**BỘ GIÁO DỤC VÀ ĐÀO TẠO VIỆN HÀN LÂM KHOA HỌC VÀ CÔNG NGHỆ**

 **VIỆT NAM**

**VIỆN TOÁN HỌC**

 TRẦN THỊ ĐÔNG

**PHÂN TÍCH CÔNG TÁC ĐÀO TẠO**

**TẠI TRƯỜNG CAO ĐẲNG KINH TẾ KỸ THUẬT THƯƠNG MẠI**

**BẰNG PHƯƠNG PHÁP THỐNG KÊ**

**LUẬN VĂN THẠC SĨ TOÁN HỌC**

**Hà Nội - 2014**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO VIỆN HÀN LÂM KHOA HỌC VÀ CÔNG NGHỆ**

 **VIỆT NAM**

**VIỆN TOÁN HỌC**

 TRẦN THỊ ĐÔNG

**PHÂN TÍCH CÔNG TÁC ĐÀO TẠO**

**TẠI TRƯỜNG CAO ĐẲNG KINH TẾ KỸ THUẬT THƯƠNG MẠI**

**BẰNG PHƯƠNG PHÁP THỐNG KÊ**

Chuyên ngành: **Lý thuyết Xác suất và Thống kê toán học**

Mã số: **60 46 0106**

**LUẬN VĂN THẠC SĨ TOÁN HỌC**

Người hướng dẫn khoa học:

**PGS. TS. HỒ ĐĂNG PHÚC**

**Hà Nội - 2014**

**MỤC LỤC**

[LỜI MỞ ĐẦU 1](#_Toc383969516)

[Chương 1. MỘT SỐ PHƯƠNG PHÁP PHÂN TÍCH HIỆU QUẢ THÔNG DỤNG 3](#_Toc383969517)

[1.1. So sánh hai giá trị trung bình – Phân phối T và kiểm định T-Student 3](#_Toc383969518)

[1.2. So sánh nhiều giá trị trung bình – Mô hình Phân tích phương sai 7](#_Toc383969519)

[1.3. Mô hình hồi quy đơn 10](#_Toc383969520)

[1.4. Mô hình hồi quy bội 13](#_Toc383969521)

[Chương 2. MÔ HÌNH HỒI QUY NHIỀU MỨC 17](#_Toc383969522)

[2.1 Mô hình hồi quy hai mức cơ bản 17](#_Toc383969523)

[2.2. Tính toán các tham số ước lượng và chiến lược phân tích. 22](#_Toc383969524)

[Chương 3. PHÂN TÍCH SỐ LIỆU ĐÀO TẠO 30](#_Toc383969525)

[TẠI TRƯỜNG CAO ĐẲNG KINH TẾ - KỸ THUẬT THƯƠNG MẠI 3](#_Toc383969526)0

[3.1. Số liệu đào tạo cao đẳng dùng trong phân tích 3](#_Toc383969527)0

[3.2. Phân tích hiệu quả đào tạo theo phương pháp đơn biến 32](#_Toc383969528)

[3.3. Phân tích hiệu quả theo mô hình hồi quy tuyến tính cổ điển 40](#_Toc383969529)

[3.4. Phân tích hiệu quả theo mô hình hồi quy tuyến tính nhiều mức 45](#_Toc383969530)

[3.5. Kết luận 46](#_Toc383969531)

[TÀI LIỆU THAM KHẢO 49](#_Toc383969532)

**LỜI MỞ ĐẦU**

Trường Cao đẳng Kinh tế - Kỹ thuật Thương mại (1961) là một trường Cao đẳng công lập do bộ Công thương làm chủ quản, nằm ở trung tâm Phường Phú Lãm, Quận Hà Đông, Thành phố Hà Nội. Trường có sứ mệnh đào tạo nguồn nhân lực chất lượng cao, nghiên cứu, ứng dụng và chuyển giao công nghệ trong lĩnh vực công nghiệp, thương mại và du lịch; hợp tác quốc tế về đào tạo và nghiên cứu khoa học phục vụ sự nghiệp xây dựng và bảo vệ Tổ quốc. Mục tiêu của Trường là đào tạo đa ngành, đa cấp, nghiên cứu, ứng dụng, chuyển giao công nghệ và dịch vụ tư vấn có uy tín trong lĩnh vực công nghiệp, thương mại; sau năm 2015 trở thành trường đại học đáp ứng yêu cầu đào tạo nguồn nhân lực chất lượng cao của sự nghiệp công nghiệp hoá, hiện đại hoá và hội nhập quốc tế. Cơ cấu về trình độ đào tạo của Trường gồm Cao đẳng, Trung cấp chuyên nghiệp, Cao đẳng nghề, Trung cấp nghề, Sơ cấp nghề và Cao đẳng liên thông Trung cấp.

Bài luận văn này là một thử nghiệm nghiên cứu định lượng về công tác đào tạo tại Trường từ năm 2004 đến năm 2010, với bộ số liệu thu thập từ 6 chuyên ngành chủ đạo: Kế toán Doanh nghiệp (CKT), Quản trị Kinh doanh Tổng hợp (CKD), Quản trị Khách sạn (CKS), Tài chính Doanh nghiệp (CTD), Tin học Kế toán (CTK), Quản trị Kinh doanh xuất nhập khẩu (CKN). Bằng các phương pháp thống kê thích hợp, luận văn này sẽ phác thảo đôi nét về tình hình giáo dục và hơi hướng thay đổi cơ cấu chuyên ngành của Trường.

Sau lời mở đầu, luận văn này sẽ có ba chương và danh mục tài liệu tham khảo. Chương 1 sẽ trình bày về một số phương pháp phân tích hiệu quả thông dụng, đó là phương pháp kiểm định Student, mô hình phân tích phương sai, mô hình hồi quy đơn và hồi quy bội. Chương 2 giới thiệu về mô hình tuyến tính nhiều mức. Chương 3 đưa ra các kết quả phân tích công tác đào tạo, diễn giải các yếu tố ảnh hưởng đến thành tích học tập của sinh viên trong từng chuyên ngành cũng như giữa các chuyên ngành với nhau. Và để chỉ ra tính thuyết phục, tính khoa học của phương pháp thống kê được áp dụng. Phần cuối của chương 3 đề xuất một số ý kiến về xây dựng các chiến lược đào tạo của Trường phù hợp với nhu cầu và yêu cầu thực tế của xã hội Việt Nam

Luận văn này được hoàn thành dưới sự hướng dẫn của Phó Giáo sư Hồ Đăng Phúc, Viện Toán học - Viện Hàn Lâm Khoa Học Và Công Nghệ Việt Nam. Tôi xin được bày tỏ lòng biết ơn sâu sắc đối với sự quan tâm chỉ dẫn tận tụy của thầy.

Tôi xin chân thành cảm ơn các thầy cô trong Ban lãnh đạo Viện Toán học, các thầy cô tham gia giảng dạy lớp cao học khóa 18, cùng các thầy cô ở Trung tâm đào tạo sau đại học của Viện Toán học đã nhiệt thành giúp đỡ tôi trong thời gian học tập. Tôi cũng xin gửi lời cảm ơn tới các anh chị em cao học khóa 18, nhóm Semina Xác suất Thống kê, cùng các bạn đồng nghiệp và gia đình đã nhiệt tình đóng góp ý kiến, động viên, giúp đỡ tôi trong suốt quá trình học tập và hoàn thành luận văn.

Đặc biệt, tôi xin gửi lời cảm ơn sâu sắc tới Ban giám hiệu trường Cao đẳng Kinh tế - Kỹ thuật Thương mại và các anh em trong phòng Đào tạo, phòng Công tác Học sinh – Sinh viên đã nhiệt tình và nghiêm túc cung cấp những dữ liệu chính xác quý báu, mà nếu thiếu nguồn số liệu này thì nghiên cứu của tôi không thể thực hiện được.

Tuy đã có nhiều cố gắng nhưng bản luận văn này cũng không tránh khỏi những thiếu sót, tác giả rất mong có được sự tham gia đóng góp ý kiến của các thầy cô giáo, các nhà nghiên cứu Xác suất Thống kê, nghiên cứu kinh tế và các độc giả quan tâm đến bản luận văn này.

Hà Nội, ngày 26 tháng 03 năm 2014

 Trần Thị Đông

Chương 1.

MỘT SỐ PHƯƠNG PHÁP

PHÂN TÍCH HIỆU QUẢ THÔNG DỤNG

Trong thực tế, người ta hay gặp phải bài toán đánh giá hiệu quả của một hoặc nhiều nhân tố lên giá trị của một đại lượng nào đó. Đại lượng cần đánh giá thường là một biến định lượng, tức là một biến số nhận giá trị liên tục có thể phủ kín một đoạn trên trục số. Các nhân tố có thể là biến định lượng hoặc biến định tính, tức là biến nhận hữu hạn các giá trị tính trạng, mỗi tính trạng thường được gọi là mức của nhân tố. Nếu nhân tố là biến định lượng, phương pháp hồi quy thường được dùng để đánh giá hiệu quả của nhân tố đó lên đại lượng cần quan tâm. Nếu nhân tố là một biến định tính có *k* mức tính trạng, chia mẫu quan sát thành *k* nhóm, thì việc đánh giá tác động của nhân tố đó lên đại lượng đang xét chính là việc so sánh giá trị trung bình của đại lượng nói trên giữa các nhóm xác định bởi các mức của nhân tố. Trường hợp đơn giản nhất khi nhân tố chỉ có hai mức, đó chính là bài toán so sánh hai giá trị trung bình. Bài toán này được giải quyết thông qua phép kiểm định T-Student.

**1.1. So sánh hai giá trị trung bình – Phân phối *T* và kiểm định T-Student**

Để so sánh hai giá trị trung bình, chúng ta thường sử dụng phương pháp kiểm định *T* (hay T-Student). Kiểm định *T* dựa căn bản trên phân phối xác suất Student. Phân phối Student với *n* bậc tự do là phân phối của biến ngẫu nhiên *T* xác định bởi



Trong đó



là trung bình của mẫu  gồm các biến ngẫu nhiên độc lập có cùng phân phối chuẩn với kỳ vọng  đã biết và phương sai  chưa biết,



là phương sai mẫu hiệu chỉnh,



là độ lệch tiêu chuẩn.

Phân phối Student với *n* bậc tự do có hàm mật độ xác suất:



trong đó

 

Hàm mật độ của phân phối *T* đối xứng qua trục tung, có dạng hình chuông giống hàm mật độ của phân phối chuẩn.

Để so sánh hai giá trị trung bình bằng kiểm định *T*, hai mẫu ứng với hai mức của nhân tố phải thỏa mãn các giả định:

i) Tuân theo quy luật phân phối chuẩn

ii) Độc lập

iii) Các đối tượng trong mỗi mẫu được chọn một cách ngẫu nhiên

iv) Phân phối của hai mẫu có phương sai bằng nhau.

**Bảng 1.1: Thông tin về mẫu trong bài toán so sánh hai giá trị trung bình**

|  |  |  |
| --- | --- | --- |
|  | Mẫu 1 | Mẫu 2 |
| N (cỡ mẫu) |  |  |
| Mean (Trung bình mẫu) |  |  |
| SD (Độ lệch chuẩn mẫu ) |  |  |

Ta chọn hai mẫu ngẫu nhiên (mẫu 1) rút ra từ biến ngẫu nhiên *X*  (ứng với mức thứ nhất của nhân tố) và (mẫu 2) rút ra từ biến ngẫu nhiên *Y*  (ứng với mức thứ hai của nhân tố) thỏa mãn các giả định nêu trên. Hai mẫu đó có các thông tin đặc trưng mô tả trong Bảng 1.1. Số liệu mẫu được lấy ra để suy luận về đám đông với các thông tin đặc trưng chưa biết là kỳ vọng và độ lệch chuẩn lý thuyết như trong Bảng 1.2.

