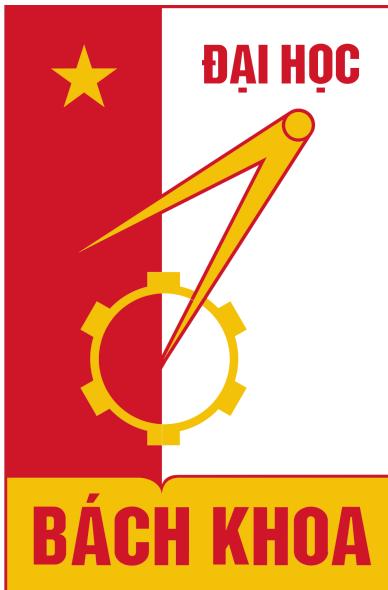


TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI  
KHOA TOÁN – TIN



TIÊU LUẬN CUỐI KỲ  
HỌC PHẦN SUY LUẬN THỐNG KÊ

CHỦ ĐỀ

ĐỊNH GIÁ Ô TÔ THEO CÔNG NGHỆ VÀ THƯƠNG HIỆU  
PHÂN TÍCH HỒI QUY TUYẾN TÍNH

Giảng viên hướng dẫn: PGS.TS. Nguyễn Thị Thu Thủy

Sinh viên thực hiện: Nguyễn Ngọc Bích

MSSV: 20237418

Mã học phần: MI3031

Mã lớp học: 163641

HÀ NỘI – 2025

## LỜI CẢM ƠN

Lời đầu tiên, em xin gửi lời cảm ơn chân thành và sâu sắc nhất tới cô **Nguyễn Thị Thu Thủy**, giảng viên phụ trách học phần *Suy luận thống kê*. Nhờ sự tận tâm trong giảng dạy cùng những hướng dẫn nhiệt tình của cô, em đã tiếp thu và học được rất nhiều kiến thức quan trọng, tạo nền tảng để em hoàn thành báo cáo này.

Mặc dù đã cố gắng hoàn thiện, bài báo cáo chắc chắn không tránh khỏi những thiếu sót. Em kính mong nhận được sự góp ý của cô để bài làm được hoàn thiện hơn.

Một lần nữa, em xin gửi lời cảm ơn chân thành tới cô. Kính chúc cô luôn mạnh khỏe, hạnh phúc và tiếp tục đạt được nhiều thành công trong sự nghiệp giảng dạy cũng như nghiên cứu khoa học.

Trân trọng.

Hà Nội, tháng 12 năm 2025

Tác giả báo cáo

**Nguyễn Ngọc Bích**

# Mục lục

Lời cảm ơn . . . . .	2
<b>Chương 1. BÀI TOÁN ĐẶT RA . . . . .</b>	<b>5</b>
1.1 Bối cảnh . . . . .	5
1.2 Câu hỏi đặt ra . . . . .	5
1.3 Định hướng nghiên cứu . . . . .	6
1.4 Phạm vi và cách tiếp cận . . . . .	6
<b>Chương 2. CƠ SỞ LÝ THUYẾT . . . . .</b>	<b>7</b>
2.1 Hàm hồi quy lý thuyết . . . . .	7
2.2 Mô hình hồi quy tuyến tính bội . . . . .	7
2.3 Các giả thiết cơ bản của mô hình . . . . .	8
2.4 Phương pháp ước lượng bình phương nhỏ nhất . . . . .	8
2.5 Đánh giá và kiểm định mô hình hồi quy . . . . .	9
2.6 Đánh giá mức độ phù hợp của mô hình . . . . .	10
<b>Chương 3. DỮ LIỆU VÀ CÁC BIẾN TRONG MÔ HÌNH . . . . .</b>	<b>11</b>
3.1 Thông tin bộ dữ liệu . . . . .	11
3.2 Xử lý dữ liệu và tạo biến . . . . .	11
3.3 Biến phụ thuộc . . . . .	12
3.4 Các biến độc lập . . . . .	13
3.5 Trực quan hóa dữ liệu . . . . .	13
<b>Chương 4. MÔ HÌNH HỒI QUY TUYẾN TÍNH BỘI . . . . .</b>	<b>16</b>
4.1 Mô hình hồi quy tuyến tính bội cơ bản . . . . .	16
4.2 Mô hình hồi quy mở rộng với biến thương hiệu (Make) . . . . .	17
4.3 Ước lượng hệ số hồi quy trong mô hình hồi quy bội . . . . .	17
4.4 Kiểm định ý nghĩa của hồi quy . . . . .	19
4.5 Hệ số xác định và hệ số xác định hiệu chỉnh . . . . .	20
4.6 Kiểm định giả thuyết về các hệ số hồi quy riêng . . . . .	21
4.7 So sánh vai trò của yếu tố thương hiệu . . . . .	23
4.8 Phân tích phần dư . . . . .	23
<b>Chương 5. KẾT LUẬN . . . . .</b>	<b>26</b>
5.1 Kết quả đạt được . . . . .	26
5.2 Ý nghĩa và đóng góp . . . . .	26
5.3 Hạn chế và hướng nghiên cứu tiếp theo . . . . .	27
<b>TÀI LIỆU THAM KHẢO . . . . .</b>	<b>27</b>

# Danh sách hình vẽ

3.2	Phân phối của biến log(MSRP) . . . . .	13
3.3	Mối quan hệ giữa ln(MSRP) và Engine HP . . . . .	14
3.4	Mối quan hệ giữa ln(MSRP) và Year . . . . .	14
3.5	Ma trận tương quan giữa các biến . . . . .	15
4.1	Đồ thị phần dư theo giá trị dự đoán (Residuals vs Fitted) . . . . .	24
4.2	Phân phối của phần dư . . . . .	24
4.3	Q-Q plot của Phần dư . . . . .	25

# Danh sách bảng

3.1	Mô tả các biến số trong tập dữ liệu . . . . .	11
4.1	Bảng phân tích phương sai (ANOVA) - Mô hình cơ sở . . . . .	20
4.2	Bảng phân tích phương sai (ANOVA) - Mô hình mở rộng . . . . .	20
4.3	Kết quả kiểm định . . . . .	22
4.4	Kết quả kiểm định - Mô hình mở rộng . . . . .	22
4.5	So sánh hệ số EV/Hybrid và $R^2$ giữa mô hình cơ sở và mô hình mở rộng . . . . .	23

# Chương 1

## BÀI TOÁN ĐẶT RA

### 1.1 Bối cảnh

Thị trường ô tô toàn cầu đang dần phục hồi sau đại dịch, tuy nhiên tốc độ tăng trưởng có xu hướng chậm lại và chuyển trọng tâm sang các dòng xe điện (EV) và xe Hybrid. Trong bối cảnh cạnh tranh ngày càng gay gắt, đặc biệt thông qua các cuộc điều chỉnh giá, chiến lược định giá trở thành một công cụ quan trọng để các hãng xe định vị sản phẩm và định hình kỳ vọng của người tiêu dùng đối với công nghệ mới.

Trong bối cảnh đó, giá bán lẻ đề xuất (MSRP) không chỉ phản ánh chi phí sản xuất và biên lợi nhuận, mà còn hàm chứa các yếu tố chiến lược như định vị thương hiệu, phân khúc thị trường và kỳ vọng về giá trị công nghệ. Đặc biệt đối với các dòng xe EV và hybrid, giá bán lẻ đề xuất (MSRP) thường cao hơn so với xe sử dụng động cơ đốt trong truyền thống (ICE), tạo ra một mức giá chênh lệch đáng kể.

Tuy nhiên, bản chất của mức giá chênh lệch này vẫn chưa được làm rõ. Liệu sự chênh lệch giá chủ yếu xuất phát từ các yếu tố kỹ thuật và công nghệ có thể quan sát và đo lường được (công suất của động cơ, loại nhiên liệu mà động cơ sử dụng,...), hay phần lớn phản ánh chiến lược định vị và giá trị thương hiệu của nhà sản xuất? Việc phân tích cấu trúc định giá bán lẻ đề xuất (MSRP) của EV/Hybrid so với ICE, sau khi đã kiểm soát các yếu tố kỹ thuật và phân khúc, là cần thiết để làm rõ logic cạnh tranh và chiến lược định giá của các hãng xe trong giai đoạn chuyển dịch sang công nghệ xanh.

### 1.2 Câu hỏi đặt ra

Nghiên cứu này nhằm trả lời câu hỏi trọng tâm:

**“Các nhà sản xuất ô tô đang định vị và định giá cho công nghệ Hybrid và EV như thế nào thông qua giá bán lẻ đề xuất (MSRP), và cấu trúc giá đó được hình thành chủ yếu bởi yếu tố công nghệ thực hay bởi chiến lược thương hiệu và marketing?”**

Để làm rõ câu hỏi trên, nghiên cứu được cụ thể hóa thành hai câu hỏi sau:

- **Câu 1:** Liệu có tồn tại một mức chênh lệch giá có ý nghĩa thống kê trong giá bán lẻ đề xuất (MSRP) dành cho xe Hybrid và EV so với xe xăng truyền thống, sau khi đã kiểm soát các yếu tố về hiệu năng và đặc tính kỹ thuật cơ bản?
- **Câu 2:** Trong mô hình hồi quy MSRP, mức độ đóng góp của công nghệ truyền động (Hybrid, EV) và các yếu tố thương hiệu (được đại diện bằng biến giả thương hiệu) vào mức chênh lệch giá ra sao?

