BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC KINH TẾ - TÀI CHÍNH THÀNH PHỐ HỒ CHÍ MINH**  
 **---------------------------------------**



**BÁO CÁO NHÓM**

**“CÁC YẾU TỐ ẢNH HƯỞNG ĐẾN GIÁ CẢ CỦA RƯỢU VANG”**

**Môn: Kinh tế lượng**

**Mã lớp: ECO1104 – Nhóm B**

**Giảng viên: ThS. Nguyễn Phú Quới**

*Thành phố Hồ Chí Minh, tháng 01 năm 2026.*

**DANH SÁCH THÀNH VIÊN**

|  |  |  |  |
| --- | --- | --- | --- |
| **Họ và Tên** | **MSSV** | **Nhiệm vụ** | **Mức độ đóng góp** |
| 1. Dương Gia Mẩn | 225013921 | Viết mở đầu, báo cáo MH3, MH4, chỉnh sửa báo cáo và đánh giá lại mô hình hồi quy cuối cùng, slide thuyết trình. | 100% |
| 2. Hồ Văn Như Ý | 245191885 | Viết code và đánh giá MH1, MH2, MH3, MH4 và mô hình tối ưu (MH5) Viết báo cáo MH1, MH2, Chỉnh sửa , định dạng báo cáo và đánh giá lại mô hình hồi quy cuối cùng, sửa báo cáo, tối ưu hóa mô hình tuyến tính và hồi quy tuyến tính, slide thuyết trình, Viết kết luận. | 100% |
| 3 Âu Hoàng Duy | 215025174 | Làm slide thuyết trình | 10% |
| Trần Gia Huy | 225084294 | Làm slide thuyết trình | 10% |

**MỤC LỤC**

[I. MỞ ĐẦU 1](#_Toc219333014)

[1. Lý do chọn đề tài 1](#_Toc219333015)

[2. Mục tiêu nghiên cứu 1](#_Toc219333016)

[3. Câu hỏi nghiên cứu 1](#_Toc219333017)

[II. MÔ TẢ DỮ LIỆU 1](#_Toc219333018)

[1. Nguồn và đặc điểm dữ liệu 1](#_Toc219333019)

[2. Mô tả biến 1](#_Toc219333020)

[Mô tả biến định lượng 2](#_Toc219333021)

[Mô tả biến định tính 3](#_Toc219333022)

[III. MÔ HÌNH 1: DẠNG TUYẾN TÍNH – TUYẾN TÍNH 4](#_Toc219333023)

[1. Chạy mô hình hồi quy mẫu 4](#_Toc219333024)

[2. Kiểm định đa cộng tuyến và loại bớt biến đa cộng tuyến 5](#_Toc219333025)

[3. Kiểm Định Thừa Biến 5](#_Toc219333026)

[4. Kiểm Tra sót biến 6](#_Toc219333027)

[5. Ý nghĩa hệ số r2 và kiểm định mô hình 6](#_Toc219333028)

[6. Xác định khoảng tin cậy và giải thích ý nghĩa 7](#_Toc219333029)

[7. Kiểm định tác động từng biến độc lập lên biến phụ thuộc 7](#_Toc219333030)

[8. Kiểm định sai số có phân phối chuẩn 7](#_Toc219333031)

[9. Kiểm định hiện tượng phương sai có sai số thay đổi 8](#_Toc219333032)

[10. Kiểm định hiện tượng tương quan 8](#_Toc219333033)

[11. Khắc phục mô hình nếu có 9](#_Toc219333034)

[12. Đánh giá mô hình hồi quy 9](#_Toc219333035)

[IV. MÔ HÌNH 2: DẠNG LOGARIT – TUYẾN TÍNH 9](#_Toc219333036)

[1. Chạy mô hình hồi quy mẫu 9](#_Toc219333037)

[2. Kiểm định đa cộng tuyến 11](#_Toc219333038)

[3. Kiểm tra thừa biến 11](#_Toc219333039)

[4. Kiểm tra sót biến: 11](#_Toc219333040)

[5. Ý nghĩa hệ số R2 và kiểm định sự phù hợp của mô hình 12](#_Toc219333041)

[6. Xác định khoảng tin cậy 95% và giải thích ý nghĩa 12](#_Toc219333042)

[7. Kiểm định tác động từng biến độc lập lên biến phụ thuộc 12](#_Toc219333043)

[8. Kiểm định sai số có phân phối chuẩn 13](#_Toc219333044)

[9. Kiểm định hiện tượng phương sai sai số thay đổi 13](#_Toc219333045)

[10. Kiểm định hiện tượng tương quan 14](#_Toc219333046)

[11. Đánh giá chung về mô hình hồi quy 15](#_Toc219333047)

[12. Khắc phục mô hình 15](#_Toc219333048)

[V. MÔ HÌNH 3: MÔ HÌNH HỒI QUY (LOG-LOG) 15](#_Toc219333049)

[1. Chạy hồi quy mẫu và giải thích ý nghĩa các hệ số hồi quy mẫu 15](#_Toc219333050)

[2. Kiểm tra đa cộng tuyến 16](#_Toc219333051)

[3. Kiểm định biến thừa 16](#_Toc219333052)

[4. Giải thích ý nghĩa các hệ số hồi quy mẫu 16](#_Toc219333053)

[5. Ý nghĩa hệ số R2 và kiểm định mô hình 17](#_Toc219333054)

[6. Kiểm định mô hình: 17](#_Toc219333055)

[7. Xác định khoảng tin cậy và giải thích ý nghĩa 17](#_Toc219333056)

[8. Kiểm định tác động của từng biến độc lập lên biến phụ thuộc 17](#_Toc219333057)

[9. Kiểm định sót biến 18](#_Toc219333058)

[10. Kiểm định sai số có phân phối chuẩn 19](#_Toc219333059)

[11. Kiểm định hiện tượng phương sai có sai số thay đổi 19](#_Toc219333060)

[12. Kiểm định hiện tượng tương quan: 20](#_Toc219333061)

[13. Đánh giá mô hình hồi quy: 20](#_Toc219333062)

[14. Khắc phục mô hình nếu có: 21](#_Toc219333063)

[VI. MÔ HÌNH 4: MÔ HÌNH HỒI QUY TUYẾN TÍNH- LOGARIT 21](#_Toc219333064)

[1. Chạy mô hình hồi quy mẫu 21](#_Toc219333065)

[2. Kiểm định đa cộng tuyến 22](#_Toc219333066)

[3. Kiểm định thừa biến 22](#_Toc219333067)

[4. Giải thích ý nghĩa các hệ số hồi quy mẫu 22](#_Toc219333068)

[5. Giải thích ý nghĩa hệ số hồi quy mẫu 22](#_Toc219333069)

[6. Ý nghĩa hệ số R2 và kiểm định mô hình: 23](#_Toc219333070)

[7. Xác định khoảng tin cậy và giải thích ý nghĩa 23](#_Toc219333071)

[8. Kiểm định tác động từng biến độc lập lên biến phụ thuộc 24](#_Toc219333072)

[9. Kiểm định bỏ sót biến 24](#_Toc219333073)

[10. Kiểm định sai số có phân phối chuẩn 25](#_Toc219333074)

[11. Kiểm định hiện tượng phương sai có sai số thay đổi 25](#_Toc219333075)

[12. Kiểm định hiện tương quan: 26](#_Toc219333076)

[13. Đánh giá mô hình: 27](#_Toc219333077)

[14. Khắc phục mô hình nếu có 27](#_Toc219333078)

[VII. ĐÁNH GIÁ LẠI MÔ HÌNH HỒI QUY CUỐI CÙNG. 27](#_Toc219333079)

[1. Chọn mô hình tốt nhất trong 4 dạng 27](#_Toc219333080)

[VIII. THẢO LUẬN VÀ KIẾN NGHỊ 30](#_Toc219333081)

[2. Kiến nghị quản trị 30](#_Toc219333082)

[VIII. Cải tiến tối ưu MH2: 31](#_Toc219333083)

[IX. KẾT LUẬN 34](#_Toc219333084)

[1. Tổng kết quá trình thực hiện 34](#_Toc219333085)

[2. Các phát hiện chính 34](#_Toc219333086)

[3. Ý nghĩa thực tiễn và hạn chế: 34](#_Toc219333087)

# 

# MỞ ĐẦU

## Lý do chọn đề tài

Trong những năm gần đây, thị trường rượu vang trên thế giới phát triển mạnh mẽ với sự đa dạng về chất lượng, xuất xứ và mức giá. Giá rượu vang không chỉ phản ánh chi phí sản xuất mà còn chịu tác động bởi các yếu tố như điểm đánh giá của chuyên gia và người tiêu dùng, độ tuổi của rượu, mức độ phổ biến trên thị trường cũng như quốc gia sản xuất. Việc phân tích định lượng các yếu tố này có ý nghĩa quan trọng đối với người tiêu dùng trong việc lựa chọn sản phẩm, đồng thời hỗ trợ nhà sản xuất và nhà phân phối trong chiến lược định giá.   
Kinh tế lượng cung cấp các công cụ khoa học để lượng hóa mức độ ảnh hưởng của từng yếu tố đến giá rượu vang thông qua các mô hình hồi quy và hệ thống kiểm định thống kê. Do đó, đề tài “Phân tích các yếu tố ảnh hưởng đến giá rượu vang” được lựa chọn nhằm vận dụng lý thuyết kinh tế lượng vào một bộ dữ liệu thực tế, qua đó đánh giá mức độ phù hợp của các dạng mô hình hồi quy khác nhau.

