

Ứng dụng Microsoft Excel vào phân tích hồi quy trong giảng dạy Xác suất thống kê tại Trường Đại học Tài nguyên và Môi trường Hà Nội

Nguyễn Ngọc Linh*, Lê Thị Hương*, Đoàn Thị Thanh Huyền*, Đàm Thanh Tuấn*

*ThS. Khoa Khoa học đại cương, Trường Đại học Tài nguyên và Môi trường Hà Nội

Received: 03/10/2024; Accepted: 14/10/2024; Published: 30/10/2024

Abstract: Data analysis is one of the important steps to have an accurate basis for statements and reports in scientific research. In the field of teaching statistical probability and data analysis, Microsoft Excel has proven its useful role. Excel is not only a simple calculation tool but also a powerful platform for data analysis problems, especially regression analysis. Regression analysis is a popular statistical method, helping to study the relationship between variables and forecast trends based on real data. This article presents the application of Microsoft Excel to regression analysis in teaching statistical probability at Hanoi University of Natural Resources and Environment.

Keywords: Regression analysis, Microsoft Excel, teaching statistical probability and data analysis, Hanoi University of Natural Resources and Environment.

1. Đặt vấn đề

Phân tích dữ liệu là một trong những bước quan trọng để có căn cứ chính xác cho những nhận định, báo cáo trong nghiên cứu khoa học. Thống kê là lĩnh vực thu thập và phân tích số liệu cung cấp những thông tin quan trọng cho những đánh giá, nhận định, dự đoán kết luận... Trong thời đại công nghệ thông tin phát triển như hiện nay, việc ứng dụng phần mềm và công cụ hỗ trợ vào giảng dạy ngày càng trở nên phổ biến. Đặc biệt, trong lĩnh vực giảng dạy xác suất thống kê và phân tích dữ liệu, Microsoft Excel đã chứng minh được vai trò hữu ích của mình. Excel không chỉ là một công cụ tính toán đơn thuần mà còn là nền tảng mạnh mẽ cho các bài toán phân tích dữ liệu, đặc biệt là phân tích hồi quy. Phân tích hồi quy là một phương pháp thống kê phổ biến, giúp nghiên cứu mối quan hệ giữa các biến và dự báo xu hướng dựa trên dữ liệu thực tế. Vì vậy vận dụng kết hợp công cụ Data Analysis trong Microsoft Excel trong giảng dạy phân tích hồi quy không chỉ giúp sinh viên tiếp cận dễ dàng hơn với môn học, mà còn tạo điều kiện thực hành với công cụ phổ biến trong công việc sau này.

2. Nội dung nghiên cứu

2.1. Thực trạng giảng dạy xác suất thống kê tại trường Đại học tài nguyên & môi trường Hà Nội.

Hiện nay, tại trường Đại học Tài nguyên và Môi trường Hà Nội, môn học Xác suất thống kê được giảng dạy 30 tiết, thuộc khối kiến thức cơ sở, nhằm

trang bị cho sinh viên tri thức khoa học, phương pháp luận nghiên cứu, các kỹ năng, kỹ xảo của môn Xác suất thống kê, góp phần nâng cao khả năng kết nối với môn học chuyên ngành về kinh tế, quản trị kinh doanh, và nhiều lĩnh vực khác. Tuy nhiên, không phải sinh viên nào cũng có nền tảng tốt về toán học, dẫn đến khó khăn trong việc hiểu và vận dụng các kiến thức thống kê phức tạp. Trên thực tế, nhiều sinh viên cho rằng môn học này khô khan, thiên về lý thuyết và khó áp dụng vào thực tế. Điều này tạo ra một khoảng cách giữa kiến thức lý thuyết và kỹ năng thực hành, khiến việc dạy và học trở nên kém hiệu quả. Việc sử dụng các phần mềm chuyên sâu về thống kê như SPSS, R hay Stata có thể giúp ích cho việc giảng dạy, nhưng đòi hỏi sinh viên phải có thời gian làm quen và có thể gây khó khăn đối với những người mới tiếp cận. Trong bối cảnh đó, Microsoft Excel với giao diện trực quan, dễ sử dụng, và phổ biến trong môi trường làm việc thực tế đã trở thành một giải pháp khả thi để hỗ trợ giảng dạy.