### 1.3 Định hướng nghiên cứu

Dựa trên các câu hỏi đặt ra, báo cáo này tập trung vào ba mối quan hệ chính giữa các yếu tố kỹ thuật, thương hiệu và giá niêm yết của xe ô tô:

- **Ảnh hưởng công nghệ:** Xe Hybrid và EV có mức MSRP cao hơn xe xăng truyền thống một cách có ý nghĩa thông kê, ngay cả sau khi đã kiểm soát yếu tố kỹ thuật cơ bản.
- **Vai trò thương hiệu:** Biến thương hiệu có ảnh hưởng đáng kể và độc lập đến MSRP, cho thấy yếu tố thương hiệu đóng vai trò quan trọng trong cấu trúc định giá, bên cạnh các yếu tố kỹ thuật và công nghệ.
- **Phân tầng công nghệ:** Giá của xe điện (EV) cao hơn đáng kể so với xe Hybrid, phản ánh sự khác biệt trong chiến lược định giá giữa hai loại công nghệ điện hóa.

### 1.4 Phạm vi và cách tiếp cận

- **Đối tượng nghiên cứu:** giá bán lẻ đề xuất (MSRP) của các mẫu ô tô mới trên thị trường Mỹ, bao gồm ba nhóm công nghệ truyền động: xe xăng (ICE), xe Hybrid và xe điện thuần túy (EV).
- **Cách tiếp cận:** Báo cáo sử dụng mô hình hồi quy tuyến tính để phân tích tác động của các nhóm yếu tố đến MSRP, bao gồm đặc tính kỹ thuật và hiệu năng cơ bản, công nghệ truyền động, và yếu tố thương hiệu. Mục tiêu là tách biệt và định lượng mức chênh lệch giá của xe.

# Chương 2

## CƠ SỞ LÝ THUYẾT

### 2.1 Hàm hồi quy lý thuyết

Cơ sở của mô hình hồi quy tuyến tính bội bắt đầu từ giả định rằng giá trị trung bình của biến ngẫu nhiên  $Y$  có mối quan hệ tuyến tính với các biến  $x_1, x_2, \dots, x_k$ . Mỗi quan hệ này được mô tả thông qua hàm hồi quy lý thuyết:

$$E(Y | x_1, x_2, \dots, x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k. \quad (2.1)$$

Trong đó:

- $\beta_0$  là hệ số chặn, biểu thị giá trị trung bình của  $Y$  khi tất cả các biến độc lập  $x_j$  bằng 0;
- $\beta_j$  ( $j = 1, \dots, k$ ) là các hệ số hồi quy riêng, phản ánh mức thay đổi kỳ vọng của  $Y$  khi biến  $x_j$  tăng một đơn vị, trong điều kiện các biến độc lập khác được giữ không đổi.

Hàm hồi quy lý thuyết chỉ mô tả xu hướng trung bình của  $Y$  và không phản ánh đầy đủ các biến động ngẫu nhiên trong dữ liệu thực tế.

### 2.2 Mô hình hồi quy tuyến tính bội

Trong thực tế, các giá trị quan sát  $y$  không bao giờ nằm chính xác trên siêu phẳng hồi quy lý thuyết do sự tác động của các yếu tố ngoài mô hình. Vì vậy, giá trị thực tế của  $Y$  được xác định bằng hàm giá trị trung bình cộng với một sai số ngẫu nhiên  $\epsilon$ .

Dạng tổng quát của mô hình hồi quy tuyến tính bội được biểu diễn như sau:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon = \beta_0 + \sum_{j=1}^k \beta_j x_j + \epsilon \quad (2.2)$$

trong đó:

- $\beta_0$  là hệ số chặn, biểu thị giá trị kỳ vọng của  $Y$  khi tất cả các biến độc lập bằng 0;
- $\beta_j$  ( $j = 1, 2, \dots, k$ ) là các hệ số hồi quy riêng, phản ánh mức thay đổi trung bình của  $Y$  khi biến  $x_j$  tăng một đơn vị, trong điều kiện các biến độc lập khác được giữ không đổi;
- $\epsilon$  là sai số ngẫu nhiên, đại diện cho các yếu tố không quan sát được hoặc không được đưa vào mô hình.

## 2.3 Các giả thiết cơ bản của mô hình

Để việc ước lượng và suy luận thống kê có ý nghĩa, mô hình hồi quy tuyến tính bội dựa trên các giả thiết cơ bản sau:

- Mẫu dữ liệu  $(\mathbf{X}, \mathbf{y})$  là mẫu ngẫu nhiên
- Sai số ngẫu nhiên có kỳ vọng bằng 0:  $E(\epsilon) = 0$ ;
- Sai số có phương sai không đổi:  $Var(\epsilon) = \sigma^2$ ;
- Các sai số không có tương quan với nhau;
- Không tồn tại hiện tượng đa cộng tuyến hoàn hảo giữa các biến độc lập, hay nói cách khác, ma trận  $\mathbf{X}^T \mathbf{X}$  khả nghịch.

## 2.4 Phương pháp ước lượng bình phương nhỏ nhất

Để ước lượng các hệ số  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  trong mô hình hồi quy bội, ta sử dụng phương pháp bình phương nhỏ nhất.

Nguyên lý: tìm ra bộ giá trị ước lượng  $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)$  sao cho tổng bình phương của các sai số (phản dư) giữa giá trị thực tế quan sát được và giá trị dự đoán từ mô hình là nhỏ nhất:

$$L = \sum_{i=1}^n e_i^2. \quad (2.3)$$

Giả sử có  $k$  biến hồi quy và  $n > k$  quan sát,  $(x_{i1}, x_{i2}, \dots, x_{ik}; y_i)$ ,  $i = 1, \dots, n$ . Ta có mô hình:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \epsilon_i$$

Hệ phương trình này có thể biểu diễn dưới dạng ma trận như sau:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Trong đó:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

- $\mathbf{y}$  là một vectơ  $n \times 1$  của các quan sát,
- $\mathbf{X}$  là một ma trận  $n \times (k + 1)$  của các mức của các biến độc lập,
- $\boldsymbol{\beta}$  là một vectơ  $(k + 1) \times 1$  của các hệ số hồi quy,
- $\boldsymbol{\epsilon}$  là một vectơ  $n \times 1$  của các sai số ngẫu nhiên.

Ước lượng bình phương nhỏ nhất của vector tham số được xác định bởi:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (2.4)$$

## 2.5 Đánh giá và kiểm định mô hình hồi quy

Sau khi ước lượng các tham số của mô hình hồi quy tuyến tính bội bằng phương pháp bình phương tối thiểu, cần tiến hành đánh giá và kiểm định mô hình nhằm xác định mức độ phù hợp cũng như ý nghĩa thống kê của các biến giải thích. Các kiểm định và chỉ tiêu đánh giá dưới đây được sử dụng phổ biến trong phân tích hồi quy.

### Kiểm định ý nghĩa tổng thể của mô hình

Kiểm định Fisher được sử dụng để xác định sự tồn tại của mối quan hệ tuyến tính giữa biến phụ thuộc  $Y$  và tập hợp các biến độc lập  $x_1, x_2, \dots, x_k$  trong mô hình hồi quy.

#### Quy trình kiểm định

##### 1. Phát biểu cặp giả thiết cần kiểm định:

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0, \\ H_1 : \exists \beta_j \neq 0, \quad j = 1, \dots, k. \end{cases}$$

##### 2. Chọn tiêu chuẩn kiểm định:

$$F_0 = \frac{SS_R/k}{SS_E/(n-p)} = \frac{MS_R}{MS_E}$$

Nếu giả thiết  $H_0$  là đúng thì  $F_0$  có phân phối  $F$  với  $\nu_1 = k$  và  $\nu_2 = n - p$  bậc tự do.

##### 3. Tính giá trị quan sát của tiêu chuẩn kiểm định:

$$f_0 = \frac{SS_R/k}{SS_E/(n-p)} = \frac{MS_R}{MS_E}$$

Trong đó:

$$\begin{aligned} SST &= \mathbf{y}^T \mathbf{y} - \frac{(\sum_{i=1}^n y_i)^2}{n} \\ SSR_R &= \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y} - \frac{(\sum_{i=1}^n y_i)^2}{n} \\ SSE &= \mathbf{y}^T \mathbf{y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y} \end{aligned}$$

##### 4. Tìm miền bác bỏ $H_0$ :

$$W_\alpha = \{f_0 \mid f_0 > f_\alpha(k, n-p)\}$$

$f_\alpha(k, n-p)$  được xác định từ bảng giá trị tới hạn phân phối  $F$  với  $\nu_1 = k$  và  $\nu_2 = n - p$  bậc tự do.

##### 5. Kết luận

- Nếu  $f_0 \in W_\alpha$  thì bác bỏ giả thuyết  $H_0$
- Nếu  $f_0 \notin W_\alpha$  thì chưa có cơ sở để bác bỏ giả thuyết  $H_0$

### Kiểm định ý nghĩa của từng hệ số hồi quy

Sau khi mô hình tổng thể được xác nhận là có ý nghĩa, cần tiếp tục kiểm định vai trò của từng biến độc lập trong mô hình khi các biến còn lại đã được đưa vào.