**Bảng 1.2: Thông tin quần thể trong bài toán so sánh hai giá trị trung bình**

|  |  |  |
| --- | --- | --- |
|  | Đám đông 1 | Đám đông 2 |
| Kỳ vọng (Trung bình lý thuyết) |  |  |
| SD (Độ lệch chuẩn lý thuyết) |  |  |

Như vậy, việc đánh giá hiệu quả của nhân tố hai mức lên giá trị của một đại lượng về thực chất là việc so sánh hai kỳ vọng  và , được đưa về bài toán so sánh hai giá trị trung bình mẫu và , tức là so sánh hiệu số  với 0. Nhân tố có hiệu quả thực sự nếu hiệu số đó khác 0 một cách có ý nghĩa thống kê. Nếu ngược lại, ta có thể kết luận sự khác biệt giữa hai trung bình mẫu chỉ mang tính chất ngẫu nhiên. Cuối cùng, từ bài toán kiểm định giả thiết thống kê

*H: , đối thuyết* *K: *

dẫn đến các bước cần tiến hành như sau:

Bước 1: Tính độ sai lệch giữa hai trung bình mẫu: 

Bước 2: Ước lượng hai phương sai mẫu, vì có giả thiết phương sai lý thuyết của hai mẫu là bằng nhau nên có thể sử dụng hai ước lượng đó để đưa ra ước lượng phương sai chung của hai mẫu (đó chính là phương sai của *d*):

:





Bước 3: Thành lập thống kê:



Bước 4: Với  cho trước, tìm giá trị tới hạn  là phân vị phải mức  của phân phối Student với  bậc tự do.

Bước 5: So sánh giá trị tuyệt đối của thống kê *t* với giá trị tới hạn  :

* Nếu  thì bác bỏ giả thuyết *H*, kết luận , tức là nhân tố ảnh hưởng một cách có ý nghĩa lên đại lượng cần xem xét;
* Nếu  thì chấp nhận giả thuyết *H*, kết luận , và cho rằng nhân tố không ảnh hưởng đáng kể đến giá trị của đại lượng cần nghiên cứu.

*Chú ý:* Ở Bước 4, thay vì tính giá trị tới hạn người ta có thể tính xác suất ý nghĩa



trong đó *T* là biến ngẫu nhiên có phân phối Student với bậc tự do . Tiếp đó, so sánh xác suất ý nghĩa *p* với mức ý nghĩa :

* Nếu  thì bác bỏ giả thuyết *H*, kết luận , nhân tố ảnh hưởng một cách có ý nghĩa thống kê lên đại lượng cần xem xét;
* Nếu  thì chấp nhận giả thuyết *H*, kết luận , nhân tố không tác động lên giá trị của đại lượng cần nghiên cứu.

*Ví dụ 1* (xem tài liệu[2]): Người ta thí nghiệm hai phương pháp chăn nuôi gà khác nhau, sau một tháng kết quả tăng trọng như sau:

Phương pháp I:  con,  kg, 

Phương pháp II:  con,  kg, 

Với mức ý nghĩa  có thể kết luận hai phương pháp có hiệu quả khác nhau hay không? Giả thiết mức tăng trọng của gà tuân theo quy luật chuẩn.

Ta có:



Với , tra bảng phân phối Student ta có , ta bác bỏ *H* và chấp nhận *K*, tức là việc thay đổi phương pháp chăn nuôi sẽ tác động lên mức tăng trọng của gia cầm.

***Chú giải lịch sử***: W.S Gosset là nhà thống kê người Ailen có nhiều đóng góp quan trọng cho sự phát triển của Thống kê toán học, trong đó đáng kể nhất là khái niệm phân phối T - Student. Ông từng làm việc tại hãng bia nổi tiếng Guinness ở Dublin, Ai len từ năm 1899 sau khi tốt nghiệp khoa Hóa tại Đại học Oxford lúc 23 tuổi. Nghiên cứu về thống kê của ông xuất phát từ việc phải bảo đảm chất lượng bia khi thực hiện quá trình lên men. Công trình nghiên cứu này được công bố trên tờ Biometrica vào năm 1907 với tên giả là ***“Student”*** có tựa đề “Sai số đếm của máy đếm tế bào” (On the Error of Counting With a Hemacytometer). Tiếp đó, ông thực hiện một nghiên cứu khác về phân phối *T* với tựa đề “Sai số có thể của giá trị trung bình” (The Probable Error of a Mean). Công trình nổi tiếng này cũng được công bố trên tờ Biometrika vào năm 1908.

 Phép kiểm định T-Student trình bày trên đây cho phép đánh giá hiệu quả tác động của một nhân tố hai mức lên một đại lượng cần nghiên cứu. Khi cần đánh giá hiệu quả của một nhân tố có nhiều hơn hai mức lên một đại lượng nào đó ta cần áp dụng phương pháp phân tích phương sai, được trình bày tiếp sau đây.

**1.2. So sánh nhiều giá trị trung bình – Mô hình Phân tích phương sai**

Phân tích phương sai cho phép so sánh nhiều giá trị trung bình của biến định lượng *Y* trên các nhóm khác nhau, mỗi nhóm ứng với một mức của một biến định tính *X*, biến ngẫu nhiên *X* đó còn được gọi là nhân tố.

Giả sử các giá trị quan sát của biến định lượng *Y* trên k nhóm của nhân tố *X* (nhóm thứ *i* có  quan sát) có dạng:

 hay 

trong đó  là quan sát thứ *j* của mẫu *Y* trên nhóm thứ *i*,

 là hiệu quả của nhóm thứ *i*,

là sai số ngẫu nhiên,



*m* là cỡ mẫu *Y*

Các giả định ban đầu cho mẫu ngẫu nhiên bao gồm :

là các biến ngẫu nhiên độc lập

 **

có nghĩa là các biến ngẫu này có phân phối chuẩn với kỳ vọng bằng 0 và phương sai bằng nhau :

 

Từ mẫu số liệu trên đây, ta cần ước lượng tham số  và , từ đó kiểm định giả thuyết



Kí hiệu

 ( giá trị trung bình của *Y* trong nhân tố thứ *i* )

 ( giá tị trung bình chung )

Khi đó lời giải của bài toán ước lượng là:



Khi các giả định ban đầu của mô hình phân tích phương sai được thỏa mãn, ta thấy:

+ Ước lượng  của  có phân phối chuẩn 

+ Các ước lượng trên là độc lập với nhau,  với 

+ Ước lượng  có phân phối  với *m-k* bậc tự do:



Để kiểm định giả thuyết  ta tính toán và so sánh:

“Độ biến động giữa các nhóm” của nhân tố:



“Độ biến động nội tại các nhóm” của nhân tố



Ta thấy độ biến động nội tại các nhóm chính bằng ước lượng của phương sai chung của các sai số ngẫu nhiên ở tất cả các nhóm, thể hiện mức độ sai lệch của dữ liệu gây ra bởi nhiễu ngẫu nhiên chứ không phải do tác động của nhân tố. Trong khi đó, độ biến động giữa các nhóm phản ánh tác động của nhân tố gây ra các sai lệch so với trung bình chung, là giá trị có thể có của đại lượng cần nghiên cứu khi chưa có tác động của nhân tố.

Khi nhiễu ngẫu nhiên lớn mà tác động của nhân tố lại nhỏ, ta có thể coi ảnh hưởng của nhân tố là không đáng kể. Ngược lại nếu độ biến động do nhân tố gây ra lớn so với độ biến động của nhiễu ngẫu nhiên, ta kết luận nhân tố ảnh hưởng một cách có ý nghĩa đối với giá trị của đại lượng cần nghiên cứu. Do vậy, ta có thể lấy tỷ số



để làm tiêu chuẩn kiểm định giả thuyết đánh giá sự ảnh hưởng của nhân tố lên giá trị của đại lượng cần nghiên cứu. Người ta chứng minh được rằng, nếu giả thuyết  là đúng thì tỷ số *F* có phân phối Fisher

Để kiểm định giả thuyết  với mức ý nghĩa  ta tra bảng phân phối Fisher tìm giá trị  (phân vị phải mức  của phân phối Fisher với *k-1* và *m-k* bậc tự do) và thực hiện

Bác bỏ  nếu *F*

Chấp nhận  nếu *F*

*Chú ý:* Trong thủ tục kiểm định giả thuyết trên đây, thay vì tính giá trị tới hạn người ta có thể tính xác suất ý nghĩa



trong đó *T* là biến ngẫu nhiên có phân phối Fisher với bậc tự do *k-1* và *m-k*. Tiếp đó, so sánh xác suất ý nghĩa *p* với mức ý nghĩa :

* Nếu  thì bác bỏ giả thuyết *H*, kết luận , nhân tố ảnh hưởng một cách có ý nghĩa thống kê lên đại lượng cần xem xét;
* Nếu  thì chấp nhận giả thuyết *H*, kết luận , nhân tố không tác động lên giá trị của đại lượng cần nghiên cứu.

***Chú giải lịch sử***: Nói đến phân tích phương sai, người ta nhắc đến Ronald Aylme Fisher (1890-1962), một nhà di truyền học nổi tiếng người Anh. Theo lời Anders Hanld mô tả, ông là “ một thiên tài gần như tự mình tạo nền tảng cho ngành Khoa học Thống kê hiện đại” và Richard Dawkins mô tả ông là “học trò vĩ đại nhất của Dawwin”

Phép kiểm định T-Student và mô hình phân tích phương sai trình bày trên đây chỉ giúp đánh giá hiệu quả tác động của một biến định tính (nhân tố tính trạng) lên một đại lượng cần nghiên cứu. Khi cần đánh giá hiệu quả của một biến định lượng (nhân tố liên tục) lên một đại lượng nào đó ta cần áp dụng phương pháp phân tích hồi quy đơn, được trình bày tiếp sau đây.

**1.3. Mô hình hồi quy đơn**

Ta sử dụng mô hình hồi quy tuyến tính



để đánh giá tác động của nhân tố định lượng *X* lên giá trị của đại lượng *Y*. Trong mô hình này,  là sai số ngẫu nhiên thỏa mãn một số giả thiết sẽ được nêu trong phần tiếp theo, đại lượng *Y* còn được gọi là *biến phụ thuộc* hay biến được giải thích, nhân tố *X* còn được gọi là *biến độc lập* hoặc biến giải thích. Đồng thời, được gọi là hệ số chặn,  được gọi là hệ số dốc, là hai tham số cần được ước lượng của mô hình . Nếu  thì rõ ràng nhân tố *X* không ảnh hưởng đến giá trị của đại lượng *Y.* Ngược lại, nếu  thì khi giá trị của biến *X* thay đổi, giá trị của biến *Y* cũng phải thay đổi theo, do đó nhân tố *X* tác động lên đại lượng *Y* một cách đáng kể. Như vậy để đánh giá vai trò của nhân tố *X* đối với đại lượng *Y*, ta cần ước lượng các hệ số của mô hình hồi quy, rồi kiểm tra xem hệ số dốc của mô hình có bằng 0 hay không.

 Để ước lượng các tham số của mô hình, ta tiến hành lấy mẫu gồm n cặp là các giá trị quan sát được của cặp biến (*X,Y*). Ta ước lượng bằng phương pháp bình phương bé nhất, tìm  là các giá trị ước lượng của  sao cho tổng bình phương các sai lệch



đạt cực tiểu. Điều kiện cần để hàm  đạt cực tiểu là đạo hàm riêng của nó theo từng đối số triệt tiêu, tức là

 ,



Đây là hệ phương trình với nghiệm

 (1.7)

 (1.8)

Ta thấy  là điểm dừng của hàm *S* . Để khẳng định đó là điểm cực tiểu của hàm *S*, cần lấy đạo hàm riêng bậc hai, có ma trận Hessian

****

Rõ ràng ma trận này xác định dương, do đó điểm dừng của *S* cũng chính là điểm cực tiểu.

Mô hình hồi quy tuyến tính thường được nghiên cứu với năm giả thiết sau đây:

1. Các giá trị của biến độc lập *X* là không ngẫu nhiên (tất định) và được xác định từ trước,
2. Sai số ngẫu nhiên có kỳ vọng bằng 0



1. Sai số ngẫu nhiên có phương sai cố định:



1. Các sai số không tương quan với nhau:



1. Sai số là đại lượng ngẫu nhiên có phân phối chuẩn:



 Để đánh giá vai trò của nhân tố *X* đối với đại lượng *Y* thông qua phương trình hồi quy, ta cần kiểm tra giả thuyết

 

Nếu giả thuyết được chấp nhận, , thì rõ ràng nhân tố *X* không ảnh hưởng đến sự thay đổi giá trị của đại lượng *Y*. Ngược lại, nếu giả thuyết bị bác bỏ, , thì khi giá trị của *X* thay đổi sẽ kéo theo sự thay đổi giá trị của đại lượng *Y*.

 Việc kiểm tra giả thuyết trên đây được tiến hành bằng cách sử dụng tiêu chuẩn kiểm định

 , trong đó 

Nếu năm giả thiết của mô hình hồi quy trình bày phía trên được thỏa mãn thì thống kê *t* xác định như trên là một biến ngẫu nhiên có phân phối Student với (*n-2*) bậc tự do. Do đó bài toán kiểm định giả thuyết có thể tiếp tục thực hiện thông qua một trong ba thủ tục sau đây:

1. *Kiểm tra bằng khoảng tin cậy*: Khoảng tin cậy  của  là



trong đó  là phân vị phải mức  của phân phối Student với (*n-2*) bậc tự do. Nếu khoảng tin cậy này chứa 0 ta phải chấp nhận giả thuyết và kết luận nhân tố *X* không ảnh hưởng đến giá trị của đại lượng *Y*. Ngược lại, nếu khoảng tin cậy không chứa điểm 0, giả thuyết bị bác bỏ, ta có thể khẳng định nhân tố *X* tác động một cách có ý nghĩa đến giá trị của đại lượng *Y*.

