

NGHIÊN CỨU ẢNH HƯỞNG CỦA MỘT SỐ YẾU TỐ ĐẾN MỨC CHI CHO SẢN PHẨM THỜI TRANG CỦA NAM TỪ 25 - 55 TUỔI TẠI THÀNH PHỐ HÀ NỘI

STUDY THE INFLUENCE OF A NUMBER OF FACTORS TO THE LEVEL OF SPENDING ON MEN'S FASHION PRODUCTS FROM 25 TO 55 YEARS OLD IN HANOI CITY

Phạm Thị Thảo¹, Nguyễn Thu Hà¹, Nguyễn Thị Thu Phương¹,
Phạm Hồng Thu¹, Vũ Thị Vân¹, Nguyễn Thị Lệ^{2,*}

TÓM TẮT

Mức chi cho sản phẩm thời trang là số tiền một người chi cho việc mua các sản phẩm thời trang cho bản thân trong một khoảng thời gian xác định. Có nhiều yếu tố tác động đến mức chi cho sản phẩm thời trang nhưng vẫn chưa được đề cập tới một cách đầy đủ, đặc biệt là từng nhóm khách hàng mục tiêu. Bài báo trình bày những kết quả nghiên cứu xác định ảnh hưởng của một số yếu tố thuộc về người tiêu dùng tới mức chi cho sản phẩm thời trang của nam giới trong độ tuổi từ 25 - 55 tuổi tại thành phố Hà Nội.

Từ khóa: Mức chi cho thời trang, kỹ thuật BMA, phân tích PCA, K-means cluster.

ABSTRACT

The level spent on fashion products is the amount of money a person spends on buying fashion products for themselves over a specified period of time. There are many factors that affect the level of spending on fashion products but have not been fully mentioned, especially each target group of customers. The paper presents the results of the study to determine the influence of a number of consumer factors to the level of spending on fashion products of men aged 25-55 years in Hanoi city.

Keywords: Levels spent on fashion, BMA engineering, PCA analysis, K-means cluster.

¹Lớp ĐH Công nghệ May 01 - K13, Khoa Công nghệ May và Thiết kế thời trang, Trường Đại học Công nghiệp Hà Nội

²Khoa Công nghệ May và Thiết kế thời trang, Trường Đại học Công nghiệp Hà Nội

*Email: le.nguyenthi@hau.edu.vn

1. GIỚI THIỆU

Mức chi cho sản phẩm thời trang là số tiền một người chi cho việc mua các sản phẩm thời trang cho bản thân trong một khoảng thời gian xác định như tháng hoặc năm. Việc xác định được mối quan hệ và mức độ giữa mức chi cho sản phẩm thời trang và các yếu tố ảnh hưởng đến mức chi của người tiêu dùng rất quan trọng đối với các doanh nghiệp, cho phép doanh nghiệp hiểu rõ hơn các phân nhóm khách hàng mục tiêu. Từ đó, có thể tìm ra các giải pháp đáp ứng tốt

hơn trong sản xuất và kinh doanh phù hợp với nhu cầu tiêu dùng sản phẩm thời trang của nhóm khách hàng. Đã có các nghiên cứu đề cập đến mức ảnh hưởng của tiếp thị sản phẩm thời trang đến nhu cầu người tiêu dùng, nhu cầu thị trường của sản phẩm thời trang, đặc điểm tiêu dùng sản phẩm thời trang của nhóm khách hàng mục tiêu như Samiyah Jamal và cộng sự đã nghiên cứu về ảnh hưởng của tiếp thị thời trang đến nhu cầu của người tiêu dùng [1]. Tác giả Nguyễn Thị Mai Hoa đã nghiên cứu đặc điểm tiêu dùng sản phẩm thời trang công sở của phụ nữ từ 25 - 55 tuổi trên thị trường thành phố Hà Nội [2, 3]. Các yếu tố ảnh hưởng tới hành vi mua hàng, thị hiếu của người tiêu dùng cũng đã được nghiên cứu [4, 5, 6]. Việc xác định các yếu tố ảnh hưởng đến mức chi cho sản phẩm thời trang là rất cần thiết trong bối cảnh xuất khẩu hàng may mặc đình trệ bởi dịch Covid 19 và hướng tới phát triển sản phẩm phục vụ thị trường nội địa tại Việt nam nhưng vẫn chưa được đề cập một cách đầy đủ, đặc biệt là với nhóm khách hàng mục tiêu là nam giới trên thị trường thành phố Hà nội có nhiều tiềm năng. Nghiên cứu này nhằm xác định những đặc trưng quan trọng để phân nhóm và khám phá đặc tính phân nhóm dữ liệu đối tượng khảo sát về ảnh hưởng của một số yếu tố liên quan tới người tiêu dùng tới mức chi cho sản phẩm thời trang của nam từ 25 - 55 tuổi tại thành phố Hà Nội.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu của đề tài là mức chi cho sản phẩm thời trang của nam giới từ 25 - 55 tuổi trên địa bàn thành phố Hà nội và các yếu tố có liên quan tới cá nhân người tiêu dùng như mức thu nhập, nghề nghiệp, tình trạng gia đình, trình độ, giới tính,... được lựa chọn để nghiên cứu thông qua phỏng vấn trực tiếp và điều tra khảo sát mức chi cho sản phẩm thời trang từ 1/5/2020 đến 30/4/2021. Thời gian khảo sát khách hàng: 1/5-15/5/2021.

