Exercices d'économétrie

TD 4. Régression multiple : inférence.

Janvier, 2020

Exercice 4.1

Le modèle suivant peut être utilisé pour étudier si les dépenses de campagne influent sur les résultats des élections :

$$\texttt{voteA}_i = \beta_0 + \beta_1 \log(\texttt{expendA}_i) + \beta_2 \log(\texttt{expendB}_i) + \beta_3 \texttt{prtystrA}_i + u_i$$

où `voteA` est le pourcentage de vote obtenu par le candidat A, `expendA` et `expendB` les dépenses de campagne des candidats A et B, et `prtystrA` une mesure de l'importance du partie pour le candidat A (le pourcentage de votes en faveur du parti du candidat A lors des élections les plus récentes).

- i Quelle est l'interprétation du paramètre β_1 ?
- ii En fonction des paramètres du modèle, spécifiez l'hypothèse nulle selon laquelle l'impact sur le vote d'une augmentation de 1% des dépenses de A est compensée par une augmentation de 1% des dépenses de B.
- iii Estimez le modèle en utilisant la base de données `vote1`. Les dépenses du candidat A influencent-elles le résultat ? Qu'en est-il des dépenses du candidat B ? Pouvez-vous utiliser ces résultats pour tester l'hypothèse de la question (ii) ?
- iv Estimez un modèle donnant directement la statistique de Student adéquate pour tester l'hypothèse décrite en question (ii). Que concluez-vous ?

Exercice 4.2

On utlise la base de données `hprice1` pour estimer le modèle :

$$\log(\texttt{price}_i) = \beta_0 + \beta_1 \texttt{sqrft}_i + \beta_2 \texttt{bdrms}_i + u_i$$

où `price` est le prix d'une maison (en milliers de dollars), `sqrft` mesure sa superficie (en pieds carrés), et `bdrms` représente le nombre de chambres.

- i On s'intéresse à l'estimation et l'obtention d'un intervalle de confiance pour le taux de variation en pourcentage de la variable `price` lorsqu'une pièce de 150 pieds carrés est ajoutée à la configuration existante d'une maison donnée. Sous forme décimale, ceci s'écrit $\theta = 150\beta_1 + \beta_2$. Estimez θ.
- ii Écrivez β_2 en fonction de β_1 et θ et introduisez ce paramètre dans l'équation expliquant log(`price`).
- iii Utilisez les éléments de la question (ii) afin d'obtenir un écart-type estimé pour $\hat{\theta}$, puis calculez à partir de cet écart-type un intervalle de confiance à 95%.

Exercice 4.3

On utilise la base de données `mlb1` pour expliquer les salaires des joueurs de baseball de la ligue professionnelle américaine à l'aide du modèle suivant :

$$\log(\texttt{salary}_i) = \beta_0 + \beta_1 \texttt{years}_i + \beta_2 \texttt{gamesyr}_i + \beta_3 \texttt{bavg}_i + \beta_4 \texttt{hrunsyr}_i + \beta_5 \texttt{rbisyr}_i + u_i$$

où `salary` est le salaire total en 1993, `years` le nombre d'années passées au sein de la ligue, `gamesur` le nombre moyen de rencontre disputées dans l'année, `bavg` la « moyenne à la batte » sur l'ensemble de la carrière du joueur, `hrunsyr` le nombre de « *home runs* » par an, et `rbisyr` le nombre de points produits par un frappeur au cours de l'année.

- i Estimez le modèle avec et sans la variable `rbisyr`. Qu'advient-il de la significativité statistique de `hrunsyr` ? Qu'en est-il de la taille du coefficient de `hrunsyr` ?

ii Ajoutez les variables `runsyr` (nombre de points par an), `fldperc` (pourcentage de mise en service), et `sbasesyr` (base « volées » par an) au second modèle de régression de la question (i). Lequel de ces facteurs est-il individuellement statistiquement significatif ?

iii Dans le modèle de la question (ii), testez la significativité jointe de `bavg`, `fldperc`, et `sbasesyr`.

