UED Journal of Sciences, Humanities & Education – ISSN 1859 - 4603

TẠP CHÍ KHOA HỌC XÃ HỘI, NHÂN VĂN VÀ GIÁO DỤC

SỬ DỤNG PHÂN TÍCH NHÂN TỐ NGHIÊN CỨU CÁC NHÂN TỐ ẢNH HƯỞNG ĐẾN KẾT QUẢ HỌC TẬP CỦA HỌC SINH

Nhận bài:
08 – 01 – 2016
Chấp nhận đăng:
23 – 06 – 2016
http://jshe.ued.udn.vn/

Lê Văn Dũng[a*], Nguyễn Thị Huyền My[b], Lê Thị Tuyết Nhung[b]

Tóm tắt: Nội dung của bài báo này nghiên cứu các nhân tố ảnh hưởng đến kết quả học tập của học sinh khối 12 bậc trung học phổ thông. Số liệu dùng để phân tích là kết quả học tập cả năm của học sinh Trường Trung học phổ thông Lương Văn Can - TP. Hồ Chí Minh (số liệu được cung cấp ở địa chỉ web của nhà trường: http://thptluongvancan.hcm.edu.vn/DataEschool/DiemTongKetLopm.aspx). Phương pháp sử dụng để phân tích là thống kê mô tả và phân tích nhân tố. Kết quả cho thấy có 3 nhân tố ảnh hưởng đến kết quả học tập của học sinh là: "nỗ lực" của học sinh cuối cấp, nhân tố khoa học tự nhiên, nhân tố khoa học xã hội. Nghiên cứu cũng chỉ ra rằng các môn học Ngữ Văn, Lịch sử, Địa lí chịu ảnh hưởng tích cực bởi năng lực khoa học xã hội. Môn Toán và tiếng Anh chịu ảnh hưởng tích cực của cả khoa học tự nhiên và khoa học xã hội.

Từ khóa: thống kê nhiều chiều; phân tích thành phần chính; phân tích nhân tố; phân bố chuẩn nhiều chiều; vectơ ngẫu nhiên.

1. Giới thiệu

Ý tưởng đầu tiên về phân tích nhân tố đã được Pearson [3] và Spearman [4] nêu ra trong những năm cuối thế kỉ 20. Ngày nay với sự hỗ trợ của các phần mềm thống kê, phân tích nhân tố nói riêng và phân tích thống kê nhiều chiều nói chung ngày càng có nhiều ứng dụng mạnh mẽ trong các nghiên cứu về kinh tế, xã hội và các ngành khoa học. Trong bài báo này, chúng tôi nêu một ứng dụng phân tích nhân tố nghiên cứu các nhân tố ảnh hưởng đến kết quả học tập của học sinh khối 12.

2. Cơ sở lý thuyết và phương pháp nghiên cứu (xem [1, 2])

2.1. Cơ sở lý thuyết

2.1.1. Mô hình phân tích nhân tố trực giao

Cho vectơ ngẫu nhiên có thể quan sát được $X=(X_1,X_2,...,X_p)$ có vectơ kì vọng $E(X)=\mu$ và ma trận hiệp phương sai $Var(X)=\Sigma$. Mô hình nhân tố giả định rằng X là tổ hợp tuyến tính của một số ít các biến ngẫu nhiên không quan sát được F_1, F_2,..., F_m ($m<p$) gọi là các nhân tố chung và p biến ngẫu nhiên cộng thêm ε_1, ε_2 ,..., ε_p. Tức là

$$\begin{cases} X_1-\mu_1=l_{11}F_1+l_{12}F_2+...+l_{1m}F_m+\varepsilon_1 \\ X_2-\mu_2=l_{21}F_1+l_{22}F_2+...+l_{2m}F_m+\varepsilon_2 \\ ... \\ X_p-\mu_p=l_{p1}F_1+l_{p2}F_2+...+l_{pm}F_m+\varepsilon_p \end{cases}$$

Hoặc dưới dạng ma trận

$$X-\mu=L\times F+\varepsilon.$$

Phần tử l_{ij} của ma trận L được gọi là tải trọng của biến X_i đặt lên nhân tố F_j.

Các giả thiết của mô hình.

