RÉPUBLIQUE DU SÉNÉGAL

Un peuple- un But- une Foi



Agence nationale de la Statistique et de la Démographie



Ećole nationale de la Statistique et de l'Analyse Ećonomique Pierre Ndiaye



Projet d'analyse de variance multidimensionnelle

TRAVAUX PRATIQUE: ECONOMETRIE DU MODELE LINEAIRE

Rédigé par :

Crépin MEDEHOUIN

ÉLÈVE INGÉNIEUR STATISTICIEN ÉCONOMISTES

Sous la supervision de :

M. Ahmadou Dicko

INGÉNIEUR STATISTICIEN ÉCONOMISTE

©Février-2024

SOMMAIRE

A۱	VANT-PROPOS		
RESUME			
I	PRO	PROBLEME 1	
	I.1	Nuage de point (log(totexp), pclothes) et l'élasticité ponctuelle des dépenses de vêtements par rapport aux dépenses totales aux moyennes	1
	I.2	Nuage de point [log(totexp), log(clothes)]. Calculez un intervalle de confiance à 95% de l'élasticité des dépenses de vêtements par rapport aux dépenses totales. L'élasticité calculée en partie (a) est-elle dans cet intervalle ?	2
	I.3	Nuage de point [totexp, clothes]	3
	I.4	Testez la présence de l'hétéroscédasticité dans chaque modèle dans les parties (a)–(b). Utilisez le niveau de signification de 1% pour vos tests. Quelles sont vos conclusions ? Pour quelle spécification	
		l'hétéroscédasticité semble-t-elle être un problème ?	4
	I.5	Répétez l'analyse dans laquelle l'hétéroscédasticité était significative au niveau de 1%	7
П	PRO	DBLEME 2	10
	II.1	Visualiser la tendance de consommation d'essence et celle du prix de détail de l'essence. Commentez.	10
	II.2	Pour la période 1950-1972, estimez les modèles (1) et (2) :	11
	II.3	Quelles restrictions les β doivent-ils satisfaire dans le modèle (1) pour aboutir aux γ dans le modèle (2) ?	12
	II.4	Comparez les estimations et les erreurs standard correspondantes des modèles (1) et (2)	13
	II.5	Calculez les corrélations simples parmi les regresseurs du modèle (1). Que observez-vous ?	14
	II.6	Estimez les équations (1) et (2) maintenant en utilisant l'ensemble de données complet de 1950 à 1987. Discutez brièvement des effets sur les estimations des paramètres individuels et leurs erreurs standard	
		sur l'ensemble de données plus large.	15
	II.7	A l'aide du test F de Chow, testez l'hypothèse selon laquelle la demande d'essence par voiture a diminué	
		de facon permanente pour le modèle (2) suite à l'embargo pétrolier arabe de 1973	16

LISTE DE FIGURES

1	Nuage de point (log(totexp), log(clothes))
2	Nuage de point (totexp, clothes)
3	Spread-Level Plot for modèle (a)
4	Spread-Level Plot for modèle (b)
5	Spread-Level Plot for modèle (c)
6	Spread-Level Plot for modèle (c) corrigé
7	Visualisation: Consommation - Prix
8	Matrice des coefficients de corrélation

ECONOMETRIE DU MODELE LINEAIRE ISE2

AVANT-PROPOS

L'objectif de ce document est de concrétiser les concepts enseignés lors du cours d'économétrie du modèle linéaire en mettant en avant les principes de la régression linéaire simple et multiple. Il vise à tester des hypothèses et à élaborer des modèles pour représenter les relations économiques.

Le but est de modéliser, quantifier et analyser les relations entre les variables économiques au sein d'une structure linéaire, afin d'améliorer la compréhension et l'interprétation des phénomènes économiques.

RESUME

Les différents exercices constitutifs de la présente étude ont permis de mettre en pratique toutes les notions apprises au cours de l'économetrie du modèle linéaire.

