100e+03
- DROITSDETIMBRE	1	2.8755e+08	3.0730e+09	1.6130e+03
- RECETTESNONBUDGETAIRES	1	3.3529e+08	3.1207e+09	1.6140e+03
- IMPOTSALACONSOMMATION	1	1.0489e+09	3.8343e+09	1.6330e+03
- IMPOTSOCIETES	1	8.9854e+09	1.1771e+10	1.7340e+03

Step: AIC= 1604.03

RECETTES TOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
 IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDENREGISTREM + TAXESDENREGISTREM +
 DROITSDETIMBRE + TAXEFORJEUX + RECETTESMINJUSTICE + TAXESFORESTIERES +
 TAXESDELEVAGE + RECETTESNONBUDGETAIRES + TAXESMINIERES

	Df	Sum of Sq	RSS	AIC
- TAXESMINIERES	1	6.6631e+06	2.7944e+09	1.6010e+03
- TAXESDENREGISTREM	1	1.3680e+07	2.8014e+09	1.6010e+03
- TAXESDELEVAGE	1	3.6033e+07	2.8238e+09	1.6010e+03
- RECETTESMINJUSTICE	1	3.7105e+07	2.8249e+09	1.6020e+03
- DROITSDENREGISTREM	1	5.0733e+07	2.8385e+09	1.6020e+03
- TAXEFORJEUX	1	8.0332e+07	2.8681e+09	1.6030e+03
- TAXESFORESTIERES	1	8.9761e+07	2.8775e+09	1.6030e+03
<none>			2.7878e+09	1.6040e+03
- IMPSUREVPERSPHYS	1	1.3726e+08	2.9250e+09	1.6050e+03
- TAXESASSIMAUXXID	1	1.7790e+08	2.9657e+09	1.6060e+03
- DROITSDETIMBRE	1	3.0347e+08	3.0912e+09	1.6100e+03
- RECETTESNONBUDGETAIRES	1	3.3435e+08	3.1221e+09	1.6110e+03
- IMPOTSALACONSOMMATION	1	1.0528e+09	3.8405e+09	1.6290e+03

- IMPOTSOCIETES 1 9.0004e+09 1.1788e+10 1.7300e+03

Step: AIC= 1600.55

RECETTES TOTALES ~ IMPURSUREVPERSPHYS + IMPOTSALACONSOMMATION +
IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDENREGISTREM + TAXESDENREGISTREM +
DROITSDETIMBRE + TAXEFORJEUX + RECETTESMINJUSTICE + TAXESFORESTIERES +
TAXESDELEVAGE + RECETTESNONBUDGETAIRES

	Df	Sum of Sq	RSS	AIC
- TAXESDENREGISTREM	1	1.1540e+07	2.8060e+09	1.5970e+03
- RECETTESMINJUSTICE	1	3.4094e+07	2.8285e+09	1.5980e+03
- DROITSDENREGISTREM	1	4.6691e+07	2.8411e+09	1.5980e+03
- TAXESDELEVAGE	1	5.1413e+07	2.8458e+09	1.5990e+03
- TAXEFORJEUX	1	7.6433e+07	2.8709e+09	1.5990e+03
- TAXESFORESTIERES	1	8.3322e+07	2.8777e+09	1.6000e+03
<none>			2.7944e+09	1.6010e+03
- IMPURSUREVPERSPHYS	1	1.5511e+08	2.9495e+09	1.6020e+03
- TAXESASSIMAUXXID	1	1.7133e+08	2.9657e+09	1.6020e+03
- DROITSDETIMBRE	1	2.9722e+08	3.0916e+09	1.6060e+03
- RECETTESNONBUDGETAIRES	1	3.5583e+08	3.1502e+09	1.6080e+03
- IMPOTSALACONSOMMATION	1	1.0528e+09	3.8472e+09	1.6260e+03
- IMPOTSOCIETES	1	9.2056e+09	1.2000e+10	1.7280e+03

Step: AIC= 1597.23

RECETTES TOTALES ~ IMPURSUREVPERSPHYS + IMPOTSALACONSOMMATION +
IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDENREGISTREM + DROITSDETIMBRE +
TAXEFORJEUX + RECETTESMINJUSTICE + TAXESFORESTIERES + TAXESDELEVAGE +
RECETTESNONBUDGETAIRES

	Df	Sum of Sq	RSS	AIC
- RECETTESMINJUSTICE	1	3.8434e+07	2.8444e+09	1.5950e+03
- DROITSDENREGISTREM	1	5.2814e+07	2.8588e+09	1.5950e+03
- TAXESDELEVAGE	1	5.5346e+07	2.8613e+09	1.5950e+03
- TAXESFORESTIERES	1	7.7149e+07	2.8831e+09	1.5960e+03
- TAXEFORJEUX	1	8.4345e+07	2.8903e+09	1.5960e+03
<none>			2.8060e+09	1.5970e+03
- IMPURSUREVPERSPHYS	1	1.6388e+08	2.9698e+09	1.5990e+03
- TAXESASSIMAUXXID	1	1.6571e+08	2.9717e+09	1.5990e+03
- DROITSDETIMBRE	1	3.4443e+08	3.1504e+09	1.6040e+03
- RECETTESNONBUDGETAIRES	1	3.5908e+08	3.1650e+09	1.6040e+03
- IMPOTSALACONSOMMATION	1	1.0769e+09	3.8828e+09	1.6230e+03
- IMPOTSOCIETES	1	1.1713e+10	1.4519e+10	1.7410e+03

Step: AIC= 1594.77

RECETTES TOTALES ~ IMPURSUREVPERSPHYS + IMPOTSALACONSOMMATION +
IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDENREGISTREM + DROITSDETIMBRE +
TAXEFORJEUX + TAXESFORESTIERES + TAXESDELEVAGE + RECETTESNONBUDGETAIRES

