

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC KINH TẾ - TÀI CHÍNH
THÀNH PHỐ HỒ CHÍ MINH



BÁO CÁO NHÓM (10%)
“Các yếu tố ảnh hưởng đến dự đoán giá xe Ford ”
Môn: Kinh tế lượng
Mã lớp: ECO1104 – Nhóm B04
Giảng viên: ThS. Nguyễn Phú Quới

Họ và Tên	MSSV	Nhiệm vụ	Mức độ đóng góp
1. Trần Thị Minh Nhung	235022881	MH1+Viết báo cáo, chỉnh sửa báo cáo.	100%
2. Huỳnh Hạo Dân	235022424	MH2 + Viết báo cáo, chỉnh sửa báo cáo	100%
3. Lưu Đức Huy	235623356	MH3 + Viết báo cáo, chỉnh sửa báo cáo	100%
4. Đỗ Bảo Thiên Ân	245024341	Đánh giá lại mô hình hồi quy cuối cùng.	100%
5. Nguyễn Thị Thuỷ Trâm	235087446	Đánh giá lại mô hình hồi quy cuối cùng.	100%

Thành phố Hồ Chí Minh, tháng 01 năm 2026

MỤC LỤC

1. Thống kê mô tả biến định tính.....	5
2. Thống kê mô tả biến định lượng.....	5
I. MÔ HÌNH 1: MÔ HÌNH HỒI QUY TUYẾN TÍNH – TUYẾN TÍNH.....	7
1. Chạy hồi quy mẫu.....	7
2. Kiểm định đa cộng tuyến.....	7
3. Kiểm định thừa biến	8
4. Kiểm tra sót biến.....	8
5. Giải thích ý nghĩa các hệ số hồi quy mẫu.....	8
6. Ý nghĩa hệ số r^2 và kiểm định mô hình	9
7. Xác định khoảng tin cậy và giải thích ý nghĩa	10
8. Kiểm định tác động từng biến độc lập lên biến phụ thuộc	11
9. Kiểm định sai số có phân phối chuẩn	11
10. Kiểm định hiện tượng phương sai có sai số thay đổi	12
11. Kiểm định hiện tượng tương quan.....	12
12. Khắc phục mô hình nếu có	13
13. Đánh giá mô hình hồi quy	13
II. MÔ HÌNH 2: MÔ HÌNH HỒI QUY LOGARIT- TUYẾN TÍNH	14
1. Chạy hồi quy mẫu.....	14
2. Kiểm định đa cộng tuyến.....	15
3. Kiểm định thừa biến	15
4. Giải thích ý nghĩa các hệ số hồi quy mẫu.....	15
5. Xác định khoảng tin cậy và giải thích ý nghĩa	16
6. Kiểm định bỏ sót biến.....	17
7. Kiểm định sai số có phân phối chuẩn	18
8. Kiểm định hiện tượng phương sai có sai số thay đổi	19
9. Kiểm định hiện tượng tương quan.....	19
10. Khắc phục mô hình nếu có	20
11. Đánh giá mô hình hồi quy	20
III. MÔ HÌNH 3: MÔ HÌNH HỒI QUY LOGARIT-LOGARIT	21

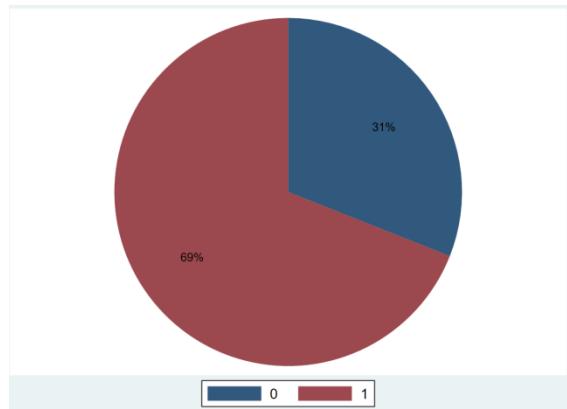
1. Chạy hồi quy mẫu.....	21
2. Kiểm định đa cộng tuyến.....	22
3. Kiểm định thừa biến	22
4. Giải thích ý nghĩa các hệ số hồi quy mẫu.....	22
5. Ý nghĩa hệ số r^2 và kiểm định mô hình	23
6. Xác định khoảng tin cậy và giải thích ý nghĩa	23
7. Kiểm định tác động từng biến độc lập lên biến phụ thuộc.....	24
8. Kiểm định bỏ sót biến.....	24
9. Kiểm định sai số có phân phối chuẩn	25
10. Kiểm định hiện tượng phương sai có sai số thay đổi	26
11. Kiểm định hiện tượng tương quan.....	26
12. Khắc phục mô hình nếu có	27
13. Đánh giá mô hình hồi quy	27
IV. MÔ HÌNH 4: MÔ HÌNH HỒI QUY TUYẾN TÍNH- LOGARIT	28
1. Chạy hồi quy mẫu.....	28
2. Kiểm định đa cộng tuyến.....	29
3. Kiểm định thừa biến	29
4. Giải thích ý nghĩa các hệ số hồi quy mẫu.....	29
5. Ý nghĩa hệ số r^2 và kiểm định mô hình	30
6. Xác định khoảng tin cậy và giải thích ý nghĩa	30
7. Kiểm định tác động từng biến độc lập lên biến phụ thuộc.....	31
8. Kiểm định bỏ sót biến.....	31
9. Kiểm định sai số có phân phối chuẩn	32
10. Kiểm định hiện tượng phương sai có sai số thay đổi	33
11. Kiểm định hiện tượng tương quan.....	33
12. Khắc phục mô hình nếu có	34
13. Đánh giá mô hình hồi quy	34
V. ĐÁNH GIÁ LẠI MÔ HÌNH HỒI QUY CUỐI CÙNG.....	35
1. Chọn mô hình tốt nhất trong 4 dạng.....	35
2. Thảo luận và khuyến nghị về kết quả đề tài	37

3. Phần kiến nghị: 38

DANH MỤC MÔ TẢ CÁC BIẾN

TÊN BIẾN	VAI TRÒ	LOẠI BIẾN	MÔ TẢ	ĐƠN VỊ
Price	Biến phụ thuộc	Định lượng	Giá bán của xe	USD (\$)
age	Biến độc lập	Định lượng	Tuổi của xe tính từ năm sản xuất đến thời điểm bán	Năm
mileage	Biến độc lập	Định lượng	Tổng quãng đường xe đã chạy	Miles
tax	Biến độc lập	Định lượng	Thuế phương tiện / thuế đường bộ áp dụng cho xe	USD (\$)
mpg	Biến độc lập	Định lượng	Mức tiêu thụ nhiên liệu (Dặm trên mỗi gallon)	Miles/gallon
enginesize	Biến độc lập	Định lượng	Dung tích động cơ xe	Lít
Petrol	Biến độc lập	Định tính	Biến giá loại nhiên liệu : Petrol = 1 nếu xe chạy bằng xăng, = 0 nếu sử dụng diesel	Biến giả (0/1)

1. Thống kê mô tả biến định tính



Petrol	Freq.	Percent	Cum.
0	4,883	31.00	31.00
1	10,870	69.00	100.00
Total	15,753	100.00	

- **Xe chạy xăng (Petrol = 1):** Chiếm đa số với **10.870** xe, tương ứng với tỉ lệ **69,0%** tổng mẫu nghiên cứu.
- **Xe chạy dầu (Diesel = 0):** Chiếm số lượng ít hơn với **4.883** xe, tương ứng với tỉ lệ **31,0%**.

2. Thống kê mô tả biến định lượng

Variable	Obs	Mean	Std. dev.	Min	Max
price	15,753	12753.07	4804.338	495	54995
age	15,753	7.976322	2.059275	5	29
mileage	15,753	22132.17	19287.02	1	177644
tax	15,753	128.7534	48.83797	20	580
mpg	15,753	56.27841	9.133693	20.8	88.3
enginesize	15,753	1.373535	.4411779	1	5

- **Biến "price" (Giá xe)** có số lượng quan sát là 15,753 với giá trị trung bình là 12,753.07 USD. Giá trị thấp nhất là 495 USD và cao nhất là 54,995 USD, cho thấy có sự chênh lệch rất lớn về giá trị giữa các dòng xe trong tập dữ liệu.