2.2. Ứng dụng Microsoft Excel vào phân tích hồi quy

Phân Thống kê ở môn học Xác suất thống kê gồm các nội dung về lý thuyết chọn mẫu, tính toán các tham số đặc trưng của mẫu, ước lượng các tham số thống kê, kiểm định giả thiết thống kê, phân tích tương quan và hồi quy và sâu hơn là phân tích phương sai (Anova). Microsoft Excel là một công cụ mạnh mẽ và phổ biến, giúp thực hiện các phép tính

phức tạp, bao gồm cả phân tích hồi quy. Dưới đây là một số ứng dụng cụ thể của Excel vào phân tích hồi quy trong giảng dạy Xác suất thống kê tại trường Đại học Tài nguyên và Môi trường Hà Nội:

Ví dụ 1: Phân tích hồi quy tuyến tính đơn giản

Giả sử giảng viên muốn dạy sinh viên về mối quan hệ giữa số giờ học và kết quả thi của một nhóm sinh viên. Dữ liệu được thu thập như sau:

Bảng 2.1. Dữ liệu về số giờ học và điểm thi tương ứng:

Số giờ học	2	3	5	7	9	4	6	8	10	1	11	12	13	14	15
Điểm thi	50	55	65	70	85	60	68	75	90	45	92	95	97	99	100

Trong Excel, sinh viên sẽ nhập dữ liệu này vào một bảng tính. Sau đó, sử dụng công cụ **“Data Analysis”** và chọn **“Regression”**, Excel sẽ tính toán và cho ra kết quả về hệ số hồi quy, hệ số xác định R^2 , và các giá trị kiểm định ý nghĩa thống kê. Hệ số hồi quy cho biết sự thay đổi của điểm thi khi số giờ học tăng thêm một giờ. Giá trị R^2 cho biết mức độ giải thích của mô hình đối với sự biến thiên của biến phụ thuộc (điểm thi). Bài học này giúp sinh viên hiểu được mối liên hệ giữa các biến và cách dự báo điểm thi dựa trên số giờ học.

Các kết quả chi tiết từ phân tích hồi quy được trình bày dưới đây:

OLS Regression Results

```

=====
Dep. Variable: ExamScore R-squared: 0.977
Model: OLS Adj. R-squared: 0.976
Method: Least Squares F-statistic: 561.4
Date: Wed,09Oct2024 Prob(F-statistic): 4.43e-12
Time: 10:48:16 Log-Likelihood: -36.431
No. Observations: 15 AIC: 76.86
Df Residuals: 13 BIC: 78.28
Df Model: 1
Covariance Type: nonrobust
=====
coef std err t P>|t| [0.025 0.975]
-----
const 43.0000 1.602 26.839 0.000 39.539 46.461
Study Hours 4.1750 0.176 23.693 0.000 3.794 4.556
=====
Omnibus: 0.254 Durbin-Watson: 1.996
Prob(Omnibus): 0.881 Jarque-Bera(JB): 0.104
Skew: 0.159 Prob(JB): 0.949
Kurtosis: 2.746 Cond. No. 19.3
=====

```

Dựa trên phân tích hồi quy, phương trình hồi quy nhận được là:

$$\text{Điểm Thi} = 43.000 + 4.175 * \text{Số Giờ Học}$$

Điều này có nghĩa là mỗi giờ học thêm sẽ làm tăng điểm thi trung bình khoảng 4.18 điểm.

Ví dụ 2: Phân tích hồi quy tuyến tính bội

Giả sử giảng viên muốn minh họa một ví dụ phức tạp hơn, như phân tích mối quan hệ giữa lợi nhuận của một công ty và các yếu tố ảnh hưởng như giá, chi phí quảng cáo, chi phí quản lý, chi phí bán hàng. Dữ liệu mẫu có thể như sau:

Bảng 2.2. Lợi nhuận Công ty.