## Mục tiêu nghiên cứu

• Xác định các yếu tố có ảnh hưởng đáng kể đến doanh thu.

• Đo lường mức độ tác động thông qua các dạng hàm khác nhau (Lin-Lin, Lin-Log, Log-Lin, Log-Log).

• Kiểm định các giả định kinh tế lượng để đảm bảo tính tin cậy của mô hình.

• Lựa chọn mô hình tối ưu và đề xuất kiến nghị quản trị.

## Câu hỏi nghiên cứu

• Yếu tố nào tác động mạnh nhất đến giá rượu vang?

• Quốc gia nào có xu hướng giá rượu cao hơn trong mẫu nghiên cứu?

# MÔ TẢ DỮ LIỆU

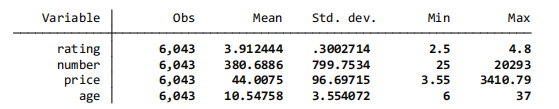
## Nguồn và đặc điểm dữ liệu

Bộ dữ liệu được sử dụng trong nghiên cứu gồm 6.043 quan sát, được xử lý bằng phần mềm Stata. Dữ liệu phản ánh thông tin về giá rượu vang và một số đặc điểm liên quan đến chất lượng, mức độ phổ biến và xuất xứ quốc gia. Đây là dữ liệu chéo (cross–section), do đó không đặt nặng vấn đề tự tương quan theo thời gian.

## Mô tả biến

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TÊN BIẾN** | **VAI TRÒ** | **LOẠI BIẾN** | **MÔ TẢ** | **ĐƠN VỊ** |
| price | Biến phụ thuộc | Định lượng | Giá rượu vang | EUR |
| rating | Biến độc lập | Định lượng | Điểm đánh giá trung bình của rượu vang | Điểm |
| number | Biến độc lập | Định lượng | Số lượng lượt đánh giá của rượu vang | Lượt đánh giá |
| age | Biến độc lập | Định lượng | Độ tuổi của rượu vang | Năm |
| country | Biến độc lập | Định lượng | Quốc gia sản xuất rượu vang:  Nếu country = 1 (nếu là nước Pháp), country = 2 (nếu là nước Ý), country = 3 (nếu là nước Tây Ban Nha) | Biến giả (1/2/3) |

### Mô tả biến định lượng



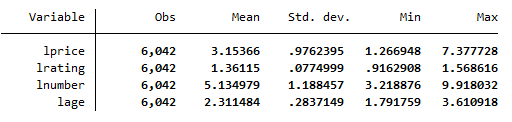
**Biến "rating" (Điểm đánh giá):** Có số lượng quan sát là 6.043 chai với giá trị trung bình là 3,91 điểm. Điểm đánh giá thấp nhất là 2,5 và cao nhất là 4,8. Điều này cho thấy chất lượng các chai rượu trong mẫu khá đồng đều và ở mức tốt, không có chai nào bị đánh giá quá thấp.

**Biến "number" (Số lượng đánh giá):** Có 6.043 quan sát với giá trị trung bình là 380,69 lượt đánh giá mỗi chai. Giá trị nhỏ nhất là 25 và lớn nhất lên tới 20.293 lượt. Độ lệch chuẩn rất cao (799,75) cho thấy sự chênh lệch cực kỳ lớn về mức độ phổ biến hoặc độ nhận diện giữa các dòng rượu trên thị trường.

**Biến "price" (Giá rượu):** Có giá trị trung bình là 44,01 EUR với mức giá thấp nhất là 3,55 EUR và cao nhất là 3.410,79 EUR. Sự khác biệt khổng lồ giữa giá trị nhỏ nhất và lớn nhất cùng độ lệch chuẩn cao (96,69) phản ánh mẫu dữ liệu bao gồm cả những chai rượu bình dân lẫn những chai rượu cao cấp, xa xỉ.

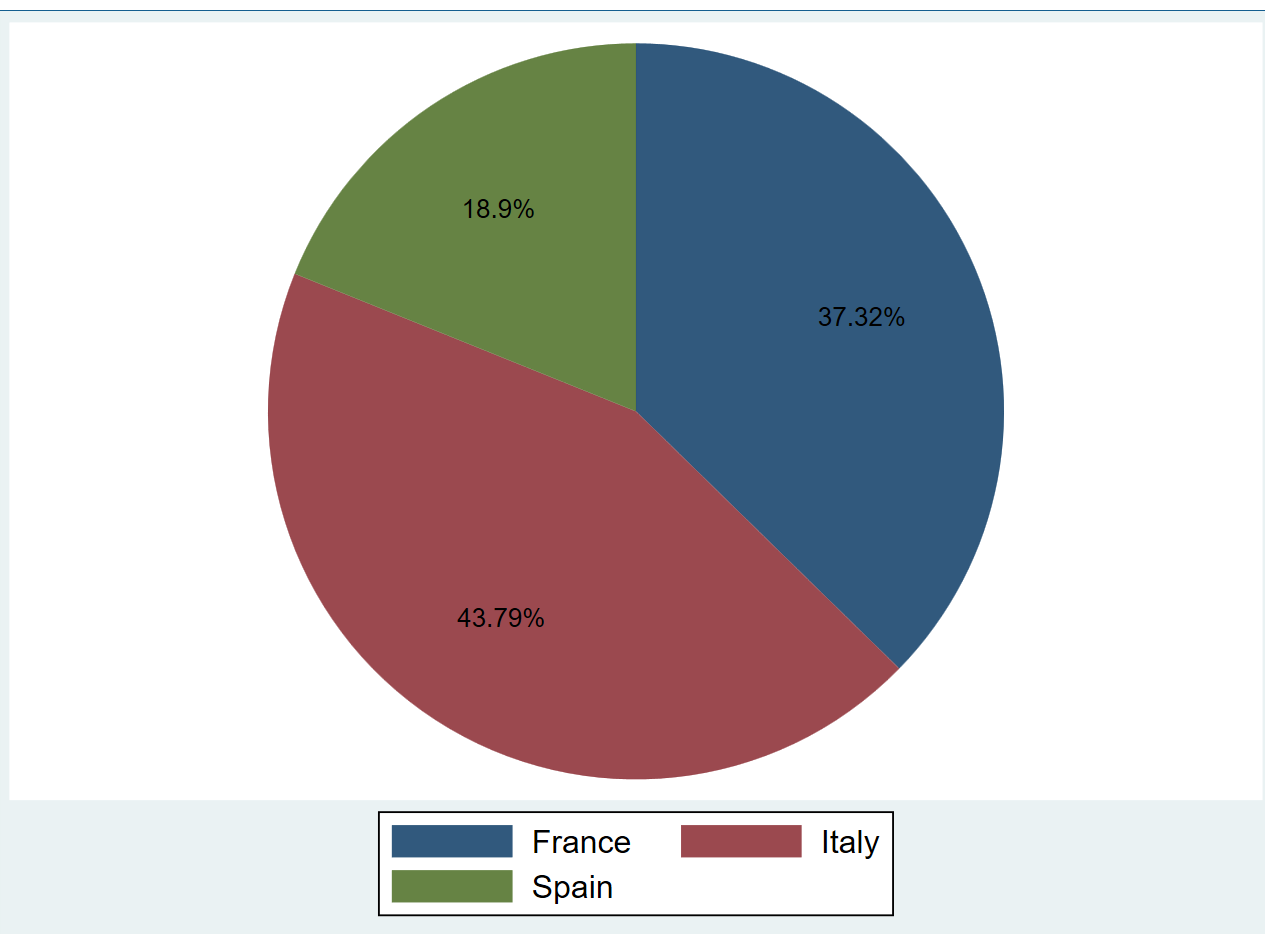
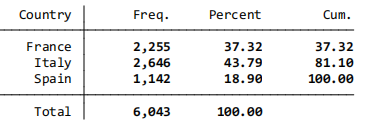
**Biến "age" (Số năm ủ/Tuổi rượu):** Có trung bình là 10,55 năm với giá trị nhỏ nhất là 6 năm và lớn nhất là 37 năm. Chỉ số này cho thấy tập dữ liệu tập trung vào các dòng rượu đã có thời gian trưởng thành nhất định, từ những chai rượu trẻ cho đến những chai có tuổi đời lâu năm và giá trị sưu tầm cao.