1. *Kiểm tra bằng giá trị tới hạn*: Phân vị phải mức  của phân phối Student với (*n-2*) bậc tự do  còn được gọi là giá trị tới hạn của phép kiểm định. Có thể so sánh thống kê t với giá trị tới hạn này để đưa ra kết luận đối với phép kiểm định. Cụ thể, nếu  thì giả thuyết bị bác bỏ. Ngược lại, nếu  thì phải chấp nhận giả thuyết.
2. *Kiểm tra bằng xác suất ý nghĩa*: Xét *T* là một biến ngẫu nhiên có phân phối Student với (*n-2*) bậc tự do. Xác suất ý nghĩa ứng với thống kê *t* của phép kiểm định được xác định bằng

.

So sánh xác suất ý nghĩa *p* để đưa ra kết luận về phép kiểm định. Nếu  thì phải chấp nhận giả thuyết. Ngược lại, nếu  thì có thể bác bỏ giả thuyết.

***Chú giải lịch sử***: Vào năm 1885, Francis Galton đưa ra khái niệm "Hồi quy" trong một nghiên cứu chứng tỏ rằng con cái không có khuynh hướng theo trạc người của cha mẹ, mà hướng về mức trung bình so với cha mẹ. Tuy nhiên, phương pháp hồi quy có một lịch sử lâu hơn. Thực tế, nhà toán học huyền thoại người Pháp tên là Adrien Marie Legendre xuất bản tác phẩm đầu tiên về hồi quy (dù ông không dùng  từ này) vào năm 1805. Tuy nhiên, công lao khám phá phương pháp bình phương tối thiểu thường được quy cho  Carl Friedrich Gauss (một nhà toán học huyền thoại khác người Đức), người sử dụng phương thức này trong phần đầu của thế kỉ 19.

 Cả ba phương pháp kiểm định Student, phân tích phương sai và hồi quy đơn đều là các phương pháp đơn biến và chỉ có thể đánh giá một cách đơn lẻ tác động của một nhân tố lên giá trị của một đại lượng. Muốn đánh giá đồng thời tác động của nhiều nhân tố lên một đại lượng, người ta phải dùng các phương pháp đa biến. Một trong những phương pháp đa biến thường được dùng là phương pháp hồi quy bội, được trình bày tiếp sau đây

**1.4. Mô hình hồi quy bội**

Mô hình hồi quy bội



được sử dụng để đánh giá tác động của nhiều nhân tố lên giá trị của đại lượng *Y,* trong đó  là *k+*1 tham số và  là sai số ngẫu nhiên. Lấy mẫu gồm *n* quan sát , khi đó mô hình được biểu diễn thành hệ *n* phương trình như sau:





. . .



Các phương trình trên có thể đưa về dạng ma trận như sau:



Tức là



Trong đó *Y* là véc tơ cột gồm *n* giá trị quan sát được, *X* là ma trận gồm *n* hàng và *k+*1 cột, trong đó các giá trị của cột đầu tiên luôn bằng 1, *b* là véc tơ cột gồm *k+*1 phần tử, *e* là véc tơ cột gồm *n* số hạng sai số.

Ta có thể giả thiết



với *I* là ma trận đơn vị cấp *n* mà *n* phần tử trên đường chéo chính bằng 1, các phần tử còn lại bằng 0. Khi đó, mô hình hồi quy được đưa về dạng



 Véc tơ tham số *b* được chọn sao cho tổng bình phương các sai lệch



đạt cực tiểu, được gọi là ước lượng bình phương bé nhất của các tham số hồi quy và được xác định qua tính toán ma trận bằng công thức



Cũng giống như đối với mô hình hồi quy đơn, mô hình hồi quy ttuyến tính bội cũng được nghiên cứu với năm giả thiết sau:

1. Giá trị của các biến độc lập không ngẫu nhiên và được xác định từ trước,
2. Sai số ngẫu nhiên có kỳ vọng bằng 0, 
3. Sai số ngẫu nhiên có phương sai cố định, 
4. Các sai số không tương quan với nhau, 
5. Sai số là đại lượng ngẫu nhiên có phân phối chuẩn, 

 Cũng tương tự như cách làm đối với mô hình hồi quy đơn, ta có thể đánh giá vai trò trong phương trình hồi quy bội của từng nhân tố  (*i=1,...,k*) đối với đại lượng *Y*, bằng cách kiểm tra giả thuyết

 

Nếu giả thuyết được chấp nhận, , thì rõ ràng nhân tố  không ảnh hưởng đến sự thay đổi giá trị của đại lượng *Y*. Ngược lại, nếu giả thuyết bị bác bỏ,  thì khi giá trị của  thay đổi sẽ kéo theo sự thay đổi giá trị của đại lượng *Y*.

 Ta sử dụng tiêu chuẩn kiểm định

 , trong đó 

và  là phần tử thứ *i*+1 trên đường chéo chính của ma trận  để kiểm tra giả thuyết trên đây. Nếu năm giả thiết của mô hình hồi quy trình bày phía trên được thỏa mãn thì thống kê  là một biến ngẫu nhiên có phân phối Student với (*n-k-*1) bậc tự do. Bài toán kiểm định giả thuyết có thể tiếp tục thực hiện thông qua một trong ba thủ tục sau đây:

1. *Kiểm tra bằng khoảng tin cậy*: Khoảng tin cậy  của  là



trong đó  là phân vị phải mức  của phân phối Student với (*n-k-*1) bậc tự do. Nếu khoảng tin cậy này chứa 0 ta phải chấp nhận giả thuyết và kết luận nhân tố  không ảnh hưởng đến giá trị của đại lượng *Y*. Ngược lại, nếu khoảng tin cậy không chứa điểm 0, giả thuyết bị bác bỏ, ta có thể khẳng định nhân tố  tác động một cách có ý nghĩa đến giá trị của đại lượng *Y*.

1. *Kiểm tra bằng giá trị tới hạn*: Phân vị phải mức  của phân phối Student với (*n-k-*1) bậc tự do  còn được gọi là giá trị tới hạn của phép kiểm định. Có thể so sánh thống kê t với giá trị tới hạn này để đưa ra kết luận đối với phép kiểm định. Cụ thể, nếu  thì giả thuyết bị bác bỏ. Ngược lại, nếu  thì phải chấp nhận giả thuyết.
2. *Kiểm tra bằng xác suất ý nghĩa*: Xét *T* là một biến ngẫu nhiên có phân phối Student với (*n-k-1*) bậc tự do. Xác suất ý nghĩa ứng với thống kê ** của phép kiểm định được xác định bằng

.

So sánh xác suất ý nghĩa *p* để đưa ra kết luận về phép kiểm định. Nếu  thì phải chấp nhận giả thuyết. Ngược lại, nếu  thì có thể bác bỏ giả thuyết.

 Thông thường các nhân tố ( các biến độc lập trong mô hình hồi quy bội ) đều là các biến định lượng. Tuy nhiên, cũng có thể đưa vào mô hình một số biến định tính nhị phân ( nhận hai giá trị 0 và 1 ). Bằng cách này, có thể đánh giá tác động của cả các nhân tố định tính cũng như các nhân tố định lượng lên giá trị của đại lượng *Y* cần xem xét.

 Các phép kiểm định của mô hình hồi quy bội chỉ có hiệu lực nếu năm giả thiết của mô hình được thỏa mãn, đặc biệt là điều kiện các sai số là độc lập với nhau và có phương sai không đổi. Trong thực tế, đó là hai điều kiện khó được thỏa mãn, do đó có thể dẫn đến những kết luận sai lầm về vai trò của các nhân tố. Để khắc phục hiện tượng này, có thể sử dụng mô hình hồi quy tuyến tính nhiều mức. Đó là nội dung được trình bày trong chương tiếp theo.

.

Chương 2.

MÔ HÌNH HỒI QUY NHIỀU MỨC

Khi thu thập số liệu trong điều tra xã hội học, chúng ta thường gặp cấu trúc số liệu có thứ bậc hay cấu trúc số liệu lồng nhóm, nói cách khác số liệu được thu thập ở các mức khác nhau của đơn vị quan sát. Chẳng hạn, trong điều tra về khả năng tiếp thu kiến thức của học sinh, số liệu về thành tích học tập của học sinh được thu thập ở các cá nhân từng học sinh, song kết luận có thể được đưa ra cho các trường và khu vực. Ở đây các cá thể học sinh (đơn vị mức 1) được xếp trong phạm vi các lớp, sau đó các lớp (đơn vị mức 2) được xếp lồng nhóm trong phạm vi các trường và các trường (đơn vị mức 3) được lồng nhóm trong phạm vi các khu vực (đơn vị mức 4).

Từ những năm cuối của thập kỷ 80 của thế kỷ 20, mô hình nhiều mức đã được xây dựng và sử dụng rộng rãi để nghiên cứu các dữ liệu có cấu trúc lồng nhóm như vậy. Mô hình nhiều mức đã cho thấy nhiều ưu điểm so với các phương pháp phân tích thống kê truyền thống khác, chẳng hạn như khắc phục được hạn chế của giả thiết về tính độc lập giữa các quan sát. Trong ví dụ trên, thành tích học tập của các học sinh (đơn vị mức 1) không hoàn toàn độc lập nhau (vì các học sinh cùng trường, lớp có thể ảnh hưởng nhau bởi cùng thầy dạy, hoặc có sự trao đổi kiến thức giữa các học sinh với nhau). Khi đó giả thiết về tính độc lập giữa các quan sát mà các phương pháp thống kê cổ điển đòi hòi sẽ bị vi phạm.

**2.1 Mô hình hồi quy hai mức cơ bản**

Mô hình hồi quy nhiều mức đã được biết đến trong các nghiên cứu với các tên khác nhau như mô hình hệ số ngẫu nhiên, mô hình thành tố phương sai, mô hình tuyến tính có thứ bậc. Những mô hình mô tả trên không hoàn toàn giống nhau (đặc biệt khi các tính toán chi tiết được quan tâm) tuy nhiên chúng khá giống nhau. Chúng ta sẽ xem xét đến các mô hình này một cách chung như những “mô hình hồi quy nhiều mức”.

Thuật ngữ thông thường để ký hiệu mức thấp nhất của thứ bậc là mức 1, mức thấp tiếp theo là mức 2, v.v… Chẳng hạn, những sinh viên được xếp lồng nhóm trong phạm vi các lớp, và các lớp lại được xếp lồng trong phạm vi các trường khi nghiên cứu hiệu quả đào tạo. Khi ấy ta dùng các thuật ngữ tương ứng: sinh viên là đơn vị mức 1, lớp là đơn vị mức 2 và trường là đơn vị mức 3.

Mô hình hồi quy nhiều mức đầy đủ giả sử rằng có một bộ dữ liệu thứ bậc với một biến phụ thuộc đo được ở mức thấp nhất và một số biến giải thích đo được ở tất cả các mức hiện có. Cụ thể, mô hình đó có thể được xem xét như một hệ thống có thứ bậc các công thức hồi quy. Ví dụ, giả thiết rằng chúng ta có số liệu thu thập ở *J* trường, với số liệu từ một số lượng học sinh khác nhauở mỗi trường. Ở mức học sinh chúng ta có biến phụ thuộc “tác động của quá trình đào tạo của trường” (*Y*) và biến giải thích “điều kiện kinh tế - xã hội” của học sinh (*X*), còn ở mức trường chúng ta có biến giải thích “quy mô của trường” (*Z*). Do vậy chúng ta có thể sắp đặt một công thức hồi quy riêng biệt cho mỗi trường riêng biệt để dự báo biến phụ thuộc *Y* bằng biến độc lập *X* như sau:



Trong công thức hồi quy này  là hệ số chặn,  là hệ số hồi quy (độ dốc hồi quy),  là số hạng sai số chưa tính được (sai số ngẫu nhiên). Chỉ số dưới j là chỉ trường thứ j và chỉ số dưới i là chỉ những cá thể học sinh . Sự khác biệt của mô hình này so với mô hình hồi quy thông thường là chúng ta đã giả sử rằng mỗi trường được đặc trưng bởi một hệ số chặn  khác nhau và một hệ số dốc khác nhau. Cũng như trong mô hình hồi quy nhiều mức thông thường, những sai số ngẫu nhiên  ở mỗi trường được giả định có kì vọng 0 và phương sai ; hầu hết các mô hình nhiều mức giả định rằng phương sai của sai số ngẫu nhiên là giống nhau ở tất cả các trường và chỉ rõ phương sai của sai số này bằng .