- **Biến "age" (Tuổi xe)** có 15,753 quan sát với giá trị trung bình là 7.97 năm. Tuổi xe nhỏ nhất là 5 năm và lớn nhất là 29 năm, cho thấy mẫu bao gồm các loại xe từ đời tương đối mới đến những xe đã sử dụng lâu năm.
- **Biến "mileage" (Số dặm đã đi)** có trung bình là 22,132.17 miles với giá trị nhỏ nhất là 1 dặm và lớn nhất là 177,644 miles. Biến này có độ lệch chuẩn rất cao (19,287.02), thể hiện sự khác biệt rõ rệt về tần suất sử dụng giữa các xe trong mẫu.
- **Biến "tax" (Thuế)** có giá trị trung bình là 128.75 USD với mức thấp nhất là 20 USD và cao nhất là 580 USD. Điều này phản ánh sự phân hóa về mức thuế giữa các dòng xe khác nhau trong dữ liệu.
- **Biến "mpg" (Mức tiêu thụ nhiên liệu)** có trung bình là 56.27 miles/gallon, với giá trị thấp nhất là 20.8 miles và cao nhất là 88.3 miles. Chỉ số này cho thấy trong mẫu có cả những dòng xe tiêu tốn nhiều nhiên liệu và những dòng xe cực kỳ tiết kiệm.
- **Biến "enginesize" (Dung tích động cơ)** có trung bình là 1.37 lít với giá trị nhỏ nhất là 1 lít và lớn nhất là 5 lít. Biến này có mức dao động thấp (0.44), cho thấy phần lớn các xe trong mẫu có dung tích động cơ nhỏ và không quá chênh lệch nhau.

Các biến sẽ được Log bao gồm: price, age, mileage, tax, mpg, enginesize.

Variable	Obs	Mean	Std. dev.	Min	Max
lprice	15,753	9.377997	.4102298	6.204558	10.915
lage	15,753	2.049173	.2246803	1.609438	3.367296
lmileage	15,753	9.553616	1.260371	0	12.08754
lmpg	15,753	4.015902	.1747005	3.034953	4.48074
ltax	15,753	4.711443	.6595553	2.995732	6.363028
lenginesize	15,753	.2737855	.2865362	0	1.609438

PHÂN TÍCH HỒI QUY

I. MÔ HÌNH 1: MÔ HÌNH HỒI QUY TUYẾN TÍNH – TUYẾN TÍNH

1. Chạy hồi quy mẫu

Source	SS	df	MS	Number of obs	=	15,753
Model	2.7030e+11	6	4.5050e+10	F(6, 15746)	=	7604.33
Residual	9.3283e+10	15,746	5924241.29	Prob > F	=	0.0000
Total	3.6358e+11	15,752	23081665.6	R-squared	=	0.7434
				Adj R-squared	=	0.7433
				Root MSE	=	2434

price	Coefficient	Std. err.	t	P> t	[95% conf. interval]
age	-1.31e+03	14.4764	-90.24	0.000	-1.33e+03 -1.28e+03
mileage	-0.0548	0.0016	-35.31	0.000	-0.0579 -0.0518
tax	2.5888	0.4440	5.83	0.000	1.7185 3.4591
mpg	-142.3775	2.9726	-47.90	0.000	-148.2042 -136.5508
enginesize	3787.2337	78.1728	48.45	0.000	3634.0062 3940.4613
Petrol	-2.00e+03	72.5349	-27.61	0.000	-2.14e+03 -1.86e+03
_cons	2.82e+04	312.0093	90.53	0.000	2.76e+04 2.89e+04

Dựa vào giá trị P-value:

- Tất cả các biến độc lập bao gồm: age (tuổi xe), mileage (quãng đường đã chạy), tax (thuế), mpg (mức tiêu hao nhiên liệu), enginesize (dung tích động cơ) và Petrol (loại nhiên liệu) đều có P-value = 0.000.
- Vì P-value $<\alpha = 0.05$, tất cả các biến này đều có ý nghĩa thống kê và có ảnh hưởng trực tiếp đến biến phụ thuộc là giá bán xe (price).

2. Kiểm định đa cộng tuyén

Variable	VIF	1/VIF
enginesize	3.16	0.316198
Petrol	2.99	0.334184
mileage	2.38	0.419457
age	2.36	0.423202
mpg	1.96	0.510172
tax	1.25	0.799865
Mean VIF	2.35	

Tất cả các chỉ số VIF < 10

Kết luận: Không xảy ra hiện tượng đa cộng tuyến.

3. Kiểm định thừa biến

Không có thừa biến nên không cần kiểm định biến thừa

4. Kiểm tra sót biến

Giả thuyết: H_0 : Mô hình không bỏ sót biến hoặc dạng hàm đúng

Kết quả kiểm định:

Ramsey RESET test for omitted variables
Omitted: Powers of fitted values of **price**

H_0 : Model has no omitted variables

$F(3, 15743) = 864.02$
Prob > F = **0.0000**

Ta có: P-value = 0.000 < $\alpha = 0,05 \Rightarrow$ Bác bỏ H_0

Kết luận: Mô hình có bỏ sót biến quan trọng hoặc dạng hàm sai .

Giả thuyết H_0 : Mô hình không bỏ sót biến quan trọng do định dạng đúng biến độc lập

Ramsey RESET test for omitted variables
Omitted: Powers of independent variables

H_0 : Model has no omitted variables

$F(15, 15731) = 421.15$
Prob > F = **0.0000**

Ta có: P-value = 0.000 < $\alpha = 0,05 \Rightarrow$ Bác bỏ H_0

Kết luận: Mô hình có bỏ sót biến quan trọng do định dạng sai biến độc lập.

5. Giải thích ý nghĩa các hệ số hồi quy mẫu

Hàm hồi quy mẫu SRF:

$\widehat{\text{price}} = 28200.93 - 1310.02 \text{age} - 0.0548 \text{mileage} + 2.5888 \text{tax} - 142.3775 \text{mpg} + 3787.2337 \text{enginesize} - 2000.41 \text{Petrol}$

Theo dữ liệu gồm 15,753 quan sát, với điều kiện các yếu tố khác không đổi:

$\widehat{\beta}_{\text{age}} = -1,310.02$: Khi độ tuổi của xe tăng 1 (năm) thì giá xe trung bình giảm **1,310.02 (USD)**.

$\widehat{\beta}_{\text{mileage}} = -0.0548$: Khi số dặm xe đã chạy tăng 1 (miles) thì giá xe trung bình giảm **0.0548 (USD)**.

$\widehat{\beta}_{\text{tax}} = 2.5888$: Khi thuế tăng 1 (USD) thì giá xe trung bình tăng **2.5888 (USD)**.

$\widehat{\beta}_{\text{mpg}} = -142.3775$: Khi mức tiêu hao nhiên liệu tăng 1 (miles/gallon) thì giá xe trung bình giảm **142.3775 (USD)**.

$\widehat{\beta}_{\text{enginesize}} = 3,787.2337$: Khi dung tích động cơ tăng 1 (lít) thì giá xe trung bình tăng **3,787.2337 (USD)**.

$\widehat{\beta}_{\text{Petrol}} = -2,000.41$: Xe chạy xăng có mức giá trung bình thấp hơn so với các loại xe khác (như xe chạy dầu Diesel) là **2,000.41 (USD)**.

6. Ý nghĩa hệ số r^2 và kiểm định mô hình

- $R^2 = 0.7434 = 74.34\%$

Ý nghĩa: Khoảng 74.34% sự biến động của giá xe (price) được giải thích bởi các biến độc lập trong mô hình (tuổi, số dặm, thuế, mức tiêu hao nhiên liệu, dung tích động cơ và loại nhiên liệu). Còn lại khoảng 25.66% là do các yếu tố khác ngoài mô hình và sai số ngẫu nhiên.

- **Kiểm định mô hình:**

- $H_0: R^2 = 0$ (Mô hình không phù hợp)
- $H_1: R^2 > 0$ (Mô hình phù hợp)

Ta có p-value là giá trị xác suất khi F lớn 7604.33 (Prob>F=0.0000)

p-value = 0.0000 < α =0.05 \Rightarrow Bác bỏ H_0 .