Các biến logarit được: rating, number, price và age trong đó:



### Mô tả biến định tính

**Biến Country**

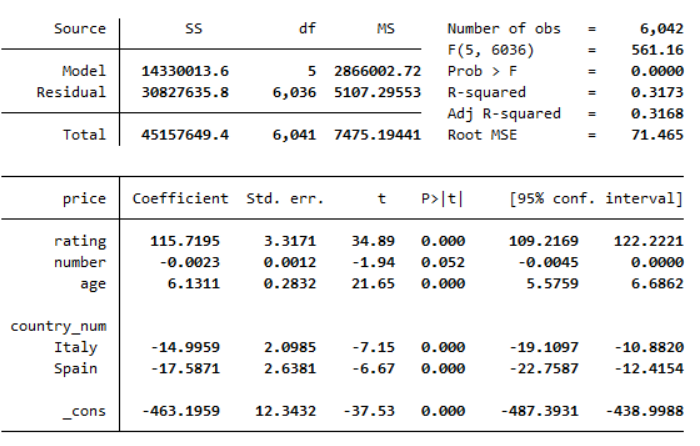
Biến quốc gia (country) có sự phân bố trong mẫu :

* Rượu đến từ Pháp (France =1): Chiếm 2,255 chai, tương đương 37,32% trên tổng số rượu là 6,943 chai.
* Rượu đến từ Ý (Italy = 1): Chiếm số lượng nhiều nhất với 2,646 chai, chiếm tỷ lệ 43.79%.
* Rượu đến từ Tây Ban Nha (Spain =1): Có số lượng ít nhất với 1.142 chai, chiếm 18.9%.

→ Kết quả thống kê mô tả cho thấy cơ cấu sản phẩm theo quốc gia có sự tập trung rõ rệt. Trong đó, rượu Ý là dòng sản phẩm chủ đạo với 2.646 chai (43,79%). Tiếp theo là rượu Pháp với 2.255 chai (37,32%). Thấp nhất là rượu Tây Ban Nha chỉ với 1.142 chai, tương ứng 18,9% tổng số chai rượu được khảo sát.

# MÔ HÌNH 1: DẠNG TUYẾN TÍNH – TUYẾN TÍNH

## Chạy mô hình hồi quy mẫu



**Phương trình hồi quy mẫu SRF:**

= −480.5589 + 120.7570⋅rating − 0.0025.number + 6.0480.age − 16.4884.Italy − 18.5874⋅

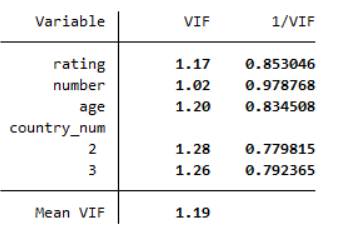
**Theo dữ liệu gồm 6,043 quan sát, với điều kiện các yếu tố khác không đổi:**

* rating = 120.7570: Khi điểm đánh giá của rượu tăng thêm 1 điểm, thì giá rượu trung bình tăng khoảng 120.7570 EUR. Hệ số này có ý nghĩa thống kê (p-value = 0.000), cho thấy điểm đánh giá có tác động cùng chiều và mạnh đến giá rượu.
* number = –0.0025: Khi số lượng đánh giá tăng thêm 1 lượt, thì giá rượu trung bình giảm khoảng 0.0025 EUR, trong điều kiện các yếu tố khác không đổi. Tuy nhiên, hệ số này không có ý nghĩa thống kê ở mức 5% (p-value = 0.064), do đó tác động của biến này chưa đủ mạnh để khẳng định chắc chắn.
* age = 6.0480: Khi tuổi của rượu tăng thêm 1 năm, giá rượu trung bình tăng khoảng 6.0480 EUR. Hệ số này có ý nghĩa thống kê ở mức 1%, phản ánh rượu có tuổi đời cao thường có giá trị lớn hơn.
* Italy (–16.4884): Trong điều kiện các yếu tố khác không đổi, rượu có xuất xứ từ Italy có giá trung bình thấp hơn khoảng 16.4884 EUR so với rượu từ France (nhóm tham chiếu). Sự khác biệt này có ý nghĩa thống kê.
* Spain ( –18.5874): Tương tự, rượu có xuất xứ từ Spain có giá trung bình thấp hơn khoảng 18.5874 EUR so với rượu từ France, khi các yếu tố khác không đổi. Hệ số này có ý nghĩa thống kê ở mức 1%.

**Dựa vào giá trị P-value:**

* T Các biến độc lập bao gồm: rating (đánh giá), age (tuổi), và các biến giả quốc gia Italy, Spain đều có P-value = 0.000 < 0.05. Điều này cho thấy các biến này đều có ý nghĩa thống kê và ảnh hưởng đến biến phụ thuộc là giá (price).
* Biến number có P-value = 0.052. Vì 0.052 > α = 0.05, nên ở mức ý nghĩa 5%, biến này không có ý nghĩa thống kê.
* Mô hình tổng thể có Prob > F = 0.0000, cho thấy mô hình phù hợp và có ý nghĩa thống kê.

## Kiểm định đa cộng tuyến và loại bớt biến đa cộng tuyến



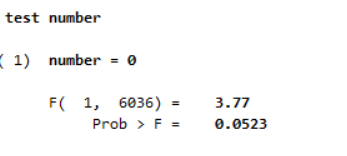
**Nhận xét:** Tất cả VIF < 10. Mô hình không xảy ra hiện tượng đa cộng tuyến.

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

## Kiểm Định Thừa Biến

P > |t| number = 0.052 > 0.05, nên biến number là biến bị coi thừa, không có ý nghĩa thống kê trong mô hình này ở mức ý nghĩa 5%.

Tiến hành kiểm tra thừa biến:

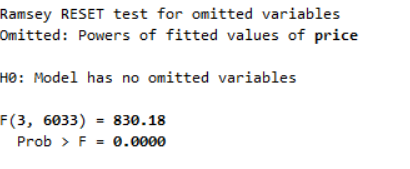


Kết luận: Ở mức tin cậy 95%, biến number là biến thừa trong mô hình này vì có Prob > F = 0.0523 > 0.05.

## Kiểm Tra sót biến

**Giả thuyết:** H0: Mô hình không bỏ sót biến hoặc dạng hàm đúng

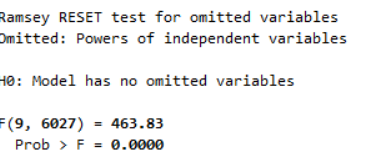
**Kết quả kiểm định:**



Ta có: P-value = 0.000 < α = 0,05 => 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**



Ta có: P-value = 0.000 < α = 0,05 => Bác bỏ H0

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

## Ý nghĩa hệ số r2 và kiểm định mô hình

* **Hệ số xác định R²** (R-squared = 0.3173=31.73%)
* **Ý nghĩa:** Khoảng 31.73% sự biến động của giá (price) được giải thích bởi các biến độc lập trong mô hình (rating, number, age, country). Còn lại khoảng 68.27% là do các yếu tố khác ngoài mô hình và sai số ngẫu nhiên.
* Kiểm Định mô hình:
  + - * **H0:** R2 = 0 (Mô hình không phù hợp)
      * **H1:** R2 >0 (Mô hình phù hợp)

Ta có p-value là giá trị xác suất khi F lớn 561.16 (Prob > F = 0.0000).

p-value = 0.0000< α=0.05 => Bác bỏ H0.

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

## Xác định khoảng tin cậy và giải thích ý nghĩa

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

* *β*rating  ∈(109.2169;122.2221): Khi điểm đánh giá tăng thêm 1 điểm, giá rượu trung bình tăng trong khoảng từ 109.2169 đến 122.2221 EUR.
* number ∈(-0.0045;0.0000): Khi Số lượng lượt đánh giá của rượu vang tăng thêm 1 lượt đánh giá, thì trung bình giá giảm từ 0.0045 đến 0.
* *β*age  ∈(5.5759;6.6862):Khi tuổi tăng 1 năm, mức giá rượu trung bình của các quốc gia tăng từ 5.58 đến 6.69 EUR.
* βItaly ∈ (−19.1097;−10.8820): Tại Italy, mức giá trung bình thấp hơn rượu đến từ Pháp (France) từ 10.88 đến 19.11 EUR.
* *βSpain* ∈ (−22.7587;−12.4154): Tại Spain, mức giá trung bình thấp hơn rượu đến từ Pháp (France) từ 12.42 đến 22.76 EUR.

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

**Giả thuyết:** H0: = 0 (Các biến rating, number, country, age không ảnh hưởng đến biến price)

|  |  |  |
| --- | --- | --- |
| **Tên biến** | **p\_value** | Chấp nhận/Bác bỏ giả thuyết H0 |
| Rating | 0.000 < α = 0.05 | Bác bỏ |
| Age | 0.000 < α = 0.05 | Bác bỏ |
| Number | 0.052 > α = 0.05 | Chấp nhận H0 |
| Country | 0.000 < α = 0.05 | Bác bỏ |

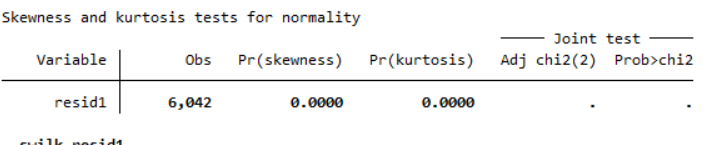
**Kết luận**:  Các biến rating, age, Italy, Spain đều ảnh hưởng tới biến phụ thuộc price. Riêng biến number chưa đủ cơ sở khẳng định tác động ở mức ý nghĩa 5%.

## 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:

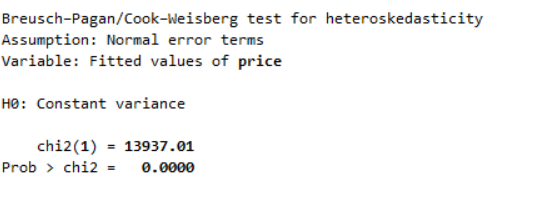


Ta có: **P- value= (Prob > chi2 )** = 0.00000 < α = 0.05 => Bác bỏ Ho.