- Đối với nhân tố F:

$$E(F)=0, cov(F)=E(FF^T)=I$$

[a] Trường Đại học Sư phạm - Đại học Đà Nẵng
[b] Học viên cao học K29 Phương pháp Toán sơ cấp - ĐHĐN
* **Liên hệ tác giả**
Lê Văn Dũng
Email: lvdung@ud.edu.vn

- Đối với sai số ngẫu nhiên ε:

$$E(\varepsilon)=0, cov(\varepsilon)=E(\varepsilon\varepsilon^T)=\psi=diag(\psi_1,...,\psi_p)$$

- F và ε không tương quan:

$$cov(F;\varepsilon)=0.$$

Nếu các giả thiết trên được thỏa mãn thì

$$cov(X)=\Sigma=LL^T+\psi.$$

Ta có

$$Var(X_i)=\sigma_{ii}=l_{i1}^2+l_{i2}^2+...+l_{im}^2+\psi_i.$$

Đại lượng $h_i^2=l_{i1}^2+l_{i2}^2+...+l_{im}^2$ gọi là phương sai chung, còn ψ_i được gọi là phương sai xác định. Như vậy:

$$\sigma_{ii}=h_i^2+\psi_i.$$

2.1.2. Phương pháp ước lượng dựa trên thành phần chính

Giả sử $(\lambda_1;e_1)$, $(\lambda_2;e_2)$, ..., (λ_p,e_p) là p cặp giá trị riêng - vectơ riêng của Σ. Do Σ là ma trận xác định dương nên $\lambda_1,\ldots,\lambda_p$ là các số dương và giả sử

$$\lambda_1>\lambda_2>...>\lambda_p.$$

Nếu $p-m$ giá trị riêng λ_{m+1}, λ_{m+2},..., λ_p có tổng

$$\lambda_{m+1}+\lambda_{m+2}+...+\lambda_p$$

là nhỏ thì có thể bỏ qua $p-m$ nhân tố cuối. Khi đó

$$L=\left[\sqrt{\lambda_1}e_1 \text{M} \sqrt{\lambda_2}e_2 \text{M} ... \text{ M} \sqrt{\lambda_m}e_m\right]_{p\times m}$$

Đặt $\psi=diag(\psi_1,...,\psi_p)$ với $\psi_i=\sigma_{ii}-\sum_{i=1}^{m}l_{ii}$ trong đó l_{ii} là các phần tử nằm trên đường chéo chính của ma trận LL^T ta được $\Sigma\approx L\times L^T+\psi.$ Ta cũng có thể chuẩn hóa vectơ ngẫu nhiên $X=(X_1,X_2,...,X_p)$:

$$Z_i=\frac{X_i-\mu_i}{\sqrt{\sigma_{ii}}}.$$

Khi đó, ta thực hiện tương tự như trên đối với ma trận tương quan ρ của vectơ ngẫu nhiên X.

2.1.3. Phương pháp ước lượng hợp lý cực đại

Nếu các nhân tố chung F và nhân tố ò có phân bố đồng thời chuẩn thì ta có thể sử dụng phương pháp hợp lý cực đại để ước lượng ma trận tải trọng L và ma trận phương sai xác định ψ.

Giả sử ta có phân tích nhân tố $X-\mu=LF+$ò.

Khi đó n quan sát $X_1,X_2,...,X_n$ của X cũng có phân tích

$$X_j-\mu=LF_j+\text{ò}_j, \quad =\overline{1,n}$$

Ta có hàm hợp lý:

$$L(\mu,\Sigma)=2\pi^{-nk/2}|\Sigma|^{-n/2}\times$$

$$exp\{\frac{-1}{2}tr[\Sigma^{-1}\sum_{j=1}^{n}(X_j-\bar{X})(X_j-\bar{X})^T$$

$$+n(\bar{X}-\mu)(\bar{X}-\mu)^T]\}$$

phụ thuộc vào L và ψ qua $\Sigma=LL^T+\psi$.

Mô hình đó còn chưa xác định vì L được xác định sai khác một ma trận trực giao nhân với nó. Vì vậy để tiện cho việc tính toán, người ta còn buộc thêm điều kiện

$$L^T\mu\psi^{-1}L=\Delta$$

là một ma trận chéo.