	Df	Sum of Sq	RSS	AIC
- TAXESDELEVAGE	1	4.3272e+07	2.8877e+09	1.5920e+03
- TAXEFORJEUX	1	7.5340e+07	2.9197e+09	1.5930e+03
- DROITSDENREGISTREM	1	8.7935e+07	2.9323e+09	1.5940e+03
- TAXESFORESTIERES	1	9.1654e+07	2.9360e+09	1.5940e+03
<none>			2.8444e+09	1.5950e+03
- IMPURSUREVPERSPHYS	1	1.3651e+08	2.9809e+09	1.5950e+03
- TAXESASSIMAUXXID	1	1.6581e+08	3.0102e+09	1.5960e+03
- DROITSDETIMBRE	1	3.1923e+08	3.1636e+09	1.6010e+03
- RECETTESNONBUDGETAIRES	1	3.6662e+08	3.2110e+09	1.6020e+03
- IMPOTSALACONSOMMATION	1	1.1119e+09	3.9563e+09	1.6210e+03
- IMPOTSOCIETES	1	1.1684e+10	1.4529e+10	1.7380e+03

Step: AIC= 1592.44

RECETTES TOTALES ~ IMPURSUREVPERSPHYS + IMPOTSALACONSOMMATION +
IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDENREGISTREM + DROITSDETIMBRE +
TAXEFORJEUX + TAXESFORESTIERES + RECETTESNONBUDGETAIRES

	Df	Sum of Sq	RSS	AIC
- DROITSDENREGISTREM	1	6.7177e+07	2.9548e+09	1.5910e+03
- TAXEFORJEUX	1	6.7525e+07	2.9552e+09	1.5910e+03
- TAXESFORESTIERES	1	8.5045e+07	2.9727e+09	1.5910e+03
<none>			2.8877e+09	1.5920e+03
- IMPURSUREVPERSPHYS	1	1.2675e+08	3.0144e+09	1.5930e+03
- TAXESASSIMAUXXID	1	1.8193e+08	3.0696e+09	1.5940e+03
- DROITSDETIMBRE	1	2.8568e+08	3.1733e+09	1.5970e+03
- RECETTESNONBUDGETAIRES	1	3.4731e+08	3.2350e+09	1.5990e+03
- IMPOTSALACONSOMMATION	1	1.0854e+09	3.9731e+09	1.6170e+03
- IMPOTSOCIETES	1	1.1677e+10	1.4565e+10	1.7340e+03

Step: AIC= 1590.82

RECETTES TOTALES ~ IMPURSUREVPERSPHYS + IMPOTSALACONSOMMATION +
IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDETIMBRE + TAXEFORJEUX +
TAXESFORESTIERES + RECETTESNONBUDGETAIRES

	Df	Sum of Sq	RSS	AIC
- TAXEFORJEUX	1	4.4149e+07	2.9990e+09	1.5880e+03
- TAXESFORESTIERES	1	1.0482e+08	3.0597e+09	1.5900e+03
<none>			2.9548e+09	1.5910e+03
- IMPURSUREVPERSPHYS	1	1.2957e+08	3.0844e+09	1.5910e+03

```

- TAXESASSIMAUXXID      1 2.2284e+08 3.1777e+09 1.5940e+03
- RECETTESNONBUDGETAIRES 1 2.8415e+08 3.2390e+09 1.5950e+03
- DROITSDETIMBRE        1 3.9211e+08 3.3470e+09 1.5980e+03
- IMPOTSALACONSOMMATION 1 1.4059e+09 4.3608e+09 1.6220e+03
- IMPOTSOCIETES         1 1.2694e+10 1.5648e+10 1.7370e+03

```

Step: AIC= 1588.47

```

RECETTESTOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
  IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDETIMBRE + TAXESFORESTIERES +
  RECETTESNONBUDGETAIRES

```

	Df	Sum of Sq	RSS	AIC
- TAXESFORESTIERES	1	1.1411e+08	3.1131e+09	1.5880e+03
- IMPSUREVPERSPHYS	1	1.2030e+08	3.1193e+09	1.5880e+03
<none>			2.9990e+09	1.5880e+03
- TAXESASSIMAUXXID	1	2.4694e+08	3.2459e+09	1.5920e+03
- RECETTESNONBUDGETAIRES	1	3.3504e+08	3.3340e+09	1.5940e+03
- DROITSDETIMBRE	1	3.7881e+08	3.3778e+09	1.5950e+03
- IMPOTSALACONSOMMATION	1	1.5038e+09	4.5028e+09	1.6210e+03
- IMPOTSOCIETES	1	1.2664e+10	1.5663e+10	1.7340e+03

Step: AIC= 1588.14

```

RECETTESTOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
  IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDETIMBRE + RECETTESNONBUDGETAIRES

```

	Df	Sum of Sq	RSS	AIC
<none>			3.1131e+09	1.5880e+03
- TAXESASSIMAUXXID	1	3.0317e+08	3.4163e+09	1.5930e+03
- IMPSUREVPERSPHYS	1	3.1746e+08	3.4306e+09	1.5930e+03
- RECETTESNONBUDGETAIRES	1	3.8409e+08	3.4972e+09	1.5950e+03
- DROITSDETIMBRE	1	3.9340e+08	3.5065e+09	1.5950e+03
- IMPOTSALACONSOMMATION	1	1.3936e+09	4.5067e+09	1.6180e+03
- IMPOTSOCIETES	1	1.3286e+10	1.6399e+10	1.7340e+03