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

## Kiểm định hiện tượng phương sai có sai số thay đổi

* Kiểm định giả thuyết:
* H0: Phương sai sai số không đổi
* H1: Phương sai sai số thay đổi
* Kết quả kiểm định:

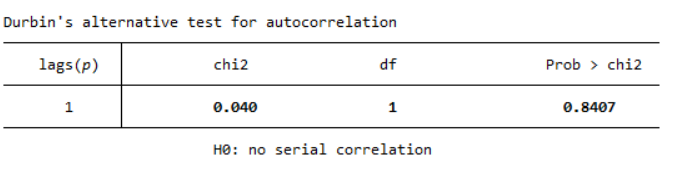


Ta có: P-value = 0.0000 < α= 0,05 => Bác bỏ H0.

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

## Kiểm định hiện tượng tương quan

* Kiểm định giả thuyết:
* H0: không có tự tương quan
* H1: 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

****

Ta có: P-value =0.0000 > α = 0,05 => Chấp nhận H0.

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

## Khắc phục mô hình nếu có

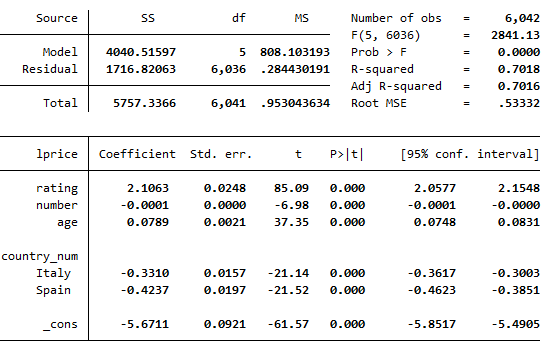
* Dùng lệnh regress price rating number age i.country\_num, vce(robust) để khắc phục hiện tương tương quan sai số thay đổi của mô hình.

## Đánh giá mô hình hồi quy

* Về ý nghĩa thống kê của các biến: Các biến độc lập rating (điểm đánh giá), age (tuổi rượu) và biến giả quốc gia (Italy, Spain) đều có ý nghĩa thống kê ở mức 5% (P-value = 0.000). Tuy nhiên, biến number (số lượng) có P-value = 0.052 > 0.05, do đó không có ý nghĩa thống kê ở mức ý nghĩa 5%.
* Mức độ giải thích của mô hình: Các biến độc lập trong mô hình giải thích được 31.73% sự biến động của giá rượu.
* Đa cộng tuyến: Mô hình không xảy ra hiện tượng đa cộng tuyến (tất cả VIF < 10, Mean VIF = 1.19).
* Dạng hàm/Sót biến: Kiểm định Ramsey RESET cho thấy mô hình bị bỏ sót biến quan trọng hoặc dạng hàm sai (P-value = 0.0000).
* Phân phối chuẩn của sai số: Sai số ngẫu nhiên của mô hình không có phân phối chuẩn.
* Phương sai sai số: Mô hình có hiện tượng phương sai của sai số thay đổi.
* Tự tương quan: Mô hình không xảy ra hiện tượng tự tương quan

# MÔ HÌNH 2: DẠNG LOGARIT – TUYẾN TÍNH

## Chạy mô hình hồi quy mẫu



Dựa vào giá trị P-value:

* Tất cả các biến độc lập bao gồm: rating (đánh giá), number (số lượng), age (tuổi), và các biến giả quốc gia Italy, Spain đều có P-value = 0.000 < 0.05.
* Điều này khẳng định rằng tất cả các biến đưa vào mô hình đều có ý nghĩa thống kê ở mức ý nghĩa 5% và có ảnh hưởng thực sự đến logarit giá bán (lprice).

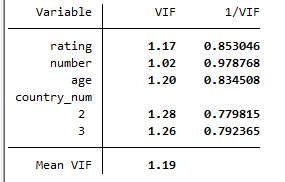
**Giải thích ý nghĩa các hệ số hồi quy mẫu**

Dựa trên bảng kết quả hồi quy, phương trình hồi quy có dạng:

= −5.6711 + 2.1063.rating − 0.0001.number + 0.0789.age − 0.3310.Italy − 0.4237.Spain

* Theo dữ liệu gồm **6,042** chai rượu, với điều kiện các yếu tố khác không đổi:
* **βrating** = 2.1063: Khi điểm đánh giá tăng 1 điểm, giá bán trung bình tăng khoảng **210.63%.**
* **βnumber =** −0.0001: Khi số lượng tăng 1 lượt đánh giá, giá bán trung bình giảm khoảng **0.01%.**
* **βage = 0.0789:** Khi tuổi của sản phẩm tăng 1 năm, giá bán trung bình tăng khoảng **7.89%.**
* **βItaly = −0.3310:** Sản phẩm tại Italy có mức giá trung bình thấp hơn khoảng **33.10%** so với nhóm tham chiếu.
* **ΒSpain = -0.4237:** Sản phẩm tại Spain có mức giá trung bình thấp hơn khoảng **42.37%** so với nhóm tham chiếu.

## Kiểm định đa cộng tuyến



Tất cả các chỉ số VIF<10

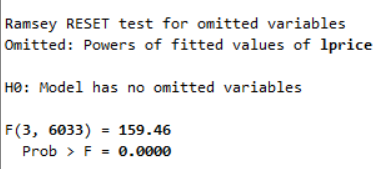
**Kết luận:** Không xảy ra hiện tượng đa cộng tuyến.

## Kiểm tra thừa biến

Không có thừa biến nên không cần kiểm định biến thừa

## Kiểm tra sót biến:

* H0: Mô hình không bỏ sót biến hoặc dạng hàm đúng.
* H1: Mô hình bỏ sót biến hoặc dạng hàm sai

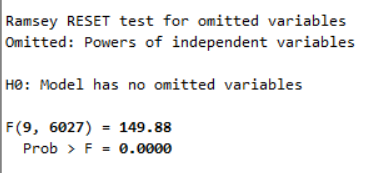


Ta có: F(3, 6033) = 159.46; Prob > F = 0.0000.

Vì P-value = 0.0000 < α = 0.05 => 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

****

Vì: p-value = 0.0000 < α=0.05 => Bác bỏ H0.

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

## Ý nghĩa hệ số R2 và kiểm định sự phù hợp của mô hình

* **Hệ số xác định R2 = 0.7018 = 70.18%:** Khoảng 70.18% sự biến động của logarit giá (lprice) được giải thích bởi các biến độc lập trong mô hình. Đây là mức giải thích khá cao, tốt hơn nhiều so với mô hình tuyến tính ban đầu (31.73%).
* **Kiểm định F (Sự phù hợp của mô hình):**
  + **H0:** R2 = 0 (Mô hình không phù hợp)
  + **H1:** R2 >0 (Mô hình phù hợp)

Ta có Prob > F = 0.0000 < α = 0.05 => Bác bỏ H0.

* + **Kết luận:** **Mô hình phù hợp dữ liệu với độ tin cậy 95%**

## Xác định khoảng tin cậy 95% và giải thích ý nghĩa

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

* βrating ∈ (2.0577;2.1548): Khi rating tăng 1 điểm, mức giá trung bình tăng từ 205.77% đến 215.48%
* βnumber ∈ (-0.0001; 0.0000): Khi lượt đánh giá đi từ âm đến 0, nên không có ý nghĩa trên thực tế.
* βage ∈ (0.0748;0.0831): Khi tuổi tăng 1 năm, mức giá trung bình tăng từ 7.48% đến 8.31%.
* βItaly ∈ (−0.3617;−0.3003): Tại Italy, mức giá trung bình thấp hơn rượu đến từ Pháp (France) từ 30.03% đến 36.17%.
* βSpain ∈ (−0.4623;−0.3851): Tại Spain, mức giá trung bình thấp hơn rượu đến từ Pháp (France) từ 38.51% đến 46.23%.

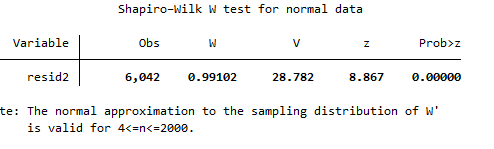
## Kiểm định tác động từng biến độc lập lên biến phụ thuộc

H0: Các biến độc lập không ảnh hưởng đến lprice () = 0

|  |  |  |
| --- | --- | --- |
| Tên biến | P\_value | Chấp nhận/Bác bỏ giả thuyết H0 |
| rating | 0.000 < α = 0.005 | Bác bỏ |
| number | 0.000 < α = 0.005 | Bác bỏ |
| age | 0.000 < α = 0.005 | Bác bỏ |
| Italy | 0.000 < α = 0.005 | Bác bỏ |
| Spain | 0.000 < α = 0.005 | Bác bỏ |

**Kết luận**: Các biến độc lập rating, number, Italy, Spain đều ảnh hưởng tới biến phụ thuộc price.

## Kiểm định sai số có phân phối chuẩn



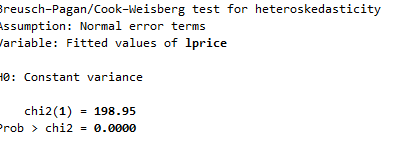
Sử dụng kiểm định Shapiro-Wilk và kiểm định Skewness/Kurtosis:

* Kiểm định Shapiro-Wilk: Prob > z = 0.00000.
* Kiểm định Skewness/Kurtosis: Prob > chi2 = 0.0000.