#### 1. Phát biểu cặp giả thuyết cần kiểm định:

$$\begin{cases} H_0 : H_0 : \beta_j = \beta_{j0}, \\ H_1 : H_1 : \beta_j \neq \beta_{j0} \end{cases}$$

#### 2. Tiêu chuẩn kiểm định:

$$T_0 = \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{\sigma^2 C_{jj}}}$$

Trong đó:  $C_{jj}$  là phần tử trên đường chéo chính của ma trận  $(\mathbf{X}^T \mathbf{X})^{-1}$

#### 3. Tính giá trị quan sát của tiêu chuẩn kiểm định:

$$t_0 = \frac{\hat{\beta}_j - \beta_{j0}}{se(\hat{\beta}_j)}$$

Nếu giả thuyết  $H_0 : \beta_j = \beta_{j0}$  là đúng thì  $T_0 \sim t(n-p)$

#### 4. Tìm miền bắc bỏ giả thiết $H_0$ :

$$W_\alpha = \left\{ t_0 \in \mathbb{R} \mid |t_0| > t_{\alpha/2, n-p} \right\}$$

#### 5. Kết luận:

- Nếu  $t_0 \in W_\alpha$  thì bác bỏ giả thuyết  $H_0$
- Nếu  $t_0 \notin W_\alpha$  thì chưa có cơ sở để bác bỏ giả thuyết  $H_0$

## 2.6 Đánh giá mức độ phù hợp của mô hình

Mức độ phù hợp của mô hình hồi quy tuyến tính bội được đánh giá thông qua các chỉ tiêu sau:

**Hệ số xác định ( $R^2$ )** Hệ số xác định  $R^2$  đo lường tỷ lệ biến thiên của biến phụ thuộc được giải thích bởi mô hình:

$$R^2 = \frac{SS_R}{SS_T}.$$

Giá trị  $R^2$  càng lớn thì mô hình càng có khả năng giải thích tốt dữ liệu mẫu.

**Hệ số xác định hiệu chỉnh ( $R_{adj}^2$ )** Trong hồi quy tuyến tính bội, việc bổ sung thêm biến độc lập luôn làm tăng giá trị  $R^2$ , kể cả khi biến đó không thực sự có ý nghĩa. Do đó, hệ số xác định hiệu chỉnh được sử dụng để so sánh các mô hình có số lượng biến khác nhau:

$$R_{adj}^2 = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)}.$$

Chỉ số này chỉ tăng khi biến mới được đưa vào giúp cải thiện đáng kể khả năng giải thích của mô hình, qua đó hạn chế hiện tượng quá khóp.

## Chương 3

# DỮ LIỆU VÀ CÁC BIẾN TRONG MÔ HÌNH

### 3.1 Thông tin bộ dữ liệu

**Tên bộ dữ liệu:** Car Features and MSRP

**Nguồn dữ liệu:** Kaggle

**Cơ sở dữ liệu:** Bộ dữ liệu sử dụng trong phân tích này gồm **7951** quan sát.

Bộ dữ liệu này bao gồm các thông tin như: Hãng xe, kiểu dáng, loại động cơ, năm sản xuất và các đặc tính khác của ô tô để dự đoán giá của ô tô. Trong bài báo cáo này, chúng ta chỉ quan tâm tới các thông tin sau:

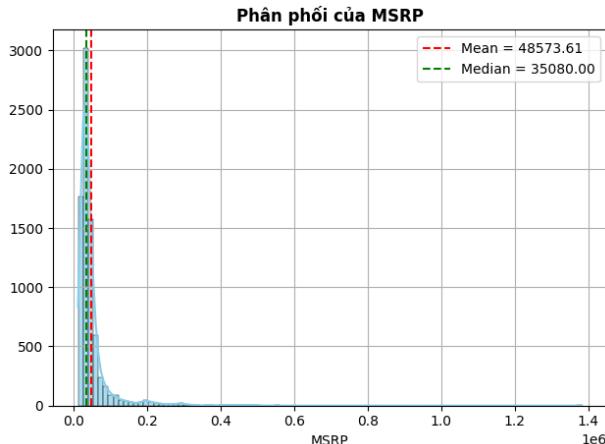
Tên biến	Mô tả
MSRP	Giá bán lẻ đề xuất ban đầu của xe tại thị trường Mỹ, tính bằng USD.
Make	Hãng sản xuất ô tô (ví dụ: BMW, Audi, Ford...).
Year	Năm sản xuất của mẫu xe.
Powertrain	Loại hệ truyền động của xe, phân thành 3 nhóm: Xăng (ICE), Hybrid, và Điện (EV).
Engine HP	Công suất động cơ – đơn vị mã lực (HP).

**Bảng 3.1.** Mô tả các biến số trong tập dữ liệu

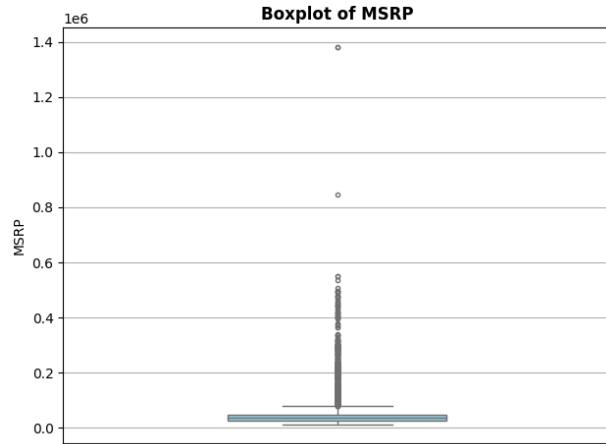
### 3.2 Xử lý dữ liệu và tạo biến

Quá trình xử lý dữ liệu và tạo biến được thực hiện nhằm chuẩn hóa bộ dữ liệu, đảm bảo các biến đưa vào mô hình hồi quy phù hợp với giả định tuyến tính và có thể diễn giải rõ ràng về mặt kinh tế. Các bước xử lý chính bao gồm tạo biến phụ thuộc, xây dựng biến công nghệ truyền động và mã hóa biến thương hiệu.

## Biến MSRP



(a) Phân phối của MSRP



(b) Boxplot của MSRP

Hình 3.1a, 3.1b minh họa phân phối của MSRP. Có thể nhận thấy phân phối của biến này xuất hiện nhiều giá trị ngoại lai và bị lệch phải nghiêm trọng. Đặc điểm này thường gây ra hiện tượng phương sai sai số thay đổi và làm giảm hiệu quả ước lượng của mô hình hồi quy.

Chính vì vậy, biến MSRP được biến đổi logarit để tạo thành biến  $\ln(\text{MSRP})$ . Việc biến đổi này giúp thu hẹp khoảng cách giữa các quan sát, giảm thiểu tác động của các giá trị cực biên, đồng thời cho phép diễn giải các hệ số hồi quy theo độ co giãn (mức thay đổi phần trăm), giúp mô hình có ý nghĩa kinh tế phù hợp hơn.

## Biến Powertrain

Từ biến phân loại Powertrain, hai biến giả *EV* và *Hybrid* được tạo ra, với nhóm ICE đóng vai trò làm nhóm tham chiếu trong mô hình hồi quy.

## Biến Make

Biến thương hiệu Make được xử lý theo hai bước nhằm kiểm soát ảnh hưởng của thương hiệu trong mô hình hồi quy, đồng thời đảm bảo độ ổn định của các ước lượng.

Trước hết, các hãng xe có số lượng quan sát lớn nhất trong bộ dữ liệu được giữ nguyên (lấy top 10 hãng đầu tiên), trong khi các hãng còn lại được gộp chung vào một nhóm thương hiệu tổng hợp, ký hiệu là *Other*. Cách xử lý này giúp giảm số lượng biến giả cần đưa vào mô hình, hạn chế hiện tượng mô hình quá phức tạp và cải thiện độ tin cậy của các hệ số ước lượng.

Từ biến thương hiệu đã được nhóm lại này (ký hiệu là *Make\_grouped*), một tập hợp các biến giả được tạo ra, trong đó hãng *Toyota* được lựa chọn làm nhóm tham chiếu. Việc lựa chọn Toyota dựa trên tiêu chí hãng này có số lượng quan sát lớn và đại diện cho phân khúc phổ thông của thị trường, qua đó cho phép các hệ số của biến giả thương hiệu phản ánh mức chênh lệch giá tương đối so với một hãng “chuẩn” của thị trường ô tô.

### 3.3 Biến phụ thuộc

Biến phụ thuộc trong mô hình hồi quy là  $\ln(\text{MSRP})$ , được tạo ra từ biến MSRP gốc thông qua phép biến đổi logarit tự nhiên.

Biến này phản ánh mức độ định giá của nhà sản xuất đối với từng mẫu xe và được sử dụng làm đại diện cho giá trị thị trường của xe.