En outre, le problème 1 portait la regression linéaire simple à savoir la relation entre les dépenses totales des ménages et les dépenses en vêtements. En utilisant les données du fichier clothes.csv, notre analyse s'est basé trois modèles :

PCLOTHES =
$$\beta_1 + \beta_2 ln(TOTEXP) + \varepsilon$$
 (a), $ln(CLOTHES) = \alpha_1 + \alpha_2 ln(TOTEXP) + v$ (b) et $CLOTHES = \gamma_1 + \gamma_2 TOTEXP + \mu$ (c)

Le problème 2 qui portait sur la regression linéaire multiple en considérant les données sur l'essence des États-Unis de 1950 à 1987 obtenues à partir du fichier usgas.csv. Cela a partir de tester l'hypothèse selon laquelle la demande d'essence par voiture a diminué de façon permanente pour le modèle (2) suite à l'embargo pétrolier arabe de 1973.

$$\text{mod\`ele } (2): log\big(\frac{GAS}{CAR}\big) = \gamma_1 + \gamma_2 log\big(\frac{GNP}{POP}\big) + \gamma_3 log\big(\frac{CAR}{POP}\big) + \gamma_4 log\big(\frac{PRICE}{DEFLATOR}\big) + \upsilon$$

I PROBLEME 1

Dans cet exercice, nous explorons la relation entre les dépenses totales des ménages (TOTEXP) et les dépenses en vêtements (PCLOTHES). Utilisez le fichier de données clothes.csv. Nous considérons trois modèles :

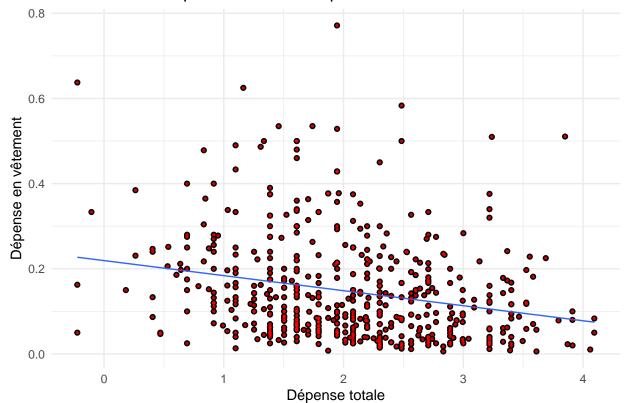
$$PCLOTHES = \beta_1 + \beta_2 ln(TOTEXP) + \varepsilon$$
 (a)

$$ln(CLOTHES) = \alpha_1 + \alpha_2 ln(TOTEXP) + v$$
 (b)

$$CLOTHES = \gamma_1 + \gamma_2 TOTEXP + \mu \tag{c}$$

I.1 Nuage de point (log(totexp), pclothes) et l'élasticité ponctuelle des dépenses de vêtements par rapport aux dépenses totales aux moyennes





Calculez l'élasticité ponctuelle des dépenses de vêtements par rapport aux dépenses totales aux moyennes

	Х
(Intercept)	0.2192588
log(totexp)	-0.0353208

L'élasticité ponctuelle des dépenses de vêtements par rapport aux dépenses totales aux moyennes est : elasticite = -0.2391922

Ainsi lorsque les dépenses totales des ménages augmentent de 1%, les dépenses en vêtement dimunient de 23,9%.

I.2 Nuage de point [log(totexp), log(clothes)]. Calculez un intervalle de confiance à 95% de l'élasticité des dépenses de vêtements par rapport aux dépenses totales. L'élasticité calculée en partie (a) est-elle dans cet intervalle ?