Ta có P-value < 0.05 => Bác bỏ H0.

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

## Kiểm định hiện tượng phương sai sai số thay đổi



Sử dụng kiểm định Breusch-Pagan/Cook-Weisberg:

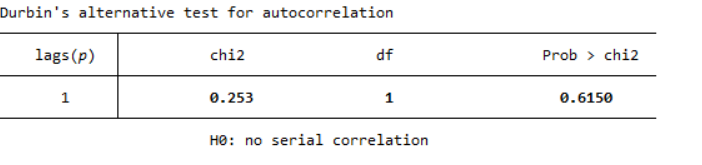
* H0: Phương sai sai số không đổi.
* H1: Phương sai sai số thay đổi.
* Ta có: P-value (Prob > chi2) = 0.0000 < α = 0.05 => Bác bỏ H0.

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

## Kiểm định hiện tượng tương quan

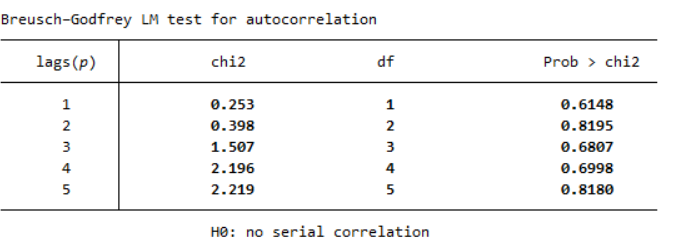
**Kiểm định giả thuyết:**

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



Ta có: P-value =0.0000 > α = 0,05 => Chấp nhận H0.

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

****

Theo quan sát: Tất cả các giá trị p-value từ bậc 1 đến bậc 5 đều nhỏ hơn mức ý nghĩa α=0,05.

**Kết luận: Mô hình không xảy ra hiện tượng tự tương quan từ bậc 1 đến bậc 5. (đạt BLUE).**



* H0: Không có tự tương quan bậc 1.
* H1: Có tự tương quan bậc 1.

Giá trị thống kê Durbin-Watson

**Kết luận:** **Kết luận: Không xảy ra hiện tượng tự tương quan bậc 1**.

## Đánh giá chung về mô hình hồi quy

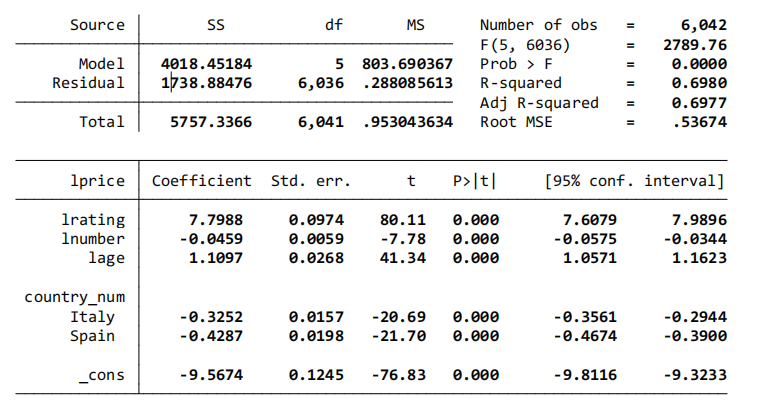
* Về ý nghĩa thống kê của các biến: Tất cả các biến độc lập bao gồm rating (điểm đánh giá), number (số lượng), age (tuổi rượu) và các biến giả quốc gia (Italy, Spain) đều có ý nghĩa thống kê ở mức 5% (P-value = 0.000). Biến number đã trở nên có ý nghĩa thống kê rất cao.
* Các biến độc lập trong mô hình giải thích được 70.18% sự biến động của Logarit giá rượu (lprice).
* Đa cộng tuyến: Mô hình không xảy ra hiện tượng đa cộng tuyến (Tất cả VIF < 10, Mean VIF = 1.19).
* Tự tương quan: Mô hình không xảy ra hiện tượng tự tương quan.
* Dạng hàm: Kiểm định Ramsey RESET cho thấy mô hình vẫn bị bỏ sót biến quan trọng hoặc dạng hàm sai.
* Phương sai sai số: Mô hình có hiện tượng phương sai của sai số thay đổi

## Khắc phục mô hình

Chạy lệnh: regress lprice rating number age i.country\_num, vce(robust) để khắc phục hiện tượng phương sai thay đổi, giúp kết quả đáng tin cậy hơn.

# MÔ HÌNH 3: MÔ HÌNH HỒI QUY (LOG-LOG)

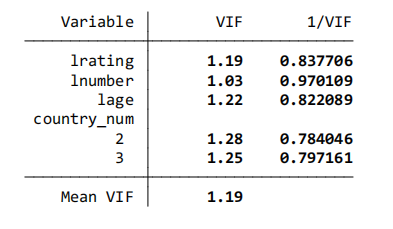
## Chạy hồi quy mẫu và giải thích ý nghĩa các hệ số hồi quy mẫu



**Dựa vào P-value:**

* Các biến lrating,lnumber, lage, country\_num (Italy,Spain) có P-value = 0.0000 <= 0.05 -> có ý nghĩa thông 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 rượu vang.

## Kiểm tra đa cộng tuyến



Tất cả các chỉ số VIF>10

Kết Luận: Mô hình không xảy ra hiện tượng đa cộng tuyến

## Kiểm định biến thừa

Không có thừa biến nên không cần kiểm định biến thừa

## Giải thích ý nghĩa các hệ số hồi quy mẫu

Hàm hồi quy Mẫu (SRF):

-9.5674 + 7.7988 -0.0459lnumber + 1.1097lage – 0.3252Italy – 0.4287Spain

Ý nghĩa: Các hệ số này cho biết sự thay đổi của biến phụ thuộc khi biến độc lập thay đổi, trong điều kiện các yếu tố khác không đổi.

* 7.7988: Khi đánh giá tăng 1%, thì giá bán trung bình của rượu vang khoảng 7.79%
* -0.0459: Khi lượt đánh giá tăng 1%, thì giá bán trung bình của rượu vang giảm khoảng 0.045%
* 1.1097: Khi độ tuổi của rượu tăng 1%, thì giá bán trung bình của rượu vang tăng khoảng 1.11%.
* = -0.3252: Sản phẩm tại Italy có mức giá bán trung bình thấp hơn 32.52% so với các nhóm tham chiếu
* -0.4287: Sản phẩm tại Spain có mức giá bán trung bình của rượu thấp hơn 42.87% so với các nhóm tham chiếu

## Ý nghĩa hệ số R2 và kiểm định mô hình

R2=0.6980=69.80%

Ý nghĩa: cho biết 69.80% sự biến thiên của logarit giá rượu vang (lprice) được giải thích bởi các biến lrating, lnumber,lage có trong mô hình,. Còn lại 30.20% do sự tác động của các yếu tố khác bên ngoài mô hình hoặc do sai số ngẫu nhiên.

## Kiểm định mô hình:

* H0: R2=0 (mô hình không phù hợp)
* H1: R2>0 (mô hình phù hợp)

Ta có p-value là giá trị xác suất khi F lớn 2789.76(Prob>0.0000)

p-value = 0.0000<a=0.05=> bác bỏ H0

**Kết luận: mô hình phù hợp (ĐTC 95 %)**

## Xác định khoảng tin cậy và giải thích ý nghĩa

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

* lrating [7.6079, 7.9896] : khin rating tăng 1%, giá sẽ tăng trong khoảng từ 7.61% đến 7.99%.
* lnumber [-0.0575, -0.0344] : khi số lượt đánh giá tăng 1%, giá sẽ giảm trong khoảng từ 0.034% đến 0.058%.
* lage [1.0571, 1.1623] : khi số tuổi của rượu vang tăng 1%, giá sẽ tăng trong khoảng từ 1.06% đến 1.16%.
* Italy [-0.3561, -0.2944] : So với nhóm cơ sở, giá tại Ý thấp hơn trong khoảng từ 29.4% đến 35.6%.
* Spain [-0.4674, -0.3900] : So với nhóm cơ sở, giá tại Tây Ban Nha thấp hơn trong khoảng từ 39.0% đến 46.7%.