Exercice 4.4

On utilise la base de données `discrim` qui contient des données de prix pour différents produits vendus dans des enseignes de restauration rapide.

i Estimez le modèle suivant par les MCO :

$$\log(\texttt{psoda}_i) = \beta_0 + \beta_1 \texttt{prpblck}_i + \beta_2 \log(\texttt{income}_i) + \beta_3 \texttt{prppov}_i + u_i$$

où `psoda` est le prix du soda, `prpblck` la proportion locale de personnes d'origine afro-américaine, `income` le revenu local moyen, et `prppov` la proportion locale de personnes vivant sous le niveau de pauvreté. Le paramètre estimé $\hat{\beta}_1$ est-il statistiquement différent de zéro au seuil de 5% contre l'hypothèse alternative bilatérale ? Qu'en est-il au seuil de 1% ?

ii Quelle est la corrélation entre log(`income`) et `prppov` ? Les variables apparaissent-elles chacune statistiquement significative ?

iii Reprenez le modèle de la question (i) et ajoutez-y la variable log(`hseval`) (la valeur médiane des logements dans la commune). Interprétez son coefficient et indiquez la p-valeur relative au test bilatéral pour $\mathrm{H}_0 : \beta_4 = 0$.

iv Dans la régression décrite en question (iii), que se passe-t-il pour la sigificativité statistique individuelle des variables log(`income`) et `prppov` ? Ces variables sont-elles conjointement significatives ? Qu'advient-il de vos précédentes réponses ?

v Compte tenu des résultats des régressions précédentes, laquelle seriez-vous prêt à reporter avec le plus de confiance afin d'analyser si la composition ethnique d'une localité influence les prix des fast-foods ?

Solution 4.1

(i) La variable `voteA` est comprise entre 0 et 100. $\beta_1/100$ s'interprète comme le point de pourcentage sur `voteA` induit par une augmentation de 1% des dépenses de campagne du candidat A (*ceteris paribus*). En effet, si nous reprenons le modèle en faisant varier `expendA` et laissant le reste inchangé :

$$\Delta\texttt{voteA} = \beta_1 \Delta \log(\texttt{expendA})$$

soit de façon équivalente :

$$\Delta\texttt{voteA} = \frac{\beta_1}{100} \times \underbrace{100\Delta \log(\texttt{expendA})}_{\substack{\text{Taux de croissance de } \texttt{expendA} \\ \text{en pourcentage}}}$$

(ii) Pour que l'augmentation de 1% de A soit compensée par une augmentation de 1% de B il faudrait poser une contrainte sur les paramètres β_1 et β_2. L'hypothèse nulle serait alors $\beta_1 + \beta_2 = 0$, ou de façon équivalente $\beta_1 = -\beta_2$. (iii) L'estimation par les MCO du modèle décrit plus haut donne :

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  voteA   R-squared:                       0.793
Model:                            OLS   Adj. R-squared:                  0.789
Method:                 Least Squares   F-statistic:                     215.2
Date:                Sun, 08 Mar 2020   Prob (F-statistic):           1.80e-57
Time:                        19:27:59   Log-Likelihood:                -596.88
No. Observations:                 173   AIC:                             1202.
Df Residuals:                     169   BIC:                             1214.
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept         45.0879      3.927     11.482      0.000      37.336      52.840
```