Trong mô hình trên hệ số chặn và hệ số dốc được giả định biến đổi qua các trường. Vì lý do đó, chúng được thường được xem xét đến như là các hệ số ngẫu nhiên. Trong ví dụ của chúng ta, mỗi trường được đặc trưng bởi giá trị riêng biệt của hệ số chặn và hệ số dốc của biến “điều kiện kinh tế - xã hội” của học sinh. Đối với những học sinh có giá trị bằng nhau ở biến giải thích “điều kiện kinh tế - xã hội”, một trường với hệ số chặn cao hơn sẽ được dự báo đem lại một kết quả đào tạo cao hơn so với trường có hệ số chặn thấp. Tương tự, sự khác nhau về hệ số dốc của “điều kiện kinh tế - xã hội” có thể được giải thích theo nghĩa là mối quan hệ điều kiện kinh tế xã hội của học sinh và kết quả học tập là không giống nhau trong tất cả các trường. Một số trường có giá trị của hệ số dốc của biến “điều kiện kinh tế - xã hội” cao, như vậy trong những trường này “điều kiện kinh tế - xã hội” có ảnh hưởng lớn đến quá trình đào tạo của nhà trường. Các trường khác có giá trị của hệ số dốc của “điều kiện kinh tế - xã hội” thấp, thể hiện trong những trường này “điều kiện kinh tế - xã hội” ít có ảnh hưởng đối với quá trình đào tạo của nhà trường.

Qua tất cả các trường, các hệ số hồi quy  có một phân bố với kì vọng và phương sai nào đó. Bước tiếp theo ở mô hình hồi quy thứ bậc là dự báo sự biến thiên của hệ số hồi quy  bằng cách đưa vào các biến giải thích ở mức trường như dưới đây:



Công thức (2.2) nói rõ rằng mức độ hiệu quả của quá trình đào tạo chung của một trường (điểm chặn ) có thể được dự báo bởi quy mô của trường (*Z*). Như vậy nếu  là dương, chúng ta phát biểu rằng kết quả của quá trình đào tạo của những trường lớn thì cao hơn ở những trường nhỏ. Ngược lại, nếu  là âm, chúng ta phát biểu rằng kết quả quá trình đào tạo của các trường lớn là thấp hơn ở những trường nhỏ. Sự diễn giải của công thức (2.3) phức tạp hơn. Công thức (2.3) nói rõ rằng mối quan hệ (diễn tả bởi hệ số dốc ) giữa quá trình đào tạo của trường (*Y*) và “điều kiện kinh tế - xã hội” (*X*) của học sinh phụ thuộc vào quy mô nhà trường (*Z*). Một trường có giá trị của  cao hay thấp, phụ thuộc vào quy mô trường (ít nhất là phần nào đó). Nếu  là dương, những trường lớn có khuynh hướng có giá trị của cao hơn những trường nhỏ. Ngược lại, nếu  là âm, những trường lớn có khuynh hướng có giá trị của  thấp hơn là những trường nhỏ. Do đó, quy mô trường đóng vai trò của biến trung gian cho mối quan hệ giữa quá trình đào tạo của trường và “điều kiện kinh tế - xã hội”; mối quan hệ này biến đổi theo giá trị của biến trung gian.

Các số hạng như  và  trong công thức (2.2) và (2.3) là các sai số (phần dư) chưa tính được (ngẫu nhiên) ở mức trường. Phần dư  được giả định có kỳ vọng 0, độc lập với các sai số ngẫu nhiên  ở mức cá thể (mức học sinh). Phương sai của các phần dư theo lý thuyết là , và phương sai của các sai số theo lý thuyết là . Hiệp phương sai  giữa các phần dư  và nói chung không được giả thiết bằng 0.

Chú ý rằng trong công thức (2.2) và (2.3) không giả định các hệ số hồi quy  được thay đổi giữa các trường (nên chúng không có chỉ số dưới *j* để biểu thị một trường nào mà chúng được áp dụng chung cho tất cả các trường). Vì thế các hệ số này được nói đến như các hệ số cố định, sự biến đổi còn lại giữa tất cả các trường được đưa vào hệ số . Sau khi dự báo các hệ số này với biến trường , được giả định là thay đổi theo các trường và được xác định lại qua phần dư  (bởi vậy chúng có chỉ số dưới *j* để chỉ trường nào chúng được áp dụng).

Khi xem xét đơn lẻ một học sinh ở một trường, biến giải thích có thể được viết như một công thức hồi quy đơn phức tạp bằng cách thế công thức (2.2) và (2.3) vào công thức (2.1). Xắp xếp lại các số hạng được:



Phần  trong công thức (2.4) bao gồm tất cả các hệ số cố định, vì thế phần này thường được gọi là phần cố định (hay tất định) của mô hình. Phần  trong công thức (2.4) bao gồm tất cả các số hạng sai số ngẫu nhiên, do đó phần này thường được gọi là phần ngẫu nhiên (bất định) của mô hình. Số hạng  là một số hạng chỉ sự tương tác, xuất hiện trong mô hình như một hệ quả của việc mô hình hoá sự biến động của hệ số dốc  của biến  (mức học sinh) theo biến  (mức trường). Do đó, ảnh hưởng trung gian của *Z* tới mối quan hệ giữa biến phụ thuộc *Y* và biến độc lập *X* được biểu diễn như một tương tác chéo giữa các mức. Việc diễn giải mối tương tác giữa các số hạng trong phân tích hồi quy nhiều mức có thể sẽ phức tạp. Nói chung, việc diễn giải một cách riêng biệt các hệ số trong mô hình với những tương tác là đơn giản hơn nếu các biến tạo ra sự tương tác được biểu diễn như độ sai lệch so với giá trị trung bình tương ứng của chúng. Chú ý rằng, số hạng sai lệch ngẫu nhiên được đi kèm với giá trị. Vì số hạng sai lệch  được nhân với biến dự báo  nên sai lệch tổng cộng sẽ khác nhau đối với các giá trị khác nhau của , là tình thế xảy ra trong mô hình hồi quy bội cổ điển được gọi là hiện tượng “phương sai sai số thay đổi”.

Như đã giải thích ở trên, những mô hình nhiều mức là cần thiết vì khi số liệu được phân nhóm, các quan sát trong cùng một nhóm nói chung giống nhau hơn các quan sát ở những nhóm khác, điều này đã vi phạm giả thiết về tính độc lập của của quan sát. Sự không độc lập của các quan sát trong cùng một nhóm có thể được biểu diễn qua một hệ số tương quan: hệ số tương quan nội tại của nhóm. Các tài liệu về phương pháp luận đưa ra một số công thức khác nhau để tính , hệ số tương quan nội tại của nhóm. Ví dụ, nếu chúng ta dùng phương pháp phân tích phương sai một lối để kiểm tra xem có ảnh hưởng nhóm đáng kể hay không, hệ số tương quan nội tại của nhóm được ước lượng bởi

,

trong đó, *MS(A)* là trung bình bình phương sai lệch toàn phần, *MS(error)* là trung bình bình phương sai lệch của sai số, còn *k* là số nhóm.

Mô hình hồi quy nhiều mức cũng có thể được sử dụng để ước lượng tương quan nội tại nhóm. Để đạt được đích này ta dùng một mô hình không có biến giải thích, là mô hình chỉ có điểm chặn. Điều này có thể được suy ra từ công thức (2.1) và (2.2) như sau: Nếu không có biến giải thích X ở mức thấp nhất, công thức (2.1) rút lại thành

 (2.5)

Tương tự, nếu không có biến giải thích Z ở mức cao nhất, công thức (2.2) còn lại là

  (2.6)

Khi đó, chúng ta tìm mô hình công thức đơn bằng cách thay thế (2.6) vào (2.5):

  (2.7)

Chúng ta cũng có thể tìm được công thức (2.7) bằng cách đơn giản hóa công thức (2.4), bỏ đi tất cả các số hạng chứa biến X hay Y. Mô hình của công thức (2.7) tách phương sai thành hai thành phần độc lập, đó là phương sai  của sai số ngẫu nhiên (sai số mức thấp nhất)  và  là phương sai của sai số mức cao nhất . Dùng mô hình này chúng ta có thể ước lượng tương quan nội tại lớp  bằng công thức:



 Công thức trên phát biểu một cách đơn giản là tương quan nội tại trong các nhóm bằng tỷ số của phương sai mức nhóm so với phương sai toàn bộ ước lượng được.

**2.2. Tính toán các tham số ước lượng và chiến lược phân tích.**

Mô hình tổng quát sẽ có nhiều hơn một biến giải thích ở mức thấp nhất và cũng nhiều hơn một biến giải thích ở mức cao nhất. Giả sử rằng chúng ta có *P* biến giải thích *X* ở mức thấp nhất, được đánh dấu bằng chỉ số dưới *p* (*p=1,…,P*). Tương tự như vậy, chúng ta có *Q* biến giải thích *Z* ở mức cao nhất, được đánh dấu bằng chỉ số dưới *q* (*q=1,…,Q*). Khi đó, (2.4) trở thành công thức chung hơn như sau:

  (2.9)

trong đó



còn  là hằng số, và  là các sai số ngẫu nhiên. Sai số ngẫu nhiên ở mức thấp nhất  được giả định có phân bố chuẩn với kì vọng 0 và phương sai  trong tất cả các nhóm. Các số hạng  và trong công thức (2.9) là các phần dư ở mức cao nhất. Chúng được giả định là độc lập với sai số ngẫu nhiên  ở mức cá thể, và có phân bố chuẩn nhiều chiều với kỳ vọng 0. Phương sai của phần dư  là sự sai lệch của các điểm chặn giữa các nhóm, phương sai này được xác định là  Phương sai của các phần dư  là những sự sai lệch của các độ dốc giữa các nhóm, chúng được xác định là . Hiệp phương sai giữa các phần dư  nói chung không được giả định bằng 0.

Lý thuyết thống kê chứa đựng phía sau mô hình hồi quy nhiều mức là rất phức tạp. Trên cơ sở dữ liệu quan sát được trên các đối tượng của mẫu, chúng ta muốn ước lượng các tham số của mô hình hồi quy nhiều mức, cụ thể là các hệ số hồi quy và các thành phần phương sai. Phương pháp ước lượng Hợp lý Cực đại (ML) là một trong những thường được dùng hiện nay trong phân tích hồi quy nhiều mức.

Phương pháp hợp lý cực đại cũng cung cấp các sai số tiêu chuẩn cho hầu hết các ước lượng. Các sai số chuẩn này có thể được sử dụng cho phép kiểm định ý nghĩa của các tham số thống kê, thống kê dùng để kiểm định có dạng



và có thể được nói đến như phân bố chuẩn tắc được dùng để xác định giá trị xác suất ý nghĩa p cho giả thuyết *H* không, nói rằng quần thể có tham số đặc trưng bằng 0.

 Trong các mô hình thực tế, các sai số tiêu chuẩn mang tính tiệm cận, chúng thực sự có giá trị cho những mẫu lớn. Thông thường, người ta không được biết một cách chính xác khi nào một mẫu là đủ lớn để tin chắc về độ chính xác của ước lượng. Trong phân tích hồi quy thông thường, một quy tắc kinh nghiệm phổ biến là đòi hỏi mười quan sát cho mỗi hệ số hồi quy được ước lượng. Đối với hồi quy nhiều mức, chúng ta nên nhớ rằng các hệ số của mô hình và các thành tố phương sai được ước lượng một cách tổng hợp từ nhiều nhóm mẫu, thường không có cỡ lớn lắm. Tuy nhiên do thông tin được xử lý một cách tổng hợp từ nhiều nguồn nên vẫn có được mô hình tốt mà không đòi hỏi cỡ mẫu phải lớn như đối với mô hình hồi quy thông thường.

Thủ tục hợp lý cực đại cũng cung cấp một thống kê gọi là độ lệch, thống kê này biểu thị mức độ mô hình phù hợp với số liệu. Nói chung, các mô hình với độ lệch nhỏ phù hợp hơn các mô hình có độ lệch lớn. Nếu hai mô hình được lồng nhóm (có nghĩa rằng một mô hình cụ thể có thể được suy ra từ một mô hình tổng quát hơn bằng cách lại bớt một số tham số từ mô hình tổng quát) thì sự khác biệt của độ lệch từ hai mô hình có phân bố với bậc tự do bằng hiệu của số các tham số được ước lượng trong hai mô hình. Như vậy, có thể sử dụng tiêu chuẩn kiểm định Khi-bình phương để kiểm định xem mô hình chung hơn có thực sự thích hợp hơn so với mô hình đơn giản hay không.