2.2. Phương pháp nghiên cứu định tính

Phương pháp nghiên cứu định tính được áp dụng trong nghiên cứu này nhằm thu thập dữ liệu xác định sơ bộ các

yếu tố ảnh hưởng tới mức chi cho sản phẩm thời trang của nam giới tại thành phố Hà Nội, được thực hiện thông qua việc phỏng vấn trực tiếp và trực tuyến với 10 chuyên gia về lĩnh vực sản xuất và kinh doanh thời trang trên địa bàn Hà Nội. Kết quả khảo sát định tính cho thấy, hầu hết những người được phỏng vấn đều cho rằng các yếu tố như mức thu nhập, nghề nghiệp, độ tuổi, tình trạng hôn nhân gia đình, số con, phong cách sống,... có ảnh hưởng đến mức chi cho sản phẩm thời trang của nam giới.

2.3. Phương pháp nghiên cứu định lượng

Để đạt được mục tiêu của nghiên cứu này là xác định được ảnh hưởng của một số yếu tố liên quan tới các nhân người tiêu dùng tới mức chi cho sản phẩm thời trang của nam giới ở nhóm đã chọn, phiếu khảo sát đã được thiết lập với các câu hỏi dựa trên cơ sở kết quả phỏng vấn ở nghiên cứu định tính. Phiếu hỏi được hoàn thiện sau khi tiến hành khảo sát thử và kiểm tra dữ liệu thu được trên 20 người thuộc nhóm đối tượng nghiên cứu. Phiếu khảo sát chính thức được phát và thu thập dữ liệu tại các địa điểm thuộc thành phố Hà Nội như trên bảng 1. Cỡ mẫu khảo sát được tính theo công thức [7]:

$$n = \frac{N}{1+N(e^2)}$$

Trong đó, n là số lượng mẫu tối thiểu cần khảo sát

N là tổng số nam giới trong độ tuổi 25 - 55 tại TP Hà Nội

e là độ lệch chuẩn; chọn e = 0,1; khi đó:

$$n = \frac{N}{1+N(e^2)} = \frac{3653000}{1+3653000(0,1^2)} = 99$$

Bảng 1. Các địa điểm khảo sát

STT	Đơn vị	Địa chỉ	Số lượng
1	Trường ĐHCN Hà Nội	Số 298 Cầu Diễn, Bắc Từ Liêm, Hà Nội	15
2	Vincom Mega Mall Royal City	72A Nguyễn Trãi, Thượng Đình, Thanh Xuân, Hà Nội	36
3	Chợ Nhà Xanh	Dịch Vọng Hậu, Cầu Giấy, Hà Nội	13
4	Aeon Mall Hà Đông	Phường Dương Nội, Hà Đông, Hà Nội	20
5	Indochina Plaza Hanoi Residences	241 Xuân Thủy, Dịch Vọng Hậu, Cầu Giấy, Hà Nội	14
6	Chợ Cầu Diễn	Cầu Diễn, Phường Cầu Diễn, Nam Từ Liêm, Hà Nội	12
Tổng			110

Các biến số được khảo sát của đối tượng là biến liên tục (như mức thu nhập, mức chi cho thời trang, tuổi), hoặc biến phân loại (trình độ, tình trạng gia đình, nghề nghiệp,...). Áp dụng thang đo Likert 5 điểm (với 1 là không ảnh hưởng và 5 là rất ảnh hưởng) để nghiên cứu định lượng về các yếu tố ảnh hưởng đến mức chi cho sản phẩm thời trang của nam giới trên thị trường thành phố Hà Nội. Các biến số phân loại được mã hóa như trên bảng 2.