En régressant sur les variables sélectionnées par la méthode backward, on a les résultats suivants :

```

summary(step.reg)

Call:
lm(formula = RECETTESTOTALES ~ IMPSUREVPERSPHYS + IMPOTSALACONSOMMATION +
  IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDETIMBRE + RECETTESNONBUDGETAIRES,
  data = data1)

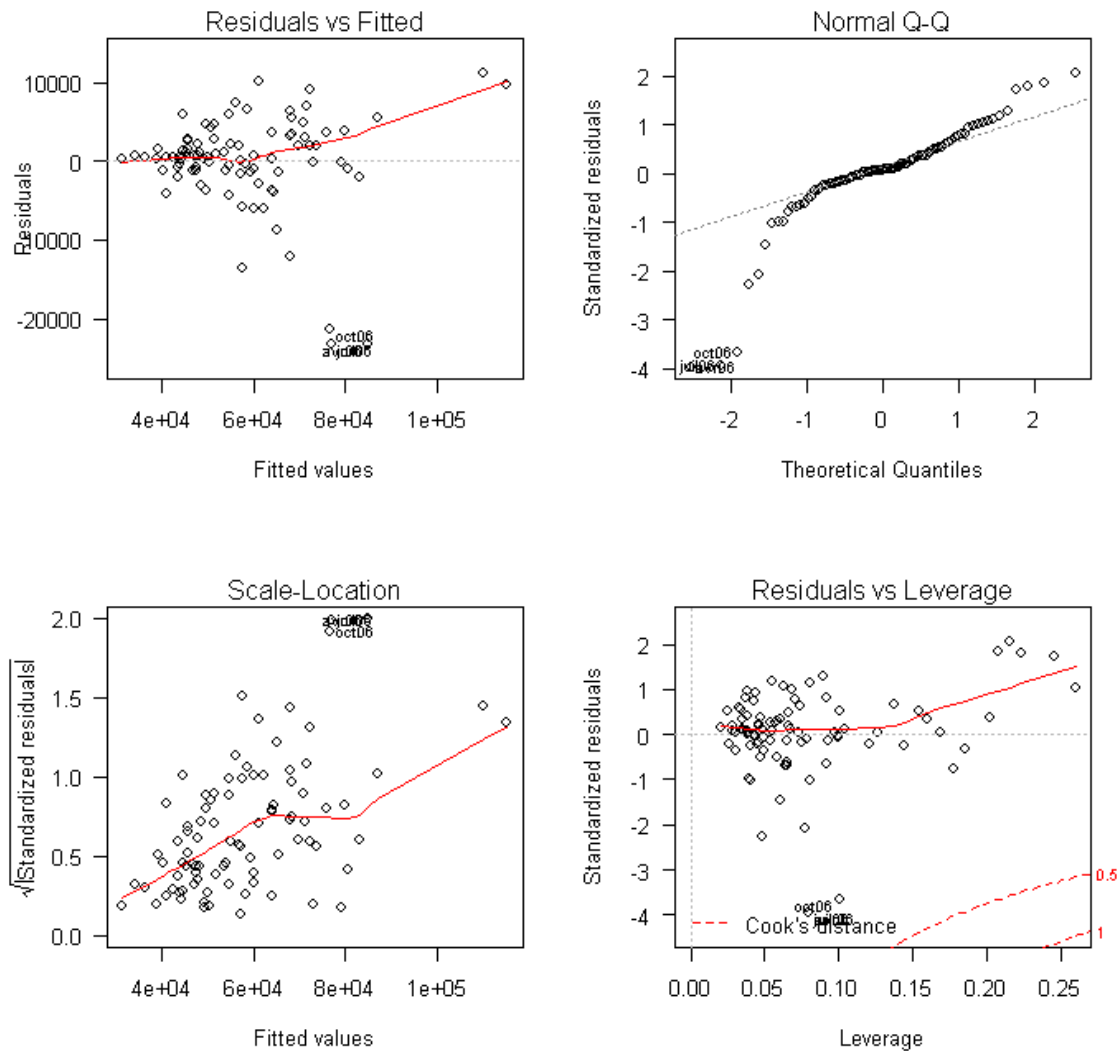
Residuals:
    Min       1Q   Median       3Q      Max
-23305.3 -1190.3   457.6   2805.8  11239.3

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -945.55716  4811.21073   -0.197  0.84467
IMPSUREVPERSPHYS    1.25517    0.43143    2.909  0.00465 **
IMPOTSALACONSOMMATION  0.91661    0.15037    6.096  3.29e-08 ***
IMPOTSOCIETES      0.88000    0.04676   18.821 < 2e-16 ***
TAXESASSIMAUXXID  1.04870    0.36887    2.843  0.00562 **
DROITSDETIMBRE     3.31715    1.02425    3.239  0.00173 **
RECETTESNONBUDGETAIRES 1.27822    0.39944    3.200  0.00195 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6124 on 83 degrees of freedom
Multiple R-Squared:  0.8665,    Adjusted R-squared:  0.8569
F-statistic: 89.81 on 6 and 83 DF,  p-value: < 2.2e-16

```

On ici les coefficients Presque tous hautement significatifs et le $\overline{R^2}=0.8569$ qui est proche de 1. Regardons dès à présent ce qui se passe du côté des résidus.