Kiểm định Khi-bình phương cho các độ lệch cũng có hiệu quả tốt khi khảo sát tầm quan trọng của ảnh hưởng ngẫu nhiên, bằng cách so sánh một mô hình có chứa những ảnh hưởng này với một mô hình mà ảnh hưởng ngẫu nhiên bị loại trừ. Nếu những mô hình được so sánh không là những mô hình lồng nhóm thì nguyên tắc cơ bản là ta sẽ chọn mô hình nào đơn giản hơn.

Trong thực hành, phương pháp ước lượng hợp lý cực đại đòi hỏi phải tính toán thông qua một quy trình lặp. Lúc bắt đầu, ta phải cung cấp cho máy tính các giá trị ban đầu hợp lý của các tham số cần ước lượng (đối với mô hình phân tích hồi quy nhiều mức, những giá trị ban đầu này thường được xác định dựa trên kết quả của mô hình hồi quy một mức đơn giản). Ở bước tiếp theo, phương pháp tính toán sẽ cố gắng khéo léo cải thiện giá trị ban đầu để đưa ra ước lượng tốt hơn. Bước thứ hai này được lặp lại nhiều lần. Sau mỗi bước lặp, chương trình kiểm tra xem tham sô được ước lượng đã thay đổi bao nhiêu so với bước trước. Nếu sự thay đổi này rất nhỏ, thì thủ tục này đã hội tụ và chương trình kết thúc, và việc ước lượng được hoàn thành.

 Hiện nay, một phương pháp ước lượng khác thường được sử dụng trong phân tích hồi quy nhiều mức là phương pháp Ước lượng lặp bình phương bé nhất suy rộng (IGLS – Interative Generalised Least Squares). Chẳng hạn, xét mô hình tổng hợp phương sai hai mức đơn giản



 Nếu chúng ta đã biết giá trị của các phương sai và như vậy có thể xây dựng ngay được ma trận hiệp phương sai V của Y có dạng đường chéo khối. Lúc đó ta có thể áp dụng ngay phương pháp bình phương bé nhất thông thường để ước lượng các tham số cố định.

  (2.11)

Ở đây,

  (2.12)

Với m là số lượng đơn vị nhóm ở mức 2, còn  là số lượng đơn vị mức 1.

Tuy nhiên, lúc đầu ta chưa hề biết giá trị của các phương sai, do đó phải sử dụng thủ tục ước lượng lặp. Ta bắt đầu từ việc ước lượng các tham số cố định bằng một cách hợp lý nào đó. Chẳng hạn, tạm thời giả thiết  và sử dụng phương pháp bình phương nhỏ nhất thông thường để xác định giá trị ban đầu của các tham số cố định. Từ giá trị ban đầu của các tham số cố định, chúng ta sẽ tạo ra các phần dư “thô”,

  (2.13)

 

Xem xét ma trận tích chéo ta thấy kỳ vọng của ma trận này chính là ma trận hiệp phương sai *V*. Như vậy, ta có thể sử dụng ma trận tích chéo cho vai trò của ma trận hiệp phương sai trong bước tiếp theo của quá trình lặp. Chúng ta có thể sắp xếp lại ma trận tích chéo này thành một véc tơ cột bằng cách lần lượt nối chồng cột trước lên cột tiếp theo sau của ma trận. Khi ấy mối quan hệ giữa véc tơ này và giá trị của các phương sai có thể được biểu diễn qua mô hình hồi quy tuyến tính sau:

  (2.14)

trong đó R là véc tơ phần dư.

Vế trái của (2.14) là véc tơ đáp ứng cho mô hình hồi quy tuyến tính, còn vế bên phải chứa hai biến giải thích tương ứng với hệ số  và , là các tham số cần ước lượng. Việc giải phương trình hồi quy trên đây sẽ cung cấp cho chúng ta giá trị ước lượng của các phương sai, qua đó xác định được giá trị mới của ma trận hiệp phương sai *V*. Đưa ma trận mới thu được này vào (2.11), ta lại nhận được ước lượng mới của các tham số cố định và sử dụng chúng vào chu trình tiếp theo của quá trình lặp.

Quá trình lặp sẽ được kết thúc khi giá trị của các tham số tại hai bước lặp liên tiếp không thay đổi một cách đáng kể. Cuối cùng ta sẽ thu được ước lượng xấp xỉ của các tham số cố định cùng với các tham số ngẫu nhiên của mô hình hồi quy nhiều mức. Quá trình ước lượng tham số bằng phương pháp bình phương bé nhất suy rộng được tóm tắt trong biểu đồ 1.

Nhập số liệu

Cho , ước lượng “thô” ma trận hiệp phương sai.

Ước lượng 

Giải phương trình hồi quy (2.14) để ước lượng  và 

Kiểm tra tính hội tụ của các tham số  và 

Kết thúc

Lập ma trận hiệp phương sai V từ các giá trị của  và 

**Biểu đồ 1. Thuật toán ước lượng lặp bình phương bé nhất suy rộng**

 Trong thực tế, để xây dựng được một mô hình thích hợp, người ta thường phải dựa vào các hiểu biết chuyên sâu về lĩnh vực của đối tượng cần nghiên cứu. Nếu chúng ta chưa có cơ sở lý thuyết về đối tượng cần quan tâm, chúng ta có thể dùng một quy trình để thu được mô hình thích hợp. Một quy trình khá hấp dẫn là bắt đầu với mô hình đơn giản nhất có thể, mô hình chỉ có điểm chặn, xem xét từng bước một cho việc phân loại thành các nhóm khác nhau theo từng tham số. Ở mỗi bước chúng ta kiểm tra kết quả để xem tham số nào có ý nghĩa, và phần dư còn lại ở mỗi mức riêng biệt là bao nhiêu. Các bước khác nhau của một thủ tục được lựa chọn như vậy được đưa ra như dưới đây:

***Bước 1.*** Phân tích một mô hình không có biến giải thích, *mô hình chỉ có điểm chặn*, được đưa ra bởi mô hình của công thức (2.7):

 .

Mô hình chỉ có điểm chặn cho chúng ta một ước lượng của tương quan nội tại lớp bằng cách áp dụng công thức (2.8):

 .

Mô hình chỉ có điểm chặn cũng cho chúng ta giá trị của độ lệch, đó là một độ đo của mức độ không phù hợp của mô hình.

***Bước 2.*** Phân tích một mô hình với tất cả các biến giải thích được xác định của mức thấp nhất. Có nghĩa rằng tương ứng với các thành tố phương sai của các độ dốc đã được cố định bằng 0. Mô hình này được biểu diễn bằng công thức (2.10):

 

Ở bước này chúng ta đánh giá phần đóng góp của mỗi biến giải thích riêng lẻ. Nếu chúng ta dùng phương pháp ước lượng hợp lý cực đại (ML), chúng ta có thể kiểm định được sự cải tiến của mô hình được chọn cuối cùng trong bước này, sự khác nhau giữa độ lệch của mô hình này với mô hình trước (mô hình chỉ có điểm chặn). Sự khác nhau này gần giống đại lượng ngẫu nhiên Khi-bình phương  với bậc tự do là hiệu số của số lượng tham số của hai mô hình. Trong trường hợp này, bậc tự do thường bằng số biến giải thích được cộng vào ở bước 2.

***Bước 3.*** Đánh giá độ dốc của biến giải thích bất kỳ xem phương sai thành phần giữa các nhóm có khác 0 một cách ý nghĩa hay không. Mô hình quan tâm được cho bởi công thức (2.11):

 .

Việc đánh giá sự sai lệch của độ dốc ngẫu nhiên được tiến hành tốt nhất thông qua kiểm định lần lượt cho từng biến một. Các biến bị bỏ qua ở bước trước có thể được phân tích lại ở bước này. Rất có thể một biến giải thích không có ý nghĩa về độ dốc hồi quy (ở kiểm định bước 2), nhưng có thành tố phương sai có ý nghĩa cho độ dốc này. Sau khi xác định được độ dốc nào có sự khác biệt có ý nghĩa giữa các nhóm, chúng ta có thể thêm vào tất cả các thành tố phương sai ở mô hình cuối cùng và dùng kiểm định bình phương  dựa trên các độ lệch để kiểm định mô hình cuối cùng của bước 3 xem có thích hợp hơn là mô hình cuối cùng của bước 2 hay không.

***Bước 4.*** Cộng tất cả các biến giải thích ở mức cao nhất, như công thức (2.12):

 .

Việc này cho phép chúng ta kiểm tra xem những biến này có giải thích được sự biến thiên giữa nhóm của biến độc lập hay không. Thêm nữa, nếu chúng ta dùng ước lượng ML, chúng ta có thể dùng kiểm định bình phương  toàn bộ để chính thức kiểm định sự cải tiến của sự phù hợp của mô hình.

***Bước 5.*** Có thể bổ sung các tương tác chéo mức giữa các biến giải thích ở mức nhóm và những biến giải thích ở mức cá thể nào mà sự biến thiên của độ dốc có ý nghĩa ở bước 3. Điều này dẫn đến mô hình đầy đủ đã phát biểu ở công thức (2.9):

 

 Nếu sử dụng ước lượng ML, ta có thể dùng kiểm định bình phương  toàn bộ một lần nữa để kiểm định sự cải thiện của tính phù hợp.

Chương 3.

PHÂN TÍCH SỐ LIỆU ĐÀO TẠO TẠI TRƯỜNG

CAO ĐẲNG KINH TẾ - KỸ THUẬT THƯƠNG MẠI

**3.1. Số liệu đào tạo cao đẳng dùng trong phân tích**

Nghiên cứu này thực hiện trên bộ số liệu thu thập tại trường Cao đẳng Kinh tế - Kỹ thuật Thương mại có trụ sở chính tại Phường Phú Lãm – Hà Đông – Hà Nội. Trường Cao đẳng Kinh tế - Kỹ thuật Thương mại tiền thân là trường Cán bộ Vật tư và trường Trung học Thương nghiệp Sơn Tây, cả hai trường được thành lập vào năm 1961. Ngày 24/11/1990 Bộ Thương nghiệp (nay là Bộ Thương mại) đã quyết định hợp nhất hai trường đó thành trường Thương mại Trung ương I với nhiệm vụ đào tạo nhiều chuyên ngành cho bậc trung học và dạy nghề, phục vụ nhu cầu nhân lực cho sự phát triển của hoạt động thương mại nước ta. Năm 1998, Trường được nâng cấp thành trường cao đẳng và mang tên như ngày nay.

Với mục tiêu xác định các nhân tố có tác động ảnh hưởng đến kết quả học tập của sinh viên, số liệu được lấy từ phòng Đào tạo và phòng Công tác học sinh sinh viên, bao gồm các thông tin thu thập của tổng số 2431 sinh viên thuộc 6 chuyên ngành trong 5 khóa học với thời điểm nhập học từ năm 2004 đến 2008. Thông tin thu thập bao gồm điểm tổng kết từng môn học của từng học kỳ trong cả quá trình 3 năm đào tạo và điểm trung bình học tập toàn khóa cùng với các yếu tố có khả năng ảnh hưởng đến các điểm trung bình đó, gồm có ngày sinh, giới tính, nơi cư trú của sinh viên, chuyên ngành, khóa, khối thi đầu vào, tổng điểm thi đầu vào, điểm số của 6 môn thi tốt nghiệp và xếp loại tốt nghiệp phổ thông. Các điểm trung bình của từng học kỳ và điểm trung bình toàn khóa của sinh viên được tính theo trọng số là số tín chỉ tương ứng của mỗi môn học. Trong số liệu này chỉ bao gồm các khóa nhập học từ 2004-2008 và đã hoàn thành chương trình học tập. Một số khóa học khác không có đầy đủ thông tin đào tạo hoặc thuộc chuyên ngành mới phát sinh thì không được đưa vào phân tích

Các biến được sử dụng như sau:

1. Lop: Lớp
2. Malop: Mã lớp
3. Cnganh: Chuyên ngành
4. MaCNg: Mã chuyên ngành
5. Khoa: Khóa
6. Nam: Năm
7. Khoithi: Khối thi đầu vào
8. Ho: Họ
9. Ten: Tên
10. Gioitinh: Giới Tính
11. Ngaysinh: Ngày sinh
12. Tuoi: Tuổi
13. Noi\_sinh: Nơi sinh
14. Vung: Vùng cư trú
15. Dauvaocd: Tổng điểm đầu vào cao đẳng
16. Ky1: Trung bình kỳ 1
17. Ky2: Trung bình kỳ 2
18. Ky3: Trung bình kỳ 3
19. Ky4: Trung bình kỳ 4
20. Ky5: Trung bình kỳ 5
21. Ky6: Trung bình kỳ 6
22. TBC\_new: Trung bình chung học tập
23. Hocmuon: Sinh viên nhập học muộn

Việc nhập số liệu được tiến hành bằng công cụ Excel, tạo thành các file lưu trữ tại các phòng chức năng của Nhà trường, đồng thời được công bố trên website: http://www.ctet.edu.vn/trang-chu.html. Trước tiên, việc phân tích số liệu được thực hiện với sự trợ giúp của phần mềm SPSS để xem xét các mô hình tuyến tính cổ điển, từ đó đánh giá vai trò của từng biến độc lập để thuận lợi hơn trong việc xây dựng mô hình nhiều mức thích hợp. Sau đó, phần mềm STATA được sử dụng để xây dựng các mô hình nhiều mức phù hợp hơn với cấu trúc đặc thù của dữ liệu và khắc phục được khiếm khuyết của mô hình tuyến tính cổ điển.