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

Giả thuyết: H0: = 0 (Các biến age,number,rating,Country\_num(Italy,Spain) không ảnh hưởng đến biến price)

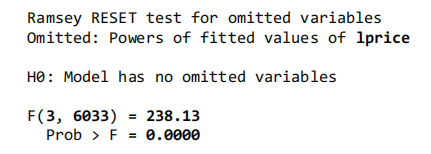
|  |  |  |
| --- | --- | --- |
| **Tên biến** | **p\_value** | Chấp nhận/Bác bỏ giả thuyết H0 |
| age | 0.000 < α = 0.05 | Bác bỏ |
| Rating | 0.000 < α = 0.05 | Bác bỏ |
| Number | 0.000 < α = 0.05 | Bác bỏ |
| Italy | 0.000 < α = 0.05 | Bác bỏ |
| Spain | 0.000 < α = 0.05 | Bác bỏ |

**Kết luận:** các biến age, rating, number, Italy, Spain đều ảnh hưởng đến biến phụ thuộc Price

## Kiểm định sót biến

Giả thuyết:

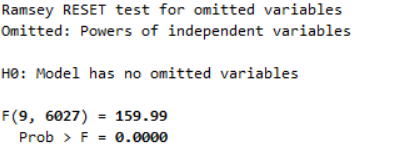
H0: Mô hình không bỏ sót biến hoặc dạng hàm đúng



Ta có P-value=0.0000< =0.05 => 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

****

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

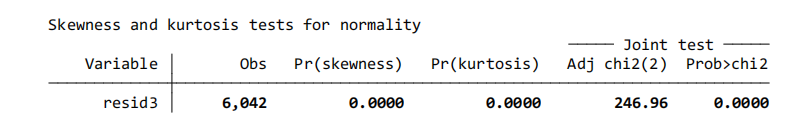
**Kết luận: Mô hình có bỏ sót biến quan trọng do định dạng sai biến độc lập.**

## 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:**



Ta có: **P- value= (Prob > chi2) = 0.0000 < α = 0.05** => Bác bỏ H0.

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

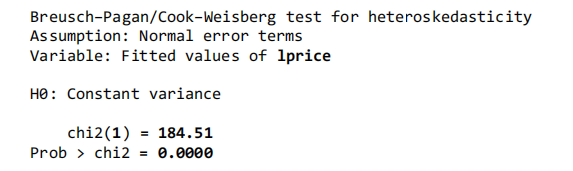
## Kiểm định hiện tượng phương sai có sai số thay đổi

**Kiểm định giả thuyết:**

**H0:** Phương sai sai số không đổi

**H1:** Phương sai sai số thay đổi

**Kết quả kiểm định:**



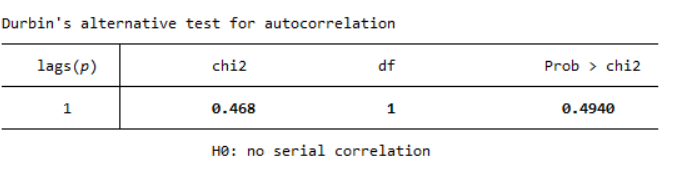
Ta có: P-value = 0.0000 < α= 0,05 => **Bác bỏ H0**.

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

## Kiểm định hiện tượng tương quan:

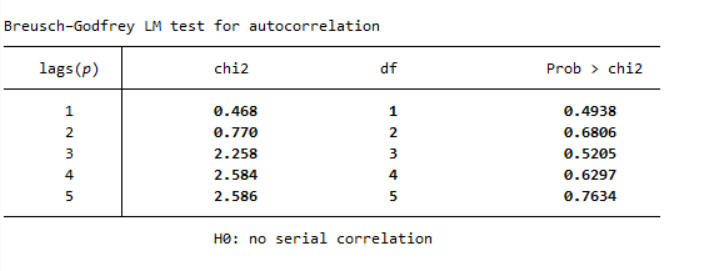
**Kiểm định giả thuyết:**

* H0: không có tự tương quan
* H1: có tự tương quan.
* Kết quả kiểm định



Ta có: P-value =0.0000 > α = 0,05 => Chấp nhận H0.

**Kết luận: Mô hình không có xảy ra hiện tượng tự tương quan ( đạt BLUE).**

****

**Kết luận: Mô hình không xảy ra hiện tượng tự tương quan từ bậc 1 đến bậc 5. (đạt BLUE).**

## Đánh giá mô hình hồi quy:

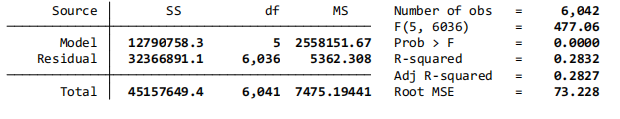
* Về ý nghĩa thống kê của các biến: Tất cả các biến độc lập bao gồm lrating (log điểm đánh giá), lnumber (log số lượng), lage (log tuổi rượu) và các biến giả quốc gia (Italy, Spain) đều có P-value = 0.000 < 0.05.
* Các biến độc lập trong mô hình giải thích được 69.80% sự biến động của logarit giá rượu.
* Mô hình không xảy ra hiện tượng đa cộng tuyến. Tất cả các hệ số VIF đều rất nhỏ (< 2) và Mean VIF = 1.19 < 10.
* Mô hình vẫn bị bỏ sót biến quan trọng hoặc dạng hàm sai.
* Sai số ngẫu nhiên không có phân phối chuẩn.
* Mô hình có hiện tượng phương sai của sai số thay đổi.
* Mô hình không xảy ra hiện tượng tự tương quan.

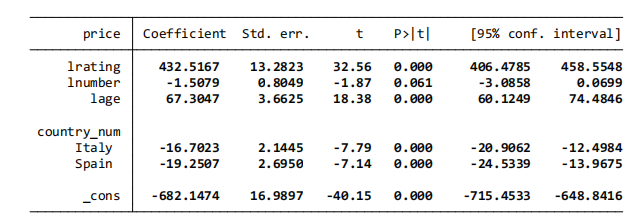
## Khắc phục mô hình nếu có:

Không có

# MÔ HÌNH 4: MÔ HÌNH HỒI QUY TUYẾN TÍNH- LOGARIT

## Chạy mô hình hồi quy mẫu

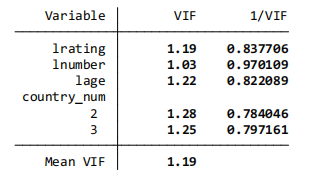




**Dựa vào giá trị P-value:**

* Các biến lrating,lnumber, lage, country\_num (Italy,Spain) có P-value = 0.0000 <= 0.05 -> có ý nghĩa thông 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 rượu vang.

## Kiểm định đa cộng tuyến



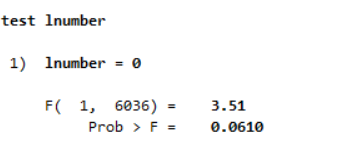
Tất cả các chỉ số VIF<10

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

## Kiểm định thừa biến

Vì P-value = 0.061> 0.05. Nên biến lnumber là biến thừa ở mức tin cậy 95%.

Tiến hành kiểm định biến lnumber:



Biến lnumber có Prob > F = 0.0610 > 0.05 nên lnumber là biến thừa trong mô hình này ở mức tin cậy 95%.

## Giải thích ý nghĩa các hệ số hồi quy mẫu

Hàm số hồi quy mẫu (SRF):

Price = -682,1474 + 432,5167.lrating - 1,5079.lnumber + 67,3047.lage - 16,7023.Italy - 19,2507.Spain

## Giải thích ý nghĩa hệ số hồi quy mẫu

Trong mô hình Log-Log, các hệ số của biến độc lập dạng log chính là hệ số co giãn.

* lrating (432,5167): Khi rating tăng 1%, thì giá (price) trung bình tăng khoảng 432,52%. (Lưu ý: Hệ số này cực kỳ lớn, trong thực tế có thể biến price ở đây đơn thuần là dạng mức, nếu là dạng mức thì khi lrating tăng 1 đơn vị, giá tăng 432,5 EUR).
* lnumber (-1,5079): Khi number tăng 1%, thì giá (price) trung bình giảm khoảng 1,51%. Tuy nhiên, hệ số này có P>|t| = 0,061 (> 0,05), nên nó không có ý nghĩa thống kê ở mức ý nghĩa 5%.
* lage (67,3047): Khi age tăng 1%, thì giá (price) trung bình tăng khoảng 67,3%.
* Biến giả (Italy & Spain):
  + Italy (-16,7023): Ở cùng điều kiện, giá tại Ý thấp hơn nhóm cơ sở khoảng 16,7 EUR (do Italy không phải dạng log).
  + Spain (-19,2507): Ở cùng điều kiện, giá tại Tây Ban Nha thấp hơn nhóm cơ sở khoảng 19,25 EUR.

## Ý nghĩa hệ số R2 và kiểm định mô hình:

* R2 = 0.2832 = 28.32%

Ý nghĩa:các biến độc lập trong mô hình logarit giá rượu vang (lprice) lrating, lnumber,lage giải thích đươc 28.32% sự biến động của giá rượu vang, Còn lại 71.68%% do sự tác động của các yếu tố khác bên ngoài mô hình hoặc do sai số ngẫu nhiên.

**Kiểm định mô hình:**

* H0: R2 = 0 (Mô hình không phù hợp)
* H1: R2 > 0 (Mô hình phù hợp)

Ta có p-value là giá trị xác suất khi F lớn 7739.62 (Prob>F=0.0000) p-value = 0.0000 < α=0.05 => Bác bỏ H0.

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

## 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:

βlrating ∈(406,4785 ; 458,5548): khi lượng đánh giá tăng 1%, giá sẽ tăng trong khoảng từ 406,48% đến 458,55%.

βlnumber ∈(-3,0858 ; 0,0699): Khoảng này chứa giá trị 0, củng cố việc biến này không có ý nghĩa thống kê ở mức 5%.

βlage ∈(−233.8229;−102.2487): khi độ tuổi của rượu tăng 1%, giá sẽ tăng trong khoảng từ 60,12% đến 74,48%.