Khi đó ước lượng hợp lý cực đại $\hat{L},\hat{\psi}$ có thể nhận được bằng cách cực đại hóa (2.2) với điều kiện (2.3).

2.2. Phương pháp nghiên cứu

2.2.1. Ước lượng vectơ trung bình, ma trận hiệp phương sai, ma trận hệ số tương quan

Giả sử $\mathbf{x}_1$, $\mathbf{x}_2$,..., $\mathbf{x}_n$ là mẫu được chọn ngẫu nhiên từ tổng thể $X^T=[X_1,X_2,...,X_p]$, trong đó

$$\mathbf{x}_i^T=[x_{i1},x_{i2},...,x_{ip}].$$

Đặt

$$\bar{x}_j=\frac{1}{n}(x_{1j}+x_{2j}+...+x_{nj}), \quad j=1,2,...,p,$$

$$s_{ij}=\frac{1}{n-1}\sum_{k=1}^{n}(x_{ki}-\bar{x}_i)(x_{kj}-\bar{x}_j),$$

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}.$$

Khi đó

- Vectơ trung bình mẫu $\bar{x}^T = [\bar{x}_1, \bar{x}_2, ..., \bar{x}_p]$ là một ước lượng không chệch của trung bình mẫu μ.

- Ma trận hiệp phương sai mẫu

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{1p} \\ \cdots & \cdots & \cdots & \cdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$$

là một ước lượng không chệch của ma trận hiệp phương sai Σ.

Ma trận hệ số tương quan mẫu

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$$

là một ước lượng không chệnh của ma trận hệ số tương quan ρ.

2.2.2. Ước lượng ma trận tải trọng

Để ước lượng L và ψ dựa trên mẫu số liệu, ta thực hiện như sau:

- Tìm p cặp giá trị riêng - vectơ riêng của ma trận hiệp phương sai mẫu S: $(\hat{\lambda}_1; \hat{e}_1)$; $(\hat{\lambda}_2; \hat{e}_2)$,..., $(\hat{\lambda}_m; \hat{e}_p)$.

- Phân tích hệ số tương quan và phân tích thành phần chính, chọn m giá trị riêng lớn nhất đầu tiên.

- Ước lượng L bởi

$$\hat{L} = \left[\hat{l}_{ij}\right]_{p\times m} = \left[\sqrt{\hat{\lambda}_1}\hat{e}_1 \quad \sqrt{\hat{\lambda}_2}\hat{e}_2 \quad \text{M.} \quad \sqrt{\hat{\lambda}_m}\hat{e}_m\right]$$

3. Kết quả và đánh giá

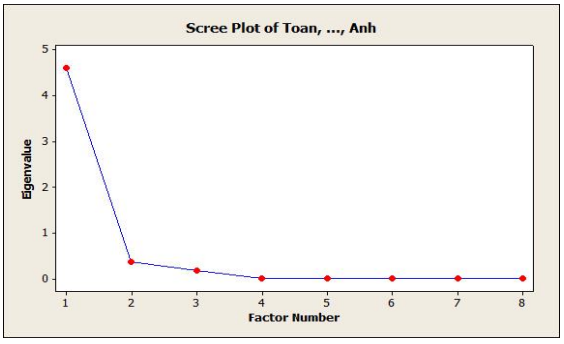
3.1. Kết quả

Trong phần này, chúng tôi nghiên cứu điểm tổng kết năm học 2015-2016 các môn Toán, Vật lý, Hóa học, Sinh học, Ngữ Văn, Lịch sử, Địa lí và Tiếng Anh của học sinh khối 12 Trường THPT Lương Văn Can (TP Hồ Chí Minh), số liệu điểm tổng kết của học sinh được Nhà trường đưa lên ở địa chỉ: http://thptluongvancan.hcm.edu.vn/ataEschool/DiemTongKetLopm.aspx. Phần mềm thống kê Minitab được chúng tôi sử dụng để xử lí số liệu.

Từ bảng hệ số tương quan (Bảng 1) ta có thể gộp các môn Toán, Vật lý, Hóa học và Sinh học vào một nhân tố, gộp các môn Ngữ Văn, Lịch sử, Địa lí và Tiếng Anh vào nhân tố tiếp theo. Kết hợp với phân tích thành phần chính (Hình 1) chúng tôi quyết định chọn 3 nhân tố.