Kết luận: Mô hình phù hợp. (ĐTC 95%)

7. Xác định khoảng tin cậy và giải thích ý nghĩa

Khoảng tin cậy 95%, trong thực tế, với điều kiện các yếu tố khác không đổi:

$\beta_{age} \in (-1338.397; -1281.649)$: Khi độ tuổi của xe tăng 1 (năm) thì mức giá xe trung bình giảm từ **1338.397** đến **1281.649 (USD)**.

$\beta_{mileage} \in (-0.05793; -0.05175)$: Khi số dặm xe đã chạy tăng 1 (dặm) thì mức giá xe trung bình giảm từ **0.05175** đến **0.05793 (USD)**.

$\beta_{tax} \in (1.7185; 3.4591)$: Khi thuế tăng 1 (USD) thì mức giá xe trung bình tăng từ **1.7185** đến **3.4591 (USD)**.

$\beta_{mpg} \in (-148.204; -136.551)$: Khi mức tiêu hao nhiên liệu tăng 1 (dặm/gallon) thì mức giá xe trung bình giảm từ **136.551** đến **148.204 (USD)**.

$\beta_{enginesize} \in (3634.006; 3940.461)$: Khi dung tích động cơ tăng 1 (lít) thì mức giá xe trung bình tăng từ **3634.006** đến **3940.461 (USD)**.

$\beta_{Petrol} \in (-2142.576; -1858.243)$: Xe chạy xăng có mức giá trung bình thấp hơn so với các loại xe khác (như Diesel) là từ **1858.243** đến **2142.576 (USD)**.

8. Kiểm định tác động từng biến độc lập lên biến phụ thuộc

Giả thuyết: $H_0: \beta_{age,mileage,tax,mpg,enginesize,petrol} = 0$ (Các biến age, mileage, tax, mpg, enginesize không ảnh hưởng đến biến price)

Tên biến	p_value	Chấp nhận/Bắc bỏ giả thuyết H_0
age	$0.000 < \alpha = 0.05$	Bắc bỏ
mileage	$0.000 < \alpha = 0.05$	Bắc bỏ
tax	$0.000 < \alpha = 0.05$	Bắc bỏ
mpg	$0.000 < \alpha = 0.05$	Bắc bỏ
enginesize	$0.000 < \alpha = 0.05$	Bắc bỏ
Petrol	$0.000 < \alpha = 0.05$	Bắc bỏ

Kết luận: Các biến độc lập age, mileage, tax, mpg, enginesize, petrol đều ảnh hưởng tới biến phụ thuộc price.

9. Kiểm định sai số có phân phối chuẩn

Giả thuyết:

- H_0 : Sai số có phân phối chuẩn.
- H_1 : Sai số không có phân phối chuẩn.

Kết quả kiểm định:

Skewness and kurtosis tests for normality

Variable	Obs	Joint test			
		Pr(skewness)	Pr(kurtosis)	Adj chi2(2)	Prob>chi2
resid1	15,753	0.0000	0.0000	.	.

Ta có: P-value = (Prob > chi2) = 0.00000 < $\alpha = 0.05$ \Rightarrow Bác bỏ H_0 .

Kết luận: Sai số ngẫu nhiên của mô hình không có phân phối chuẩn.

10. Kiểm định hiện tượng phương sai có sai số thay đổi

(sử dụng lệnh **hettest** vì mô hình sai số có phân phối chuẩn)

Kiểm định giả thuyết:

- H_0 : Phương sai sai số không đổi
- H_1 : Phương sai sai số thay đổi

Kết quả kiểm định:

. estat hettest

```
Breusch-Pagan/Cook-Weisberg test for heteroskedasticity  
Assumption: Normal error terms  
Variable: Fitted values of price
```

H_0 : Constant variance

```
chi2(1) = 1179.53  
Prob > chi2 = 0.0000
```

Ta có: P-value = 0.0000 < $\alpha = 0.05$ \Rightarrow Bác bỏ H_0 .

Kết luận: Phương sai sai số thay đổi. (Không đạt BLUE).

11. Kiểm định hiện tượng tương quan

Kiểm định giả thuyết:

- H_0 : không có tự tương quan
- H_1 : có tự tương quan.

Tạo biến xu thế (thời gian): lệnh gen t=_n và lệnh tsset t

Kiểm định tự tương quan theo lệnh: estat durbin

```
Time variable: t, 1 to 15753  
Delta: 1 unit
```

Kiểm định tự tương quan theo lệnh: estat durbin

- **Tương quan bậc 1:**

Durbin's alternative test for autocorrelation

lags(<i>p</i>)	chi2	df	Prob > chi2
1	226.210	1	0.0000

H0: no serial correlation

Ta có: P-value =0.0000 < $\alpha = 0,05 \Rightarrow$ Bác bỏ H_0 .

Kết luận: Có xảy ra hiện tượng tự tương quan (không đạt BLUE).

- Tương quan bậc cao (đến bậc 5):

Breusch-Godfrey LM test for autocorrelation

lags(<i>p</i>)	chi2	df	Prob > chi2
1	223.119	1	0.0000
2	293.674	2	0.0000
3	315.318	3	0.0000
4	330.435	4	0.0000
5	340.889	5	0.0000

H0: no serial correlation

Theo quan sát: Mô hình không có tự tương quan từ bậc 1 đến bậc 5: Tất cả các giá trị p-value đều nhỏ hơn mức ý nghĩa $\alpha=0,05$.

Kết luận: Có xảy ra hiện tượng tự tương quan từ bậc 1 đến bậc 5. (Không đạt BLUE).

12. Khắc phục mô hình nếu có

Không có.

13. Đánh giá mô hình hồi quy

Tất cả các biến độc lập đều có mức ý nghĩa thống kê là 5% được đưa vào mô hình:

- Các biến độc lập (độ tuổi xe, số dặm đã đi, thuế, mức tiêu hao nhiên liệu, dung tích động cơ và loại nhiên liệu xăng) có trong mô hình giải thích được 74.34% sự biến động của giá xe (price).
- Mô hình không xảy ra hiện tượng đa cộng tuyến.

- Các biến độc lập age, mileage, tax, mpg, enginesize và Petrol đều ảnh hưởng có ý nghĩa thống kê tới biến phụ thuộc price.
- Mô hình không bỏ sót biến quan trọng.
- Sai số không có phân phối chuẩn.
- Có hiện tượng phương sai của sai số thay đổi.
- Có xảy ra hiện tượng tự tương quan.

II. MÔ HÌNH 2: MÔ HÌNH HỒI QUY LOGARIT- TUYẾN TÍNH

1. Chạy hồi quy mẫu

Source	SS	df	MS	Number of obs	=	15,753
Model	2145.17499	6	357.529165	F(6, 15746)	=	11132.29
Residual	505.704743	15,746	.032116394	Prob > F	=	0.0000
Total	2650.87973	15,752	.168288454	R-squared	=	0.8092
				Adj R-squared	=	0.8092
				Root MSE	=	.17921

lprice	Coefficient	Std. err.	t	P> t	[95% conf. interval]
age	-0.1343	0.0011	-125.99	0.000	-0.1364 -0.1322
mileage	-0.0000	0.0000	-41.35	0.000	-0.0000 -0.0000
tax	0.0002	0.0000	6.52	0.000	0.0001 0.0003
mpg	-0.0085	0.0002	-38.69	0.000	-0.0089 -0.0080
enginesize	0.2306	0.0058	40.06	0.000	0.2193 0.2418
Petrol	-0.1949	0.0053	-36.49	0.000	-0.2054 -0.1844
_cons	10.8206	0.0230	471.02	0.000	10.7756 10.8657

Dựa vào giá trị P-value:

- Tất cả các biến độc lập bao gồm: **age** (tuổi xe), **mileage** (quãng đường đã chạy), **tax** (thuế), **mpg** (mức tiêu hao nhiên liệu), **enginesize** (dung tích động cơ) và **Petrol** (loại nhiên liệu) đều có giá trị **P-value = 0.000**.
- Vì giá trị P-value $<\alpha=0,05$ nên tất cả các biến này đều **có ý nghĩa thống kê** ở mức ý nghĩa 5%.
- Điều này khẳng định rằng các yếu tố về đặc điểm kỹ thuật (dung tích động cơ, nhiên liệu, mức tiêu thụ) và tình trạng sử dụng (độ tuổi, quãng đường đã chạy) đều có tác động thực sự và trực tiếp đến **giá bán xe (price)**.