Bảng 2. Mã hóa các biến phân loại

Yếu tố	Các phương án lựa chọn	Tên biến và thang đo
Trình độ học vấn		lev
	Trung học phổ thông	1
	Cao đẳng, đại học	2
Nghề nghiệp	Trên đại học	3
		oc
	Buôn bán, lao động tự do	1
	Công nhân, lái xe	2
	Giáo viên, công nhân viên chức	3
Nhân viên văn phòng, kinh doanh	4	
Kỹ sư	5	
Số con		con
	0 con	0
	1 con	1
	2 con	2
	3 con	3
Nhiều hơn 3 con	4	
Tình trạng gia đình		mar
	Độc thân	1
Đã kết hôn	2	
Tuổi	Tuổi của người được khảo sát	age
Thu nhập	Thu nhập một năm của người được khảo sát (triệu đồng)	inc
Mức chi	Mức chi cho thời trang cá nhân một năm của người được khảo sát (triệu đồng)	outc

2.4. Xác định mô hình tối ưu bằng kỹ thuật BMA

Xác định mô hình đa biến tuyến tính tối ưu thể hiện mối quan hệ giữa mức chi cho thời trang (outc) với các biến tuổi (age), thu nhập (inc), số con (con), tình trạng hôn nhân (mar), nghề nghiệp (oc), trình độ học vấn (lev) của đối tượng bằng kỹ thuật BMA trên phần mềm R.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Trong đó, Y là biến phụ thuộc (mức chi cho thời trang), X_i là biến độc lập (tuổi, thu nhập, nghề nghiệp, số con,...); β₀, β₁, β₂,... là các hệ số hồi quy được xác định từ dữ liệu thực nghiệm.

Để chọn mô hình tối ưu có thể dựa vào các tiêu chuẩn như xem xét hệ số xác định R², RMS (Residual Mean Square), chỉ số Mallows' Cp hoặc AIC/BIC (Akaike Information Criterion/Bayesian Information Criterion) của các mô hình. Tuy nhiên, không có tiêu chuẩn nào được coi là tốt nhất [8].

Khi xem xét hệ số xác định R², các nghiên cứu đã cho thấy rằng hệ số xác định sẽ tăng khi có thêm biến trong mô hình. Mô hình có hệ số xác định R² cao nhất là mô hình có tất cả

các biến đầu vào. Tuy nhiên, khi đó sẽ rất tốn kém khi xác định tất cả các biến đầu vào. Điều này còn có thể dẫn đến tình trạng quá nhiều biến đầu vào hơn là cần thiết. Mặc dù vậy, hệ số xác định R^2 có thể được dùng để so sánh giữa các mô hình có cùng số biến đầu vào để lựa chọn mô hình và khi đó, cần điều chỉnh R^2 để phạt đối với những mô hình phức tạp có quá nhiều biến.

Chỉ số AIC (Akaike Information Criterion) và BIC (Bayesian Information Criterion) là những chỉ số được khuyến cáo là nên dùng để đánh giá và lựa chọn mô hình tối ưu. AIC là chỉ số được kỹ sư Akaike đề xuất, dùng để đánh giá mức độ tối ưu của các mô hình hồi quy tuyến tính tìm thấy trên cơ sở phân tích dữ liệu. BIC là chỉ số tương tự như AIC. Các chỉ số này được coi là những thước đo để lựa chọn mô hình tối ưu thể hiện mối quan hệ tuyến tính giữa các biến đầu vào và đầu ra và được khuyến cáo nên dùng bởi các nhà thống kê. Đây là thước đo để cân đối tính phức tạp (số biến) và tính tối ưu của mô hình (thông qua RSS). Các chỉ số AIC và BIC càng thấp có nghĩa là mô hình càng tốt.

BMA (Bayesian Model Average) là một kỹ thuật lựa chọn mô hình tối ưu đa biến. Phương pháp tìm kiếm mô hình tối ưu dựa trên BMA đang thu hút được rất nhiều sự quan tâm ứng dụng trong xử lý số liệu và thống kê. Phương pháp này dùng BIC làm tiêu chuẩn chọn mô hình tốt nhất nhưng quá trình tính toán dùng hết nhiều dung lượng bộ nhớ của máy tính nếu số biến đầu vào lớn. Khi sử dụng BMA, kết quả thu được nhiều mô hình khác nhau chứ không chỉ là 1 mô hình duy nhất. Người dùng có thể lựa chọn mô hình phù hợp nhất để giải thích dữ liệu thu được trong các mô hình được đưa ra.

2.5. Phân tích thành phần PCA và phân nhóm K-means clustering

Phân tích thành phần PCA

Để xác định mức độ ảnh hưởng của các biến số khảo sát tới sự phân nhóm, phân tích thành phần PCA đã được áp dụng. Phương pháp này cho phép tìm ra ma trận W "tốt nhất" gồm p vector trực giao từng đôi một dựa trên phép biến đổi trực giao. Việc này thực chất có mục đích chuyển đổi dữ liệu từ không gian ban đầu với p trục tọa độ sang một hệ trục mới có ít chiều hơn (d trục tọa độ, $d < p$) sao cho thông tin được bảo toàn tốt nhất.