**3.2. Phân tích hiệu quả đào tạo theo phương pháp đơn biến**

Trước tiên ta sử dụng phương pháp phân tích đơn biến để đánh giá tác động của từng nhân tố lên kết quả đào tạo (thể hiện qua điểm trung bình của các học kỳ và điểm trung bình của toàn khóa). Cụ thể, ta sẽ dùng phép kiểm đinh T-Test nếu nhân tố là một biến định tính nhị phân nhận hai giá trị, sử dụng phương pháp phân tích phương sai nếu nhân tố là một biến định tính nhận nhiều hơn hai giá trị, phương pháp hồi quy đơn nếu nhân tố là một biến định lượng. Trong các phân tích đó, các nhân tố đóng vai trò của các biến độc lập, còn kết quả đào tạo là biến phụ thuộc.

Xét nhân tố “khối thi đầu vào”, được mã hóa thành biến Thi khoi D1, nhận hai giá trị 0 (khối A) và 1 (khối D1). Bảng 3.1 cho ta thấy điểm trung bình các học kỳ của sinh viên thi đầu vào khối A cao hơn so với điểm của sinh viên thi đầu vào khối D1, trừ trường hợp điểm trung bình của kỳ 1 của hai khối này lại có giá trị xấp xỉ nhau. Phép kiểm đinh T-Test trình bày trong bảng 3.2 cho thấy đánh giá trên là đảm bảo về mặt thống kê, khi xác suất ý nghĩa của phép kiểm định so sánh giá trị trung bình của hai nhóm tại các học kỳ 2 – 6 đều nhỏ hơn 5%, chỉ riêng đối với học kỳ 1 thì xác suất ý nghĩa bằng 0.504 lớn hơn 5%. Như vậy nhân tố khối thi đầu vào ảnh hưởng rõ ràng tới điểm trung bình của các học kỳ và điểm trung bình trung toàn khóa, loại trừ ở học kỳ 1.

**Bảng 3.1. Điểm số trung bình các học kỳ và toàn khóa phân theo khối thi đầu vào**

|  | Thi khoi D1 | Số sinh viên | Trung bình | Độ lệch tiêu chuẩn |
| --- | --- | --- | --- | --- |
| TB ky 1 | 0 | 2090 | 5,9326 | 0,51644 |
| 1 | 341 | 5,9122 | 0,55149 |
| TB ky 2 | 0 | 2090 | 5,7871 | 0,49115 |
| 1 | 341 | 5,6893 | 0,46915 |
| TB ky 3 | 0 | 2090 | 6,3389 | 0,57269 |
| 1 | 341 | 6,2397 | 0,53147 |
| TB ky 4 | 0 | 2090 | 6,6796 | 0,71192 |
| 1 | 341 | 6,3114 | 0,57262 |
| TB ky 5 | 0 | 2090 | 6,9824 | 0,72560 |
| 1 | 341 | 6,8042 | 0,63545 |
| TB ky 6 | 0 | 2089 | 7,5926 | 0,74805 |
| 1 | 341 | 7,2688 | 0,68299 |
| TBC hoc tap  | 0 | 2071 | 6,7083 | 0,50644 |
| 1 | 337 | 6,4688 | 0,43028 |

 **Bảng 3.2. Kiểm định T-Student so sánh trung bình hai nhóm thi đầu vào khối A và D1**

|  |  | Kiểm định so sánh phương sai  | Kiểm đinh so sánh trung bình |
| --- | --- | --- | --- |
|  |  | F | Xác suất ý nghĩa P | T | Bậc tự do | Xác suất ý nghĩa | Chênh lệch trung bình |
| TB ky 1 |  | 1.665 | 0.197 | 0.668 | 2429 | 0.504 | 0.020 |
| TB ky 2 |  | 1.916 | 0.166 | 3.430 | 2429 | **0.001** | 0.098 |
| TB ky 3 |  | 3.716 | 0.054 | 2.995 | 2429 | **0.003** | 0.099 |
| TB ky 4 |  | 27.962 | 0.000 | 9.084 | 2429 | **0.000** | 0.368 |
| TB ky 5 |  | 8.437 | 0.004 | 4.275 | 2429 | **0.000** | 0.178 |
| TB ky 6 |  | 0.434 | 0.510 | 7.500 | 2428 | **0.000** | 0.324 |
| TBC hoc tap  |  | 13.172 | 0.000 | 8.215 | 2406 | **0.000** | 0.240 |
|  |  |  |  |  |  |  |

**Bảng 3.3. Điểm số trung bình các học kỳ và toàn khóa phân theo độ tuổi nhập học**

|  | Học muộn | Số SV | Trung Bình | Độ lệch tiêu chuẩn |
| --- | --- | --- | --- | --- |
| TB ky 1 | 0 | 1616 | 5,9457 | 0,51309 |
| 1 | 814 | 5,8973 | 0,53632 |
| TB ky 2 | 0 | 1616 | 5,7884 | 0,49680 |
| 1 | 814 | 5,7435 | 0,47296 |
| TB ky 3 | 0 | 1616 | 6,3681 | 0,57761 |
| 1 | 814 | 6,2387 | 0,53852 |
| TB ky 4 | 0 | 1616 | 6,6911 | 0,71106 |
| 1 | 814 | 6,5028 | 0,67849 |
| TB ky 5 | 0 | 1616 | 6,9986 | 0,72879 |
| 1 | 814 | 6,8764 | 0,68377 |
| TB ky 6 | 0 | 1615 | 7,5942 | 0,72291 |
| 1 | 814 | 7,4537 | 0,78703 |
| TBC hoc tap  | 0 | 1602 | 6,7153 | 0,50732 |
| 1 | 805 | 6,5941 | 0,48588 |

Xét nhân tố “Học muộn” nhận hai giá trị 0 (học đúng tuổi, tuổi tại thời điểm nhập học nhỏ hơn 19) và 1 (học muộn, tuổi tại thời điểm nhập học từ 19 trở lên), Bảng 3.3. cho ta thấy điểm trung bình các học kỳ và điểm trung bình chung học tập toàn khóa của sinh viên học đúng tuổi cao hơn so với điểm của sinh viên học muộn. Phép kiểm định T-Test trình bày trong bảng 3.4. cho thấy đánh giá trên là đảm bảo về mặt thống kê, khi xác suất ý nghĩa của phép kiểm định so sánh giá trị trung bình của hai nhóm đều nhỏ hơn 5%. Như vậy có thể nói việc sinh viên có nhập học ngay sau khi tốt nghiệp phổ thông hay không sẽ ảnh hưởng rõ ràng đến hiệu quả đào tạo trong suốt sáu học kỳ.

**Bảng 3.4. Kiểm định T-Student so sánh trung bình hai nhóm học đúng tuổi và học muộn**

|  |  | Kiểm định Levene về phương sai bằng nhau | Kiểm đinh T-Student so sánh trung bình |
| --- | --- | --- | --- |
|  |  | F | Xác suất ý nghĩa P | t | Bậc tự do | Xác suất ý nghĩa | Chênh lệch trung bình |
| TB ky 1 |  | 1,564 | 0,211 | 2,161 | 2428 | 0,031 | 0,048 |
| TB ky 2 |  | 1,386 | 0,239 | 2,139 | 2428 | 0,033 | 0,045 |
| TB ky 3 |  | 3,255 | 0,071 | 5,331 | 2428 | 0,000 | 0,129 |
| TB ky 4 |  | 3,035 | 0,082 | 6,254 | 2428 | 0,000 | 0,188 |
| TB ky 5 |  | 2,134 | 0,144 | 3,983 | 2428 | 0,000 | 0,122 |
| TB ky 6 |  | 2,802 | 0,094 | 4,386 | 2427 | 0,000 | 0,140 |
| TBC hoc tap  |  | 1,326 | 0,250 | 5,609 | 2405 | 0,000 | 0,121 |

Xét nhân tố “giới tính”, được mã hóa thành biến Nu sinh vien, nhận hai giá trị 0 (Nam) và 1 (Nữ). Bảng 3.5. cho ta thấy điểm số trung bình các học kỳ của nữ sinh viên cao hơn hẳn so với điểm số của nam sinh viên. Phép kiểm đinh T-Test trình bày trong bảng 3.6. cho thấy đánh giá trên là đảm bảo về mặt thống kê, khi xác suất ý nghĩa của phép kiểm định so sánh giá trị trung bình của hai nhóm tại các học kỳ và trung bình chung học tập đều nhỏ hơn 5%.

**Bảng 3.5. Điểm số trung bình các học kỳ và toàn khóa phân theo giới tính**

|  | Nu sinh vien | Số SV | Trung bình | Độ lệch tiêu chuẩn |
| --- | --- | --- | --- | --- |
| TB ky 1 | 0 | 431 | 5,7351 | 0,48673 |
| 1 | 2000 | 5,9717 | 0,51925 |
| TB ky 2 | 0 | 431 | 5,6209 | 0,43739 |
| 1 | 2000 | 5,8062 | 0,49364 |
| TB ky 3 | 0 | 431 | 6,0210 | 0,52583 |
| 1 | 2000 | 6,3905 | 0,55545 |
| TB ky 4 | 0 | 431 | 6,2025 | 0,63283 |
| 1 | 2000 | 6,7197 | 0,68681 |
| TB ky 5 | 0 | 431 | 6,5221 | 0,72067 |
| 1 | 2000 | 7,0512 | 0,67981 |
| TB ky 6 | 0 | 430 | 7,0644 | 0,89355 |
| 1 | 2000 | 7,6510 | 0,66854 |
| TBC hoc tap  | 0 | 419 | 6,3233 | 0,47404 |
| 1 | 1989 | 6,7489 | 0,47743 |

Xét nhân tố “vùng cư trú” xác định theo biến “nơi sinh”, chia số liệu thành 5 vùng. Vùng 1 gồm các tỉnh trung du miền núi: Bắc Ninh, Bắc Giang, Thái Nguyên, Lào Cai, Yên Bái, Đắc Lắc, Quảng Ninh. Vùng 2 gồm các tỉnh đồng bằng bắc bộ: Hải Phòng, Hải Dương, Hưng Yên, Thái Bình, Hà Nam, Nam Định, Ninh Bình. Vùng 3 là Hà Nội, tính theo khu vực địa danh trước khi mở rộng. Vùng 4 là khu vực Hà Tây cũ, trước khi sáp nhập vào Hà Nội. Vùng 5 là các tỉnh thuộc Bắc Trung bộ: Thanh Hóa, Nghệ An, Hà Tĩnh, Quảng Bình. Bảng 3.7 trình bày một số thống kê cơ bản của điểm số trung bình các học kỳ và điểm số trung bình chung toàn khóa của sinh viên phân theo vùng, còn Bảng 3.8 cung cấp kết quả của các phép kiểm định ANOVA so sánh điểm trung bình giữa các vùng.