```
logged_expendA     6.0814      0.382     15.915      0.000       5.327       6.836
logged_expendB    -6.6156      0.379    -17.461      0.000      -7.364      -5.868
prtystrA           0.1520      0.062      2.451      0.015       0.030       0.274
==============================================================================
Omnibus:                        8.908   Durbin-Watson:                   1.604
Prob(Omnibus):                  0.012   Jarque-Bera (JB):                8.841
Skew:                           0.493   Prob(JB):                       0.0120
Kurtosis:                       3.505   Cond. No.                         344.
==============================================================================
```

Les dépenses des deux candidats infuencent les résultats du candidat A ; on trouve bien les signes attendus $\hat{\beta}_1 > 0$ et $\hat{\beta}_2 < 0$ et les deux paramètres sont significativement différents de zéro puisque les statistiques de Student sont largement plus grandes que 1,96. D'après cette estimation une augmentation de 10% des dépenses de campagne du candidat A induirait une augmentation de $6,08/100 \times 10 = 0,608$ points de poucentage de votes pour le candidat A. Les estimations suggèrent que l'hypothèse nulle évoquée en (ii) n'est pas rejetée par les données, mais on ne peut pas le tester formellement avec cette estimation. (iv) Pour tester formellement l'hypothèse nous allons estimer un modèle cohérent avec cette hypothèse en reparamétrisant le modèle. On pose $\theta = \beta_1 + \beta_2$ et on introduit $\beta_2 \log(\texttt{expendA}_i) - \beta_2 \log(\texttt{expendA}_i)$ dans l'équation du modèle, ce qui nous permet de réécrire le modèle sous la forme :

$$\texttt{voteA}_i = \beta_0 + \theta \log(\texttt{expendA}_i) + \beta_2 \left(\log(\texttt{expendB}_i) - \log(\texttt{expendA}_i)\right) + \beta_3 \texttt{prtystrA}_i + u_i$$

Tester l'hypothèse nulle revient alors à tester la nullité de θ. L'estimation du nouveau modèle par les MCO donne :

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  voteA   R-squared:                       0.793
Model:                            OLS   Adj. R-squared:                  0.789
Method:                 Least Squares   F-statistic:                     215.2
Date:                Sun, 08 Mar 2020   Prob (F-statistic):           1.80e-57
Time:                        19:27:59   Log-Likelihood:                -596.88
No. Observations:                 173   AIC:                             1202.
Df Residuals:                     169   BIC:                             1214.
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept         45.0879      3.927     11.482      0.000      37.336      52.840
logged_expendA    -0.5343      0.533     -1.002      0.318      -1.587       0.518
expend_gap        -6.6156      0.379    -17.461      0.000      -7.364      -5.868
prtystrA           0.1520      0.062      2.451      0.015       0.030       0.274
==============================================================================
Omnibus:                        8.908   Durbin-Watson:                   1.604
Prob(Omnibus):                  0.012   Jarque-Bera (JB):                8.841
Skew:                           0.493   Prob(JB):                       0.0120
Kurtosis:                       3.505   Cond. No.                         343.
==============================================================================
```

Le student associé à $\hat{\theta}$ est inférieur à 1,96 en valeur absolue, on peut ne peut donc pas rejeter l'hypothèse nulle $\theta = 0$ contre $\theta \neq 0$ au seuil de 5%. Les données ne rejettent pas l'hypothèse d'une compensation exacte entre les dépenses des candidats A et B.

Solution 4.2

(i) L'estimation du modèle par les MCO donne :

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           logged_price   R-squared:                       0.588
Model:                            OLS   Adj. R-squared:                  0.579
Method:                 Least Squares   F-statistic:                     60.73
Date:                Mon, 09 Mar 2020   Prob (F-statistic):           4.17e-17
```

```
Time:                        18:23:09   Log-Likelihood:                 19.592
No. Observations:                  88   AIC:                            -33.18
Df Residuals:                      85   BIC:                            -25.75
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      4.7660      0.097     49.112      0.000       4.573       4.959
sqrft          0.0004   4.32e-05      8.781      0.000       0.000       0.000
bdrms          0.0289      0.030      0.974      0.333      -0.030       0.088
==============================================================================
Omnibus:                       11.244   Durbin-Watson:                   1.807
Prob(Omnibus):                  0.004   Jarque-Bera (JB):               24.634
Skew:                          -0.332   Prob(JB):                     4.47e-06
Kurtosis:                       5.506   Cond. No.                     9.85e+03
==============================================================================
```

L'estimateur de θ est $\hat{\theta} = 150\hat{\beta}_1 + \hat{\beta}_2 \approx 0,0858$. Selon cette estimation, l'ajout d'une pièce supplémentaire augmenterait le prix de 8,6%. (ii) On a directement $\beta_2 = \theta - 150\beta_1$. En substituant dans le modèle, il vient :

$$\log(\texttt{price}_i) = \beta_0 + \beta_1\left(\texttt{sqrft}_i - 150\texttt{bdrms}_i\right) + \theta\texttt{bdrms}_i + u_i$$

(iii) L'estimation du modèle reparamétrisé donne :

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           logged_price   R-squared:                       0.588
Model:                            OLS   Adj. R-squared:                  0.579
Method:                 Least Squares   F-statistic:                     60.73
Date:                Mon, 09 Mar 2020   Prob (F-statistic):           4.17e-17
Time:                        18:23:09   Log-Likelihood:                 19.592
No. Observations:                  88   AIC:                            -33.18
Df Residuals:                      85   BIC:                            -25.75
Df Model:                           2
Covariance Type:            nonrobust
============================================================================================
                               coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
Intercept                    4.7660      0.097     49.112      0.000       4.573       4.959
sqrft_minus_150_bdrms        0.0004   4.32e-05      8.781      0.000       0.000       0.000
bdrms                        0.0858      0.027      3.205      0.002       0.033       0.139
==============================================================================
Omnibus:                       11.244   Durbin-Watson:                   1.807
Prob(Omnibus):                  0.004   Jarque-Bera (JB):               24.634
Skew:                          -0.332   Prob(JB):                     4.47e-06
Kurtosis:                       5.506   Cond. No.                     7.40e+03
==============================================================================
```