Trong nghiên cứu này, vì dữ liệu khảo sát được về mức chi cho thời trang và các yếu tố chưa phân nhóm, cho nên PCA là một trong những giải thuật phù hợp để trích chọn các đặc trưng và được thực hiện như sau [9]:

Bước 1: Tiến xử lý dữ liệu bằng phương pháp Z-Scores.

Bước 2: Tính ma trận hiệp phương sai $C(p \times p)$ của các thuộc tính trong X

$$C_{ij} = \frac{1}{n-1} \sum_{t=1}^n (X_t^i - \mu_i)(X_t^j - \mu_j)$$

Sau khi được chuẩn hóa bằng Z-Scores, các giá trị trung bình trên từng cột của X đều bằng 0. Do vậy, công thức tính C_{ij} sẽ trở thành:

$$C_{ij} = \frac{1}{n-1} \sum_{t=1}^n X_t^i X_t^j = \frac{1}{n-1} X^T X$$

Bước 3: Tìm tất cả các giá trị riêng $\lambda_1, \lambda_2, \dots, \lambda_p$ và các vector riêng tương ứng u_1, u_2, \dots, u_p của ma trận C (các giá trị riêng λ_i được sắp giảm dần).

Bước 4: Chọn ra d véc tơ riêng đầu tiên u_1, u_2, \dots, u_d để thu được tập W thuộc $R^{d \times p}$

Bước 5: Tính $Y = X * W^T$ để thu được bộ dữ liệu mới.

Phân cụm K-means clustering

K-means clustering là quá trình phân nhóm đối tượng đã cho vào K cụm (K là số các cụm được xác định trước, K nguyên dương) sao cho tổng bình phương khoảng cách giữa các đối tượng đến tâm là nhỏ nhất. Thuật toán phân cụm này khá phổ biến và có độ phức tạp tính toán phù hợp với các bài toán kích thước lớn. Thuật toán phân nhóm K-means bao gồm các bước sau [10]:

Bước 1: Chọn ngẫu nhiên K điểm làm tâm cho K nhóm từ tập n điểm đã cho. Mỗi nhóm được đại diện bằng tâm các nhóm.

Bước 2: Tính khoảng cách từ mỗi điểm $A(a_x, a_y)$ đến K tâm $C(c_x, c_y)$ của từng nhóm theo công thức:

$$d(A, C)^2 = (a_x - c_x)^2 + (a_y - c_y)^2$$

Bước 3: Đối với mỗi điểm A , sẽ được gán vào cụm có khoảng cách ngắn nhất từ trọng tâm cụm đó đến A .

Bước 4: Xác định lại tâm mới cho các nhóm bằng cách lấy trung bình cộng tọa độ tất cả các điểm đã được gán vào nhóm. Tính lại tọa độ trọng tâm cho mỗi nhóm:

$$c_x = \frac{1}{m} \sum_{i=1}^m (a_{1x} + a_{2x} + \dots + a_{mx})$$

$$c_y = \frac{1}{m} \sum_{i=1}^m (a_{1y} + a_{2y} + \dots + a_{my})$$

Trong đó, m là số lượng phần tử của cụm.

Bước 5: Thực hiện lại bước 2 cho đến khi nào không có sự thay đổi nhóm nào của các đối tượng thì thuật toán dừng lại.

Chọn số cụm tối ưu

Để phân nhóm mức chi cho thời trang dựa trên những tương đồng của các đặc tính có liên quan đến mức chi là các yếu tố ảnh hưởng, cần xác định K tối ưu hay số cụm tối ưu. Việc chọn số lượng cụm tối ưu mà dữ liệu có thể được nhóm lại là quan trọng với kỹ thuật K-means clustering và thuật toán học không giám sát.

Có nhiều phương pháp để xác định K tối ưu. 2 phương pháp thường được áp dụng là Elbow method và Gap Statistic. Trong đó, phương pháp Gap Statistic được nhiều nhà nghiên cứu sử dụng bởi khách quan hơn. Phân nhóm K-means clustering được thực hiện sao cho tổng bình phương khoảng cách trong cụm WSS (Within-cluster Sum of Square) là nhỏ nhất. Phương pháp Elbow method chọn số cụm k tối ưu sao cho khi thêm vào một cụm khác thì không làm cho WSS thay đổi nhiều bằng cách:

- Xác định độ méo (*average distortion*) bằng giá trị trung bình của các khoảng cách bình phương từ các tâm cụm của các cụm tương ứng.

- Xác định quán tính là tổng bình phương khoảng cách của các mẫu đến trung tâm cụm gần nhất của chúng.