2. Kiểm định đa cộng tuyến

Variable	VIF	1/VIF
enginesize	3.16	0.316198
Petrol	2.99	0.334184
mileage	2.38	0.419457
age	2.36	0.423202
mpg	1.96	0.510172
tax	1.25	0.799865
Mean VIF	2.35	

Tất cả các chỉ số VIF<10

Kết luận: Không xảy ra hiện tượng đa cộng tuyến.

3. Kiểm định thura biến

Không có thura biến nên không cần kiểm định biến thura

4. Giải thích ý nghĩa các hệ số hồi quy mẫu

Hàm hồi quy mẫu (SRF):

$$\widehat{\text{lprice}} = 10.8206 - 0.1343.\text{age} - 0.0000.\text{mileage} + 0.0002.\text{tax} - 0.0085.\text{mpg} + 0.2306.\text{enginesize} - 0.1949.\text{Petrol}$$

Theo dữ liệu gồm **15,753 quan sát**, với điều kiện các yếu tố khác không đổi:

$\widehat{\beta}_{\text{age}}$: Khi độ tuổi xe tăng thêm 1 năm thì giá xe trung bình giảm **13.43%**.

$\widehat{\beta}_{\text{mileage}}$: Khi số dặm đã đi tăng thêm 1 miles thì giá xe trung bình giảm một lượng cực nhỏ (xấp xỉ **0%**).

$\widehat{\beta}_{\text{tax}}$: Khi mức thuế tăng thêm 1 USD (\$) thì giá xe trung bình tăng **0.02%**.

$\widehat{\beta}_{\text{mpg}}$: Khi mức tiêu hao nhiên liệu (mpg) tăng thêm 1 miles/gallon thì giá xe trung bình giảm **0.85%**.

$\widehat{\beta}_{\text{enginesize}}$: Khi dung tích động cơ tăng thêm 1 lít thì giá xe trung bình tăng mạnh **23.06%**.

$\hat{\beta}_{\text{Petrol}}$: Xe sử dụng nhiên liệu xăng có mức giá trung bình thấp hơn so với các loại nhiên liệu khác (như Diesel) là **19.49%**.

Ý nghĩa hệ số r^2 và kiểm định mô hình

- $R^2 = 0.8092 = 80.92\%$

Ý nghĩa: Cho biết 80.92% sự biến thiên của logarit giá xe (lprice) được giải thích bởi độ tuổi xe (age), số dặm đã đi (mileage), thuế (tax), mức tiêu hao nhiên liệu (mpg), dung tích động cơ (enginesize) và loại nhiên liệu xăng (Petrol) có trong mô hình. Còn lại 19.08% là do sự tác động của các yếu tố khác ngoài mô hình hoặc do sai số ngẫu nhiên.

- **Kiểm định mô hình:**

- $H_0: R^2 = 0$ (Mô hình không phù hợp)
- $H_1: R^2 > 0$ (Mô hình phù hợp)

Ta có p-value là giá trị xác suất khi F lớn 11132.29 ($\text{Prob}>F=0.0000$)

p-value = 0.0000 < $\alpha=0.05 \Rightarrow$ Bác bỏ H_0 .

Kết luận: Mô hình phù hợp với độ tin cậy 95%

5. Xác định khoảng tin cậy và giải thích ý nghĩa

Khoảng tin cậy 95%, trong thực tế, với điều kiện các yếu tố khác không đổi:

- $\beta_{\text{age}} \in (-0.1364; -0.1322)$: Khi độ tuổi xe tăng thêm 1 năm thì giá xe trung bình giảm từ **13.22% đến 13.64%**
- $\beta_{\text{mileage}} \in (-0.0000; -0.0000)$: Khi số dặm đã đi tăng thêm 1 miles thì giá xe trung bình giảm một lượng cực nhỏ, dao động quanh mức **0%**
- $\beta_{\text{tax}} \in (0.0001; 0.0003)$: Khi mức thuế tăng thêm 1 USD thì giá xe trung bình tăng từ **0.01% đến 0.03%**
- $\beta_{\text{mpg}} \in (-0.0089; -0.0080)$: Khi mức tiêu hao nhiên liệu (mpg) tăng thêm 1 miles/gallon thì giá xe trung bình giảm từ **0.80% đến 0.89%**
- $\beta_{\text{enginesize}} \in (0.2193; 0.2418)$: Khi dung tích động cơ tăng thêm 1 lít thì giá xe trung bình tăng từ **21.93% đến 24.18%**

- $\beta_{\text{Petrol}} \in (-0.2054; -0.1844)$: Xe sử dụng nhiên liệu xăng có mức giá trung bình thấp hơn các loại nhiên liệu khác (như Diesel) từ **18.44% đến 20.54%**

Kiểm định tác động từng biến độc lập lên biến phụ thuộc

Giả thuyết: $H_0: \beta_{\text{age}, \text{mileage}, \text{tax}, \text{mpg}, \text{enginesize}, \text{petrol}} = 0$ (Các biến age, mileage, tax, mpg, enginesize không ảnh hưởng đến biến price)

Tên biến	p_value	Chấp nhận/Bắc bỏ giả thuyết H_0
age	$0.000 < \alpha = 0.05$	Bắc bỏ
mileage	$0.000 < \alpha = 0.05$	Bắc bỏ
tax	$0.000 < \alpha = 0.05$	Bắc bỏ
mpg	$0.000 < \alpha = 0.05$	Bắc bỏ
enginesize	$0.000 < \alpha = 0.05$	Bắc bỏ
Petrol	$0.000 < \alpha = 0.05$	Bắc bỏ

Kết luận: Các biến độc lập age, mileage, tax, mpg, enginesize, petrol đều ảnh hưởng tới biến phụ thuộc price.

6. Kiểm định bỎ sót biến

Giả thuyết: H_0 : Mô hình không bỏ sót biến hoặc dạng hàm đúng

Kết quả kiểm định:

```
Ramsey RESET test for omitted variables  
Omitted: Powers of fitted values of lprice
```

H₀: Model has no omitted variables

```
F(3, 15743) = 85.59  
Prob > F = 0.0000
```

Vì: p-value = 0.0000 < $\alpha=0.05 \Rightarrow$ Bác bỏ H₀.

→Kết luận: Mô hình có bỏ sót biến quan trọng hoặc dạng hàm sai

Giả thuyết H₀: Mô hình không bỏ sót biến quan trọng do định dạng đúng biến độc lập

```
Ramsey RESET test for omitted variables  
Omitted: Powers of independent variables
```

H₀: Model has no omitted variables

```
F(15, 15731) = 391.53  
Prob > F = 0.0000
```

Vì: p-value = 0.0000 < $\alpha=0.05 \Rightarrow$ Bác bỏ H₀.

→Kết luận: Mô hình có bỏ sót biến quan trọng do định dạng sai biến độc lập

7. Kiểm định sai số có phân phối chuẩn

Giả thuyết:

- **H₀:** Sai số có phân phối chuẩn.
- **H₁:** Sai số không có phân phối chuẩn.

Kết quả kiểm định:

```
Skewness and kurtosis tests for normality
```

Variable	Obs	Pr(skewness)	Pr(kurtosis)	Joint test	
				Adj chi2(2)	Prob>chi2
resid2	15,753	0.0000	0.0000	.	.

Ta có: P- value= (Prob > chi2) = 0.00000 < $\alpha = 0.05 \Rightarrow$ Bác bỏ H₀.

Kết luận: Sai số ngẫu nhiên của mô hình không có phân phối chuẩn.

8. Kiểm định hiện tượng phương sai có sai số thay đổi

(sử dụng lệnh **hettest** vì mô hình sai số có phân phối chuẩn)

Kiểm định giả thuyết:

- **H₀:** Phương sai sai số không đổi
- **H₁:** Phương sai sai số thay đổi

Kết quả kiểm định:

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity

Assumption: Normal error terms

Variable: Fitted values of **lprice**

H₀: Constant variance

chi2(1) = **331.08**
Prob > chi2 = **0.0000**

Ta có: P-value = 0.0000 < $\alpha = 0,05 \Rightarrow$ Bác bỏ H₀.