II.4- Procédure de sélection automatique mixte

Elle combine en effet les procédures forward et backward, compare les modèles suivant les AIC et sélectionne enfin un modèle statistiquement significatif

```
reg <- lm(RECETTESTOTALES~ 1,data=data1)
> step.reg=step(reg,direction="both",k=log(40),scope=list(lower=-1, upper=~IMPSURREVPERSPHYS +
IMPOTSALACONSOMMATION +
+ IMPOTSOCIETES + TAXESASSIMAUXXID + PRELEVEMENTSDIV + DROITSDENREGISTREM +
+ TAXESDENREGISTREM + DROITSDETIMBRE + TAXEFORJEUX + RECETTESMINJUSTICE +
+ TAXESFORESTIERES + TAXESDELEVAGE + RECETTESDOMANIALES + RECETTESNONBUDGETAIRES +
+ DROITSDETRANSIT + TAXESMINIERES),trace=TRUE)
Start: AIC= 1747.25
RECETTESTOTALES ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ IMPOTSOCIETES	1	1.6465e+10	6.8579e+09	1.6410e+03
+ TAXESDENREGISTREM	1	5.0588e+09	1.8264e+10	1.7290e+03
+ IMPOTSALACONSOMMATION	1	3.9291e+09	1.9394e+10	1.7340e+03
+ TAXESMINIERES	1	2.2260e+09	2.1097e+10	1.7420e+03
+ IMPSURREVPERSPHYS	1	1.9353e+09	2.1388e+10	1.7430e+03

+ RECETTESNONBUDGETAIRES	1	1.5820e+09	2.1741e+10	1.7450e+03
+ TAXESFORESTIERES	1	1.5719e+09	2.1751e+10	1.7450e+03
+ TAXESDELEVAGE	1	1.4158e+09	2.1908e+10	1.7450e+03
+ DROITSDENREGISTREM	1	1.2818e+09	2.2042e+10	1.7460e+03
<none>			2.3323e+10	1.7470e+03
+ PRELEVEMENTSDIV	1	4.7978e+08	2.2844e+10	1.7490e+03
+ DROITSDETIMBRE	1	3.2562e+08	2.2998e+10	1.7500e+03
+ DROITSDETRANSIT	1	2.8569e+08	2.3038e+10	1.7500e+03
+ TAXESASSIMAUXXID	1	2.0948e+08	2.3114e+10	1.7500e+03
+ RECETTESDOMANIALES	1	1.9496e+08	2.3128e+10	1.7500e+03
+ TAXEFORJEUX	1	1.2770e+08	2.3196e+10	1.7500e+03
+ RECETTESMINJUSTICE	1	3.0422e+07	2.3293e+10	1.7510e+03

Step: AIC= 1640.78
RECETTESTOTALES ~ IMPOTSOCIETES

	Df	Sum of Sq	RSS	AIC
+ IMPOTSALACONSOMMATION	1	2.2002e+09	4.6577e+09	1.6100e+03
+ IMPSUREVPERSPHYS	1	1.3013e+09	5.5566e+09	1.6260e+03
+ TAXESASSIMAUXXID	1	5.5431e+08	6.3036e+09	1.6370e+03
+ TAXESDELEVAGE	1	4.8933e+08	6.3686e+09	1.6380e+03
+ TAXESFORESTIERES	1	4.7542e+08	6.3825e+09	1.6380e+03
+ RECETTESNONBUDGETAIRES	1	3.2495e+08	6.5330e+09	1.6400e+03
+ TAXEFORJEUX	1	2.9946e+08	6.5584e+09	1.6400e+03
+ DROITSDENREGISTREM	1	2.7858e+08	6.5793e+09	1.6410e+03
<none>			6.8579e+09	1.6410e+03
+ TAXESDENREGISTREM	1	1.8727e+08	6.6706e+09	1.6420e+03
+ DROITSDETIMBRE	1	1.8375e+08	6.6742e+09	1.6420e+03
+ TAXESMINIERES	1	1.5227e+08	6.7056e+09	1.6420e+03
+ RECETTESMINJUSTICE	1	1.2139e+08	6.7365e+09	1.6430e+03
+ RECETTESDOMANIALES	1	1.1790e+08	6.7400e+09	1.6430e+03
+ DROITSDETRANSIT	1	2.0719e+07	6.8372e+09	1.6440e+03
+ PRELEVEMENTSDIV	1	1.8427e+07	6.8395e+09	1.6440e+03
- IMPOTSOCIETES	1	1.6465e+10	2.3323e+10	1.7470e+03

Step: AIC= 1609.64
RECETTESTOTALES ~ IMPOTSOCIETES + IMPOTSALACONSOMMATION

	Df	Sum of Sq	RSS	AIC
+ RECETTESNONBUDGETAIRES	1	6.1113e+08	4.0466e+09	1.6010e+03
+ TAXESFORESTIERES	1	5.1550e+08	4.1422e+09	1.6030e+03
+ DROITSDETIMBRE	1	4.5529e+08	4.2024e+09	1.6040e+03
+ IMPSUREVPERSPHYS	1	3.2986e+08	4.3278e+09	1.6070e+03
+ TAXESASSIMAUXXID	1	2.7612e+08	4.3816e+09	1.6080e+03
<none>			4.6577e+09	1.6100e+03
+ TAXEFORJEUX	1	1.3736e+08	4.5203e+09	1.6110e+03
+ RECETTESMINJUSTICE	1	8.6143e+07	4.5716e+09	1.6120e+03
+ TAXESDENREGISTREM	1	5.9029e+07	4.5987e+09	1.6120e+03
+ TAXESMINIERES	1	3.8757e+07	4.6189e+09	1.6130e+03
+ RECETTESDOMANIALES	1	3.4472e+07	4.6232e+09	1.6130e+03
+ PRELEVEMENTSDIV	1	1.4422e+07	4.6433e+09	1.6130e+03
+ TAXESDELEVAGE	1	9.8473e+06	4.6479e+09	1.6130e+03
+ DROITSDENREGISTREM	1	1.3998e+06	4.6563e+09	1.6130e+03
+ DROITSDETRANSIT	1	1.2864e+06	4.6564e+09	1.6130e+03
- IMPOTSALACONSOMMATION	1	2.2002e+09	6.8579e+09	1.6410e+03
- IMPOTSOCIETES	1	1.4737e+10	1.9394e+10	1.7340e+03

Step: AIC= 1600.67
RECETTESTOTALES ~ IMPOTSOCIETES + IMPOTSALACONSOMMATION + RECETTESNONBUDGETAIRES

	Df	Sum of Sq	RSS	AIC
+ TAXESFORESTIERES	1	3.7823e+08	3.6683e+09	1.5960e+03
+ DROITSDETIMBRE	1	2.9803e+08	3.7485e+09	1.5970e+03
+ TAXESASSIMAUXXID	1	2.9474e+08	3.7518e+09	1.5980e+03
+ IMPSUREVPERSPHYS	1	2.6013e+08	3.7864e+09	1.5980e+03
+ DROITSDENREGISTREM	1	2.4598e+08	3.8006e+09	1.5990e+03
<none>			4.0466e+09	1.6010e+03
+ TAXESMINIERES	1	1.1007e+08	3.9365e+09	1.6020e+03
+ DROITSDETRANSIT	1	7.0291e+07	3.9763e+09	1.6030e+03
+ RECETTESDOMANIALES	1	6.9638e+07	3.9769e+09	1.6030e+03
+ TAXESDENREGISTREM	1	6.6224e+07	3.9803e+09	1.6030e+03
+ TAXEFORJEUX	1	5.8208e+07	3.9884e+09	1.6030e+03
+ RECETTESMINJUSTICE	1	1.6467e+07	4.0301e+09	1.6040e+03
+ PRELEVEMENTSDIV	1	4.2941e+06	4.0423e+09	1.6040e+03
+ TAXESDELEVAGE	1	4.2571e+06	4.0423e+09	1.6040e+03
- RECETTESNONBUDGETAIRES	1	6.1113e+08	4.6577e+09	1.6100e+03
- IMPOTSALACONSOMMATION	1	2.4864e+09	6.5330e+09	1.6400e+03
- IMPOTSOCIETES	1	1.3124e+10	1.7170e+10	1.7270e+03

Step: AIC= 1595.53
RECETTESTOTALES ~ IMPOTSOCIETES + IMPOTSALACONSOMMATION + RECETTESNONBUDGETAIRES + TAXESFORESTIERES

	Df	Sum of Sq	RSS	AIC
+ DROITSDETIMBRE	1	3.1675e+08	3.3516e+09	1.5910e+03
+ TAXESASSIMAUXXID	1	2.1380e+08	3.4545e+09	1.5940e+03
+ DROITSDENREGISTREM	1	1.7493e+08	3.4934e+09	1.5950e+03