Để xác định số lượng cụm tối ưu, chọn giá trị của K là "khuyết tay", tức là điểm mà sau đó biến dạng/quán tính bắt đầu giảm tuyến tính.

Phương pháp Gap Statistic cho kết quả khách quan hơn nhưng phức tạp hơn. Để tìm k tối ưu, với mỗi giá trị K thì giá trị $G = WSS/WSS$ mong đợi (kỳ vọng) được tính dựa vào giả thuyết vô hiệu (không có cluster nào). K tối ưu khi giá trị G được tối đa hóa (đạt giá trị lớn nhất). Để xác định K tối ưu bằng phương pháp Gap Statistic, dùng giải thuật Bootstrap để tính W của dữ liệu và từ đó tính G. K tối ưu được chọn khi G nằm trong 1 độ lệch chuẩn của tương ứng với giá trị tính được với K và K+1 [11].

3. KẾT QUẢ VÀ BÀN LUẬN

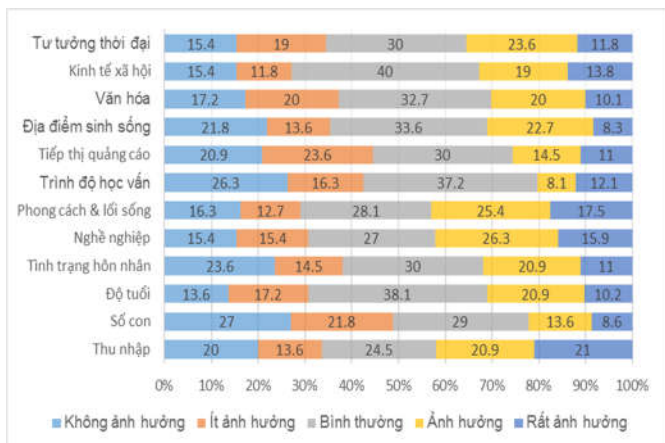
3.1. Kết quả khảo sát về các yếu tố ảnh hưởng đến mức chi cho sản phẩm thời trang của nam giới từ 25 - 55 tuổi tại thành phố Hà Nội.

Kết quả khảo sát được tổng hợp và thể hiện trên hình 1.

- Về mức độ không ảnh hưởng, yếu tố trình độ học vấn và số con có phần trăm lớn nhất là 26,3% và 27%, yếu tố độ tuổi có phần trăm ít nhất là 13,6%, 3 yếu tố tư tưởng thời đại; kinh tế xã hội và nghề nghiệp đều có 15,4%.

- Về mức độ ít ảnh hưởng, yếu tố tiếp thị quảng cáo có phần trăm lớn nhất với 23,6%, yếu tố kinh tế xã hội có phần trăm nhỏ nhất với 11,8%.

- Về mức độ ảnh hưởng bình thường, yếu tố kinh tế xã hội có phần trăm lớn nhất với 40%, chứng tỏ yếu tố này làm ảnh hưởng nhiều đến mức chi cho sản phẩm thời trang của đối tượng khảo sát, sau đó là yếu tố độ tuổi và trình độ học vấn với 38,1% và 37,2%, có phần trăm ảnh hưởng ít nhất là thu nhập và nghề nghiệp với 24,5% và 27%.



Hình 1. Mức độ ảnh hưởng của các yếu tố đến mức chi cho sản phẩm thời trang theo đánh giá của người được khảo sát

- Về mức độ ảnh hưởng, yếu tố nghề nghiệp và phong cách lối sống có phần trăm ảnh hưởng lớn nhất là 26,3% và

25,4%, 3 yếu tố tình trạng hôn nhân; độ tuổi và thu nhập đều có 20,9%, phần trăm ảnh hưởng ít nhất là trình độ học vấn với 8,1%.

- Về mức độ rất ảnh hưởng, yếu tố ảnh hưởng có phần trăm lớn nhất với 21%, chứng tỏ yếu tố này ảnh hưởng lớn nhất tới mức chi cho sản phẩm thời trang, yếu tố địa điểm sinh sống và số con có phần trăm ít nhất là 8,3% và 8,6%.

3.2. Mối quan hệ giữa mức chi cho thời trang và các yếu tố ảnh hưởng

Kết quả khảo sát về các thông số sau khi mã hóa được trình bày với 7 biến có giá trị số gồm tuổi (age), nghề nghiệp (oc), trình độ học vấn (lev), tình trạng gia đình (mar), mức thu nhập (inc), số con (con) và mức chi cho SPTT (outc) và thể hiện trên bảng 3.