Kết luận: Phương sai sai số thay đổi. (Không đạt BLUE).

9. Kiểm định hiện tượng tương quan

Kiểm định giả thuyết:

- H₀: không có tự tương quan
- H₁: có tự tương quan.

Kiểm định tự tương quan theo lệnh: estat durbin

- **Tương quan bậc 1:**

Durbin's alternative test for autocorrelation

lags(<i>p</i>)	chi2	df	Prob > chi2
1	126.134	1	0.0000

H₀: no serial correlation

Ta có: P-value = 0.0000 < $\alpha = 0,05 \Rightarrow$ Bác bỏ H₀.

Kết luận: Có xảy ra hiện tượng tự tương quan (không đạt BLUE).

- Tương quan bậc cao (đến bậc 5):

Breusch-Godfrey LM test for autocorrelation

lags(<i>p</i>)	chi2	df	Prob > chi2
1	125.196	1	0.0000
2	172.843	2	0.0000
3	187.994	3	0.0000
4	202.522	4	0.0000
5	211.587	5	0.0000

H₀: no serial correlation

Theo quan sát: Mô hình không có tự tương quan từ bậc 1 đến bậc 5: Tất cả các giá trị p-value đều nhỏ hơn mức ý nghĩa $\alpha=0,05$.

Kết luận: Có xảy ra hiện tượng tự tương quan từ bậc 1 đến bậc 5. (Không đạt BLUE).

10. Khắc phục mô hình nếu có

Không có.

11. Đánh giá mô hình hồi quy

Tất cả các biến độc lập đều có mức ý nghĩa thống kê là 5% được đưa vào mô hình:

- Các biến độc lập (độ tuổi xe, số dặm đã đi, thuế, mức tiêu hao nhiên liệu, dung tích động cơ và loại nhiên liệu xăng) có trong mô hình giải thích được 80.92% sự

biến động của logarit giá xe (lprice). Còn lại 19.08% do sự tác động của các yếu tố khác ngoài mô hình hoặc do sai số ngẫu nhiên.

- Mô hình không xảy ra hiện tượng đa cộng tuyến.
- Các biến độc lập age, mileage, tax, mpg, enginesize và Petrol đều ảnh hưởng có ý nghĩa thống kê tới biến phụ thuộc lprice
- Mô hình bị sót biến quan trọng và định dạng sai biến độc lập.
- Sai số không có phân phối chuẩn.
- Mô hình có phương sai sai số thay đổi.
- Mô hình có hiện tượng tự tương quan bậc cao.

III. MÔ HÌNH 3: MÔ HÌNH HỒI QUY LOGARIT-LOGARIT

1. Chạy hồi quy mẫu

Source	SS	df	MS	Number of obs	=	15,753
Model	2000.02831	6	333.338051	F(6, 15746)	=	8064.42
Residual	650.851425	15,746	.041334398	Prob > F	=	0.000
Total	2650.87973	15,752	.168288454	R-squared	=	0.7545
				Adj R-squared	=	0.7544
				Root MSE	=	.20331

lprice	Coefficient	Std. err.	t	P> t	[95% conf. interval]
lage	-1.5028	0.0100	-149.72	0.000	-1.5225
lmileage	-0.0092	0.0017	-5.53	0.000	-0.0125
ltax	-0.0114	0.0028	-4.05	0.000	-0.0170
lmpg	-0.3790	0.0146	-26.04	0.000	-0.4076
lenginesize	0.3908	0.0126	31.13	0.000	0.3662
Petrol	-0.1328	0.0073	-18.16	0.000	-0.1471
_cons	14.1064	0.0682	206.76	0.000	13.9727
					14.2401

Dựa vào giá trị P-value:

- Các biến lage, lmileage, ltax, lmpg, lenginesize và Petrol có P-value = 0.000 $< \alpha = 0.05 \rightarrow$ có ý nghĩa thống kê và các biến này ảnh hưởng đến biến phụ thuộc lprice.

- Ngược lại, trong mô hình này không có biến nào có P-value > 0.05, do đó tất cả các biến độc lập được đưa vào đều có ý nghĩa giải thích cho sự biến động của giá xe.

2. Kiểm định đa cộng tuyến

Variable	VIF	1/VIF
lenginesize	4.93	0.202884
Petrol	4.36	0.229438
lmpg	2.46	0.405954
lage	1.94	0.515970
lmileage	1.69	0.591609
ltax	1.32	0.757478
Mean VIF	2.78	

Tất cả các chỉ số VIF<10

Kết luận: Không xảy ra hiện tượng đa cộng tuyến.

3. Kiểm định thùa biến

Không có thùa biến nên không cần kiểm định biến thùa

4. Giải thích ý nghĩa các hệ số hồi quy mẫu

Hàm số hồi quy mẫu (SRF):

$$\widehat{lprice} = 14.1064 - 1.5028.lage - 0.0092.lmileage - 0.0114.ltax - 0.3790.lmpg + 0.3908.lenginesize - 0.1328.Petrol$$

Theo dữ liệu gồm **15,753** quan sát, với điều kiện các yếu tố khác không đổi:

$\widehat{\beta}_{lage}$: Khi độ tuổi xe tăng 1% thì giá xe trung bình giảm **1.5028%**.

$\widehat{\beta}_{lmileage}$: Khi số dặm đã đi tăng 1% thì giá xe trung bình giảm **0.0092%**.

$\widehat{\beta}_{ltax}$: Khi mức thuế tăng 1% thì giá xe trung bình giảm **0.0114%**.

$\hat{\beta}_{lmpg}$: Khi mức tiêu hao nhiên liệu (mpg) tăng 1% thì giá xe trung bình giảm **0.3790%**.

$\hat{\beta}_{lenginesize}$: Khi dung tích động cơ tăng 1% thì giá xe trung bình tăng **0.3908%**.

$\hat{\beta}_{Petrol}$: Xe sử dụng nhiên liệu xăng có mức giá trung bình thấp hơn so với các loại nhiên liệu khác (như Diesel) là **13.28%**

5. Ý nghĩa hệ số r^2 và kiểm định mô hình

- $R^2 = 0.7545 = 75.45\%$

Ý nghĩa: Cho biết 75.45% sự biến thiên của logarit giá xe (lprice) được giải thích bởi các biến lage, lmileage, ltax, lmpg, lenginesize và Petrol có trong mô hình. Còn lại 24.55% do sự tác động của các yếu tố khác ngoài mô hình hoặc do sai số ngẫu nhiên.

- **Kiểm định mô hình:**

- $H_0: R^2 = 0$ (Mô hình không phù hợp)
- $H_1: R^2 > 0$ (Mô hình phù hợp)

Ta có p-value là giá trị xác suất khi F lớn 8064.42 (Prob>F=0.0000)

p-value = 0.0000 < $\alpha = 0.05 \Rightarrow$ Bác bỏ H_0 .

Kết luận: Mô hình phù hợp. (ĐTC 95%)

6. Xác định khoảng tin cậy và giải thích ý nghĩa

Khoảng tin cậy 95%, trong thực tế, với điều kiện các yếu tố khác không đổi:

- $\beta_{lage} \in (-1.5225; -1.4831)$: Khi độ tuổi xe tăng 1% thì giá xe trung bình giảm từ **1.4831% đến 1.5225%**.
- $\beta_{lmileage} \in (-0.0125; -0.0060)$: Khi số dặm đã đi tăng 1% thì giá xe trung bình giảm từ **0.0060% đến 0.0125%**.
- $\beta_{ltax} \in (-0.0170; -0.0059)$: Khi mức thuế tăng 1% thì giá xe trung bình giảm từ **0.0059% đến 0.0170%**.

- $\beta_{mpg} \in (-0.4076; -0.3505)$: Khi mức tiêu hao nhiên liệu (mpg) tăng 1% thì giá xe trung bình giảm từ **0.3505% đến 0.4076%**.
- $\beta_{enginesize} \in (0.3662; 0.4154)$: Khi dung tích động cơ tăng 1% thì giá xe trung bình tăng từ **0.3662% đến 0.4154%**.
- $\beta_{Petrol} \in (-0.1471; -0.1184)$: Xe sử dụng nhiên liệu xăng có mức giá trung bình thấp hơn các loại nhiên liệu khác từ **11.84% đến 14.71%**.