```

<none> 3.6683e+09 1.5960e+03
+ TAXESDENREGISTREM 1 9.6990e+07 3.5713e+09 1.5970e+03
+ DROITSDETRANSIT 1 7.0943e+07 3.5974e+09 1.5970e+03
+ RECETTESDOMANIALES 1 6.7896e+07 3.6004e+09 1.5980e+03
+ IMPSUREVPERSPHYS 1 6.6662e+07 3.6017e+09 1.5980e+03
+ TAXEFORJEUX 1 4.3442e+07 3.6249e+09 1.5980e+03
+ TAXESMINIERES 1 2.5965e+07 3.6424e+09 1.5990e+03
+ PRELEVEMENTSDIV 1 2.3381e+07 3.6450e+09 1.5990e+03
+ RECETTESMINJUSTICE 1 1.3592e+07 3.6547e+09 1.5990e+03
+ TAXESDELEVAGE 1 2.5600e+03 3.6683e+09 1.5990e+03
- TAXESFORESTIERES 1 3.7823e+08 4.0466e+09 1.6010e+03
- RECETTESNONBUDGETAIRES 1 4.7387e+08 4.1422e+09 1.6030e+03
- IMPOTSALA CONSOMMATION 1 2.4859e+09 6.1542e+09 1.6380e+03
- IMPOTSOCIETES 1 1.2446e+10 1.6115e+10 1.7250e+03

Step: AIC= 1591.09
RECETTES TOTALES ~ IMPOTSOCIETES + IMPOTSALA CONSOMMATION + RECETTESNONBUDGETAIRES +
TAXESFORESTIERES + DROITSDETIMBRE

```

```

Df Sum of Sq RSS AIC
+ TAXESASSIMAUXXID 1 2.3230e+08 3.1193e+09 1.5880e+03
<none> 3.3516e+09 1.5910e+03
+ IMPSUREVPERSPHYS 1 1.0566e+08 3.2459e+09 1.5920e+03
+ DROITSDENREGISTREM 1 7.8034e+07 3.2736e+09 1.5930e+03
+ TAXEFORJEUX 1 5.6506e+07 3.2951e+09 1.5930e+03
+ DROITSDETRANSIT 1 3.5245e+07 3.3163e+09 1.5940e+03
+ TAXESDENREGISTREM 1 3.0808e+07 3.3208e+09 1.5940e+03
+ TAXESDELEVAGE 1 2.1004e+07 3.3306e+09 1.5940e+03
+ RECETTESMINJUSTICE 1 2.0410e+07 3.3312e+09 1.5940e+03
+ PRELEVEMENTSDIV 1 1.7989e+06 3.3498e+09 1.5950e+03
+ TAXESMINIERES 1 4.7437e+05 3.3511e+09 1.5950e+03
+ RECETTESDOMANIALES 1 1.3973e+05 3.3514e+09 1.5950e+03
- DROITSDETIMBRE 1 3.1675e+08 3.6683e+09 1.5960e+03
- RECETTESNONBUDGETAIRES 1 3.3344e+08 3.6850e+09 1.5960e+03
- TAXESFORESTIERES 1 3.9695e+08 3.7485e+09 1.5970e+03
- IMPOTSALA CONSOMMATION 1 2.6863e+09 6.0379e+09 1.6400e+03
- IMPOTSOCIETES 1 1.2340e+10 1.5692e+10 1.7260e+03

```

```

Step: AIC= 1588.32
RECETTES TOTALES ~ IMPOTSOCIETES + IMPOTSALA CONSOMMATION + RECETTESNONBUDGETAIRES +
TAXESFORESTIERES + DROITSDETIMBRE + TAXESASSIMAUXXID

```

```

Df Sum of Sq RSS AIC
<none> 3.1193e+09 1.5880e+03
+ IMPSUREVPERSPHYS 1 1.2030e+08 2.9990e+09 1.5880e+03
+ DROITSDETRANSIT 1 6.0877e+07 3.0584e+09 1.5900e+03
+ DROITSDENREGISTREM 1 4.8023e+07 3.0713e+09 1.5910e+03
+ TAXESDENREGISTREM 1 3.9572e+07 3.0797e+09 1.5910e+03
+ TAXEFORJEUX 1 3.4873e+07 3.0844e+09 1.5910e+03
- TAXESASSIMAUXXID 1 2.3230e+08 3.3516e+09 1.5910e+03
+ TAXESDELEVAGE 1 1.4316e+07 3.1050e+09 1.5920e+03
+ RECETTESMINJUSTICE 1 1.0928e+07 3.1084e+09 1.5920e+03
+ TAXESMINIERES 1 1.0579e+07 3.1087e+09 1.5920e+03
+ PRELEVEMENTSDIV 1 1.9452e+06 3.1173e+09 1.5920e+03
+ RECETTESDOMANIALES 1 1.7878e+06 3.1175e+09 1.5920e+03
- TAXESFORESTIERES 1 3.1128e+08 3.4306e+09 1.5930e+03
- DROITSDETIMBRE 1 3.3525e+08 3.4545e+09 1.5940e+03
- RECETTESNONBUDGETAIRES 1 3.5268e+08 3.4720e+09 1.5940e+03
- IMPOTSALA CONSOMMATION 1 2.4126e+09 5.5319e+09 1.6360e+03
- IMPOTSOCIETES 1 1.2555e+10 1.5674e+10 1.7300e+03

```

En régressant sur les variables sélectionnées, on obtient :

```
> summary(step.reg)
```

Call:

```
lm(formula = RECETTES TOTALES ~ IMPOTSOCIETES + IMPOTSALA CONSOMMATION +
RECETTESNONBUDGETAIRES + TAXESFORESTIERES + DROITSDETIMBRE +
TAXESASSIMAUXXID, data = data1)
```

Residuals:

```

      Min       1Q   Median       3Q      Max
-27496.4 -1536.0   250.5   2708.3  13544.0

```

Coefficients:

```

      Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.813e+03  4.251e+03  1.132  0.26079
IMPOTSOCIETES  8.625e-01  4.719e-02  18.278 < 2e-16 ***
IMPOTSALA CONSOMMATION  1.103e+00  1.377e-01  8.012  6.15e-12 ***
RECETTESNONBUDGETAIRES  1.231e+00  4.019e-01  3.063  0.00295 **
TAXESFORESTIERES  1.360e+00  4.727e-01  2.878  0.00509 **
DROITSDETIMBRE  3.043e+00  1.019e+00  2.987  0.00371 **
TAXESASSIMAUXXID  9.265e-01  3.727e-01  2.486  0.01492 *
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 6130 on 83 degrees of freedom

Multiple R-Squared: 0.8663, Adjusted R-squared: 0.8566
 F-statistic: 89.6 on 6 and 83 DF, p-value: < 2.2e-16

On ici les mêmes conclusions que la méthode backward car on aboutit au même modèle

II.5- Méthode sélection à l'aide du C_p de Mallou et du R^2 ajusté

II.5.1-Méthode C_p de Mallou

Les résultats obtenus sont les suivants :

```
#Recherche des meilleurs modèles au sens du Cp
reg.choix1=leaps(data,data1[, "RECETTESTOTALES"],method="Cp",nbest=1)
reg.choix1$Cp
[1] 96.233023 39.767762 25.528263 17.477579
[5] 11.060759 6.723512 5.691244 6.518097
[9] 6.733024 7.583176 8.561888 10.255230
[13] 12.151787 14.073503 16.000000

t=(reg.choix1$Cp==min(reg.choix1$Cp))
[1] FALSE FALSE FALSE FALSE FALSE FALSE TRUE
[8] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[15] FALSE

#Liste des variables explicatives
> colnames(data)[reg.choix1$whi[t]]
[1] "IMPSURREVPERSPHYS"
[2] "IMPOTSALACONSOMMATION"
[3] "IMPOTSOCIETES"
[4] "TAXESASSIMAUXXID"
[5] "DROITSDETIMBRE"
[6] "TAXESFORESTIERES"
[7] "RECETTESNONBUDGETAIRES"
```

En procédant à la régression sur les variables sélectionnées on obtient les résultats suivants :

```
> summary(reg)

Call:
lm(formula = RECETTESTOTALES ~ RECETTESNONBUDGETAIRES + TAXESFORESTIERES +
    DROITSDETIMBRE + TAXESASSIMAUXXID + IMPOTSALACONSOMMATION +
    IMPSURREVPERSPHYS)

Residuals:
    Min       1Q   Median       3Q      Max
-16025  -7642  -3475   3671   58493

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.889e+03  1.094e+04  0.355  0.7232
RECETTESNONBUDGETAIRES  2.384e+00  8.898e-01  2.679  0.0089 **
TAXESFORESTIERES      2.329e+00  1.179e+00  1.976  0.0515 .
DROITSDETIMBRE        3.508e+00  2.299e+00  1.526  0.1308
TAXESASSIMAUXXID      2.587e-01  8.315e-01  0.311  0.7565
IMPOTSALACONSOMMATION  1.470e+00  3.410e-01  4.312  4.43e-05 ***
IMPSURREVPERSPHYS     2.723e-01  1.085e+00  0.251  0.8025
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13740 on 83 degrees of freedom
Multiple R-Squared: 0.3285, Adjusted R-squared: 0.2799
F-statistic: 6.766 on 6 and 83 DF, p-value: 7.312e-06
```

On constate que le modèle obtenu a plusieurs variables non significatives et $\bar{R}^2 = 0.2799$ qui est proche de 0. D'où il n'est pas nécessaire de tracer le graphique des résidus correspondants pour vérifier la normalité.

II.5.2- Méthode R^2 ajusté

Les résultats obtenus sont les suivants :

```
reg.choix2$adjr2
 [1] 0.7026220 0.7957073 0.8204487 0.8353164 0.8477450 0.8568751 0.8604399
 [8] 0.8607968 0.8622610 0.8626077 0.8627265 0.8615156 0.8598889 0.8581707
 [15] 0.8563967

t=(reg.choix2$adjr2==max(reg.choix2$adjr2))
> t
 [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
 [13] FALSE FALSE FALSE

#Variables explicatives du meilleur mod`ele au sens du R2 ajust`e

colnames(data)[reg.choix2$whi[t]]

 1] "IMPSURREVPERSPHYS"
 2] "IMPOTSALACONSOMMATION"
 3] "IMPOTSOCIETES"
 4] "TAXESASSIMAUXXID"
 5] "DROITSDENREGISTREM"
 6] "DROITSDETIMBRE"
 7] "TAXEFORJEUX"
 8] "RECETTESMINJUSTICE"
 9] "TAXESFORESTIERES"
[10] "TAXESDELEVAGE"
[11] "RECETTESNONBUDGETAIRES"
```

En procédant à la régression sur les variables sélectionnées on obtient les résultats suivants :

```
regR² <-
lm(RECETTES_TOTALES~IMPSURREVPERSPHYS+IMPOTSALACONSOMMATION+IMPOTSOCIETES+TAXESASSIMAUXXID+DROITSDENREGISTREM
+DROITSDETIMBRE+TAXEFORJEUX+RECETTESMINJUSTICE+TAXESFORESTIERES+TAXESDELEVAGE+RECETTESNONBUDGETAIRES)
> summary(regR²)

Call:
lm(formula = RECETTES_TOTALES ~ IMPSURREVPERSPHYS + IMPOTSALACONSOMMATION +
    IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDENREGISTREM + DROITSDETIMBRE +
    TAXEFORJEUX + RECETTESMINJUSTICE + TAXESFORESTIERES + TAXESDELEVAGE +
    RECETTESNONBUDGETAIRES)

Residuals:
    Min       1Q   Median       3Q      Max
-24512.0 -1005.3   221.6   2500.2  10969.4

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -316.36163   5153.11758   -0.061  0.95120
IMPSURREVPERSPHYS    1.04704    0.49056    2.134  0.03595 *
IMPOTSALACONSOMMATION  0.92998    0.16997    5.471 5.23e-07 ***
IMPOTSOCIETES      0.86302    0.04783   18.044 < 2e-16 ***
TAXESASSIMAUXXID   0.80010    0.37279    2.146  0.03496 *
DROITSDENREGISTREM  2.32873    1.92194    1.212  0.22930
DROITSDETIMBRE     3.32434    1.07436    3.094  0.00274 **
TAXEFORJEUX       6.71136    4.38303    1.531  0.12976
RECETTESMINJUSTICE -7.95351    7.69477   -1.034  0.30451
TAXESFORESTIERES   0.77463    0.52896    1.464  0.14709
TAXESDELEVAGE     -15.56343   12.54741   -1.240  0.21856
RECETTESNONBUDGETAIRES  1.49704    0.47384    3.159  0.00225 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5998 on 78 degrees of freedom
Multiple R-Squared:  0.8797,    Adjusted R-squared:  0.8627
F-statistic: 51.85 on 11 and 78 DF,  p-value: < 2.2e-16
```