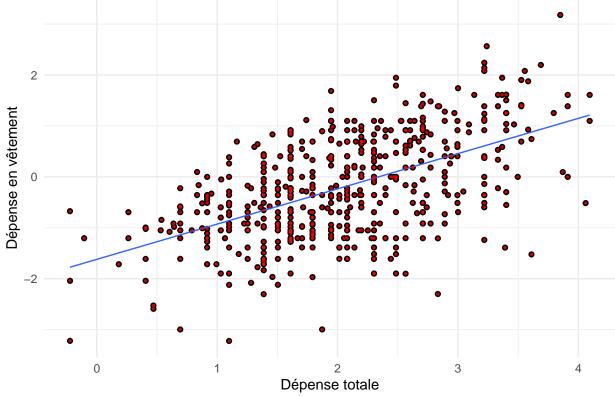


Figure 1: Nuage de point (log(totexp), log(clothes))

Calculez un intervalle de confiance à 95% de l'élasticité des dépenses de vêtements par rapport aux dépenses totales. L'élasticité calculée en partie (a) est-elle dans cet intervalle

	2.5 %	97.5 %	
(Intercept)	-1.810164	-1.4271079	
log(totexp)	0.603731	0.7798889	

Un intervalle de confiance à 95% de l'élasticité des dépenses de vêtements par rapport aux dépenses totales est :

 $IC(\alpha_2) = [0.603731, 0.7798889]$

On a bien -0.2391922 $\notin IC(\alpha_2)$, l'élasticité calculée à la 1^{me} est n'appartient donc pas à cet intervalle

I.3 Nuage de point [totexp, clothes]



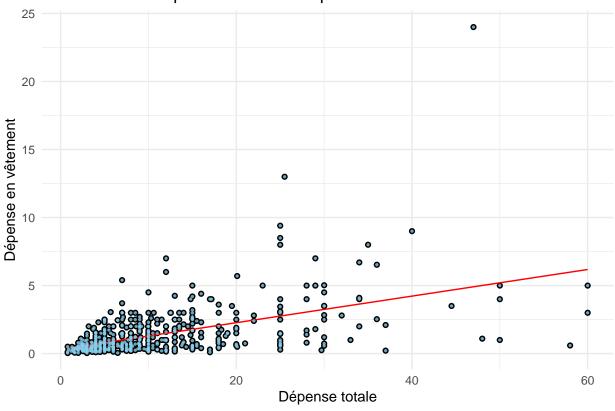


Figure 2: Nuage de point (totexp, clothes)

Calculez un intervalle de confiance à 95% de l'élasticité des dépenses de vêtements par rapport aux dépenses totales aux moyennes. L'élasticité calculée en partie (a) est-elle dans cet intervalle ?

	2.5 %	97.5 %
(Intercept)	0.1286697	0.5030276
totexp	0.0843486	0.1111655

Un intervalle de confiance à 95% de l'élasticité des dépenses de vêtements par rapport aux dépenses totales est :

```
IC(\gamma_2) = [0.6582094, 0.8674735]
```

On a bien -0.2391922 $\notin IC(\gamma_2)$, l'élasticité calculée à la 1^{me} est n'appartient donc pas à cet intervalle

I.4 Testez la présence de l'hétéroscédasticité dans chaque modèle dans les parties (a)–(b).
 Utilisez le niveau de signification de 1% pour vos tests. Quelles sont vos conclusions ?
 Pour quelle spécification l'hétéroscédasticité semble-t-elle être un problème ?

I.4.1 Test la présence de l'hétéroscédasticité dans le modèle (a)

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 8.101061, Df = 1, p = 0.0044239
```

On p-value = 0.0044239 > 0.001, les données ne fournissent pas assez d'évidence pour rejeter l'hypotèse nulle H_0 selon laquelle variance des residus est constante : **l'homoroscédasticité**

Le test de Breusch-Pagan

```
##
## studentized Breusch-Pagan test
##
## data: modele_a
## BP = 3.2438, df = 1, p-value = 0.07169
```

La p-value 0.07169 > 0.001, cela confirme donc l'homoroscédasticité

Vérification graphique

##