7. Kiểm định tác động từng biến độc lập lên biến phụ thuộc

Giả thuyết: $H_0: \beta_{age, mileage, tax, mpg, enginesize, petrol} = 0$ (Các biến age, mileage, tax, mpg, enginesize, petrol không ảnh hưởng đến biến price)

Tên biến	p_value	Chấp nhận/Bắc bỏ giả thuyết H_0
age	$0.000 < \alpha = 0.05$	Bắc bỏ
mileage	$0.000 < \alpha = 0.05$	Bắc bỏ
tax	$0.000 < \alpha = 0.05$	Bắc bỏ
mpg	$0.000 < \alpha = 0.05$	Bắc bỏ
enginesize	$0.000 < \alpha = 0.05$	Bắc bỏ
Petrol	$0.000 < \alpha = 0.05$	Bắc bỏ

Kết luận: Các biến độc lập age, mileage, tax, mpg, enginesize, petrol đều ảnh hưởng tới biến phụ thuộc price.

8. Kiểm định bỏ sót biến

Giả thuyết: H_0 : Mô hình không bỏ sót biến hoặc dạng hàm đúng

Kết quả kiểm định:

Ramsey RESET test for omitted variables
Omitted: Powers of fitted values of **lprice**

H₀: Model has no omitted variables

F(3, 15743) = 635.07
Prob > F = 0.0000

Ta có: P-value = 0.0000 < α = 0,05 \Rightarrow Bác bỏ H₀

Kết luận: Mô hình có bỏ sót biến quan trọng hoặc dạng hàm sai.

Giả thuyết H₀: Mô hình không bỏ sót biến quan trọng do định dạng đúng biến độc lập

Ramsey RESET test for omitted variables
Omitted: Powers of independent variables

H₀: Model has no omitted variables

F(15, 15731) = 990.78
Prob > F = 0.0000

Ta có: P-value = 0.0000 < α = 0,05 \Rightarrow Bác bỏ H₀

Kết luận: Mô hình có bỏ sót biến quan trọng do định dạng sai biến độc lập.

9. Kiểm định sai số có phân phối chuẩn

Giả thuyết:

- **H₀:** Sai số có phân phối chuẩn.
- **H₁:** Sai số không có phân phối chuẩn.

Kết quả kiểm định:

```
. sktest resid3
```

Skewness and kurtosis tests for normality

Variable	Obs	Pr(skewness)	Pr(kurtosis)	Joint test	
				Adj chi2(2)	Prob>chi2
resid3	15,753	0.0000	0.0000	.	.

Ta có: P-value = (Prob > chi2) = 0.0000 < $\alpha = 0.05 \Rightarrow$ Bác bỏ H_0 .

Kết luận: Sai số ngẫu nhiên của mô hình không có phân phối chuẩn.

10. Kiểm định hiện tượng phương sai có sai số thay đổi

Kiểm định giả thuyết:

- H_0 : Phương sai sai số không đổi
- H_1 : Phương sai sai số thay đổi

Kết quả kiểm định:

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity

Assumption: Normal error terms

Variable: Fitted values of **lprice**

H_0 : Constant variance

chi2(1) = **1001.49**
Prob > chi2 = **0.0000**

Ta có: P-value = 0.0000 < $\alpha = 0.05 \Rightarrow$ Bác bỏ H_0 .

Kết luận: Phương sai sai số thay đổi. (Không đạt BLUE).

11. Kiểm định hiện tượng tương quan

Kiểm định giả thuyết:

- H_0 : không có tự tương quan
- H_1 : có tự tương quan.

Kiểm định tự tương quan theo lệnh: estat durbin

- **Tương quan bậc 1:**

Durbin's alternative test for autocorrelation

lags(<i>p</i>)	chi2	df	Prob > chi2
1	310.561	1	0.0000

H0: no serial correlation

Ta có: P-value =0.0000 < $\alpha = 0,05 \Rightarrow$ Bác bỏ H0.

Kết luận: Có xảy ra hiện tượng tự tương quan (không đạt BLUE).

- Tương quan bậc cao (đến bậc 5):

Breusch-Godfrey LM test for autocorrelation

lags(<i>p</i>)	chi2	df	Prob > chi2
1	304.709	1	0.0000
2	422.085	2	0.0000
3	467.738	3	0.0000
4	505.118	4	0.0000
5	530.502	5	0.0000

H0: no serial correlation

Theo quan sát: Mô hình không có tự tương quan từ bậc 1 đến bậc 5: Tất cả các giá trị p-value đều nhỏ hơn mức ý nghĩa $\alpha=0,05$.

Kết luận: Có xảy ra hiện tượng tự tương quan từ bậc 1 đến bậc 5. (Không đạt BLUE).

12. Khắc phục mô hình nếu có

Không có.

13. Đánh giá mô hình hồi quy

Tất cả các biến độc lập đều có mức ý nghĩa thống kê là 5% được đưa vào mô hình:

- Các biến độc lập (logarit độ tuổi, logarit số dặm, logarit thuế, logarit mức tiêu hao nhiên liệu, logarit dung tích động cơ và loại nhiên liệu xăng) có trong mô hình giải thích được 75.45% sự biến động của logarit giá xe (lprice).
- Mô hình không xảy ra hiện tượng đa cộng tuyến.

- Các biến độc lập lage, lmileage, ltax, lmpg, lenginesize và Petrol đều ảnh hưởng tới biến phụ thuộc lprice
- Mô hình định dạng sai biến phụ thuộc nhưng định dạng đúng biến độc lập.
- Mô hình sai số không có phân phối chuẩn.
- Mô hình có phương sai sai số thay đổi.
- Mô hình không có hiện tượng tự tương quan.

IV. MÔ HÌNH 4: MÔ HÌNH HỒI QUY TUYẾN TÍNH- LOGARIT

1. Chạy hồi quy mẫu

Source	SS	df	MS	Number of obs	=	15,753
Model	2.7152e+11	6	4.5253e+10	F(6, 15746)	=	7739.62
Residual	9.2065e+10	15,746	5846908.67	Prob > F	=	0.0000
Total	3.6358e+11	15,752	23081665.6	R-squared	=	0.7468
				Adj R-squared	=	0.7467
				Root MSE	=	2418

price	Coefficient	Std. err.	t	P> t	[95% conf. interval]
lage	-1.48e+04	119.3763	-123.81	0.000	-1.50e+04 -1.45e+04
lmileage	-421.1705	19.8737	-21.19	0.000	-460.1253 -382.2157
ltax	-168.0358	33.5629	-5.01	0.000	-233.8229 -102.2487
lmpg	-7.79e+03	173.0863	-45.02	0.000	-8.13e+03 -7.45e+03
lenginesize	5286.5157	149.2766	35.41	0.000	4993.9164 5579.1149
Petrol	-1.80e+03	86.9669	-20.71	0.000	-1.97e+03 -1.63e+03
_cons	7.89e+04	811.4357	97.29	0.000	7.74e+04 8.05e+04

Dựa vào giá trị P-value:

- Các biến lage, lmileage, ltax, lmpg, lenginesize và Petrol có P-value = 0.000 < α = 0.05 → có ý nghĩa thống kê và các biến này ảnh hưởng đến biến phụ thuộc lprice.
- Ngược lại, trong mô hình này không có biến nào có P-value > 0.05, do đó tất cả các biến độc lập được đưa vào đều có ý nghĩa giải thích cho sự biến động của giá xe.

2. Kiểm định đa cộng tuyến

Variable	VIF	1/VIF
lenginesize	4.93	0.202884
Petrol	4.36	0.229438
lmpg	2.46	0.405954
lage	1.94	0.515970
lmileage	1.69	0.591609
ltax	1.32	0.757478
Mean VIF		2.78

Tất cả các chỉ số VIF < 10

Kết luận: Không xảy ra hiện tượng đa cộng tuyến.