Lợi nhuận y (triệu VND)	520	530	535	525	545	550	545	548	560
Giá x1	0.44	0.51	0.54	0.55	0.51	0.525	0.555	0.56	0.57
Chi phí quản lý x2	29.5	31.0	28.2	27.5	30.5	29.8	28.5	28.5	33.5
Chi phí bán hàng x3	19.5	20.5	21.8	23.8	25.8	23.0	22.0	23.5	24.0

Trong Excel, sinh viên sẽ nhập dữ liệu này vào một bảng tính. Sau đó, sử dụng công cụ **“Data Analysis”** và chọn **“Regression”**, Excel sẽ tính toán và cho ra kết quả về bảng phân tích các hệ số hồi quy:

```

=====
coef std err t P>|t| [0.025 0.975]
-----
const 314.0662 65.991 4.759 0.005 144.432 483.701
x1 164.0673 95.368 1.720 0.146 -81.083 409.217
x2 3.3377 1.731 1.928 0.112 -1.112 7.788
x3 1.7621 1.983 0.889 0.415 -3.336 6.860
=====

```

Dựa trên kết quả phân tích hồi quy cho dữ liệu lợi nhuận trên, có thể nhận thấy rằng các biến giá sản phẩm, chi phí quản lý và chi phí bán hàng đều có tác động tích cực đến lợi nhuận của công ty. Hệ số hồi quy của các biến này đều dương, cho thấy rằng khi giá sản phẩm và các chi phí này tăng, lợi nhuận cũng có xu hướng tăng theo. Mô hình hồi quy đã giải thích khá tốt biến động của lợi nhuận, và phân tích ANOVA chỉ ra rằng các biến độc lập này có ý nghĩa thống kê, phù hợp để giải thích lợi nhuận.

Với mô hình ban đầu, phương trình hồi quy là:

$$y = 314.066 + 164.067 * x_1 + 3.338 * x_2 + 1.762 * x_3.$$

Lưu ý rằng hệ số xác định $R^2 = 0.7142$, P-value của biến x1 là 0.1460.

Do P-value của biến x1 lớn hơn 0.05, chúng ta loại bỏ biến này và phân tích hồi quy rút gọn chỉ với biến x2 và x3, ra kết quả về bảng phân tích các hệ số hồi quy:

```

=====
coef std err t P>|t| [0.025 0.975]
-----
const 365.5225 67.752 5.395 0.002 199.740 531.305

```

x2	3.0623	1.985	1.542	0.174	-1.796	7.920
x3	3.6815	1.888	1.950	0.099	-0.939	8.302

Khi đó, phương trình hồi quy sau khi rút gọn cho Bảng 2 là: $y = 365.523 + 3.062 * x_2 + 3.681 * x_3$.

Ví dụ 3: Phân tích hồi quy phi tuyến tính

Giảng viên có thể dạy sinh viên về hồi quy phi tuyến tính bằng cách sử dụng Excel để phân tích mối quan hệ không tuyến tính giữa các biến. Ví dụ, mô hình hàm bậc hai giữa giá sản phẩm và lượng hàng bán được có thể như sau:

Bảng 2.3. Số lượng hàng bán

Giá sản phẩm (triệu VND)	10	15	20	25	30	35	40	45	50	55
Lượng hàng bán (số lượng)	200	180	160	150	145	140	130	125	120	115

Để phân tích mối quan hệ phi tuyến giữa giá sản phẩm và lượng hàng bán trong Excel, giảng viên hướng dẫn sinh viên thực hiện các bước sau:

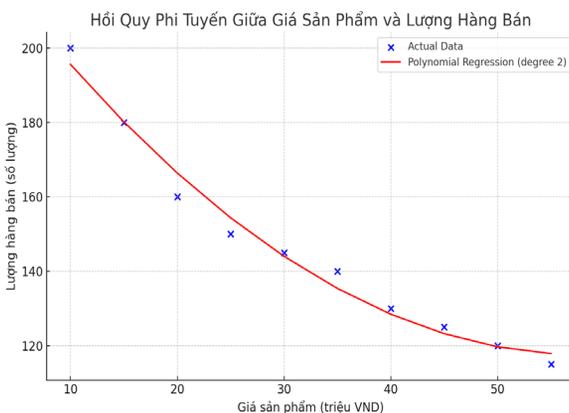
+ Nhập dữ liệu vào Excel → Vào mục “Insert” → chọn “Scatter Plot” để vẽ đồ thị phân tán của giá sản phẩm và lượng hàng bán.