$\bar{R}^2 = 0.8627$, Or toutes les coefficients ne sont pas statistiquement significatifs différents de 0 à 5%

II.5.3- Liste des meilleurs modèles pour chaque dimension

Il se sert des méthodes Cp de Mallow et \bar{R}^2 en écrivant le petit algorithme ci-dessous

```
for (i in (1 :15)) {cat(reg.choix2$adjr2[i],
reg.choix1$Cp[i],colnames(data)[reg.choix1$whi[i,]],"\n")}
```

```
0.702622 96.23302 IMPOTSOCIETES
0.7957073 39.76776 IMPOTSALACONSOMMATION IMPOTSOCIETES
0.8204487 25.52826 IMPOTSALACONSOMMATION IMPOTSOCIETES RECETTESNONBUDGETAIRES
0.8353164 17.47758 IMPOTSALACONSOMMATION IMPOTSOCIETES TAXESFORESTIERES RECETTESNONBUDGETAIRES
0.847745 11.06076 IMPOTSALACONSOMMATION IMPOTSOCIETES DROITSDETIMBRE TAXESFORESTIERES
RECETTESNONBUDGETAIRES
0.856875 6.723512 IMPSURREVPERSPHYS IMPOTSALACONSOMMATION IMPOTSOCIETES TAXESASSIMAUXXID DROITSDETIMBRE
RECETTESNONBUDGETAIRES
0.8604399 5.691244 IMPSURREVPERSPHYS IMPOTSALACONSOMMATION IMPOTSOCIETES TAXESASSIMAUXXID DROITSDETIMBRE
TAXESFORESTIERES RECETTESNONBUDGETAIRES
0.8607968 6.518097 IMPSURREVPERSPHYS IMPOTSALACONSOMMATION IMPOTSOCIETES TAXESASSIMAUXXID DROITSDETIMBRE
TAXEFORJEUX TAXESFORESTIERES RECETTESNONBUDGETAIRES
0.862261 6.733024 IMPSURREVPERSPHYS IMPOTSALACONSOMMATION IMPOTSOCIETES TAXESASSIMAUXXID DROITSDENREGISTREM
DROITSDETIMBRE TAXEFORJEUX TAXESFORESTIERES RECETTESNONBUDGETAIRES
0.8626077 7.583176 IMPSURREVPERSPHYS IMPOTSALACONSOMMATION IMPOTSOCIETES TAXESASSIMAUXXID DROITSDENREGISTREM
DROITSDETIMBRE TAXEFORJEUX TAXESFORESTIERES TAXESDELEVAGE RECETTESNONBUDGETAIRES
0.8627265 8.561888 IMPSURREVPERSPHYS IMPOTSALACONSOMMATION IMPOTSOCIETES TAXESASSIMAUXXID DROITSDENREGISTREM
DROITSDETIMBRE TAXEFORJEUX RECETTESMINJUSTICE TAXESFORESTIERES TAXESDELEVAGE RECETTESNONBUDGETAIRES
0.8615156 10.25523 IMPSURREVPERSPHYS IMPOTSALACONSOMMATION IMPOTSOCIETES TAXESASSIMAUXXID DROITSDENREGISTREM
TAXESDENREGISTREM DROITSDETIMBRE TAXEFORJEUX RECETTESMINJUSTICE TAXESFORESTIERES TAXESDELEVAGE
RECETTESNONBUDGETAIRES
0.8598889 12.15179 IMPSURREVPERSPHYS IMPOTSALACONSOMMATION IMPOTSOCIETES TAXESASSIMAUXXID DROITSDENREGISTREM
TAXESDENREGISTREM DROITSDETIMBRE TAXEFORJEUX RECETTESMINJUSTICE TAXESFORESTIERES TAXESDELEVAGE
RECETTESDOMANIALES RECETTESNONBUDGETAIRES
0.8581707 14.07350 IMPSURREVPERSPHYS IMPOTSALACONSOMMATION IMPOTSOCIETES TAXESASSIMAUXXID DROITSDENREGISTREM
TAXESDENREGISTREM DROITSDETIMBRE TAXEFORJEUX RECETTESMINJUSTICE TAXESFORESTIERES TAXESDELEVAGE
RECETTESDOMANIALES RECETTESNONBUDGETAIRES DROITSDETRANSIT
0.8563967 16 IMPSURREVPERSPHYS IMPOTSALACONSOMMATION IMPOTSOCIETES TAXESASSIMAUXXID PRELEVEMENTSDIV
DROITSDENREGISTREM TAXESDENREGISTREM DROITSDETIMBRE TAXEFORJEUX RECETTESMINJUSTICE TAXESFORESTIERES
TAXESDELEVAGE RECETTESDOMANIALES RECETTESNONBUDGETAIRES DROITSDETRANSIT
```

En parcourant toutes les méthodes de sélection des variables, à chaque fois nous obtenons s modèles différent. Or l'objectif ici c'est de construire un modèle de prédiction qui puisse correctement rendre compte du total des recettes fiscales camerounaises en un moment donné (mensuel). Pour cela, il faudrait un modèle où toutes les variables soient statistiquement significatives, avec les conditions normalité et de linéarité vérifiées (\bar{R}^2 ajusté proche de 1).

De cette analyse, il ressort que les méthodes par élimination pas à pas, backward, forward, mixte, nous fournissent des modèles où toutes les variables explicatives sont statistiquement significatives avec un bon R^2 ajusté généralement proche de 1 (plus de 85%) et avec un léger changement de celles-ci dans ces modèles.

Les variables qui reviennent le plus souvent sont :

- IMPOTSOCIETES,
- IMPOTSALACONSOMMATION,
- RECETTESNONBUDGETAIRES,
- TAXESFORESTIERES,
- DROITSDETIMBRE,
- TAXESASSIMAUXXID,
- IMPOSURREVPERSPHYS.