L'estimateur de θ (le coefficient attaché à `bdrms`) est bien identique à la valeur déduite dans la réponse à la première question. L'écart-type de l'estimateur est 0,027. L'intervalle à 95% est

$$0,0858 \pm 1,96 \times 0,027$$

c'est-à-dire $[0,0329 \quad , \quad 0,1387]$ (on peut aussi directement lire l'intervalle de confiance dans le tableau généré par la routine d'estimation). La hausse de prix est de 8,6% en moyenne et l'intervalle de confiance à 95% de la hausse de prix est $[3,3\% \quad , \quad 13,9\%]$.

Nous avons obtenu l'écart-type de $\hat{\theta}$ en reparamétrisant le modèle. Mais nous aurions pu calculer plus directement cette statistique. En effet, nous savons que :

$$\begin{aligned}
\mathbb{V}\left[\hat{\theta}\right] &= \mathbb{V}\left[150\hat{\beta}_1 + \hat{\beta}_2\right] \\
&= 150^2 \times \mathbb{V}\left[\hat{\beta}_1\right] + \mathbb{V}\left[\hat{\beta}_2\right] + 2 \times 150 \times \mathbb{C}ov\left(\hat{\beta}_1, \hat{\beta}_2\right)
\end{aligned}$$

En lisant le contenu de `results.cov_params()`, où `results` est l'output de la routine d'esimation par les MCO utilisé dans la première question, on obtient :

$$\mathbb{V}\left[\hat{\theta}\right] = 150^2 \times 1,867279 \times 10^{-9} + 8,787228 \times 10^{-4} - 300 \times -6,807887 \times 10^{-7} \approx 0,0007164999675$$

L'écart type de $\hat{\theta}$ est alors $\sqrt{0,0007164999675} \approx 0,026767517021569258$. Il n'est donc pas nécessaire de réestimer le modèle pour calculer l'écart type de $\hat{\theta}$.

Solution 4.3

(i) Les estimations des deux modèles (avec et sans la variable `rbisyr`) sont résumées par les sorties de la routine Python utilisée :

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          logged_salary   R-squared:                       0.628
Model:                            OLS   Adj. R-squared:                  0.622
Method:                 Least Squares   F-statistic:                     117.1
Date:                Tue, 10 Mar 2020   Prob (F-statistic):           2.94e-72
Time:                        06:30:23   Log-Likelihood:                -385.11
No. Observations:                 353   AIC:                             782.2
Df Residuals:                     347   BIC:                             805.4
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     11.1924      0.289     38.752      0.000      10.624      11.760
years          0.0689      0.012      5.684      0.000       0.045       0.093
gamesyr        0.0126      0.003      4.742      0.000       0.007       0.018
bavg           0.0010      0.001      0.887      0.376      -0.001       0.003
hrunsyr        0.0144      0.016      0.899      0.369      -0.017       0.046
rbisyr         0.0108      0.007      1.500      0.134      -0.003       0.025
==============================================================================
Omnibus:                        6.816   Durbin-Watson:                   1.265
Prob(Omnibus):                  0.033   Jarque-Bera (JB):               10.197
Skew:                          -0.068   Prob(JB):                      0.00610
Kurtosis:                       3.821   Cond. No.                     2.09e+03
==============================================================================
```

et

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          logged_salary   R-squared:                       0.625
Model:                            OLS   Adj. R-squared:                  0.621
Method:                 Least Squares   F-statistic:                     145.2
Date:                Tue, 10 Mar 2020   Prob (F-statistic):           6.98e-73
Time:                        06:30:23   Log-Likelihood:                -386.25
No. Observations:                 353   AIC:                             782.5
Df Residuals:                     348   BIC:                             801.8
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     11.0209      0.266     41.476      0.000      10.498      11.544
years          0.0677      0.012      5.592      0.000       0.044       0.092
gamesyr        0.0158      0.002     10.079      0.000       0.013       0.019
bavg           0.0014      0.001      1.331      0.184      -0.001       0.004
hrunsyr        0.0359      0.007      4.964      0.000       0.022       0.050
==============================================================================
```

```
Omnibus:                        7.369   Durbin-Watson:                   1.244
Prob(Omnibus):                  0.025   Jarque-Bera (JB):               11.240
Skew:                          -0.085   Prob(JB):                      0.00362
Kurtosis:                       3.858   Cond. No.                     1.90e+03
==============================================================================
```