## 3.4 Các biến độc lập

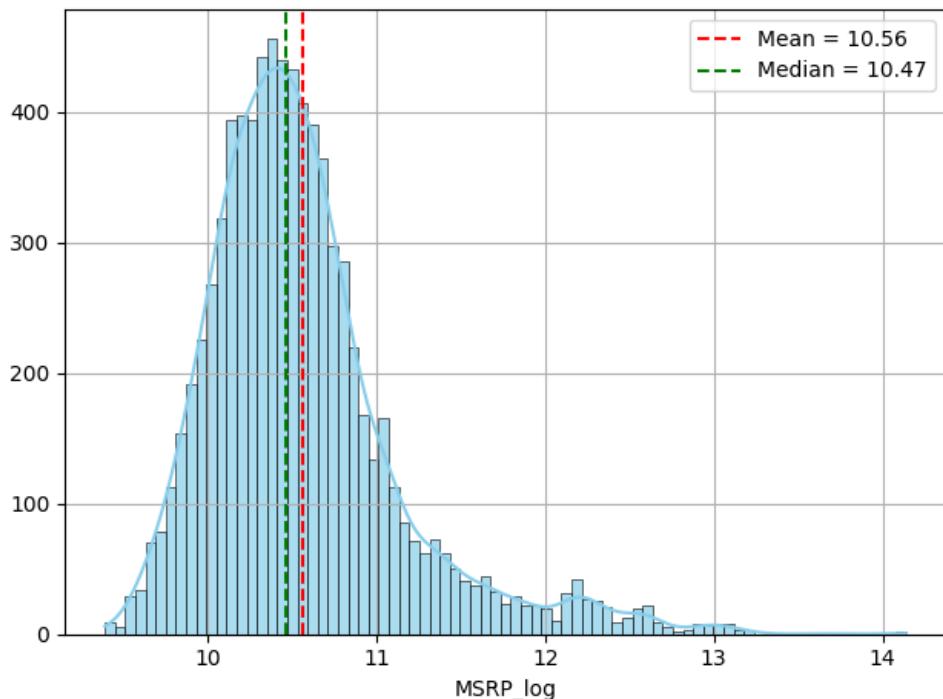
Các biến độc lập trong mô hình được lựa chọn nhằm phản ánh các yếu tố kỹ thuật, thời gian và thương hiệu có thể ảnh hưởng đến giá xe. Cụ thể, mô hình bao gồm các biến sau:

- **Engine HP**: công suất động cơ (mã lực), đại diện cho năng lực kỹ thuật và hiệu suất vận hành của xe. Biến này được kỳ vọng có tác động dương đến giá xe.
- **Year**: năm sản xuất của xe, được sử dụng để kiểm soát ảnh hưởng của tiến bộ công nghệ, lạm phát và xu hướng định giá theo thời gian.
- **EV**: biến giả nhận giá trị 1 nếu xe là xe điện thuần túy và 0 nếu không.
- **Hybrid**: biến giả nhận giá trị 1 nếu xe là xe hybrid và 0 nếu không.
- **Make\_grouped**: tập hợp các biến giả đại diện cho thương hiệu xe, với hãng Toyota được sử dụng làm nhóm tham chiếu.

## 3.5 Trực quan hóa dữ liệu

Việc trực quan hóa dữ liệu được thực hiện nhằm mô tả đặc điểm phân bố của các biến nghiên cứu cũng như mối quan hệ giữa biến phụ thuộc và các biến độc lập. Kết quả trực quan hóa giúp hỗ trợ việc lựa chọn mô hình và diễn giải kết quả hồi quy.

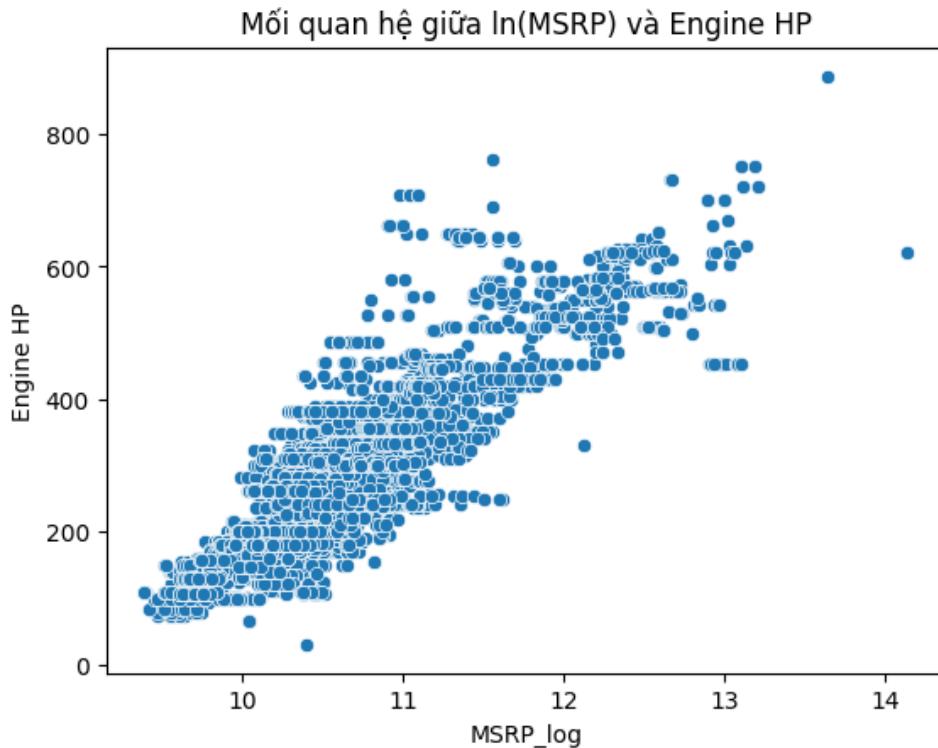
### Phân bố biến phụ thuộc



**Hình 3.2.** Phân phối của biến  $\log(\text{MSRP})$

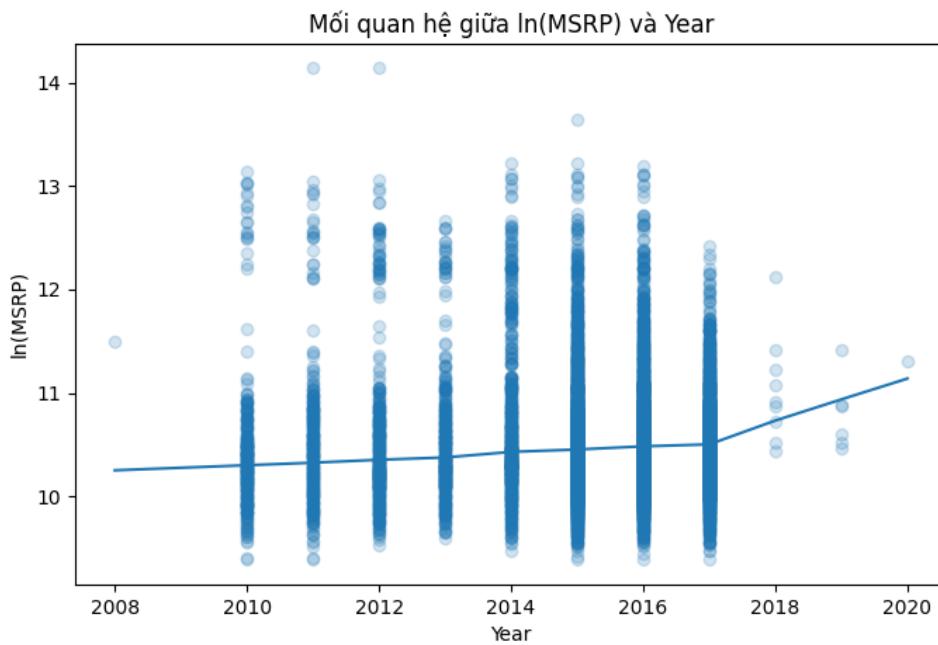
Hình 3.2 minh họa phân bố của biến phụ thuộc  $\log(\text{MSRP})$ . Có thể nhận thấy phân bố của biến này có dạng gần giống phân phối chuẩn, tuy nhiên vẫn tồn tại hiện tượng lệch nhẹ và một số giá trị ngoại lai.