Bảng 1. *Bảng hệ số tương quan*

	Toan	Ly	Hoa	Sinh	Van	Su	Dia
Ly	0.75						
Hoa	0.71	0.70					
Sinh	0.60	0.63	0.62				
Van	0.62	0.52	0.56	0.59			
Su	0.51	0.48	0.56	0.53	0.52		
Dia	0.56	0.52	0.56	0.59	0.62	0.56	
Anh	0.53	0.56	0.51	0.57	0.50	0.40	0.49



Hình 1. *Biểu đồ scree dùng để xác định số nhân tố*

Tiến hành phân tích nhân tố bằng phương pháp ước lượng hợp lí cực đại, chúng tôi thu được:

Bảng 2. *Kết quả xử lí số liệu*

Variable	Factor1	Factor2	Factor3
Toan	0.844	0.135	0.107
Ly	0.846	0.288	- 0.027
Hoa	0.812	0.063	0.092
Sinh	0.774	- 0.136	- 0.136
Van	0.726	- 0.277	0.046
Su	0.652	- 0.241	0.137
Dia	0.718	- 0.319	0.043
Anh	0.672	- 0.063	- 0.336

Ta thấy rằng tất cả hệ số tải trọng của nhân tố 1 dương và xấp xỉ bằng nhau. Chúng tôi đặt tên nhân tố này là nhân tố "nỗ lực" của học sinh cuối cấp. Nhân tố thứ 2 hệ số tải trọng các môn học Toán, Vật lý và Hóa học cùng dương, nhân tố này được gọi là nhân tố khoa học tự nhiên. Tương tự, nhân tố 3 được gọi là nhân tố khoa học xã hội.

3.2. Đánh giá

Từ kết quả phân tích nhân tố ta thấy rằng tất cả các môn học đều có hệ số tải trọng rất lớn. Như vậy, việc đưa điểm tổng kết lớp 12 có tác động rất lớn (do có hệ số tải trọng lớn) đến kết quả học tập của học sinh khối 12. Toán và tiếng Anh có ảnh hưởng tích cực của cả hai nhân tố khoa học tự nhiên và khoa học xã hội. Các môn học Ngữ Văn, Lịch sử, Địa lí chịu ảnh hưởng tích cực bởi năng lực về khoa học xã hội.

4. Kết luận

Từ nghiên cứu trên, chúng tôi có nhận xét sau: có một nhân tố mà chúng tôi đặt tên là "nỗ lực" của học sinh cuối cấp tác động rất lớn đối với kết quả học tập của học sinh lớp 12. Có thể là do điểm tổng kết năm học 12 được đưa vào xét tốt nghiệp THPT đã giúp học sinh cố gắng hết mình để đạt kết quả học tập cao.

Tài liệu tham khảo

[1] Nguyễn Văn Hữu, Nguyễn Hữu Dư (2003), Phân tích thống kê và dự báo, NXB ĐHQG Hà Nội.

[2] Johnson R. A. (2007), Applied multivariate statistical analysis, Sixth edition, Prentice Hall.

[3] Pearson, K. (1904), "On the laws of inheritance in Man: II. On the inheritance of the mental and moral characters in Man, and its comparison with the inheritance of the physical characters", Biometrika;3:131-90.

[4] Spearman, C. (1904), "The proof and measurement of association between two things", Am J Psychol, 15:72–101.

USING FACTOR ANALYSIS IN RESEARCHING FACTORS AFFECTING STUDENTS' ACADEMIC RESULTS

Abstract: This paper is aimed at studying factors affecting twelfth-grade high school students' academic results. The data for analysis were students' whole-year academic results from Luong Van Can high school - Ho Chi Minh City (its website address: http://thptluongvancan.hcm.edu.vn/DataEschool/DiemTongKetLopm.aspx). The methods used were descriptive statistics and factor analysis. The findings show that there are three factors that affect students' academic results: senior students' efforts", natural sciences and social sciences. The study also points out that Literature, History, Geography subjects are positively influenced by social sciences capacity. Mathematics and English are positively influenced by both natural sciences and social sciences.

Key words: multivariate statisticsl; principal component analysis; factor analysis; multivariate standard distribution; random vector.