βItaly ∈(−8,130;−7,450): giá tại Ý thấp hơn nhóm cơ sở từ 12,5 đến 20,9 đơn vị.

βSpain ∈(4,993.9164;5,579.1149): Giá tại Tây Ban Nha thấp hơn nhóm cơ sở từ 14,0 đến 24,5 đơn vị.

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

Giả thuyết: H0: = 0 (Các biến age,number,rating,Country\_num(Italy,Spain) không ảnh hưởng đến biến price)

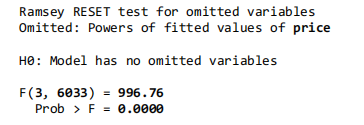
|  |  |  |
| --- | --- | --- |
| **Tên biến** | **p\_value** | Chấp nhận/Bác bỏ giả thuyết H0 |
| age | 0.000 < α = 0.05 | Bác bỏ |
| Rating | 0.000 < α = 0.05 | Bác bỏ |
| Number | 0.061 > α = 0.05 | Chấp nhận H0 |
| Italy | 0.000 < α = 0.05 | Bác bỏ |
| Spain | 0.000 < α = 0.05 | Bác bỏ |

**Kết luận**:  Các biến rating, age, Italy, Spain đều ảnh hưởng tới biến phụ thuộc price. Riêng biến number chưa đủ cơ sở khẳng định tác động ở mức ý nghĩa 5%.

## Kiểm định bỏ sót biến

**Giả thuyết:** H0:Mô hình không bỏ sót biến hoặc dạng hàm đúng

**Kết quả kiểm định:**



Vì: **p-value =0.0000 >** α =**0.05 => H0.**

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

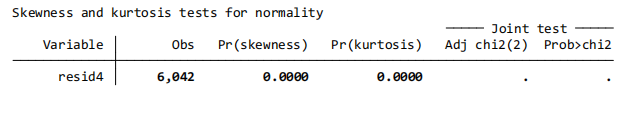
## Kiểm định sai số có phân phối chuẩn

**Giả thuyết:**

**H0:** Sai số có phân phối chuẩn.

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

**Kết quả kiểm định:**



Ta có: **P- value= (Prob > chi2) = 0.00000 < α = 0.05** => Bác bỏ H0.

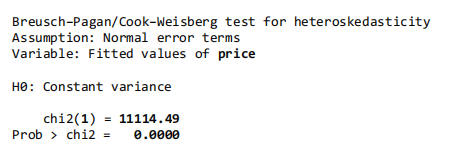
**Kết luận: Sai số ngẫu nhiên của mô hình không có phân phối chuẩn**.

## Kiểm định hiện tượng phương sai có sai số thay đổi

**Kiểm định giả thuyết:**

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

**Kết quả kiểm định:**



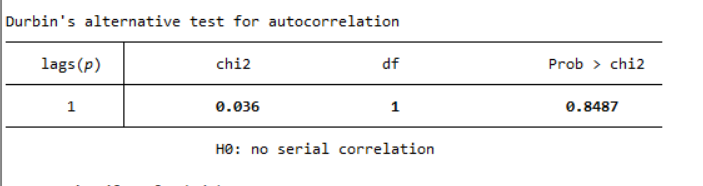
Ta có: P-value = 0.0000 < α= 0,05 => **Bác bỏ H0**.

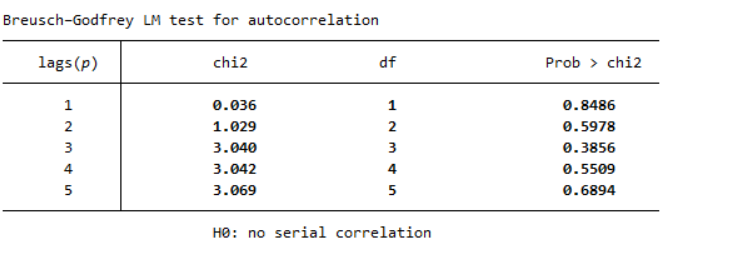
**Kết luận: Phương sai sai số thay đổi. (Không đạt BLUE).**

## Kiểm định hiện tương quan:

**Kiểm định giả thuyết:**

* H0: không có tự tương quan
* H1: có tự tương quan.
* Ta có: P-value =0.0000 > α = 0,05 => Chấp nhận H0.
* **Kết luận: Mô hình không có xảy ra hiện tượng tự tương quan ( đạt BLUE).**





**Kết luận: Mô hình không xảy ra hiện tượng tự tương quan từ bậc 1 đến bậc 5. (đạt BLUE).**

## Đánh giá mô hình:

* Về ý nghĩa thống kê của các biến: Các biến lrating (log điểm đánh giá), lage (log tuổi rượu) và biến giả quốc gia (Italy, Spain) đều có ý nghĩa thống kê ở mức 5% (P-value = 0.000). Biến lnumber (log số lượng) có P-value = 0.061, biến này không có ý nghĩa thống kê.
* Các biến độc lập trong mô hình chỉ giải thích được 28.32% sự biến động của giá rượu.
* Mô hình không xảy ra hiện tượng đa cộng tuyến. Tất cả các chỉ số VIF đều thấp (< 2).
* Mô hình định dạng sai biến phụ thuộc hoặc bỏ sót biến quan trọng.
* Sai số ngẫu nhiên không có phân phối chuẩn.
* Mô hình có hiện tượng phương sai của sai số thay đổi.
* Mô hình không xảy ra hiện tượng tự tương quan.

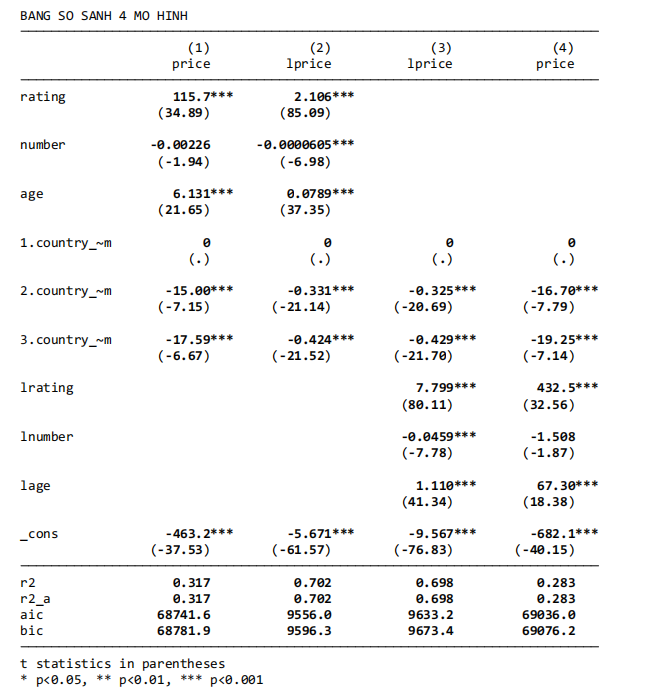
## Khắc phục mô hình nếu có

Không có

# ĐÁNH GIÁ LẠI MÔ HÌNH HỒI QUY CUỐI CÙNG.

## Chọn mô hình tốt nhất trong 4 dạng

Từ 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:



|  |  |
| --- | --- |
| **Tiêu Chí** | **MH: 2**  **L-T** |
| **R2**  ( lớn nhất ) | 0.702 |
| **R2**  ( lớn nhất ) | 0.702 |
| **Tiêu chí AIC**  ( nhỏ nhất ) | 68741.6 |
| **Tiêu chí BIC**  ( nhỏ nhất) | 68781.9 |

**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:

* **R2** cao: 0.702 Þ **mô hình có tính thích hợp.**
* hiệu chỉnh cao: 0.702
* 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 (70.2**%**) sự biến động của giá rượu 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 (9556.0)** và **BIC (9596.3)** thấp nhất trong 4 mô hình (thấp hơn rất nhiều so với MH1 là 68741.6), khẳng định lựa chọn MH2 là tối ưu nhất để giải thích sự biến động của giá rượu dựa trên bộ dữ liệu 6.042 quan sát.

Kết luận: MH2 được chọn làm mô hình cuối cùng vì không chỉ có khả năng giải thích tốt hơn (R² gấp đôi mô hình tuyến tính) mà còn đáp ứng tốt hơn các tiêu chí lựa chọn mô hình thống kê.