Bảng 3. Đặc trưng thống kê của các biến liên tục

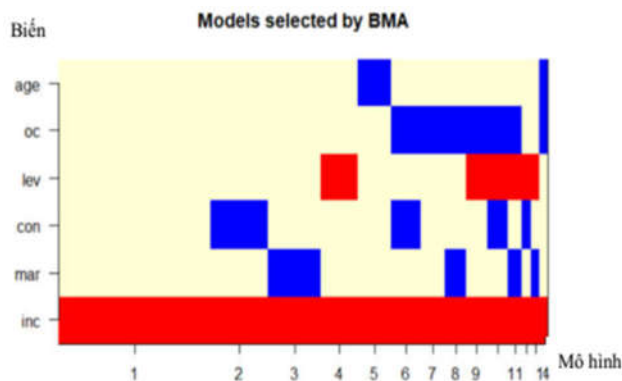
STT	Kích thước	Giá trị nhỏ nhất	Giá trị lớn nhất	Giá trị trung bình	Độ lệch chuẩn	Hệ số biến thiên	Trung vị
1	age (tuổi)	25	50	30,98	6,33	4,9	29
2	inc (triệu đồng/năm)	48	420	140,9	81,43	60,29	120
3	outc (triệu đồng/năm)	2,5	96	15,01	17,113	10,55	9,55

Tuổi trung bình của các đối tượng khảo sát là 31, có thu nhập trung bình một năm là 140,9 triệu đồng, có mức chi cho thời trang cá nhân trung bình là 15,01 triệu đồng, chiếm 10,65 % mức thu nhập.

Xử lý dữ liệu thu thập được trên phần mềm R thu được 5 mô hình tốt nhất thể hiện mối quan hệ giữa mức chi cho thời trang của nam giới *outc* và các thông số đầu vào đã được tìm thấy. Từ đó, lựa chọn mô hình tối ưu như sau:

$$outc = -8,168 - 1,63 * con + 0,171 * inc$$

$$R^2 = 0,671, BIC = -113,01$$

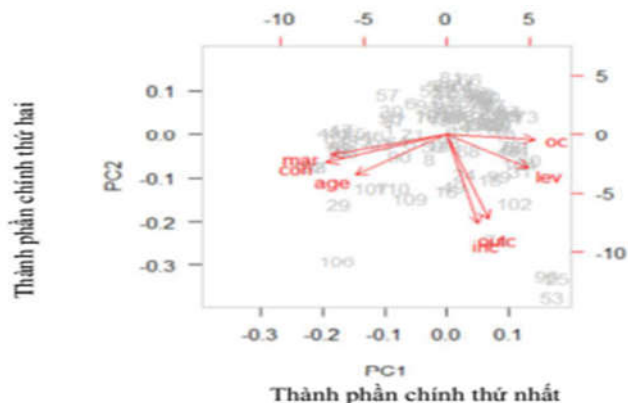


Hình 2. Biểu đồ BMA của các mô hình tối ưu

Sự biến thiên của tuổi (age), tình trạng gia đình (mar), số con (con) và thu nhập (inc) giải thích được 67,1% sự biến thiên của mức chi cho thời trang (outc) của nam giới. Mức chi cho thời trang (outc) tỷ lệ thuận với thu nhập (inc) và tỷ lệ nghịch với số con (con).

Khi thu nhập tăng thêm 1 triệu đồng thì nam giới tăng 0,171 triệu đồng chi cho thời trang. Nhưng khi tăng thêm 1 con thì mức chi cho thời trang của nam giới tại TP Hà nội có xu hướng giảm 1,63 triệu đồng.

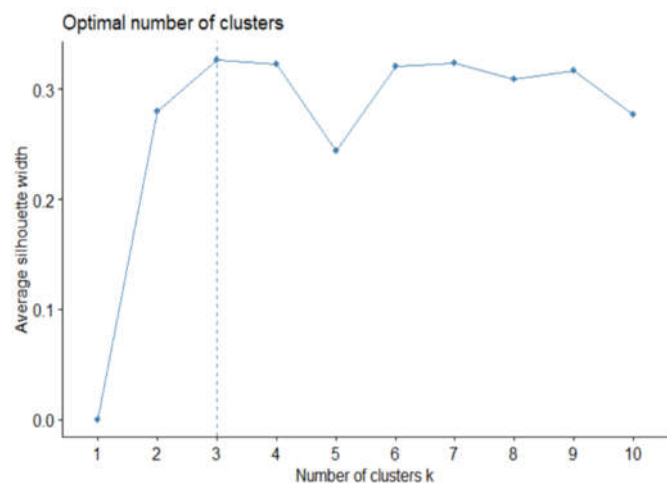
3.3. Kết quả phân tích PCA và phân nhóm Cluster



Hình 3. Kết quả phân tích thành phần chính PCA

Kết quả phân tích thành phần chính PCA cũng cho thấy các biến mức chi cho sản phẩm thời trang (outc), mức thu nhập (inc), số con (con) có mối quan hệ mạnh với thành phần chính gây nên sự khác biệt giữa các đối tượng trong nhóm. Tuổi (age), trình độ học vấn (lev) và nghề nghiệp (oc) có mối liên hệ yếu hơn (hình 4).