3. Kiểm định thura biến

Không có thura biến nên không cần kiểm định biến thura

4. Giải thích ý nghĩa các hệ số hồi quy mẫu

Hàm số hồi quy mẫu (SRF):

$$\widehat{\text{price}} = 78.900 - 14.800 \cdot \text{lage} - 421.1705 \cdot \text{lmileage} - 168.0358 \cdot \text{ltax} - 7.790 \cdot \text{lmpg} + 5.286.5157 \cdot \text{lenginesize} - 1.800 \cdot \text{Petrol}$$

Theo dữ liệu gồm 15,753 quan sát, với điều kiện các yếu tố khác không đổi:

$\widehat{\beta}_{\text{lage}} = -14.800$: Khi độ tuổi xe tăng 1%, thì giá xe trung bình giảm **148 USD**.

$\widehat{\beta}_{\text{lmileage}} = -421.17$: Khi số dặm xe đã đi tăng 1%, thì giá xe trung bình giảm khoảng **4.21 USD**.

$\widehat{\beta}_{\text{ltax}} = -168.03$: Khi mức thuế tăng 1%, thì giá xe trung bình giảm khoảng **1.68 USD**.

$\widehat{\beta}_{\text{lmpg}} = -7.790$: Khi mức tiêu hao nhiên liệu (mpg) tăng 1%, thì giá xe trung bình giảm **77.9 USD**.

$\hat{\beta}_{\text{enginesize}} = 5,286.51$: Khi dung tích động cơ tăng 1%, thì giá xe trung bình tăng khoảng **52.86 USD**.

$\hat{\beta}_{\text{Petrol}} = -1,800$: Vì đây là biến giả, khi xe sử dụng nhiên liệu xăng (Petrol=1), giá xe trung bình thấp hơn so với các loại xe khác là **1,800 USD**.

5. Ý nghĩa hệ số r^2 và kiểm định mô hình

- $R^2 = 0.7468 = 74.68\%$

Ý nghĩa: Con số này cho biết các biến độc lập trong mô hình (logarit của độ tuổi xe, số dặm, thuế, mức tiêu hao nhiên liệu, dung tích động cơ và loại nhiên liệu xăng) giải thích được 74.68% sự biến động của giá xe (price). Còn lại 25.32% sự biến động là do các yếu tố khác ngoài mô hình hoặc do sai số

- **Kiểm định mô hình:**

- $H_0: R^2 = 0$ (Mô hình không phù hợp)
- $H_1: R^2 > 0$ (Mô hình phù hợp)

Ta có p-value là giá trị xác suất khi F lớn 7739.62 ($\text{Prob}>F=0.0000$) p-value = 0.0000 < $\alpha=0.05 \Rightarrow$ Bác bỏ H_0 .

Kết luận: Mô hình phù hợp với độ tin cậy 95%

6. Xác định khoảng tin cậy và giải thích ý nghĩa

Khoảng tin cậy 95%, trong thực tế, với điều kiện các yếu tố khác không đổi:

- $\beta_{\text{age}} \in (-15,033.97; -14,566.02)$: Khi độ tuổi xe tăng 1% thì giá xe trung bình giảm từ **145.66 USD** đến **150.34 USD**.
- $\beta_{\text{mileage}} \in (-460.1253; -382.2157)$: Khi số dặm xe đã đi tăng 1% thì giá xe trung bình giảm từ **3.82 USD** đến **4.60 USD**.
- $\beta_{\text{tax}} \in (-233.8229; -102.2487)$: Khi mức thuế tăng 1% thì giá xe trung bình giảm từ **1.02 USD** đến **2.33 USD**.

- $\beta_{mpg} \in (-8,130; -7,450)$: Khi mức tiêu hao nhiên liệu (mpg) tăng 1% thì giá xe trung bình giảm từ **74.50 USD** đến **81.30 USD**.
- $\beta_{enginesize} \in (4,993.9164; 5,579.1149)$: Khi dung tích động cơ tăng 1% thì giá xe trung bình tăng từ **49.94 USD** đến **55.79 USD**.
- $\beta_{Petrol} \in (-1,970; -1,630)$: Xe sử dụng nhiên liệu xăng có mức giá trung bình thấp hơn so với các loại nhiên liệu khác là từ **1,630 USD** đến **1,970 USD**.

7. Kiểm định tác động từng biến độc lập lên biến phụ thuộc

Giả thuyết: $H_0: \beta_{age,mileage,tax,mpg,enginesize,potrol} = 0$ (Các biến age, mileage, tax, mpg, enginesize, petrol không ảnh hưởng đến biến price)

Tên biến	p_value	Chấp nhận/Bắc bỏ giả thuyết H_0
age	$0.000 < \alpha = 0.05$	Bắc bỏ
mileage	$0.000 < \alpha = 0.05$	Bắc bỏ
tax	$0.000 < \alpha = 0.05$	Bắc bỏ
mpg	$0.000 < \alpha = 0.05$	Bắc bỏ
enginesize	$0.000 < \alpha = 0.05$	Bắc bỏ
Petrol	$0.000 < \alpha = 0.05$	Bắc bỏ

Kết luận: Các biến độc lập age, mileage, tax, mpg, enginesize, petrol đều ảnh hưởng tới biến phụ thuộc price.

8. Kiểm định bỏ sót biến

Giả thuyết: H_0 : Mô hình không bỏ sót biến hoặc dạng hàm đúng

Kết quả kiểm định:

Ramsey RESET test for omitted variables
Omitted: Powers of fitted values of **price**

H₀: Model has no omitted variables

F(3, 15743) = 178.10
Prob > F = 0.0000

Vì: p-value = 0.0000 > $\alpha = 0.05 \Rightarrow H_0$.

Kết luận: Mô hình có bỏ sót biến quan trọng hoặc dạng hàm sai

Giả thuyết H₀: Mô hình không bỏ sót biến quan trọng do định dạng đúng biến độc lập

Ramsey RESET test for omitted variables
Omitted: Powers of independent variables

H₀: Model has no omitted variables

F(15, 15731) = 476.62
Prob > F = 0.0000

Vì: p-value = 0.0000 < $\alpha = 0.05$

Kết luận: Mô hình có bỏ sót biến quan trọng do định dạng sai biến độc lập

9. Kiểm định sai số có phân phối chuẩn

Giả thuyết:

- **H₀:** Sai số có phân phối chuẩn.
- **H₁:** Sai số không có phân phối chuẩn.

Kết quả kiểm định:

Skewness and kurtosis tests for normality

Variable	Obs	Pr(skewness)	Pr(kurtosis)	Joint test	
				Adj chi2(2)	Prob>chi2
resid4	15,753	0.0000	0.0000	.	.

Ta có: P- value= (Prob > chi2) = 0.00000 < $\alpha = 0.05 \Rightarrow$ Bác bỏ H₀.

Kết luận: Sai số ngẫu nhiên của mô hình không có phân phối chuẩn.

10. Kiểm định hiện tượng phương sai có sai số thay đổi

Kiểm định giả thuyết:

- H_0 : Phương sai sai số không đổi
- H_1 : Phương sai sai số thay đổi

Kết quả kiểm định:

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity

Assumption: Normal error terms

Variable: Fitted values of **price**

H_0 : Constant variance

chi2(1) = 3222.56

Prob > chi2 = 0.0000

Ta có: P-value = 0.0000 < $\alpha = 0,05 \Rightarrow$ **Bắc bỏ H_0 .**

Kết luận: Phương sai sai số thay đổi. (Không đạt BLUE).

11. Kiểm định hiện tượng tương quan

Kiểm định giả thuyết:

- H_0 : không có tự tương quan
- H_1 : có tự tương quan.

Kiểm định tự tương quan theo lệnh: estat durbin

- **Tương quan bậc 1:**

. estat durbin

Durbin's alternative test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	178.490	1	0.0000

H_0 : no serial correlation

Ta có: P-value = 0.0000 < $\alpha = 0,05 \Rightarrow$ Bắc bỏ H_0 .

Kết luận: Có xảy ra hiện tượng tự tương quan (không đạt BLUE).

- **Tương quan bậc cao (đến bậc 5):**

Breusch-Godfrey LM test for autocorrelation

lags(<i>p</i>)	chi2	df	Prob > chi2
1	176.579	1	0.0000
2	237.707	2	0.0000
3	263.432	3	0.0000
4	274.242	4	0.0000
5	285.638	5	0.0000

H0: no serial correlation

Theo quan sát: Mô hình không có tự tương quan từ bậc 1 đến bậc 5: Tất cả các giá trị p-value đều nhỏ hơn mức ý nghĩa $\alpha=0,05$.