## Mối quan hệ giữa biến phụ thuộc và các biến độc lập



**Hình 3.3.** Mối quan hệ giữa  $\ln(\text{MSRP})$  và Engine HP

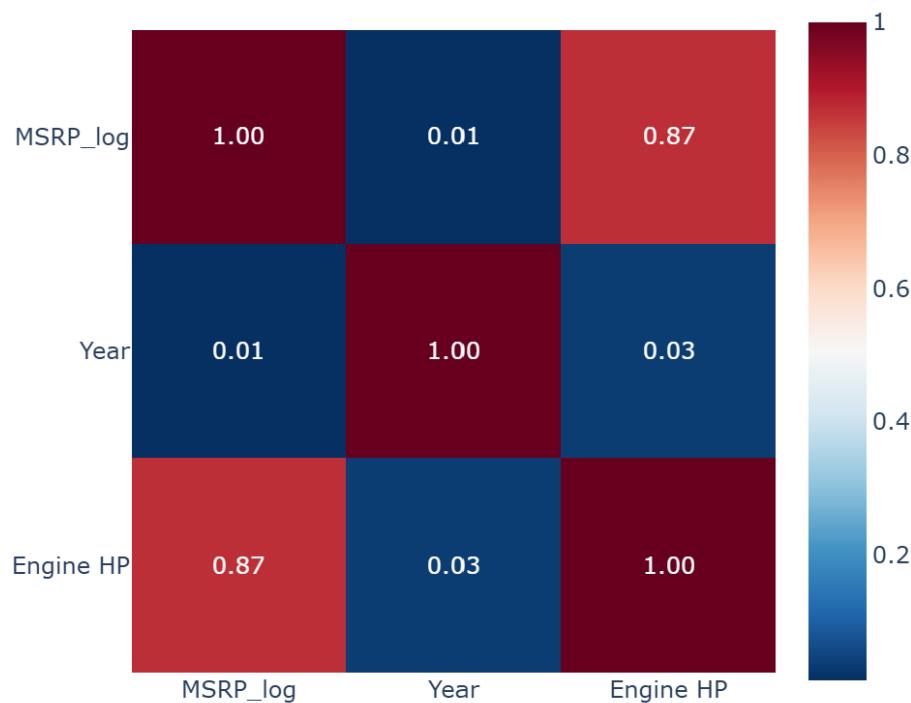
Hình 3.3 thể hiện mối quan hệ giữa  $\ln(\text{MSRP})$  và công suất động cơ. Kết quả cho thấy giá trị  $\ln(\text{MSRP})$  có xu hướng tăng khi công suất động cơ tăng.



**Hình 3.4.** Mối quan hệ giữa  $\ln(\text{MSRP})$  và Year

Hình 3.4 thể hiện mối quan hệ giữa  $\ln(\text{MSRP})$  và năm sản xuất của xe. Kết quả cho thấy mức giá xe có sự phân tán đáng kể trong cùng một năm, phản ánh sự đa dạng của các mẫu xe. Bên cạnh đó, cho thấy xu hướng tăng dần khá ổn định của  $\ln(\text{MSRP})$  theo thời gian, đặc biệt rõ rệt trong giai đoạn khoảng 2010 – 2017.

## Ma trận tương quan



**Hình 3.5.** Ma trận tương quan giữa các biến

Hình 3.5 cho thấy giữa các biến độc lập (Year và Engine HP), hệ số tương quan chỉ đạt 0,03. Mỗi liên hệ tuyến tính rất yếu, gần như không đáng kể. Điều này hàm ý rằng không tồn tại hiện tượng đa cộng tuyến giữa các biến giải thích trong mô hình.

## Kết luận

Các kết quả trực quan hóa cho thấy tồn tại mối quan hệ giữa biến phụ thuộc và các biến độc lập trong mô hình. Những quan sát này là cơ sở ban đầu cho việc xây dựng và ước lượng mô hình hồi quy tuyến tính ở các phần tiếp theo.

## Chương 4

# MÔ HÌNH HỒI QUY TUYẾN TÍNH BỘI

Dựa trên cơ sở lý thuyết và kết quả mô tả dữ liệu ở các chương trước, chương này tập trung vào việc xây dựng các mô hình hồi quy tuyến tính bội nhằm phân tích mối quan hệ giữa giá bán lẻ đề xuất của xe ô tô và các yếu tố công nghệ cũng như thương hiệu. Cụ thể, nghiên cứu xem xét tác động của các đặc tính kỹ thuật, thương hiệu và loại hệ truyền động đến giá xe, đồng thời kiểm định sự tồn tại của mức chênh lệch giá giữa xe điện (EV), xe hybrid và xe sử dụng động cơ đốt trong truyền thống (ICE).

Phân tích được triển khai theo cách tiếp cận từng bước, bắt đầu từ mô hình hồi quy cơ sở chỉ bao gồm các yếu tố kỹ thuật và thời gian, sau đó mở rộng bằng cách bổ sung yếu tố thương hiệu. Cách tiếp cận này cho phép đánh giá sự thay đổi trong mức độ và ý nghĩa của chênh lệch giá giữa các loại hệ truyền động khi kiểm soát thêm ảnh hưởng của thương hiệu.

### 4.1 Mô hình hồi quy tuyến tính bội cơ bản

Trên cơ sở câu hỏi phân tích về mức chênh lệch giá giữa các loại hệ truyền động, mục này xây dựng mô hình hồi quy tuyến tính bội cơ bản nhằm thiết lập điểm xuất phát cho phân tích thực nghiệm. Mô hình tập trung vào mối quan hệ giữa giá bán lẻ đề xuất của xe và các yếu tố kỹ thuật cốt lõi, đồng thời xem xét tác động riêng biệt của công nghệ hệ truyền động (EV và Hybrid) so với xe sử dụng động cơ đốt trong truyền thống (ICE).

Ở mục này, mô hình chưa đưa vào các yếu tố liên quan đến thương hiệu (*Make*), nhằm tách biệt và làm nổi bật vai trò của các đặc tính công nghệ, kỹ thuật. Cách tiếp cận này cho phép đánh giá sơ bộ liệu mức giá chênh lệch của xe điện (EV) và xe Hybrid có tồn tại ngay cả khi chỉ kiểm soát yếu tố kỹ thuật cơ bản, trước khi xem xét đến ảnh hưởng của thương hiệu trong mô hình mở rộng sau. Mô hình hồi quy tuyến tính bội được xây dựng như sau:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon \quad (4.1)$$

Trong đó:

- $Y = \ln(\text{MSRP})$  là logarit tự nhiên của giá bán lẻ đề xuất xe ô tô;
- $x_1$  = Year đại diện cho năm sản xuất của xe, phản ánh xu hướng giá theo thời gian;
- $x_2$  = EngineHP là công suất động cơ, đại diện cho yếu tố kỹ thuật của xe;
- $x_3$  = EV là biến giả nhận giá trị 1 nếu xe là xe điện thuần (Electric Vehicle), và 0 trong các trường hợp còn lại;
- $x_4$  = Hybrid là biến giả nhận giá trị 1 nếu xe là xe hybrid, và 0 trong các trường hợp còn lại;
- $\varepsilon$  là sai số ngẫu nhiên, phản ánh các yếu tố chưa được quan sát.

Trong mô hình này, xe sử dụng động cơ đốt trong truyền thống (ICE) được chọn làm nhóm tham chiếu. Do đó, các hệ số  $\beta_3$  và  $\beta_4$  lần lượt phản ánh mức chênh lệch giá trung bình của xe EV và xe Hybrid so với xe ICE, sau khi đã kiểm soát các yếu tố kỹ thuật và thời gian.

## 4.2 Mô hình hồi quy mở rộng với biến thương hiệu (Make)

Trên cơ sở mô hình hồi quy tuyến tính bội cơ bản đã được xây dựng ở mục trước, báo cáo tiếp tục mở rộng mô hình bằng cách bổ sung yếu tố thương hiệu thông qua các biến giả đại diện cho hãng sản xuất (Make). Việc đưa biến thương hiệu vào mô hình nhằm kiểm soát sự khác biệt trong chiến lược định vị, uy tín thương hiệu và phân khúc thị trường giữa các hãng xe, từ đó cho phép đánh giá rõ hơn vai trò của công nghệ hệ truyền động trong cấu trúc định giá.

Mô hình hồi quy tuyến tính bội mở rộng được xây dựng như sau:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \sum_{k=1}^{K-1} \gamma_k x_{4+k} + \varepsilon, \quad (4.2)$$

trong đó:

- $Y = \ln(\text{MSRP})$  là logarit tự nhiên của giá bán lẻ đề xuất xe ô tô;
- $x_1 = \text{Year}$  đại diện cho năm sản xuất của xe;
- $x_2 = \text{EngineHP}$  là công suất động cơ, phản ánh đặc tính kỹ thuật của xe;
- $x_3 = \text{EV}$  là biến giả nhận giá trị 1 nếu xe là xe điện thuần (EV), và 0 trong các trường hợp còn lại;
- $x_4 = \text{Hybrid}$  là biến giả nhận giá trị 1 nếu xe là xe Hybrid, và 0 trong các trường hợp còn lại;
- $x_{4+k} = \text{Make}_k$  là biến giả đại diện cho hãng sản xuất thứ  $k$ , với hãng Toyota được chọn làm tham chiếu.
- $\varepsilon$  là sai số ngẫu nhiên, phản ánh các yếu tố chưa được quan sát.

Trong mô hình này, xe sử dụng động cơ đốt trong truyền thống (ICE) cùng với hãng Toyota được lựa chọn làm nhóm tham chiếu đóng vai trò là cơ sở để diễn giải các hệ số hồi quy. Các hệ số  $\gamma_k$  phản ánh mức chênh lệch giá trung bình giữa từng hãng xe so với hãng tham chiếu, sau khi đã kiểm soát các yếu tố kỹ thuật và công nghệ hệ truyền động.