**Bảng 3.6. Kiểm định T-Student so sánh trung bình hai nhóm giới tính**

|  |  | Kiểm định Levene về phương sai bằng nhau | Kiểm định T-Student so sánh trung bình |
| --- | --- | --- | --- |
|  |  | F | Xác suất ý nghĩa P | T | Bậc tự do | Xác suất ý nghĩa | Chênh lệch trung bình |
| TB ky 1 |  | 3,504 | 0,061 | -8,674 | 2429 | 0,000 | -0,23659 |
| TB ky 2 |  | 10,383 | 0,001 | -7,205 | 2429 | 0,000 | -0,18525 |
| TB ky 3 |  | 3,494 | 0,062 | -12,645 | 2429 | 0,000 | -0,36956 |
| TB ky 4 |  | 11,337 | 0,001 | -14,374 | 2429 | 0,000 | -0,51720 |
| TB ky 5 |  | 0,335 | 0,563 | -14,499 | 2429 | 0,000 | -0,52916 |
| TB ky 6 |  | 14,759 | 0,000 | -15,467 | 2428 | 0,000 | -0,58659 |
| TBC hoc tap  |  | 3,110 | 0,078 | -16,604 | 2406 | 0,000 | -0,42558 |
|  |  |  |  |  |  |  |

Kết quả trong Bảng 3.7 cho ta thấy điểm trung bình học kỳ 1 và học kỳ 2 của sinh viên 5 vùng là không khác biệt một cách có ý nghĩa (xác suất ý nghĩa của phép kiểm định ANOVA trong Bảng 3.8 là 0,479 và 0.247 lớn hơn 5%). Sự khác biệt có ý nghĩa thống kê biểu hiện rõ ràng ở điểm trung bình của 4 học kỳ còn lại và điểm trung bình chung học tập toàn khóa, khi các xác suất ý nghĩa của phép kiểm định ANOVA tương ứng đều nhỏ hơn 5%.

**Bảng 3.7. Điểm số trung bình các học kỳ và toàn khóa phân theo vùng cư trú**

| Vung cu tru | Số SV | Min | Max | Trung bình | Độ lệch chuẩn |
| --- | --- | --- | --- | --- | --- |
| Mien Nui | TB ky 1 | 516 | 5,00 | 7,75 | 5,93 | 0,53 |
| TB ky 2 | 516 | 4,63 | 7,67 | 5,77 | 0,50 |
| TB ky 3 | 516 | 5,00 | 8,26 | 6,33 | 0,57 |
| TB ky 4 | 516 | 3,64 | 8,64 | 6,60 | 0,73 |
| TB ky 5 | 516 | 4,59 | 8,90 | 6,93 | 0,73 |
| TB ky 6 | 516 | 2,19 | 9,11 | 7,52 | 0,76 |
| TBC hoc tap  | 511 | 4,96 | 8,40 | 6,66 | 0,51 |
| Dong bang Bac Bo | TB ky 1 | 745 | 3,94 | 7,69 | 5,92 | 0,53 |
| TB ky 2 | 745 | 4,22 | 7,61 | 5,75 | 0,49 |
| TB ky 3 | 745 | 5,00 | 8,41 | 6,33 | 0,57 |
| TB ky 4 | 745 | 4,92 | 8,72 | 6,65 | 0,71 |
| TB ky 5 | 745 | 1,29 | 8,90 | 6,99 | 0,74 |
| TB ky 6 | 744 | 3,63 | 9,11 | 7,56 | 0,73 |
| TBC hoc tap  | 736 | 5,07 | 8,28 | 6,68 | 0,51 |
| Ha Noi | TB ky 1 | 347 | 4,50 | 7,56 | 5,97 | 0,52 |
| TB ky 2 | 347 | 4,63 | 7,61 | 5,76 | 0,47 |
| TB ky 3 | 347 | 4,28 | 8,08 | 6,24 | 0,56 |
| TB ky 4 | 347 | 4,77 | 8,20 | 6,44 | 0,69 |
| TB ky 5 | 347 | 4,85 | 8,90 | 6,79 | 0,67 |
| TB ky 6 | 347 | 4,00 | 8,89 | 7,40 | 0,70 |
| TBC hoc tap  | 341 | 5,47 | 7,97 | 6,55 | 0,48 |
| HaTay | TB ky 1 | 433 | 5,00 | 8,00 | 5,94 | 0,50 |
| TB ky 2 | 433 | 4,63 | 7,32 | 5,82 | 0,50 |
| TB ky 3 | 433 | 5,12 | 7,96 | 6,41 | 0,56 |
| TB ky 4 | 433 | 4,76 | 8,68 | 6,78 | 0,69 |
| TB ky 5 | 433 | 5,04 | 8,74 | 7,06 | 0,72 |
| TB ky 6 | 433 | 3,32 | 9,00 | 7,69 | 0,70 |
| TBC hoc tap  | 433 | 4,99 | 8,20 | 6,77 | 0,50 |
| Bac Trung bo | TB ky 1 | 390 | 3,62 | 7,38 | 5,91 | 0,51 |
| TB ky 2 | 390 | 4,17 | 7,33 | 5,78 | 0,46 |
| TB ky 3 | 390 | 5,12 | 7,88 | 6,30 | 0,56 |
| TB ky 4 | 390 | 5,12 | 8,24 | 6,63 | 0,66 |
| TB ky 5 | 390 | 5,19 | 8,62 | 6,97 | 0,67 |
| TB ky 6 | 390 | 0,95 | 9,00 | 7,54 | 0,82 |
| TBC hoc tap  | 387 | 5,28 | 7,99 | 6,67 | 0,48 |

Nhân tố “Khóa học” chia dữ liệu thành 5 khóa, từ Khóa 6 đến Khóa 10. Bảng 3.9 biểu diễn phép phân tích phương sai so sánh điểm trung bình giữa các khóa. Ta thấy yếu tố khóa học tác động rõ ràng đến hiệu quả đào tạo, khi xác suất ý nghĩa đều nhỏ hơn 5%, Bảng 3.10 trình bày một số thống kê cơ bản về điểm số trung bình các học kỳ và điểm số trung bình chung các khóa.

 **Bảng 3.8. Phân tích phương sai so sánh các nhóm sinh viên phân theo vùng cư trú**

|  |  | Tổng bình phương | Bậc tự do | Trung bình bình phương | F | Xác suất ý nghĩa |
| --- | --- | --- | --- | --- | --- | --- |
| TB ky 1 | Giữa các nhómBên trong nhómToàn phần | 0,950659,734660,685 | 424262430 | 0,2380,272 | 0,873 | 0,479 |
| TB ky 2 | Giữa các nhómBên trong nhómToàn phần | 1,296580,259581,555 | 424262430 | 0,3240,239 | 1,354 | 0,247 |
| TB ky 3 | Giữa các nhómBên trong nhómToàn phần | 5,866778,192784,058 | 424262430 | 1,4660,321 | 4,571 | **0,001** |
| TB ky 4 | Giữa các nhómBên trong nhómToàn phần | 22,9211187,0761209,996 | 424262430 | 5,7300,489 | 11,711 | **0,000** |
| TB ky 5 | Giữa các nhómBên trong nhómToàn phần | 14,9741231,4651246,438 | 424262430 | 3,7430,508 | 7,375 | **0,000** |
| TB ky 6 | Giữa các nhómBên trong nhómToàn phần | 16,6191341,1291357,747 | 424252429 | 4,1550,553 | 7,512 | **0,000** |
| TBC hoc tap | Giữa các nhómBên trong nhómToàn phần | 9,517600,247609,764 | 424032407 | 2,3790,250 | 9,525 | **0,000** |

**Bảng 3.9. Phân tích phương sai so sánh các nhóm sinh viên phân theo khóa học**

|  |  | Tổng bình phương | Bậc tự do | Trung bình bình phương | F | Xác suất ý nghĩa |
| --- | --- | --- | --- | --- | --- | --- |
| TB ky 1 | Giữa các nhóm | 45,299 | 4 | 11,325 | 44,645 | 0,000 |
| Bên trong nhóm | 615,386 | 2426 | 0,254 |  |  |
| Toàn phần | 660,685 | 2430 |  |  |  |
| TB ky 2 | Giữa các nhóm | 10,029 | 4 | 2,507 | 10,642 | 0,000 |
| Bên trong nhóm | 571,526 | 2426 | 0,236 |  |  |
| Toàn phần | 581,555 | 2430 |  |  |  |
| TB ky 3 | Giữa các nhóm | 41,784 | 4 | 10,446 | 34,141 | 0,000 |
| Bên trong nhóm | 742,273 | 2426 | 0,306 |  |  |
| Toàn phần | 784,058 | 2430 |  |  |  |
| TB ky 4 | Giữa các nhóm | 29,747 | 4 | 7,437 | 15,286 | 0,000 |
| Bên trong nhóm | 1180,250 | 2426 | 0,487 |  |  |
| Toàn phần | 1209,996 | 2430 |  |  |  |
| TB ky 5 | Giữa các nhóm | 50,716 | 4 | 12,679 | 25,724 | 0,000 |
| Bên trong nhóm | 1195,723 | 2426 | 0,493 |  |  |
| Toàn phần | 1246,438 | 2430 |  |  |  |
| TB ky 6 | Giữa các nhóm | 47,020 | 4 | 11,755 | 21,748 | 0,000 |
| Bên trong nhóm | 1310,727 | 2425 | 0,541 |  |  |
| Toàn phần | 1357,747 | 2429 |  |  |  |
| TBC hoc tap  | Giữa các nhóm | 19,562 | 4 | 4,891 | 19,912 | 0,000 |
| Bên trong nhóm | 590,201 | 2403 | 0,246 |  |  |
| Toàn phần | 609,764 | 2407 |  |  |  |

|  | **Bảng 3.10. Điểm số trung bình các học kỳ và toàn khóa phân theo khóa học**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Khoá học | Số SV | Min | Max | Trung bình | Độ lệch chuẩn |
| Khóa 6 | TB ky 1TB ky 2TB ky 3TB ky 4TB ky 5TB ky 6TBC | 304304304304304304304 | 3,624,175,124,765,053,165,23 | 7,567,448,418,528,718,688,25 | 5,925,716,476,836,997,466,76 | 0,550,470,520,680,680,770,47 |
| khóa 7 | TB ky 1TB ky 2TB ky 3TB ky 4TB ky 5TB ky 6TBC | 693693693693693692687 | 5,004,225,083,641,293,554,96 | 8,007,618,338,728,909,118,28 | 5,975,876,436,647,177,676,77 | 0,540,550,560,690,780,760,53 |
| Khóa 8 | TB ky 1TB ky 2TB ky 3TB ky 4TB ky 5TB ky 6TBC | 554554554554554554544 | 4,384,785,004,174,630,955,15 | 7,947,618,338,728,909,118,20 | 6,075,756,116,526,877,656,65 | 0,510,480,520,760,700,740,49 |
| Khóa 9 | TB ky 1TB ky 2TB ky 3TB ky 4TB ky 5TB ky 6TBC | 750750750750750750744 | 4,504,724,285,174,592,195,49 | 7,507,678,268,688,678,898,40 | 5,755,746,346,666,857,476,62 | 0,440,440,570,670,660,730,50 |
| Khóa 10 | TB ky 1TB ky 2TB ky 3TB ky 4TB ky 5TB ky 6TBC | 130130130130130130129 | 5,005,005,255,275,565,965,65 | 7,507,247,927,598,278,507,42 | 6,185,726,236,356,727,136,41 | 0,560,450,600,600,530,520,38 |

 |

Xét nhân tố “chuyên ngành” bao gồm 6 chuyên ngành: Kế toán (CKT), Tin kinh tế (CTK), Tài chính doanh nghiệp (CTD), Quản trị khách sạn (CKS), Kinh doanh xuất nhập khẩu (CKN), Quản trị kinh doanh (CKD).

Một số thống kê cơ bản của điểm số trung bình các học kỳ và điểm số trung bình chung toàn khóa của sinh viên phân theo chuyên ngành được trình bày trong Bảng 3.11, còn Bảng 3.12 được dùng để đưa ra kết quả của các phép kiểm định ANOVA so sánh điểm trung bình giữa các chuyên ngành. Sự khác biệt có ý nghĩa thống kê biểu hiện rõ ràng ở điểm trung bình của các học kỳ và điểm trung bình chung học tập, khi các xác suất ý nghĩa của phép kiểm định ANOVA tương ứng đều nhỏ hơn 5%.