```
spreadLevelPlot(modele_a, smooth = FALSE)
```

```
## Suggested power transformation: 0.3333719
```

Spread-Level Plot for modele_a

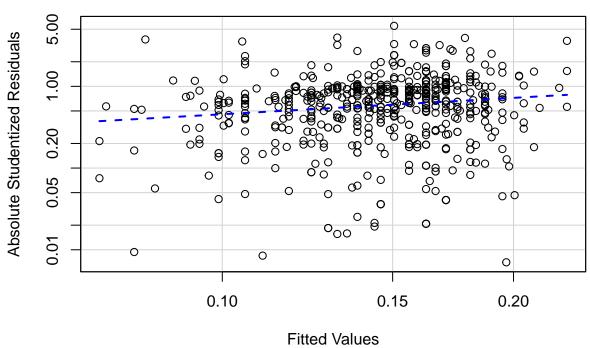


Figure 3: Spread-Level Plot for modèle (a)

L'homoroscédasticité soutend que la droite en pointier doit être horizontal, notre droite en pointier n'est pas vraiment horizontal mais nous pouvons supposer l'homoroscédasticité.

I.4.2 Test la présence de l'hétéroscédasticité dans le modèle (b)

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 11.29931, Df = 1, p = 0.00077536
```

Le test étant significatif, p-value 0.00077536 < 0.001, il y a donc hétéroscédasticité.

Le test de Breusch-Pagan

```
##
## studentized Breusch-Pagan test
##
## data: modele_b
## BP = 12.823, df = 1, p-value = 0.0003424
```

Vérification graphique

```
spreadLevelPlot(modele_b, smooth = FALSE)
```

```
## Warning in spreadLevelPlot.lm(modele_b, smooth = FALSE):
```

379 negative fitted values removed

Spread-Level Plot for modele_b

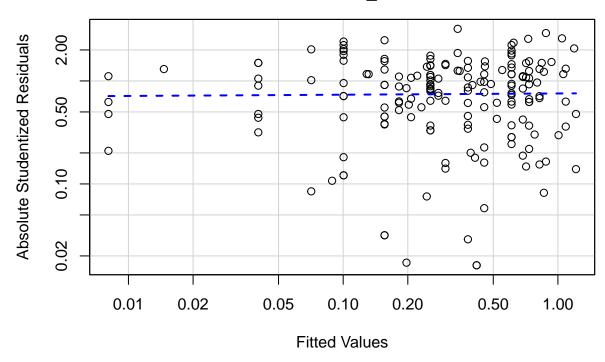


Figure 4: Spread-Level Plot for modèle (b)

##
Suggested power transformation: 0.9879058

Notre droite en pointier semble être horizontal sur cette representation graphique, on peut supposer qu'il n'y pas d'hétéroscédasticité.

I.4.3 Test la présence de l'hétéroscédasticité dans le modèle (c)

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 1204.192, Df = 1, p = < 2.22e-16</pre>
```

Le test étant significatif, il y a donc hétéroscédasticité.

Vérification graphique

spreadLevelPlot(modele_c, smooth = FALSE)

Spread-Level Plot for modele_c

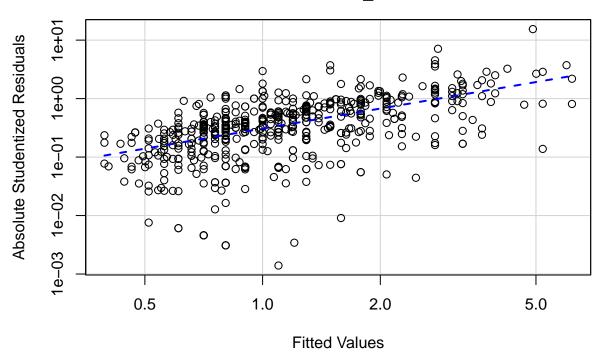


Figure 5: Spread-Level Plot for modèle (c)