## 4.3 Ước lượng hệ số hồi quy trong mô hình hồi quy bội

### Mô hình hồi quy tuyến tính bội cơ sở

Từ dữ liệu có  $n = 7951$  quan sát và  $k = 4$  biến giải thích. Với mỗi quan sát  $i$  ( $i = 1, 2, \dots, n$ ), ta thu được bộ dữ liệu  $(x_{i1}, x_{i2}, x_{i3}, x_{i4}; y_i)$ . Mô hình hồi quy tuyến tính bội được viết dưới dạng:

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (4.3)$$

Dưới dạng ma trận, mô hình được biểu diễn gọn như sau:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon, \quad (4.4)$$

trong đó:

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} & x_{14} \\ 1 & x_{21} & x_{22} & x_{23} & x_{24} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & x_{n4} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Trong báo cáo này, ma trận  $\mathbf{X}$  (ma trận của mô hình) và vector quan sát  $\mathbf{y}$  có dạng cụ thể:

$$\mathbf{X} = \begin{pmatrix} 1 & 2011 & 335 & 0 & 0 \\ 1 & 2011 & 300 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 2015 & 887 & 0 & 1 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 10.74 \\ 10.61 \\ \vdots \\ 13.65 \end{pmatrix}.$$

Theo phương pháp bình phương cực tiểu, ta cần tính các ma trận  $\mathbf{X}^T \mathbf{X}$  và  $\mathbf{X}^T \mathbf{y}$ . Từ dữ liệu quan sát, ta thu được:

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} 7951 & 16021682 & 2165396 & 75 & 347 \\ 16021682 & 32284556200 & 4363443200 & 151130 & 699088 \\ 2165396 & 4363443200 & 692018782 & 15790 & 77844 \\ 75 & 151130 & 15790 & 75 & 0 \\ 347 & 699088 & 77844 & 0 & 347 \end{pmatrix},$$

và

$$\mathbf{X}^T \mathbf{y} = \begin{pmatrix} 83966 \\ 169197398 \\ 23320707 \\ 803 \\ 3651 \end{pmatrix}.$$

Từ các ma trận  $\mathbf{X}^T \mathbf{X}$  và  $\mathbf{X}^T \mathbf{y}$  đã tính được, ước lượng bình phương nhỏ nhất của vector tham số  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T$  được xác định như sau:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} 19,4403 \\ -0,0050 \\ 0,0045 \\ 0,4361 \\ 0,1889 \end{pmatrix}.$$

Do đó, phương trình hồi quy tuyến tính bội ước lượng của mô hình cơ sở được viết như sau:

$$\hat{y} = 19,4403 - 0,0050x_1 + 0,0045x_2 + 0,4361x_3 + 0,1889x_4 \quad (4.5)$$

Trong đó:

- $x_1$  = Year đại diện cho năm sản xuất của xe, phản ánh xu hướng giá theo thời gian;
- $x_2$  = EngineHP là công suất động cơ, đại diện cho yếu tố kỹ thuật của xe;
- $x_3$  = EV là biến giả nhận giá trị 1 nếu xe là xe điện thuần (Electric Vehicle), và 0 trong các trường hợp còn lại;
- $x_4$  = Hybrid là biến giả nhận giá trị 1 nếu xe là xe hybrid, và 0 trong các trường hợp còn lại;

## Mô hình hồi quy mở rộng với biến thương hiệu (Make)

Tương tự như mô hình cơ sở, hệ số của mô hình hồi quy mở rộng được ước lượng bằng phương pháp bình phương nhô nhất.

Do đó, phương trình hồi quy tuyến tính bội ước lượng của mô hình mở rộng với biến thương hiệu được viết như sau:

$$\begin{aligned}\hat{y} = & 7,2803 + 0,0010 x_1 + 0,0044 x_2 + 0,4400 x_3 + 0,2202 x_4 \\ & + 0,1770 D_{Volkswagen} - 0,0469 D_{Chevrolet} - 0,0164 D_{Ford} - 0,0110 D_{Nissan} \\ & + 0,0898 D_{Honda} + 0,2169 D_{Cadillac} + 0,2991 D_{BMW} + 0,0465 D_{GMC} \\ & + 0,4100 D_{Audi} + 0,1831 D_{Other}.\end{aligned}\quad (4.6)$$

Trong đó:

- $x_1$  = Year là năm sản xuất của xe;
- $x_2$  = EngineHP là công suất động cơ;
- $x_3$  = EV là biến giả nhận giá trị 1 nếu xe là xe điện thuần, và 0 trong các trường hợp còn lại;
- $x_4$  = Hybrid là biến giả nhận giá trị 1 nếu xe là xe hybrid, và 0 trong các trường hợp còn lại;
- $D_{Brand}$  là biến giả đại diện cho từng nhóm thương hiệu, nhận giá trị 1 nếu xe thuộc nhóm thương hiệu tương ứng, và 0 nếu không;

## 4.4 Kiểm định ý nghĩa của hồi quy

Xét mô hình hồi quy tuyến tính bội đã được ước lượng với mức ý nghĩa  $\alpha = 5\%$ .

### Mô hình hồi quy tuyến tính bội cơ sở

#### (1) Cặp giả thuyết kiểm định

Xét cặp giả thuyết:

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0, \\ H_1 : \beta_j \neq 0 \text{ với ít nhất } 1 j, j = 1, 2, 3, 4 \end{cases}$$

#### (2) Tiêu chuẩn kiểm định

Tiêu chuẩn kiểm định được sử dụng là thống kê Fisher:

$$F_0 = \frac{SS_R/k}{SS_E/(n-p)} = \frac{MS_R}{MS_E},$$

trong đó  $k = 4$  là số biến độc lập và  $p = k + 1 = 5$  là số tham số của mô hình.

Nếu giả thuyết  $H_0$  là đúng thì  $F_0$  có phân phối  $F$  với:  $v_1 = k = 4$  và  $v_2 = n - p = 7946$  bậc tự do

#### (3) Giá trị quan sát

Từ kết quả tính toán trên bộ dữ liệu nghiên cứu, ta có:

$$SS_E = 618,5450$$

$$SS_R = 2032.2122$$

$$SS_T = 2650,7572$$

Bảng phân tích phương sai (ANOVA):

Nguồn	Tổng bình phương	Bậc tự do	Phương sai	$f_0$
Hồi quy	2032,2122	4	508,0530	
Sai số	618,5450	7946	0,0778	
Tổng	2650,7572	7950	6526,5895	

**Bảng 4.1.** Bảng phân tích phương sai (ANOVA) - Mô hình cơ sở

Vậy giá trị quan sát của tiêu chuẩn kiểm định là:  $f_0 = 6526, 5895$

#### (4) Miền bác bỏ

Với mức ý nghĩa  $\alpha = 5\%$ ,  $f_\alpha(k, n - p) = f_{0.05}(4, 7946)$ , miền bác bỏ là:

$$W_\alpha = (2, 37; +\infty)$$

#### (5) Kết luận

Do  $f_0 = 6526, 5895 \in W_\alpha$ , ta bác bỏ giả thuyết  $H_0$ . Vì vậy, mô hình hồi quy tuyến tính bội có ý nghĩa thống kê ở mức ý nghĩa 5%.

### Mô hình hồi quy mở rộng với biến thương hiệu (Make)

Kiểm định ý nghĩa thống kê chung của mô hình hồi quy mở rộng được thực hiện tương tự như đối với mô hình cơ sở, với mức ý nghĩa  $\alpha = 5\%$ .

Ta có, bảng phân tích phương sai (ANOVA):

Nguồn	Tổng bình phương	Bậc tự do	Phương sai	$f_0$
Hồi quy	2128,1503	14	152,0107	
Sai số	522,6069	7936	0,0659	
Tổng	2650,7572	7950	2308,3453	

**Bảng 4.2.** Bảng phân tích phương sai (ANOVA) - Mô hình mở rộng

Vậy giá trị quan sát của tiêu chuẩn kiểm định là:  $f_0 = 2308, 3453$ .

Với mức ý nghĩa  $\alpha = 5\%$ ,  $f_\alpha(k, n - p) = f_{0.05}(14, 7936)$ , miền bác bỏ là:

$$W_\alpha = (1, 69; +\infty)$$

Do  $f_0 = 2308, 3453 \in W_\alpha$ , ta bác bỏ giả thuyết  $H_0$ . Vì vậy, mô hình hồi quy mở rộng với biến thương hiệu có ý nghĩa thống kê ở mức ý nghĩa 5%.

## 4.5 Hệ số xác định và hệ số xác định hiệu chỉnh

### Mô hình hồi quy tuyến tính bội cơ sở

Hệ số xác định của mô hình được xác định bởi:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}.$$

Thay số ta thu được:

$$R^2 = \frac{2032,2122}{2650,7572} = 0,7666.$$

Điều này cho thấy mô hình giải thích được khoảng 76,7% sự biến thiên của ln(MSRP). Phần còn lại 23,3% biến thiên là do các yếu tố khác chưa được đưa vào mô hình và sai số ngẫu nhiên.

Hệ số xác định hiệu chỉnh:

$$R_{\text{adj}}^2 = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} = 0,7665.$$

Giá trị  $R_{\text{adj}}^2$  xấp xỉ  $R^2$  cho thấy mô hình có mức độ phù hợp tốt và không bị ảnh hưởng đáng kể bởi số lượng biến giải thích.

## Mô hình hồi quy mở rộng với biến thương hiệu (Make)

Đối với mô hình hồi quy mở rộng, hệ số xác định thu được là:

$$R^2 = 0,803.$$

Điều này cho thấy mô hình hồi quy mở rộng giải thích được khoảng 80,3% sự biến thiên của ln(MSRP), cao hơn so với mô hình cơ sở. Kết quả này phản ánh việc bổ sung biến thương hiệu giúp cải thiện khả năng giải thích của mô hình đối với biến phụ thuộc.