On note que l'estimateur du paramètre associé à `hrunsyr` est nettement supérieur dans la seconde régression (par un facteur deux). On remarque aussi que l'exclusion de `rbisyr` rend cet estimateur significativement différent de zéro (la statistique de Student est bien au dessus de 1,96). (ii) L'estimation du modèle complet donne par les MCO :

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          logged_salary   R-squared:                       0.639
Model:                            OLS   Adj. R-squared:                  0.632
Method:                 Least Squares   F-statistic:                     87.25
Date:                Tue, 10 Mar 2020   Prob (F-statistic):           1.84e-72
Time:                        06:30:23   Log-Likelihood:                -379.71
No. Observations:                 353   AIC:                             775.4
Df Residuals:                     345   BIC:                             806.3
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     10.4083      2.003      5.196      0.000       6.468      14.348
years          0.0700      0.012      5.844      0.000       0.046       0.094
gamesyr        0.0079      0.003      2.950      0.003       0.003       0.013
bavg           0.0005      0.001      0.480      0.632      -0.002       0.003
hrunsyr        0.0232      0.009      2.687      0.008       0.006       0.040
runsyr         0.0174      0.005      3.434      0.001       0.007       0.027
fldperc        0.0010      0.002      0.516      0.606      -0.003       0.005
sbasesyr      -0.0064      0.005     -1.238      0.216      -0.017       0.004
==============================================================================
Omnibus:                        1.205   Durbin-Watson:                   1.287
Prob(Omnibus):                  0.548   Jarque-Bera (JB):                0.956
Skew:                           0.090   Prob(JB):                        0.620
Kurtosis:                       3.180   Cond. No.                     5.33e+04
==============================================================================
```

Seul le paramètre associé à `runsyr` est significativement différent de zéro au seuil de 5% (la statistique de Student est plus grande que 1,96 en valeur absolue). (iii) Les estimateurs des paramètres associés à `bavg`, `fldperc`, et `sbasesyr` sont individuellement non significatifs. Pour tester la nullité jointe de ces trois paramètres, on utilise un test de Fisher. Pour cela on estime un modèle (contraint) où les variables `bavg`, `fldperc`, et `sbasesyr` sont omises. La statistique de Student est construite à partir de la somme des carrés des résidus dans le modèle contraint (SSR_c)et le modèle non contraint (SSR_{nc}) :

$$F = \frac{SSR_c - SSR_{nc}}{SSR_{nc}} \frac{n-k-1}{p}$$

où $n-k-1$ est le nombre de degré de liberté dans le modèle non contraint (n le nombre d'observations dans l'échantillon, k le nombre de variables explicatives), et p le nombre de contraintes. Dans le cas qui nous intéresse ($n = 353$, $k = 7$ et $p = 3$) on trouve $F = 0,68$. Cette statistique est à comparer avec la valeur critique 2,6 (lue dans une table de Fisher à 345 et 3 degrés de liberté au seuil de 5%). Puisque $0,68 < 2,6$, on ne peut rejeter l'hypothèse nulle de nullité jointe des coefficients au seuil de 5%. Nous parviendrions à la même conclusion en comparant la p-valeur associée à la statistique F avec le seuil 5%. La p-value associée est 0,56 (l'aire sous la densité de la distribution de Fisher F(3, 345) entre 0,68 et ∞), soit 56%. Elle est supérieure à 5%, on ne peut donc rejeter l'hypothèse de nullité jointe des coefficients au seuil de 5%.

Solution 4.4

(i) L'estimation du modèle par les MCO donne :