```
##
## Suggested power transformation: -0.1388913
```

Notre droite en pointier est loin d'être horizontal et il y a hétéroscédasticité.

Le test de Breusch-Pagan

```
##
## studentized Breusch-Pagan test
##
## data: modele_c
## BP = 44.696, df = 1, p-value = 2.301e-11
```

Le test étant significatif, il y a confirmation de hétéroscédasticité.

I.5 Répétez l'analyse dans laquelle l'hétéroscédasticité était significative au niveau de 1%

Modèle (c)

SpreadLevelPlot nous fournis le "Suggested power transformation" : -0.1388913 pour le modèle (c) qui nous permet de corriger l'hétéroscédasticité.

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 3.190638, Df = 1, p = 0.074061
```

On p-value = 0.074061 > 0.001, il y a donc desormais **l'homoroscédasticité**

Le test de Breusch-Pagan

```
##
## studentized Breusch-Pagan test
##
## data: modele_c_cor
## BP = 2.6549, df = 1, p-value = 0.1032
```

Vérification graphique

```
spreadLevelPlot(modele_c_cor, smooth = FALSE)
```

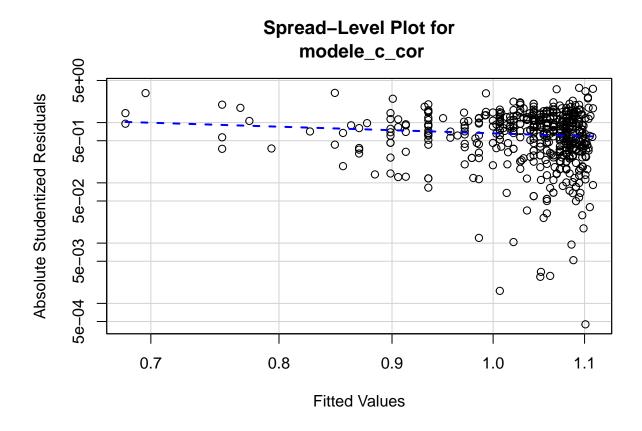


Figure 6: Spread-Level Plot for modèle (c) corrigé

```
##
## Suggested power transformation: 2.135799
```

Notre droite en pointier semble être maintenant horizontal.

	2.5 %	97.5 %
(Intercept)	1.0996705	1.1312509
totexp	-0.0083606	-0.0060983

Un intervalle de confiance à 95% de l'élasticité des dépenses de vêtements par rapport aux dépenses totales est :

$$IC_c(\gamma_2) = [\textbf{-0.0835201} \ , \textbf{-0.0609209}]$$

On a bien -0.2391922 $\notin IC_c(\gamma_2)$, l'élasticité calculée à la 1^{me} est n'appartient toujours pas à cet intervalle

II PROBLEME 2

Considérez les données sur l'essence des États-Unis de 1950 à 1987 obtenues à partir du fichier usgas.csv.

- car: Stock de voitures;
- gas: Consommation d'essence automobile (en milliers de gallons);
- price: Prix de détail de l'essence;
- pop: Population;
- pnb: Produit national brut réel (en dollars de 1982);
- deflat: Déflateur du PNB (1982 = 100);

II.1 Visualiser la tendance de consommation d'essence et celle du prix de détail de l'essence. Commentez