La régression sur ces différentes variables explicatives a été bien sélectionnée par la liste des meilleurs modèles à chaque dimension donnée par la combinaison des méthodes Cp de Mallow et R^2 ajusté. Réalisons une régression sur ces différents prédictors.

```
mod <- lm(RECETTESTOTALES~IMPSURREVPERSPHYS+IMPOTSALACONSOMMATION +IMPOTSOCIETES+ TAXESASSIMAUXXID
+DROITSDETIMBRE +TAXESFORESTIERES +RECETTESNONBUDGETAIRES)
```

```
summary(mod)
```

```
Call:
```

```
lm(formula = RECETTESTOTALES ~ IMPSURREVPERSPHYS + IMPOTSALACONSOMMATION +
IMPOTSOCIETES + TAXESASSIMAUXXID + DROITSDETIMBRE + TAXESFORESTIERES +
RECETTESNONBUDGETAIRES)
```

```
Residuals:
```

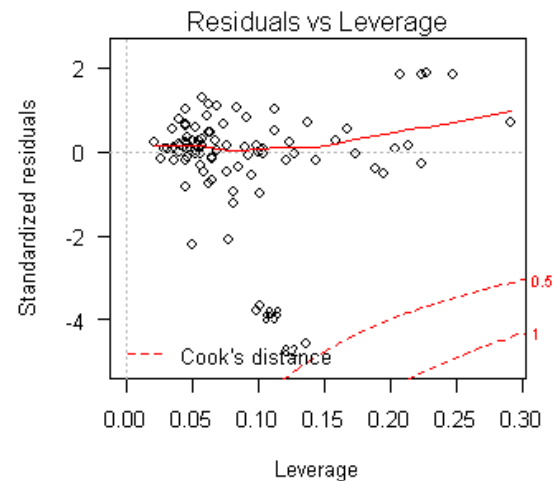
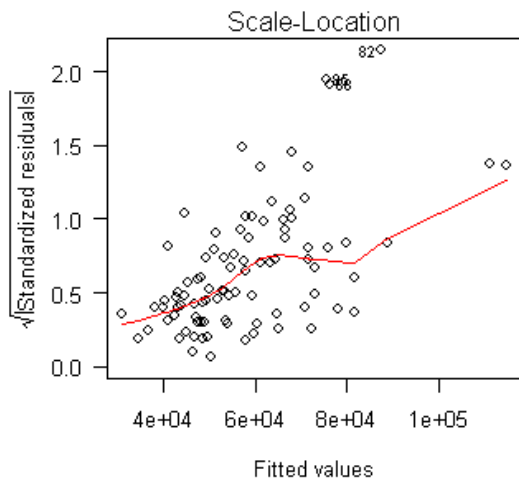
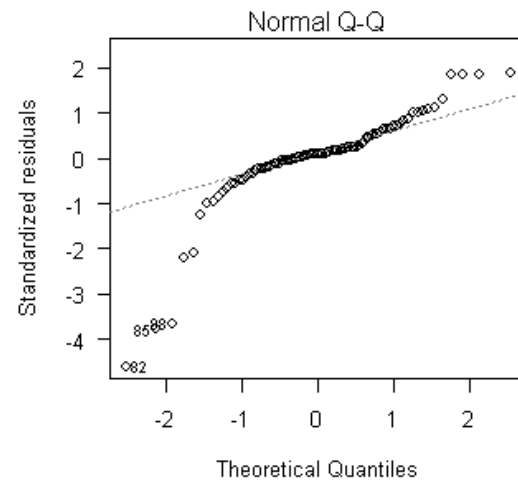
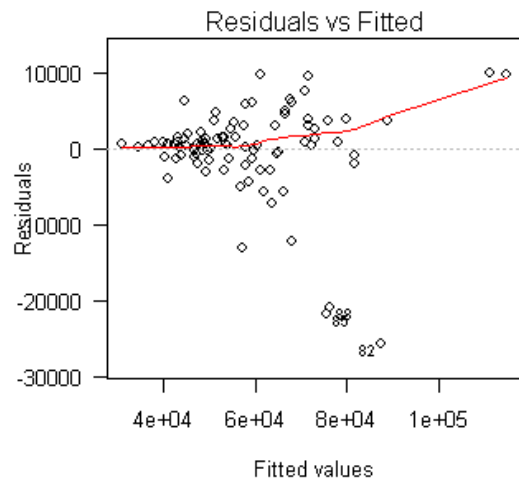
```
Min      1Q   Median      3Q      Max
-25845.1 -1154.5  522.8   2639.2 10085.2
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.007e+02	4.821e+03	0.104	0.91753
IMPSURREVPERSPHYS	8.686e-01	4.789e-01	1.814	0.07339 .
IMPOTSALACONSOMMATION	9.774e-01	1.524e-01	6.412	8.61e-09 ***
IMPOTSOCIETES	8.681e-01	4.665e-02	18.608	< 2e-16 ***
TAXESASSIMAUXXID	9.562e-01	3.680e-01	2.598	0.01110 *
DROITSDETIMBRE	3.257e+00	1.012e+00	3.218	0.00185 **
TAXESFORESTIERES	9.261e-01	5.243e-01	1.766	0.08105 .
RECETTESNONBUDGETAIRES	1.201e+00	3.968e-01	3.027	0.00330 **

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 6048 on 82 degrees of freedom
Multiple R-Squared:  0.8714,    Adjusted R-squared:  0.8604
F-statistic: 79.39 on 7 and 82 DF,  p-value: < 2.2e-16
```



On constate bien que tous les coefficients sont statistiquement différents (5%) de 0 et avec un $\bar{R}^2 = 0.8604$ proche de 1. A priori, ce modèle semble être bon. Par le graphique suivant, les résidus semblent à peu près suivre une loi normale d'où nous pouvons retenir ce modèle.

Ayant obtenu notre modèle de prédiction du montant des recettes fiscales, il ne nous reste plus qu'à le déterminer pour une période ultérieure afin que nous puissions déterminer le multiplicateur fiscal qui se calcul à l'aide de la formule :

$$\frac{\Delta PIB}{\Delta T} \text{ avec}$$

PIB qui représente le montant du revenu national en terme nominal

T le montant des impôts des impôts.