## VIII. THẢO LUẬN VÀ KIẾN NGHỊ

**Thảo luận kết quả nghiên cứu:**

Dựa vào MH2 được lựa chọn, nhóm nhận thấy rằng các yếu tố có ảnh hưởng đến dự đoán giá rượu (Price) bao gồm:

* Điểm đánh giá (rating): Tác động mạnh mẽ nhất. Hệ số hồi quy là 2.106, cho thấy khi chất lượng rượu được đánh giá tăng thêm 1 điểm, giá rượu trung bình tăng rất mạnh (khoảng 210.6%). Điều này phản ánh đúng thực tế trong thị trường rượu vang cao cấp: chỉ cần nhích lên một vài điểm trong thang đánh giá (ví dụ từ 90 lên 95), giá trị thương hiệu và đẳng cấp của chai rượu tăng lên theo cấp số nhân.
* **Tuổi rượu (age):** Tác động tích cực. Hệ số hồi quy là **0.0789**, cho thấy khi rượu được ủ thêm 1 năm (tuổi rượu tăng 1), giá rượu trung bình tăng khoảng **7.89%**. Điều này phù hợp với quy luật của rượu vang: rượu càng lâu năm (Vintage) thường càng quý và đắt đỏ do chi phí lưu kho và giá trị thời gian.
* **Xuất xứ (Country):** Với nhóm tham chiếu là **Pháp (France)** (Mã 1):
  + **Italy (Mã 2):** Hệ số **-0.331**, cho thấy rượu vang Ý có mức giá trung bình thấp hơn khoảng **33.1%** so với rượu vang Pháp cùng phẩm cấp.
  + **Spain (Mã 3):** Hệ số **-0.424**, cho thấy rượu vang Tây Ban Nha có mức giá trung bình thấp hơn khoảng **42.4%** so với rượu vang Pháp.
* Kết luận: Điều này cho thấy thị trường vẫn định giá thương hiệu rượu vang Pháp ở vị thế "chiếu trên" (premium) so với các đối thủ từ Ý và Tây Ban Nha.
* **Số lượng (number):** Hệ số là **-0.0000605**, tác động tiêu cực nhưng rất nhỏ (gần bằng 0). Tuy nhiên nó có ý nghĩa thống kê, cho thấy các loại rượu được sản xuất đại trà với số lượng lớn (hoặc số lượng mua quá nhiều, đại diện cho độ phổ biến đại chúng) có xu hướng giá thấp hơn một chút so với các loại rượu hiếm.

**Như vậy**, Điểm đánh giá (rating) và Xuất xứ (quốc gia) là những yếu tố quan trọng nhất định hình nên giá của một chai rượu.

## Kiến nghị quản trị

Từ kết quả phân tích định lượng của Mô hình 2 (MH2), nhóm nghiên cứu đề xuất các kiến nghị quản trị thực tiễn nhằm tối ưu hóa giá trị và lợi nhuận cho các bên tham gia thị trường rượu vang:

**Đối với nhà sản xuất:**

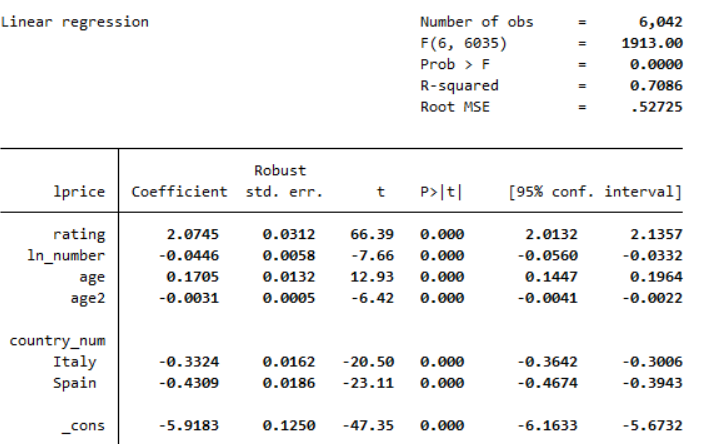
* **Chiến lược "Chất lượng hơn Số lượng" (Quality over Quantity):**  Nhà sản xuất nên tập trung nguồn lực vào việc cải tiến quy trình canh tác nho và kỹ thuật lên men để đạt điểm số cao từ các chuyên gia, thay vì chạy theo sản lượng đại trà. Một sự gia tăng nhỏ trong điểm số chất lượng có thể mang lại mức tăng giá bán theo cấp số nhân, hiệu quả hơn nhiều so với việc cố gắng tăng độ phủ thị trường bằng các sản phẩm giá rẻ.
* **Tối ưu hóa chiến lược ủ rượu (Aging Strategy):**Các nhà làm rượu cần tính toán điểm cân bằng (Break-even point) giữa chi phí lưu kho (bảo quản, vốn ứ đọng) và mức tăng giá trị kỳ vọng (7.89%/năm). Đối với các dòng vang có tiềm năng lão hóa tốt, việc giữ lại trong hầm rượu để bán ra sau 5-10 năm là một chiến lược đầu tư sinh lời hiệu quả.

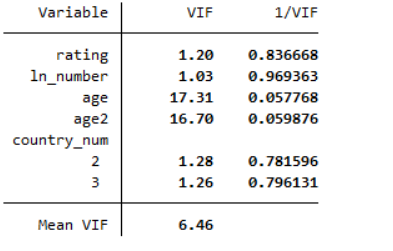
Đối với nhà nhập khẩu và phân phối:

* **Chiến lược đa dạng hóa danh mục dựa trên xuất xứ:**
  + Phân khúc cao cấp (Premium/Luxury): Nên ưu tiên nhập khẩu rượu vang Pháp để khai thác "hiệu ứng xuất xứ" (Country-of-Origin Effect), phục vụ nhóm khách hàng thượng lưu, ít nhạy cảm về giá và chú trọng thương hiệu.
  + Phân khúc tầm trung/giá trị (Value-for-money): Nên tập trung vào rượu vang Ý và Tây Ban Nha. Đây là lựa chọn chiến lược để cạnh tranh về giá, thu hút khách hàng phổ thông muốn trải nghiệm chất lượng tốt với chi phí hợp lý hơn.
* **Marketing dựa trên bằng chứng:** Vì rating là yếu tố quyết định giá, các nhà phân phối cần làm nổi bật điểm số đánh giá (nếu cao) ngay trên nhãn chai, kệ hàng hoặc các ấn phẩm truyền thông. Việc giáo dục khách hàng về ý nghĩa của các con số rating sẽ giúp củng cố niềm tin và chấp nhận mức giá cao hơn.

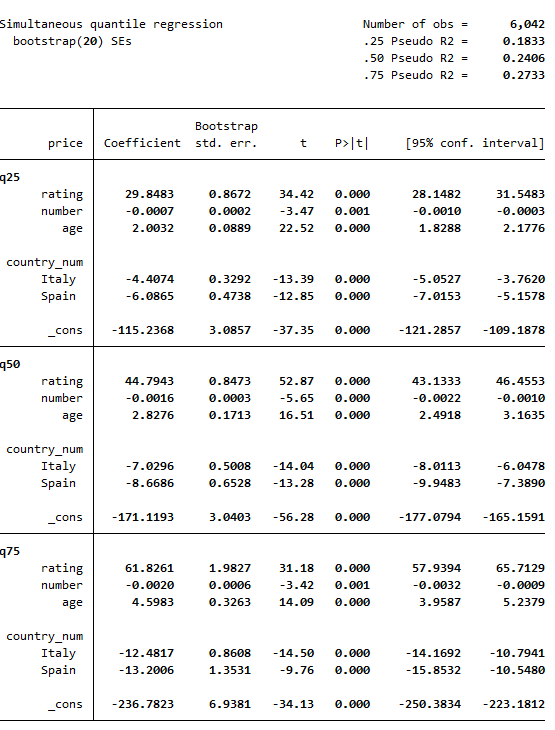
# Cải tiến tối ưu MH2:

Sau khi chạy lệnh: regress lprice rating ln\_number age age2 i.country\_num, robust. Thu được kết quả:





Mô hình có vẻ rất tốt: Không còn biến thừa, hệ số R2 = 0.7086 nhưng hai biến là age và age2 bị đa công tuyến. Vì thế em đã dùng mô hình Hồi quy phân vị (QREG) để tối ưu và được kết quả.



Kết quả cho thấy mô hình hồi quy phân vị vượt trội hơn, mọi biến đều có ý nghĩa thống kê. Có thể rút ra các nhận xét:

* Điểm số quan trọng hơn rất nhiều đối với rượu cao cấp. Người mua rượu đắt tiền cực kỳ nhạy cảm với chất lượng/điểm số (tác động gấp đôi so với rượu rẻ).
* Rượu càng cao cấp thì yếu tố thời gian ủ (tuổi rượu cao) càng tạo ra nhiều giá trị gia tăng.
* Mô hình giải thích tốt dần lên khi giá tăng.

# KẾT LUẬN

## Tổng kết quá trình thực hiện

Đề tài đã hoàn thành mục tiêu phân tích các yếu tố tác động đến giá rượu vang dựa trên bộ dữ liệu thực nghiệm gồm 6.042 quan sát. Thông qua quy trình kiểm định nghiêm ngặt và so sánh giữa 4 dạng mô hình (Tuyến tính, Log-Lin, Log-Log, Lin-Log), nhóm nghiên cứu đã xác định Mô hình Logarit – Tuyến tính (MH2) là mô hình tối ưu nhất.

## Các phát hiện chính

Mô hình được lựa chọn (MH2) có hệ số xác định R2 = 70,2%. Nghĩa là các biến độc lập giải thích được tới 70.2% sự biến thiên của giá rượu. Kết quả hồi quy mang lại những bằng chứng định lượng quan trọng.

## Ý nghĩa thực tiễn và hạn chế:

Mặc dù mô hình còn tồn tại một số hạn chế về phân phối chuẩn của sai số (điều thường gặp ở các bộ dữ liệu lớn), nhưng các khuyết tật về phương sai thay đổi đã được khắc phục bằng phương pháp sai số chuẩn vững (Robust Standard Errors).