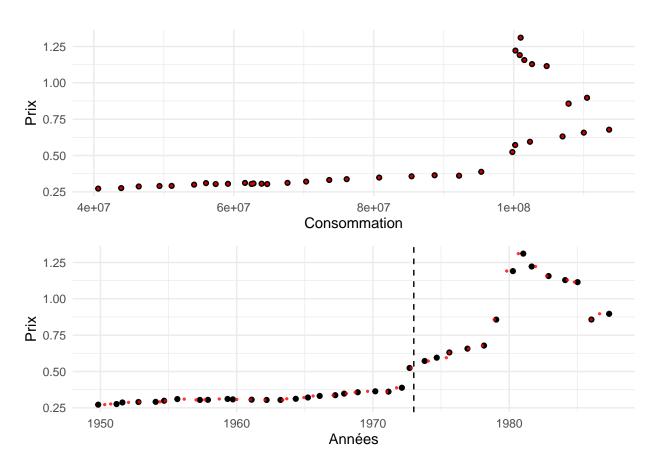


Figure 7: Visualisation: Consommation - Prix

En règle générale, le prix de l'essence est lié de manière croissante à la consommation d'essence automobiles, ce qui signifie que plus la consommation d'essence augmente, plus le prix augmente également. Néanmoins, cette

(1)

(2)

augmentation est initialement assez modérée, fluctuant entre 0.25 et 0.375, pour ensuite s'accélérer considérablement en nous présentant un pic de 1.3.

 $log(GAS) = \beta_1 + \beta_2 log(CAR) + \beta_3 log(POP) + \beta_4 log(GNP) + \beta_5 log(DEFLATOR) + \beta_6 log(PRICE) + \varepsilon$

Il est important de noter que cette transition abrupte dans les prix s'est produite à partir des années 1973.

Pour la période 1950-1972, estimez les modèles (1) et (2) :

```
## (Intercept)
                            2.7936 0.601 0.55549
                 1.6801
## log(cars)
                 0.3635
                            0.5152 0.706 0.48994
                         0.9048 1.165 0.26019
## log(pop)
                 1.0539
## log(gnp)
                 -0.3114
                            0.1625 -1.916 0.07231 .
## log(deflator) 0.1250
                            0.1580 0.791 0.44000
## log(price)
                  1.0481
                            0.2682 3.907 0.00113 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01872 on 17 degrees of freedom
## Multiple R-squared: 0.9952, Adjusted R-squared: 0.9937
```

F-statistic: 699.8 on 5 and 17 DF, p-value: < 2.2e-16

Call:

##

##

##

##

##

##

lm(formula = log(gas/cars) ~ log(gnp/pop) + log(cars/pop) + log(price/deflator),

```
data = subset(usgas, year <= 1972))
##
##
## Residuals:
##
        Min
                  1Q
                       Median
                                    3Q
                                            Max
  -0.05079 -0.03639 0.01337 0.02515 0.04526
##
## Coefficients:
##
                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)
                       -0.30653
                                   2.37844 -0.129
                                                      0.899
## log(gnp/pop)
                       -0.13972
                                   0.23851 -0.586
                                                      0.565
## log(cars/pop)
                                   0.28276
                        0.05446
                                             0.193
                                                      0.849
## log(price/deflator) 0.18527
                                   0.27882
                                             0.664
                                                      0.514
##
## Residual standard error: 0.03478 on 19 degrees of freedom
## Multiple R-squared: 0.389, Adjusted R-squared: 0.2925
## F-statistic: 4.032 on 3 and 19 DF, p-value: 0.02238
```

Le modèle 1 affiche un ajustement solide avec un R carré multiple élevé de 0,9952 et un R carré ajusté de 0,9937, ainsi qu'une analyse de variance indiquant des coefficients significativement différents de zéro. En revanche, le modèle 2 présente un ajustement moindre, avec des valeurs de R carré multiple (0,389) et R carré ajusté (0,2925) plus faibles.

II.3 Quelles restrictions les β doivent-ils satisfaire dans le modèle (1) pour aboutir aux γ dans le modèle (2) ?

$$log(\frac{GAS}{CAR}) = \gamma_1 + \gamma_2 log(\frac{GNP}{POP}) + \gamma_3 log(\frac{CAR}{POP}) + \gamma_4 log(\frac{PRICE}{DEFLATOR}) + \upsilon$$

$$log(GAS) - log(CAR) = \gamma_1 + \gamma_2 [log(GNP) - log(POP)] + \gamma_3 [log(CAR) - log(POP)] + \gamma_4 [log(PRICE) - log(DEFLATOR)] + \upsilon$$

$$log(GAS) = \gamma_1 + (1 + \gamma_3) log(CAR) - (\gamma_2 + \gamma_3) log(POP) + \gamma_2 log(GNP) - \gamma_4 log(DEFLATOR) + \gamma_4 log(PRICE) + \upsilon$$