```
                            OLS Regression Results
==============================================================================
```

```
Dep. Variable:           logged_psoda   R-squared:                       0.087
Model:                            OLS   Adj. R-squared:                  0.080
Method:                 Least Squares   F-statistic:                     12.60
Date:                Sun, 22 Mar 2020   Prob (F-statistic):           6.92e-08
Time:                        08:38:29   Log-Likelihood:                 439.04
No. Observations:                 401   AIC:                            -870.1
Df Residuals:                     397   BIC:                            -854.1
Df Model:                           3
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
Intercept        -1.4633      0.294     -4.982      0.000      -2.041      -0.886
prpblck           0.0728      0.031      2.373      0.018       0.013       0.133
logged_income     0.1370      0.027      5.119      0.000       0.084       0.190
prppov            0.3804      0.133      2.864      0.004       0.119       0.641
==============================================================================
Omnibus:                       12.002   Durbin-Watson:                   1.727
Prob(Omnibus):                  0.002   Jarque-Bera (JB):               24.056
Skew:                          -0.014   Prob(JB):                     5.97e-06
Kurtosis:                       4.200   Cond. No.                         834.
==============================================================================
```

la p-valeur de la statitique de Student associée à l'hypothèse nulle $\beta_1 = 0$ est égale à 0,018, soit 1,8%. On peut donc rejeter l'hypothèse nulle au seuil de 5% ($\hat{\beta}_1$ est significativement différent de zéro au seuil de 5%). En revanche, la valeur estimée n'est pas significativement différente de zéro au seuil de 1% (on rejette l'hypothèse nulle avec ce seuil). (ii) La corrélation entre log(`income`) et `prppov` est -0,84. Les deux variables sont donc quasi colinéaires (la corrélation est proche de 1). Néanmoins les deux paramètres estimés associés apparaissent significativement différents de zéro. (iii) L'estimation du modèle augmenté avec log(`hseval`) donne :

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           logged_psoda   R-squared:                       0.184
Model:                            OLS   Adj. R-squared:                  0.176
Method:                 Least Squares   F-statistic:                     22.31
Date:                Sun, 22 Mar 2020   Prob (F-statistic):           1.24e-16
Time:                        08:38:29   Log-Likelihood:                 461.55
No. Observations:                 401   AIC:                            -913.1
Df Residuals:                     396   BIC:                            -893.1
Df Model:                           4
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
Intercept        -0.8415      0.292     -2.878      0.004      -1.416      -0.267
prpblck           0.0976      0.029      3.334      0.001       0.040       0.155
logged_income    -0.0530      0.038     -1.412      0.159      -0.127       0.021
prppov            0.0521      0.134      0.388      0.699      -0.212       0.317
logged_hseval     0.1213      0.018      6.860      0.000       0.087       0.156
==============================================================================
Omnibus:                       16.452   Durbin-Watson:                   1.973
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               40.077
Skew:                          -0.016   Prob(JB):                     1.98e-09
Kurtosis:                       4.548   Cond. No.                     1.31e+03
==============================================================================
```

La p-value associée à l'estimateur du nouveau paramètre est nulle, cet estimateur est donc significativement différent de zéro quelque soit le seuil. Ce coefficient s'interprète comme une élasticité. Une augmentation de 1% de la valeur du logement induit une augmentation de 0,12% du prix du soda. (iv) Avec l'introduction de log(`hseval`), les estimateurs associés à log(`income`) et `prppov` ne sont plus statistiquement différents de zéro. Néanmoins, on rejette l'hypotèse nulle jointe $\beta_2 = \beta_3 = 0$ au seuil de 5% (la p-value de la statistique de Fisher est égale à 0,03). (v) On préfère le modèle augmenté car

1. La nouvelle variable est significative,
2. Même si les variables log(`income`) et `prppov` n'apparaissent plus individuellement significatives, elles sont conjointement significatives.
3. Le R^2 est beaucoup plus important.