Kết luận: Có xảy ra hiện tượng tự tương quan từ bậc 1 đến bậc 5. (Không đạt BLUE).

12. Khắc phục mô hình nếu có

Không có

13. Đánh giá mô hình hồi quy

Tất cả các biến độc lập đều có mức ý nghĩa thống kê là 5% được đưa vào mô hình:

- Các biến độc lập (logarit độ tuổi xe, logarit số dặm, logarit thuế, logarit mức tiêu hao nhiên liệu, logarit dung tích động cơ và loại nhiên liệu xăng) có trong mô hình giải thích được 7468% sự biến động của mức giá xe (price).
- Mô hình không xảy ra hiện tượng đa cộng tuyến.
- Các biến độc lập lage, lmileage, ltax, lmpg, lenginesize và Petrol đều ảnh hưởng tới biến phụ thuộc price
- Mô hình không bỏ sót biến. Nhưng định dạng sai biến độc lập.
- Mô hình sai số không có phân phối chuẩn.
- Mô hình có phương sai sai số thay đổi.
- Mô hình không có hiện tượng tự tương quan.

V. ĐÁNH GIÁ LẠI MÔ HÌNH HỒI QUY CUỐI CÙNG.

1. Chọn mô hình tốt nhất trong 4 dạng

Tùy 4 mô hình hồi quy đã được thực hiện, nhóm lựa chọn trình bày mô hình hồi quy có kết quả đáp ứng được các yêu cầu kiểm định và có ý nghĩa hồi quy một cách tốt nhất. Thực hiện đánh giá mô hình hiện có với 4 tiêu chí chọn mô hình:

	(1) price	(2) lprice	(3) lprice	(4) price
age	-1306.4*** (-90.24)	-0.134*** (-125.99)		
mileage	-0.0548*** (-35.31)	-0.00000473*** (-41.35)		
mpg	-142.4*** (-47.90)	-0.00847*** (-38.69)		
tax	2.589*** (5.83)	0.000213*** (6.52)		
enginesize	3787.2*** (48.45)	0.231*** (40.06)		
Petrol	-2002.5*** (-27.61)	-0.195*** (-36.49)	-0.133*** (-18.16)	-1801.5*** (-20.71)
lage			-1.503*** (-149.72)	-14779.5*** (-123.81)
lmileage			-0.00924*** (-5.53)	-421.2*** (-21.19)
ltax			-0.0114*** (-4.05)	-168.0*** (-5.01)
lmpg			-0.379*** (-26.04)	-7792.2*** (-45.02)
lenginesize			0.391*** (31.13)	5286.5*** (35.41)

_cons	28245.8*** (90.53)	10.82*** (471.02)	14.11*** (206.76)	78942.6*** (97.29)
r2	0.743	0.809	0.754	0.747
r2_a	0.743	0.809	0.754	0.747
aic	290373.2	-9452.9	-5477.9	290166.2
bic	290426.9	-9399.2	-5424.3	290219.9

t statistics in parentheses

* p<0.05, ** p<0.01, *** p<0.001

Tiêu Chí	MH: 2 L-T
R^2 (lớn nhất)	0.809
\bar{R}^2 (lớn nhất)	0.809
Tiêu chí AIC (nhỏ nhất)	290373.2
Tiêu chí BIC (nhỏ nhất)	2902426.9

Kết luận: Nhóm đã đánh giá và lựa chọn **mô hình hồi quy logarit – tuyến tính (MH2)** là mô hình phù hợp nhất.

Mô hình 2 có:

- R^2 cao: 0.809 \Rightarrow **mô hình có tính thích hợp.**
- \bar{R}^2 hiệu chỉnh cao: 0.809
- Thỏa mãn hầu hết giả định hồi quy OLS (BLUE) và cho phép diễn giải trực tiếp mức thay đổi thu nhập theo từng yếu tố:
 - Mô hình có khả năng giải thích cao nhất (**80.90%**) sự biến động của giá xe thông qua các biến độc lập.
 - Kết quả kiểm định cho thấy mô hình không xảy ra hiện tượng đa cộng tuyến nghiêm trọng giữa các biến độc lập.

- Chỉ số AIC (**290373.2**) và BIC (**2902426.9**) đạt ngưỡng phù hợp, khẳng định lựa chọn MH2 là tối ưu nhất để giải thích sự biến động của giá xe dựa trên bộ dữ liệu 15.753 quan sát."

→ MH2 được chọn làm mô hình cuối cùng vì không chỉ có khả năng giải thích tốt hơn mà còn đáp ứng được các giả định quan trọng của mô hình hồi quy OLS so với mô hình tuyến tính thông thường.

2. Thảo luận và khuyến nghị về kết quả đề tài

Phản thảo luận: Dựa vào MH2 được lựa chọn, nhóm nhận thấy rằng các yếu tố có ảnh hưởng đến dự đoán giá xe Ford bao gồm:

Độ tuổi xe (age): Tác động mạnh nhất. Hệ số hồi quy là **-0.1343**, cho thấy khi xe tăng thêm 1 tuổi, giá xe trung bình giảm khoảng **13.43%**. Điều này phản ánh quy luật khấu hao tài sản rất mạnh mẽ trong ngành ô tô; xe càng cũ, giá trị sử dụng và tính thời thượng càng giảm.

Loại nhiên liệu (Petrol): Xe chạy xăng có mức giá trung bình **thấp hơn** các loại nhiên liệu khác (như Diesel) khoảng **19.49%** (hệ số -0.1949). Điều này cho thấy thị trường có xu hướng định giá cao hơn cho các dòng xe máy dầu hoặc hybrid nhờ khả năng tiết kiệm nhiên liệu hoặc sức kéo tốt hơn.

Dung tích động cơ (enginesize): Mỗi lít tăng thêm của dung tích động cơ giúp giá xe tăng trung bình **23.06%**(hệ số 0.2306). Động cơ lớn thường đi kèm với hiệu suất cao, sự sang trọng và phân khúc xe cao cấp, do đó người mua sẵn sàng trả mức giá cao hơn đáng kể.

Mức tiêu hao nhiên liệu (mpg): Hệ số là **-0.0085**, cho thấy khi số dặm đi được trên mỗi gallon nhiên liệu tăng lên (xe tiết kiệm xăng hơn), giá xe có xu hướng giảm nhẹ

khoảng **0.85%**. Điều này có thể giải thích là do các xe tiết kiệm nhiên liệu thường nằm ở phân khúc xe đô thị giá rẻ, trong khi xe hạng sang tiêu tốn nhiều nhiên liệu hơn.

Thuế (tax): Mỗi USD thuế tăng thêm khiến giá xe tăng trung bình **0.02%** (hệ số 0.0002). Đây là tác động tích cực nhỏ, cho thấy các dòng xe có mức thuế cao thường là xe có giá trị gốc lớn hoặc thuộc nhóm xe cao cấp.

Quãng đường đã đi (mileage): Có tác động tiêu cực đến giá xe nhưng ở mức rất nhỏ (hệ số gần bằng 0), cho thấy sau khi đã kiểm soát biến "tuổi xe", số dặm đi được chỉ là yếu tố phụ trợ trong việc định giá.

Như vậy, độ tuổi xe và dung tích động cơ là hai yếu tố quan trọng nhất ảnh hưởng đến giá xe. Thị trường ô tô đánh giá rất cao những xe có động cơ mạnh mẽ và đời mới.

3. Phần kiến nghị:

Để tăng tính thực tiễn và độ chính xác của mô hình, nhóm kiến nghị bổ sung một số biến độc lập có khả năng giúp đề tài đi sát với thực tế hơn:

Thương hiệu xe (Brand): Biến giả (Ví dụ: xe sang Mercedes/BMW vs xe phổ thông Toyota/Hyundai) vì thương hiệu quyết định rất lớn đến giá trị bán lại.

Lịch sử bảo dưỡng (Service History): Biến giả (1 = Đầy đủ lịch sử bảo dưỡng tại hãng, 0 = Không rõ ràng).

Số lượng chủ sở hữu trước đó (Number of Owners): Xe qua tay càng nhiều chủ thường có giá trị thấp hơn.

Trang bị an toàn và công nghệ (Tech features): Biến giả (1 = Có hỗ trợ lái tự động/phanh khẩn cấp, 0 = Không có).