Par identification,

On obtient:
$$\begin{cases} \beta_{1} = \gamma_{1} & (1) \\ \beta_{2} = 1 + \gamma_{3} & (2) \\ \beta_{3} = -(\gamma_{2} + \gamma_{3}) & (3) \\ \beta_{4} = \gamma_{2} & (4) \\ \beta_{5} = -\gamma_{4} & (5) \\ \beta_{6} = \gamma_{4} & (6) \end{cases}$$

$$(2) \Rightarrow \gamma_3 = \beta_2 - 1$$

Et (3)
$$\Rightarrow \gamma_3 = -\beta_3 - \gamma_2$$

Alors

$$\gamma_3 = \gamma_3$$

$$\Rightarrow \beta_2 - 1 = -\beta_3 - \gamma_2 \qquad or \qquad \gamma_2 = \beta_4 \qquad (4)$$

$$\Rightarrow \beta_2 - 1 = -\beta_3 - \beta_4$$

$$\Rightarrow \beta_2 = 1 - \beta_3 - \beta_4$$

Par ailleurs (5) et (6) $\Rightarrow \beta_5 = -\beta_6$

Finalement:
$$\begin{cases} \beta_2 = 1 - \beta_3 - \beta_4 \\ \beta_5 = -\beta_6 \end{cases}$$

II.4 Comparez les estimations et les erreurs standard correspondantes des modèles (1) et (2)

```
## Linear hypothesis test
##
## Hypothesis:
## log_cars + log_pop + log_gnp = 1
## log_deflator + log_price = 0
##
## Model 1: restricted model
## Model 2: log_gas ~ log_cars + log_pop + log_gnp + log_deflator + log_price
##
    Res.Df
                 RSS Df Sum of Sq F
##
                                            Pr(>F)
        19 0.0229830
        17 0.0059582 2 0.017025 24.288 1.039e-05 ***
## ---
```

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

Ce test étant significatif, on va rejetter l'hypotèse selon laquelle $\beta_2 = 1 - \beta_3 - \beta_4$ et $\beta_5 = -\beta_6$

Alors on peut conclure les deux modèles diffèrent.

II.5 Calculez les corrélations simples parmi les regresseurs du modèle (1). Que observez-vous ?

```
##
                 log_cars
                            log_pop
                                      log_gnp log_deflator log_price
## log_cars
                1.0000000 0.9958819 0.9917712
                                                 0.9768628 0.9437360
                0.9958819 1.0000000 0.9809202
## log_pop
                                                 0.9628089 0.9156399
                0.9917712 0.9809202 1.0000000
                                                 0.9729462 0.9498337
## log_gnp
## log_deflator 0.9768628 0.9628089 0.9729462
                                                 1.0000000 0.9702470
                0.9437360 0.9156399 0.9498337
                                                 0.9702470 1.0000000
## log_price
```

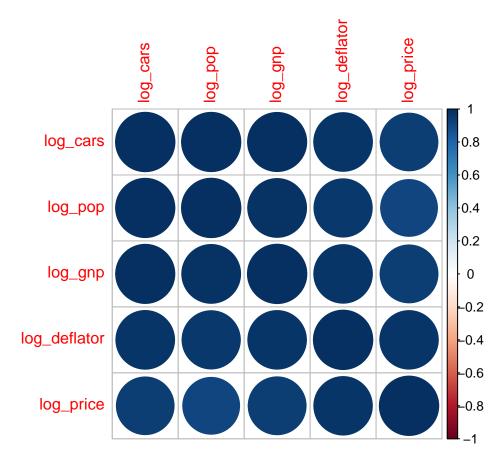


Figure 8: Matrice des coefficients de corrélation

Il existe des corrélations positives forte parmi les regresseurs du modèle (1)