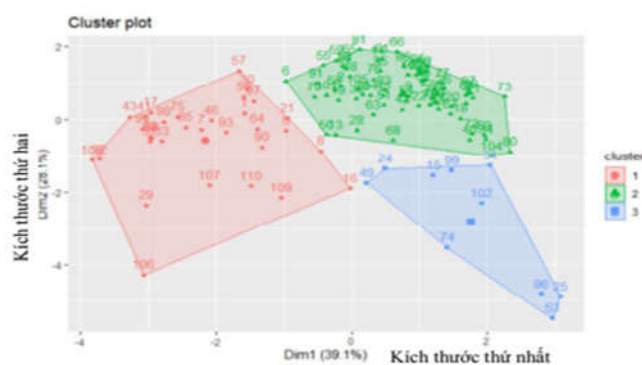
Chỉ số Hopkins của dữ liệu khảo sát = 0,759 gần với 1, do đó, dữ liệu này có thể ứng dụng phân tích cluster hay phân nhóm được một cách khả thi.



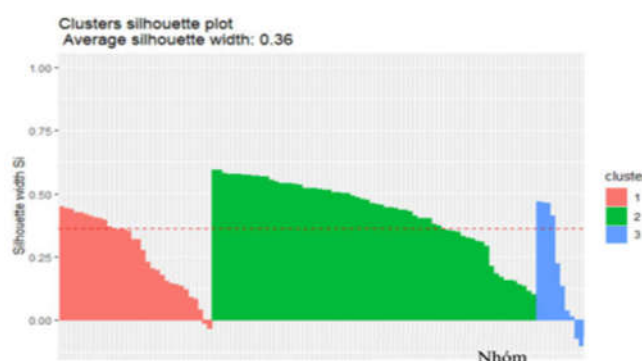
Hình 4. Kết quả xác định K tối ưu bằng phương pháp Silhouette

Từ kết quả xác định hệ số K tối ưu trên hình 4, chọn số nhóm K = 3. Các đối tượng khảo sát chia làm 3 nhóm và hiển thị theo đặc trưng chính thứ nhất và thứ 2 trên hình 5.

Các đối tượng khảo sát được chia thành 3 nhóm như trên hình 5. Thông số 1 giải thích được 39,1% tổng số giao động giữa các cluster. Thông số 2 giải thích được 28,1%. 2 thông số này giải thích được 67,2% tổng số dao động giữa các cluster. Để đánh giá sự phân nhóm này, chỉ số SW (Silhouette Withd) được tính cho các nhóm và thể hiện trên hình 6.



Hình 5. Hiển thị đối tượng theo 3 nhóm khi phân nhóm Cluster



Hình 6. Biểu đồ đánh giá sự phân nhóm

Chỉ số SW = 0,36 trên hình 6 cho thấy có thể phân nhóm cho các đối tượng để đáp ứng nhu cầu của họ tốt hơn.

Bảng 4. Các đặc trưng thống kê của 3 nhóm

Tên biến	Lớn nhất	Nhỏ nhất	Trung bình	Trung vị	Độ lệch chuẩn	Khoảng tin cậy 95%
$N_3 = 13$						
Thu nhập inc (triệu đồng/năm)	420	204	301,8	300	78,21	(259,29 - 344,31)
Tuổi age	41	25	30,54	19,00	4,39	(28,16 - 32,92)
Mức chi cho thời trang outc (triệu đồng/năm)	96	15	48,05	43	29,01	(32,28 - 63,82)
$N_2 = 66$						
Thu nhập inc (triệu đồng/năm)	300	48	110,8	108,0	48,96	(98,99 - 122,61)
Tuổi age	44	25	28,62	27,50	3,48	(27,78 - 29,46)
Mức chi cho thời trang outc (triệu đồng/năm)	27	2,5	9,31	8,50	5,52	(7,98 - 10,64)
$N_1 = 31$						
Thu nhập inc (triệu đồng/năm)	360	60	137,4	132,0	58,69	(116,74 - 158,06)
Tuổi age	50	25	36,19	37,00	8,50	(33,20 - 39,18)
Mức chi cho thời trang outc (triệu đồng/năm)	54	3	13,32	9,50	11,12	(9,41 - 17,23)

Tuổi trung bình của các đối tượng thuộc nhóm 1 là 36,19 tuổi có thu nhập trung bình một năm là 137,4 triệu đồng, có mức chi cho thời trang cá nhân trung bình là 13,32 triệu đồng, chiếm 9,69% thu nhập.