Hệ số xác định hiệu chỉnh của mô hình mở rộng là:

$$R_{\text{adj}}^2 = 0,802.$$

Giá trị  $R_{\text{adj}}^2$  chỉ thấp hơn không đáng kể so với  $R^2$ , cho thấy mức độ cải thiện của mô hình không chỉ đến từ việc gia tăng số lượng biến giải thích mà còn phản ánh đóng góp thực chất của các biến thương hiệu vào khả năng giải thích của mô hình.

## 4.6 Kiểm định giả thuyết về các hệ số hồi quy riêng

### Mô hình hồi quy tuyến tính bội cơ sở

Tiến hành kiểm định ý nghĩa của các hệ số hồi quy riêng với mức ý nghĩa  $\alpha = 5\%$ .

#### (1) Cặp giả thuyết kiểm định

Với mỗi hệ số hồi quy riêng lẻ  $\beta_j$ , xét:

$$\begin{cases} H_0 : \beta_j = 0, \\ H_1 : \beta_j \neq 0. \end{cases}$$

#### (2) Tiêu chuẩn kiểm định

Tiêu chuẩn kiểm định được sử dụng là:

$$T_0 = \frac{\hat{\beta}_j}{\sqrt{\sigma^2 C_{jj}}}.$$

Nếu giả thuyết  $H_0$  đúng thì  $T_0 \sim t(n-p)$ .

### (3) Kết quả kiểm định

	$\hat{\beta}_j$	se( $\beta_j$ )	t-value
Intercept	19.440283	3.444897	5.643212
Year	-0.005019	0.001710	-2.935406
Engine HP	0.004482	0.000028	161.479856
EV	0.436115	0.032426	13.449395
Hybrid	0.188889	0.015398	12.267346

Bảng 4.3. Kết quả kiểm định

### (4) Kết luận

Với  $t_{0.025}(7946) \approx 1,96$ , các biến đều có ý nghĩa thống kê trong mô hình hồi quy ở mức ý nghĩa 5%.

## Mô hình hồi quy mở rộng với biến thương hiệu (Make)

Tiến hành kiểm định ý nghĩa thống kê của các hệ số hồi quy riêng trong mô hình hồi quy mở rộng với mức ý nghĩa  $\alpha = 5\%$ . Giả thuyết kiểm định và tiêu chuẩn kiểm định được thực hiện tương tự như đối với mô hình cơ sở.

### (1) Kết quả kiểm định

	$\hat{\beta}_j$	se( $\beta_j$ )	t-value
Intercept	7.280273	3.255067	2.236597
Make (Volkswagen)	0.177043	0.014685	12.055757
Make (Chevrolet)	-0.046886	0.014533	-3.226224
Make (Ford)	-0.016413	0.015326	-1.070899
Make (Nissan)	-0.011020	0.016104	-0.684317
Make (Honda)	0.089841	0.016793	5.349813
Make (Cadillac)	0.216896	0.017701	12.253545
Make (BMW)	0.299056	0.018041	16.576765
Make (GMC)	0.046477	0.018371	2.529893
Make (Audi)	0.409966	0.018870	21.726329
Make (Other)	0.183067	0.011624	15.749484
Year	0.000964	0.001615	0.596972
Engine HP	0.004387	0.000027	162.839980
EV	0.439951	0.029873	14.727130
Hybrid	0.220205	0.014379	15.314609

Bảng 4.4. Kết quả kiểm định - Mô hình mở rộng

### (2) Kết luận

Với giá trị tới hạn  $t_{0.025}(7936) \approx 1,96$  kiểm định cho thấy hầu hết các hệ số hồi quy trong mô hình mở rộng đều có ý nghĩa thống kê, bao gồm công suất động cơ Engine HP), các biến giả EV và Hybrid, cũng như nhiều biến thương hiệu. Tuy nhiên, biến Year và một số biến thương hiệu (Ford, Nissan) không có ý nghĩa thống kê sau khi đã kiểm soát các yếu tố khác.

## 4.7 So sánh vai trò của yếu tố thương hiệu

Để đánh giá vai trò của yếu tố thương hiệu, chúng ta so sánh hệ số ước lượng của các biến EV và Hybrid cũng như chỉ số  $R^2$  giữa hai mô hình: mô hình cơ sở (chỉ gồm các yếu tố kỹ thuật và thời gian) và mô hình mở rộng (bao gồm biến Make\_grouped). Kết quả được trình bày trong Bảng 4.5. Từ

Biến	Mô hình cơ sở	Mô hình mở rộng (có Make)
EV	0,4361	0,4399
Hybrid	0,1889	0,2202
$R^2$	0,757	0,803

Bảng 4.5. So sánh hệ số EV/Hybrid và  $R^2$  giữa mô hình cơ sở và mô hình mở rộng

bảng trên, có thể thấy rằng việc bổ sung yếu tố thương hiệu vào mô hình:

- Làm tăng hệ số của biến Hybrid, cho thấy ảnh hưởng thực tế của xe hybrid đến giá được đánh giá cao hơn khi kiểm soát thương hiệu.
- Hệ số EV gần như không đổi, cho thấy sự chênh lệch giá của xe điện chủ yếu đến từ công nghệ.
- Nâng chỉ số  $R^2$ , chứng tỏ mô hình mở rộng giải thích tốt hơn phuơng sai của giá xe.

Kết quả này nhấn mạnh vai trò quan trọng của yếu tố thương hiệu trong việc định giá xe.

## 4.8 Phân tích phần dư

Trong phần này, chúng ta sẽ phân tích phần dư của **mô hình hồi quy mở rộng** bao gồm yếu tố thương hiệu. Lý do chỉ phân tích mô hình mở rộng là vì đây là mô hình cuối cùng mà báo cáo sử dụng để đưa ra các kết luận về ảnh hưởng của công nghệ, hệ truyền động và thương hiệu đối với giá xe. Mô hình cơ bản (chỉ gồm các yếu tố kỹ thuật và thời gian) được dùng làm tham chiếu ban đầu, nhưng phần dư của mô hình này không phục vụ trực tiếp cho các kết luận cuối cùng, nên sẽ không được trình bày ở đây.

Phần dư của mô hình hồi quy được xác định bởi:

$$e_i = y_i - \hat{y}_i.$$

Tổng bình phương sai số:

$$SS_E = \sum_{i=1}^n e_i^2 = 522,6069$$

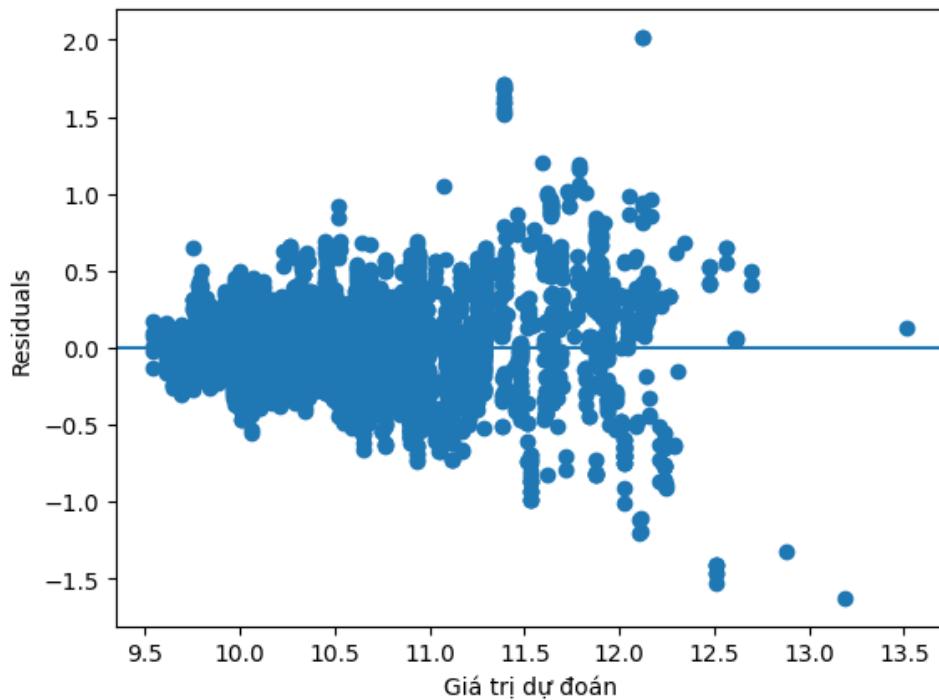
Ước lượng phuơng sai của sai số:

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} = 0,0659; \quad \hat{\sigma} = 0,2566.$$

Phần dư chuẩn hoá được xác định bởi:

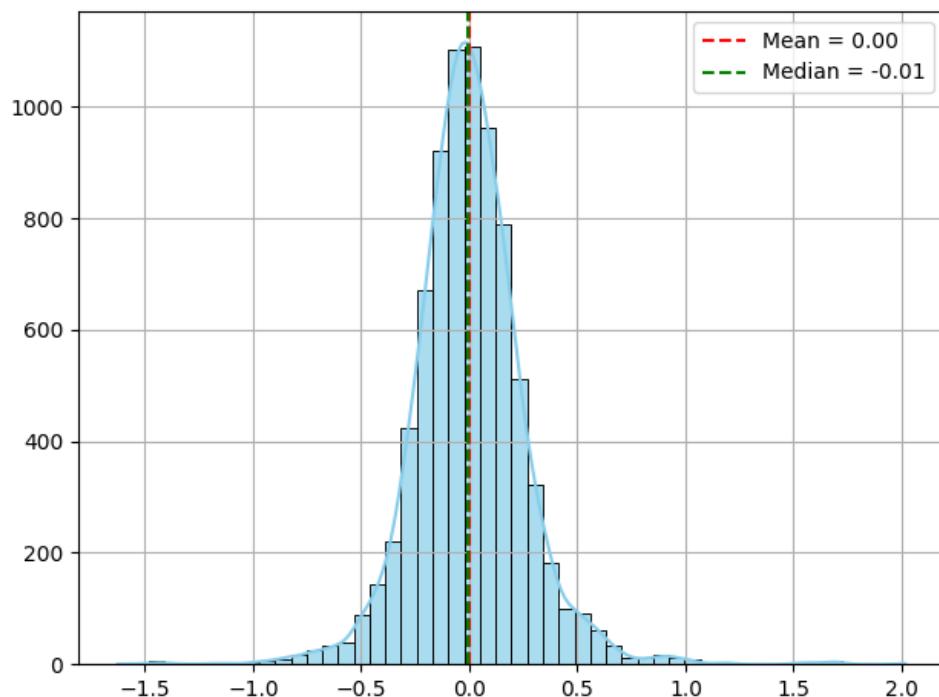
$$r_i = \frac{e_i}{\hat{\sigma}}.$$

## Đồ thị phần dư



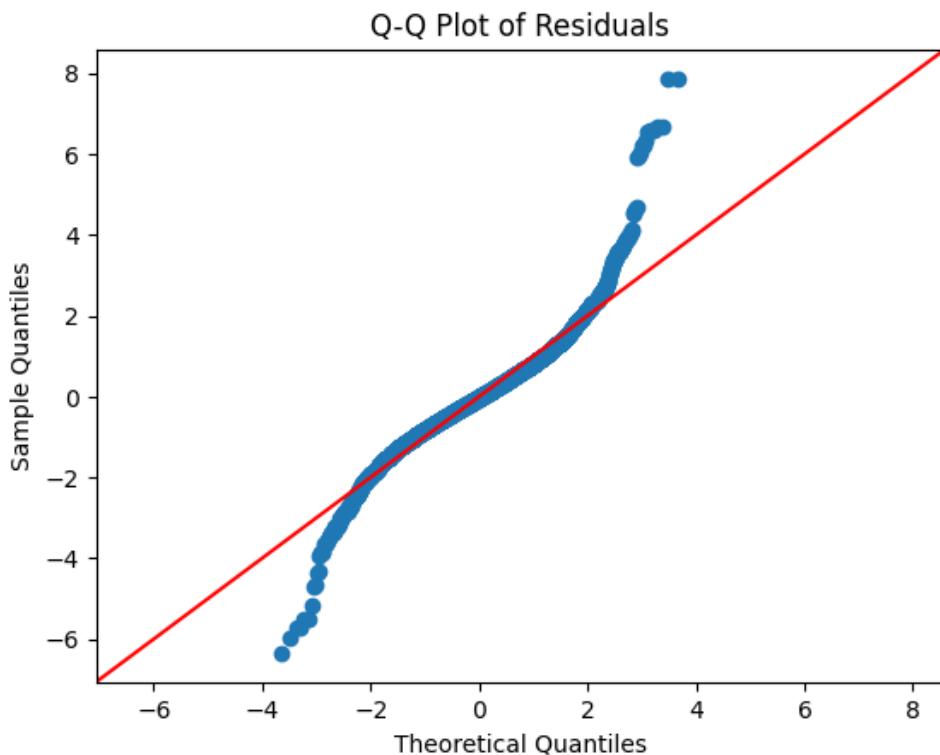
**Hình 4.1.** Đồ thị phần dư theo giá trị dự đoán (Residuals vs Fitted)

Hình 4.1 minh họa mối quan hệ giữa phần dư và giá trị dự đoán cho thấy các phần dư phân bố quanh giá trị 0, điều này cho thấy kỳ vọng của sai số xấp xỉ bằng 0. Tuy nhiên, độ phân tán của phần dư có xu hướng tăng khi giá trị dự đoán lớn hơn, điều này cho thấy tồn tại dấu hiệu phương sai thay đổi ở mức độ nhẹ. Ngoài ra, một số quan sát có phần dư lớn hơn so với phần còn lại, song số lượng các điểm này không nhiều và không gây ảnh hưởng nghiêm trọng đến kết quả ước lượng.



**Hình 4.2.** Phân phối của phần dư

Hình 4.2 thể hiện histogram của phần dư. Có thể nhận thấy phân phối của phần dư có dạng gần giống phân phối chuẩn.



**Hình 4.3.** Q-Q plot của Phần dư

Hình 4.3 là đồ thị Q – Q của phần dư. Các điểm quan sát nằm tương đối gần đường chéo, cho thấy phần dư có phân phối xấp xỉ chuẩn.

## Kết luận

Dựa trên các đồ thị phân tích phần dư, các giả định cơ bản của mô hình hồi quy tuyến tính như kỳ vọng bằng không, phương sai không đổi và phân phối xấp xỉ chuẩn của sai số được thỏa mãn ở mức chấp nhận được. Do đó, các kết luận suy luận thống kê từ mô hình là phù hợp.

# Chương 5

## KẾT LUẬN

### 5.1 Kết quả đạt được

Dựa trên các kết quả minh họa chi tiết tại Bảng 4.5 ở Chương 4, trong đó so sánh hệ số EV/Hybrid và  $R^2$  giữa mô hình cơ sở và mô hình mở rộng., ta có thể trả lời trực tiếp các câu hỏi đặt ra như sau:

#### Ảnh hưởng công nghệ (Xe EV và Hybrid)

Các hệ số ước lượng của biến EV và Hybrid trong **mô hình cơ bản** lần lượt là 0.4361 và 0.1889, đều có ý nghĩa thống kê. Điều này cho thấy, sau khi kiểm soát các yếu tố kỹ thuật cơ bản và năm sản xuất, giá của xe điện và xe hybrid cao hơn đáng kể so với xe ICE truyền thống. Như vậy, có sự khác biệt giá theo loại hệ truyền động.

#### Vai trò thương hiệu

Khi bổ sung yếu tố thương hiệu (Make\_grouped) trong mô hình mở rộng, chỉ số  $R^2$  tăng từ 0.757 lên 0.803, đồng thời  $f - test$  cho nhóm biến thương hiệu có giá trị  $f = 574.35$ ,  $p < 0.001$ . Kết quả này chứng tỏ thương hiệu có ảnh hưởng đáng kể và độc lập đến giá.

#### Phân tầng công nghệ

So sánh hệ số EV và Hybrid trong **mô hình mở rộng** cho thấy hệ số EV = 0.4399 > Hybrid = 0.2202, phản ánh rằng giá của xe điện cao hơn xe Hybrid ngay cả sau khi kiểm soát yếu tố thương hiệu và kỹ thuật. Điều này xác nhận sự phân tầng công nghệ trong chiến lược định giá.

### 5.2 Ý nghĩa và đóng góp

Báo cáo này chỉ ra rằng:

- Mức chênh lệch giá MSRP của xe EV và Hybrid xuất phát từ cả yếu tố công nghệ thực và thương hiệu.
- Yếu tố thương hiệu đóng vai trò độc lập, góp phần giải thích sự khác biệt về giá giữa các nhà sản xuất xe ô tô.
- Chiến lược định giá giữa xe EV và Hybrid là khác nhau, EV thường được định giá cao hơn Hybrid ngay cả khi kiểm soát các đặc tính kỹ thuật.

Những phát hiện này có ý nghĩa quan trọng cho các nhà hoạch định chiến lược và các nhà quản lý thương hiệu trong ngành công nghiệp ô tô, đặc biệt khi định giá các sản phẩm điện hóa mới.

### 5.3 Hạn chế và hướng nghiên cứu tiếp theo

- Dữ liệu hiện tại chỉ bao gồm các biến cơ bản về kỹ thuật và thương hiệu; chưa xem xét các yếu tố marketing, khuyến mãi hay đặc tính thị trường địa phương.
- Mô hình hồi quy tuyến tính giả định mối quan hệ tuyến tính giữa các biến; các phương pháp phi tuyến có thể khám phá thêm các tác động phức tạp.
- Nghiên cứu tiếp theo có thể mở rộng bằng cách phân tích dữ liệu longitudinal để xem sự thay đổi premium theo thời gian và công nghệ mới.

# Tài liệu tham khảo

- [1] Giảng viên Nguyễn Thị Thanh Thủy (2025). *Suy luận thống kê*, slide bài giảng môn Suy luận thống kê, Đại học Bách Khoa Hà Nội.
- [2] Kaggle. *Car Features and MSRP Dataset - Data Analysis and Price Prediction Tutorials*. Truy cập tại: <https://www.kaggle.com/datasets/CooperUnion/cardataset>.