+ Sau khi vẽ biểu đồ, nhấp chuột phải vào một điểm dữ liệu và chọn “Add Trendline”.

+ Trong cửa sổ “Format Trendline”, chọn “Polynomial” để kiểm tra mối quan hệ phi tuyến. Bạn có thể chọn bậc 2 hoặc bậc 3 tùy thuộc vào dữ liệu → chọn “Display Equation on chart” để hiển thị phương trình hồi quy phi tuyến trên biểu đồ.

+ Kiểm tra hệ số hồi quy và phương trình nhận được để đánh giá mối quan hệ giữa giá sản phẩm và lượng hàng bán.

Biểu đồ hồi quy phi tuyến nhận được là:



Phương trình hồi quy phi tuyến bậc 2 nhận được là:

Lượng hàng bán = $231.61 + -3.94 * (\text{Giá sản phẩm}) + 0.03 * (\text{Giá sản phẩm}^2)$

Ý nghĩa của các hệ số trong phương trình như sau:

- **Hệ số intercept (231.61):** Đây là giá trị lượng hàng bán khi giá sản phẩm bằng 0, nhưng trong thực tế điều này không có ý nghĩa nhiều vì giá sản phẩm không thể bằng 0.

- **Hệ số -3.94:** Mỗi khi giá sản phẩm tăng lên 1 triệu VND, lượng hàng bán giảm đi khoảng 3.94 đơn vị. Điều này phản ánh rằng khi giá sản phẩm tăng, lượng hàng bán sẽ giảm.

- **Hệ số 0.03:** Đây là hệ số bậc hai của phương trình hồi quy, biểu thị rằng có một mối quan hệ phi tuyến nhỏ giữa giá sản phẩm và lượng hàng bán. Hệ số này nhỏ, cho thấy rằng sự phi tuyến này không quá mạnh.

Từ biểu đồ và phương trình hồi quy, chúng ta thấy rằng mối quan hệ giữa giá sản phẩm và lượng hàng bán không hoàn toàn tuyến tính. Mặc dù có sự giảm dần của lượng hàng bán khi giá sản phẩm tăng, mô hình bậc hai cho thấy sự giảm dần này không hoàn toàn đều đặn.

Giá trị R-squared (0.97) cho thấy mô hình hồi quy bậc hai này có độ phù hợp tốt với dữ liệu, tức là mô hình giải thích được 97% sự thay đổi của lượng hàng bán dựa trên giá sản phẩm.

Tuy nhiên, để có dự đoán chính xác hơn, có thể cần thêm các yếu tố khác ảnh hưởng đến lượng hàng bán, như chi phí marketing, nhu cầu thị trường, hoặc các chương trình khuyến mãi.

3. Kết luận

Những ví dụ trên cho thấy rằng Excel không chỉ hữu ích trong phân tích hồi quy tuyến tính đơn giản mà còn có thể ứng dụng cho nhiều loại mô hình hồi quy phức tạp hơn, bao gồm hồi quy bội, hồi quy phi tuyến tính, ... Việc kết hợp bộ công cụ Data Analysis trong Excel trong giảng dạy Xác suất thống kê giúp sinh viên tiếp cận dễ dàng hơn với các khái niệm thống kê phức tạp và bước đầu trang bị cho họ những kỹ năng thực hành quan trọng cho công việc sau này.

Tài liệu tham khảo

[1] Nguyễn Cao Văn, Trần Thái Ninh, 1996, *Lý thuyết xác suất và thống kê toán*, NXB Khoa học kỹ thuật.
 [2] Đặng Hùng Thắng, 2008, *Thống kê và ứng dụng*, NXB Giáo dục.
 [3] Đào Hữu Hồ, 1998, *Xác suất và thống kê*, NXB ĐHQG.
 [4] Bộ công cụ Microsoft Excel