Tuổi trung bình của các đối tượng thuộc nhóm 2 là 28,62 tuổi, có thu nhập trung bình một năm là 110,8 triệu đồng, có mức chi cho thời trang cá nhân trung bình là 9,31 triệu đồng, chiếm 8,4% thu nhập.

Tuổi trung bình của các đối tượng thuộc nhóm 3 là 30,54 tuổi, có thu nhập trung bình một năm là 301,8 triệu đồng, có mức chi cho thời trang cá nhân trung bình là 48,05 triệu đồng, chiếm 15,92% thu nhập. Nhóm này có mức chi và thu nhập cao nhất trong 3 nhóm.

4. KẾT LUẬN

Qua kết quả điều tra khảo sát để tìm hiểu về các yếu tố ảnh hưởng đến mức chi cho sản phẩm thời trang có thể thấy sự biến thiên của thu nhập (inc) và số con (con) giải thích được 67,1% sự biến thiên của mức chi cho thời trang (outc) của nam giới. Mức chi cho thời trang (outc) tỷ lệ thuận với thu nhập (inc) và tỷ lệ nghịch với số con (con).

Mối liên hệ giữa mức chi cho thời trang và số con và thu nhập của nam giới được thể hiện bởi: $outc = -8,168 - 1,63*con + 0,171*inc$ với $R^2 = 0,671$.

Kết quả phân tích thành phần chính PCA cũng cho thấy các biến mức chi cho sản phẩm thời trang (outc), mức thu nhập (inc) và số con (con) có mối quan hệ mạnh với thành phần chính gây nên sự khác biệt giữa các đối tượng trong nhóm.

Có thể chia đối tượng nghiên cứu thành 3 nhóm có đặc điểm và mức chi cho thời trang trong một năm ở các mức khác nhau lần lượt chiếm 15,92%; 8,4% và 9,69% thu nhập của nam giới trong một năm, tương ứng với số tiền là 48,05 triệu; 9,31 triệu và 13,32 triệu đồng một năm. Điều này cung cấp cho doanh nghiệp có thêm những thông tin để xây dựng cho kế hoạch phát triển sản phẩm và có chiến lược sản xuất, kinh doanh sản phẩm phù hợp nhu cầu của khách hàng.

[4]. Masaaki Kanai, 2015. *Fashion and apparel*. Outlook for the Retail and Consumer Products Sector in Asia.

[5]. Abby Nakalinda, 2018. *Factors influencing consumer buying behaviour of fast fashion in the United Kingdom*. MSc. Dissertation

[6]. Tina Yinyin Wang, 2010. *Consumer Behavior Characteristics in Fast Fashion*. Thesis for the fulfillment of the Master of Fashion Management Borås, Sweden.

[7]. Trần Thị Kim Thu, 2012. *Giáo trình điều tra xã hội học*. NXB Đại học Kinh tế quốc dân.

[8]. Francesca Dominici, Julian J. Faraway, Martin Tanner, Jim Zidek, 2014. *Linear Models with R*, CHAPMAN & HALL/CRC. CRC Press Taylor & Francis Group.

[9]. MacQueen J., 1967. *Some methods for classification and analysis of multivariate observations*. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. 1, No. 14, pp. 281-297.

[10]. Joshi K D, Nalwade P. S., 2013. *Modified K-Means for Better Initial Cluster Centres*. International Journal of Computer Science and Mobile Computing 11 7 p 2.

[11]. Tibshirani R., Walther G., Hastie T., 2021. *Estimating the number of clusters in a data set via the gap statistic*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(2), 411-423.

TÀI LIỆU THAM KHẢO

[1]. Samiyah Jamal, Dr Adeel Maqbool, Dr SMTariq Zafar, 2013. *Analysis of fashion product of apparels from Consumer lifestyle perspectives: An empirical study*. International Journal of Research in Business and Technology, ISSN: 2291-2118, Volume 3 No.3.

[2]. Nguyễn Thị Mai Hoa, 2017. *Nghiên cứu đặc điểm tiêu dùng sản phẩm thời trang công sở nữ trên thị trường TP Hà Nội*. Luận văn cao học, ĐH SPKT Hùng yên.

[3]. Nguyễn Thị Mai Hoa, 2017. *Nghiên cứu đặc điểm tiêu dùng sản phẩm thời trang công sở nữ trên thị trường TP Hà Nội*. Tạp chí Khoa học và Công nghệ, Trường Đại học Công nghiệp Hà Nội.