II.6 Estimez les équations (1) et (2) maintenant en utilisant l'ensemble de données complet de 1950 à 1987. Discutez brièvement des effets sur les estimations des paramètres individuels et leurs erreurs standard sur l'ensemble de données plus large.

```
##
## Call:
## lm(formula = log(gas) ~ log(cars) + log(pop) + log(gnp) + log(deflator) +
      log(price), data = usgas)
##
##
## Residuals:
##
        Min
                  1Q
                       Median
                                    3Q
                                             Max
## -0.059240 -0.020056 -0.001106 0.027460 0.041927
##
## Coefficients:
##
               Estimate Std. Error t value Pr(>|t|)
                9.98698 2.62062 3.811 0.000594 ***
## (Intercept)
## log(cars)
                ## log(pop)
               ## log(gnp)
               -0.42927
                          0.14838 -2.893 0.006811 **
## log(deflator) -0.17887
                          0.06336 -2.823 0.008115 **
## log(price)
               -0.14111
                          0.04340 -3.252 0.002704 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 0.02825 on 32 degrees of freedom
## Multiple R-squared: 0.9927, Adjusted R-squared: 0.9915
## F-statistic: 868.8 on 5 and 32 DF, p-value: < 2.2e-16
##
## Call:
## lm(formula = log(gas/cars) ~ log(gnp/pop) + log(cars/pop) + log(price/deflator),
      data = usgas)
##
## Residuals:
##
        Min
                  1Q
                       Median
                                    3Q
                                             Max
## -0.109139 -0.045609 0.000628 0.051406 0.110555
##
## Coefficients:
```

```
##
                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)
                      -5.85398
                                  3.10248 -1.887
                                                    0.0677 .
## log(gnp/pop)
                      -0.69046
                                  0.29337 - 2.354
                                                    0.0245 *
## log(cars/pop)
                                  0.27723
                       0.28874
                                            1.041
                                                    0.3050
## log(price/deflator) -0.14313
                                  0.07488 -1.911
                                                    0.0644 .
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0605 on 34 degrees of freedom
## Multiple R-squared: 0.7415, Adjusted R-squared: 0.7187
## F-statistic: 32.51 on 3 and 34 DF, p-value: 4.249e-10
```

Le modèle 1 révèle que les coefficients estimés des variables explicatives log(gnp), log(deflator), et log(price) sont tous significativement différents de zéro. Le Rcarré multiple élevé de 0,9927 suggère que ce modèle explique efficacement la variance des données. De plus, la F-statistique est élevée avec une p-valeur très basse, confirmant ainsi la significativité globale du modèle.

En ce qui concerne le modèle 2, les coefficients des variables explicatives ne présentent pas de différences significatives de zéro, comme en témoignent les p-valeurs supérieures à 0,05 pour toutes les variables. Les valeurs de Roarré et de Roarré ajusté sont plus faibles dans ce modèle (0,389 et 0,2925 respectivement), indiquant une explication moindre de la variance des données par rapport au premier modèle. L'analyse de la variance révèle une F-statistique inférieure avec une p-valeur de 0,02238, suggérant que l'ensemble des coefficients n'est pas statistiquement significatif dans ce modèle.

II.7 A l'aide du test F de Chow, testez l'hypothèse selon laquelle la demande d'essence par voiture a diminué de façon permanente pour le modèle (2) suite à l'embargo pétrolier arabe de 1973

```
##
## Chow test
##
## data: fm
## F = 12.828, p-value = 3.344e-06
```

La valeur p associée est quasiment nulle, suggérant ainsi une différence significative entre les coefficients estimés avant et après le point de rupture. Cela indique une variation notable dans la relation étudiée entre les deux périodes.