 **Bảng 3.11. Điểm số trung bình các học kỳ và toàn khóa phân theo chuyên ngành**

|  | Chuyên Ngành | Số SV | Min | Max | Trung bình | Độ lệch chuẩn |
| --- | --- | --- | --- | --- | --- | --- |
|  | CKT | TB ky 1 | 1002 | 3,62 | 7,69 | 5,98 | 0,51 |
|  | TB ky 2 | 1002 | 4,17 | 7,67 | 5,85 | 0,52 |
|  | TB ky 3 | 1002 | 5,15 | 8,41 | 6,44 | 0,56 |
|  | TB ky 4 | 1002 | 3,64 | 8,72 | 7,06 | 0,63 |
|  | TB ky 5 | 1002 | 5,05 | 8,90 | 7,24 | 0,72 |
|  | TB ky 6 | 1002 | 3,86 | 9,11 | 7,98 | 0,53 |
|  | TBC hoc tap  | 999 | 4,96 | 8,40 | 6,98 | 0,44 |
|  | CTK | TB ky 1 | 52 | 5,00 | 6,69 | 5,66 | 0,38 |
|  | TB ky 2 | 52 | 5,24 | 6,84 | 5,83 | 0,40 |
|  | TB ky 3 | 52 | 5,35 | 7,46 | 6,34 | 0,54 |
|  | TB ky 4 | 52 | 5,28 | 7,88 | 6,53 | 0,62 |
|  | TB ky 5 | 52 | 5,18 | 7,86 | 6,47 | 0,57 |
|  | TB ky 6 | 52 | 6,42 | 8,75 | 7,50 | 0,54 |
|  | TBC hoc tap  | 52 | 5,55 | 7,38 | 6,47 | 0,39 |
|  | CTD | TB ky 1 | 50 | 4,50 | 6,50 | 5,56 | 0,39 |
|  | TB ky 2 | 50 | 5,12 | 6,64 | 5,74 | 0,38 |
|  | TB ky 3 | 50 | 5,12 | 7,64 | 6,43 | 0,65 |
|  | TB ky 4 | 50 | 5,17 | 7,87 | 6,53 | 0,65 |
|  | TB ky 5 | 50 | 5,54 | 8,04 | 6,65 | 0,59 |
|  | TB ky 6 | 50 | 6,29 | 8,75 | 7,63 | 0,55 |
|  | TBC hoc tap  | 50 | 5,74 | 7,29 | 6,57 | 0,43 |
|  | CKS | TB ky 1 | 583 | 4,38 | 8,00 | 5,95 | 0,58 |
|  | TB ky 2 | 583 | 4,22 | 7,24 | 5,63 | 0,41 |
|  | TB ky 3 | 583 | 5,00 | 7,92 | 6,21 | 0,56 |
|  | TB ky 4 | 583 | 4,77 | 8,14 | 6,32 | 0,58 |
|  | TB ky 5 | 583 | 1,29 | 8,27 | 6,73 | 0,58 |
|  | TB ky 6 | 582 | 2,19 | 8,58 | 7,04 | 0,60 |
|  | TBC hoc tap  | 569 | 5,58 | 7,64 | 6,37 | 0,38 |
|  | CKN | TB ky 1 | 149 | 5,00 | 7,25 | 5,99 | 0,51 |
|  | TB ky 2 | 149 | 5,00 | 7,32 | 5,99 | 0,50 |
|  | TB ky 3 | 149 | 5,32 | 7,64 | 6,47 | 0,52 |
|  | TB ky 4 | 149 | 5,28 | 7,76 | 6,56 | 0,59 |
|  | TB ky 5 | 149 | 5,74 | 8,61 | 7,05 | 0,61 |
|  | TB ky 6 | 149 | 3,55 | 8,45 | 7,39 | 0,62 |
|  | TBC hoc tap  | 149 | 4,99 | 7,78 | 6,66 | 0,45 |
|  | CKD | TB ky 1 | 595 | 3,94 | 7,56 | 5,87 | 0,47 |
|  | TB ky 2 | 595 | 4,72 | 7,22 | 5,73 | 0,47 |
|  | TB ky 3 | 595 | 4,28 | 8,08 | 6,19 | 0,55 |
|  | TB ky 4 | 595 | 4,17 | 8,10 | 6,24 | 0,58 |
|  | TB ky 5 | 595 | 4,59 | 8,74 | 6,75 | 0,70 |
|  | TB ky 6 | 595 | 0,95 | 8,89 | 7,35 | 0,84 |
|  | TBC hoc tap  | 589 | 5,15 | 7,75 | 6,48 | 0,46 |

 **Bảng 3.12. phân tích phương sai so sánh các nhóm sinh viên phân theo chuyên ngành**

|  |  | Tổng bình phương | Bậc tự do | Trung bình bình phương | F | Xác suất ý nghĩa |
| --- | --- | --- | --- | --- | --- | --- |
| TB ky 1 | Giữa các nhóm | 15,691 | 5 | 3,138 | 11,799 | 0,000 |
| Bên trong nhóm | 644,994 | 2425 | 0,266 |  |  |
| Toàn phần | 660,685 | 2430 |  |  |  |
| TB ky 2 | Giữa các nhóm | 26,930 | 5 | 5,386 | 23,550 | 0,000 |
| Bên trong nhóm | 554,624 | 2425 | 0,229 |  |  |
| Toàn phần | 582 | 2430 |  |  |  |
| TB ky 3 | Giữa các nhóm | 34,154 | 5 | 6,831 | 22,089 | 0,000 |
| Bên trong nhóm | 749,904 | 2425 | 0,309 |  |  |
| Toàn phần | 784 | 2430 |  |  |  |
| TB ky 4 | Giữa các nhóm | 330,668 | 5 | 66,134 | 182,382 | 0,000 |
| Bên trong nhóm | 879,329 | 2425 | 0,363 |  |  |
| Toàn phần | 1210 | 2430 |  |  |  |
| TB ky 5 | Giữa các nhóm | 151,937 | 5 | 30,387 | 67,327 | 0,000 |
| Bên trong nhóm | 1094,501 | 2425 | 0,451 |  |  |
| Toàn phần | 1246 | 2430 |  |  |  |
| TB ky 6 | Giữa các nhóm | 362,229 | 5 | 72,446 | 176,399 | 0,000 |
| Bên trong nhóm | 995,518 | 2424 | 0,411 |  |  |
| Toàn phần | 1358 | 2429 |  |  |  |
| TBC hoc tap  | Giữa các nhóm | 167,974 | 5 | 33,595 | 182,654 | 0,000 |
| Bên trong nhóm | 441,790 | 2402 | 0,184 |  |  |
| Toàn phần | 610 | 2407 |  |  |  |

**3.3. Phân tích hiệu quả theo mô hình hồi quy tuyến tính cổ điển**

Việc phân tích bằng phương pháp đơn biến như trình bày ở mục trên chỉ có thể đánh giá một cách riêng rẽ hiệu quả của từng nhân tố lên kết quả đào tạo, do đó có thể đưa ra những kết luận không thật chính xác, vì giữa các nhân tố có thể tồn tại mối liên hệ nào đó, gây ra tác động bắc cầu từ một nhân tố này đến kết quả đào tạo thông qua trung gian của một nhân tố khác. Như vậy, để đánh giá đúng vai trò của từng nhân tố ảnh hưởng đến kết quả đào tạo, ta phải dùng mô hình đa biến, trong đó phải có mặt đồng thời tất cả các nhân tố có khả năng ảnh hưởng. Tiếp sau đây ta sẽ dùng mô hình hồi quy tuyến tính cổ điển như một mô hình đa biến để thức hiện việc đánh giá đó.

  **Bảng 3.13. Mô hình hồi quy bội đánh giá tác động của các nhân tố lên kết quả học tập**

|  | Điểm Học kỳ 1 | Điểm Học kỳ 2 | Điểm Học kỳ 3 | Điểm Học kỳ 4 | Điểm Học kỳ 5 | Điểm Học kỳ 6 | ĐiểmTBC |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Biến độc lập | Hệ số hồi qui | XS ý nghĩa | Hệ số hồi qui | XS ý nghĩa | Hệ số hồi qui | XS ý nghĩa | Hệ số hồi qui | XS ý nghĩa | Hệ số hồi qui | XS ý nghĩa | Hệ số hồi qui | XS ý nghĩa | Hệ số hồi qui | XS ý nghĩa |
| (Constant) | 5,822 | **0,000** | 5,608 | **0,000** | 6,268 | **0,000** | 6,866 | **0,000** | 6,820 | **0,000** | 7,332 | **0,000** | 6,683 | **0,000** |
| CKT | 0 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| CKD | -0,106 | **0,000** | -0,090 | **0,001** | -0,121 | **0,000** | -0,732 | **0,000** | -0,381 | **0,000** | -0,574 | **0,000** | -0,428 | **0,000** |
| CKN | -0,024 | 0,611 | 0,065 | 0,143 | -0,037 | 0,462 | -0,445 | **0,000** | -0,358 | **0,000** | -0,720 | **0,000** | -0,371 | **0,000** |
| CKS | -0,064 | **0,035** | -0,248 | **0,000** | -0,150 | **0,000** | -0,688 | **0,000** | -0,454 | **0,000** | -0,979 | **0,000** | -0,583 | **0,000** |
| CTD | -0,209 | **0,005** | -0,052 | 0,467 | 0,121 | 0,127 | -0,416 | **0,000** | -0,370 | **0,000** | -0,137 | 0,128 | -0,259 | **0,000** |
| CTK | -0,141 | 0,051 | 0,017 | 0,801 | -0,062 | 0,420 | -0,517 | **0,000** | -0,628 | **0,000** | -0,357 | **0,000** | -0,425 | **0,000** |
| Thi khoi A | 0 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| Thi khoi D1 | 0,016 | 0,619 | 0,015 | 0,625 | 0,034 | 0,316 | 0,044 | 0,245 | 0,078 | 0,057 | 0,095 | **0,014** | 0,048 | 0,068 |
| Nam SV | 0 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| Nu sinh vien | 0,203 | **0,000** | 0,163 | **0,000** | 0,366 | **0,000** | 0,371 | **0,000** | 0,465 | **0,000** | 0,470 | **0,000** | 0,346 | **0,000** |
| Hoc dung t | 0 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| Hoc muon | -0,039 | 0,070 | -0,021 | 0,311 | -0,097 | **0,000** | -0,115 | **0,000** | -0,079 | **0,005** | -0,057 | **0,032** | -0,073 | **0,000** |
| ĐbsHong | 0 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| MienNui | 0,006 | 0,835 | 0,022 | 0,421 | 0,023 | 0,443 | 0,002 | 0,961 | -0,033 | 0,375 | -0,011 | 0,740 | 0,008 | 0,732 |
| HaTay | 0,011 | 0,714 | 0,038 | 0,191 | 0,077 | **0,016** | 0,077 | **0,030** | 0,021 | 0,586 | 0,046 | 0,206 | 0,044 | 0,073 |
| HaNoi | 0,023 | 0,490 | 0,048 | 0,130 | -0,007 | 0,843 | -0,075 | 0,055 | -0,070 | 0,104 | -0,024 | 0,559 | -0,019 | 0,499 |
| BacTBo | 0,006 | 0,856 | 0,059 | **0,048** | 0,000 | 0,997 | 0,016 | 0,660 | 0,042 | 0,296 | 0,027 | 0,477 | 0,028 | 0,276 |
| Khoa6 | 0 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| Khoa7 | 0,028 | 0,445 | 0,162 | **0,000** | -0,064 | 0,098 | -0,155 | **0,000** | 0,222 | **0,000** | 0,389 | **0,000** | 0,065 | **0,028** |
| Khoa8 | 0,135 | **0,000** | 0,065 | 0,062 | -0,380 | **0,000** | -0,175 | **0,000** | -0,062 | 0,191 | 0,371 | **0,000** | -0,024 | 0,418 |
| Khoa9 | -0,186 | **0,000** | 0,051 | 0,137 | -0,170 | **0,000** | -0,117 | **0,006** | -0,083 | 0,075 | 0,139 | **0,001** | -0,072 | **0,015** |
| Khoa10 | 0,235 | 0,060 | 0,175 | **0,002** | -0,189 | **0,003** | -0,102 | 0,148 | -0,025 | 0,743 | 0,356 | **0,000** | 0,019 | 0,701 |

Bảng 3.13 trình bày tổng hợp kết quả ước lượng 7 mô hình hồi quy tương ứng với biến phụ thuộc lần lượt là điểm trung bình của sáu học kỳ và điểm trung bình chung toàn khóa. Các nhân tố được đánh giá trong các mô hình bao gồm “Chuyên ngành”, “Khối thi đầu vào”, “Giới tính sinh viên”, “Nhập học đúng tuổi”, “Vùng cư trú” và “Khóa học” . Các nhân tố đó đều được mã hóa lại thành các biến nhị phân.

Cụ thể, nhân tố “Chuyên ngành” chia sinh viên thành 6 nhóm chuyên ngành, được mã hóa thành năm biến nhị phân CKD (Quản trị kinh doanh), CKN (Kinh doanh xuất nhập khẩu), CKS (Quản trị khách sạn), CTD (Tài chính doanh nghiệp) và CTK (Tin học kinh tê). Đối với nhân tố này, nhóm chuyên ngành CKT (Kế toán) được chọn làm nhóm chứng để so sánh với năm nhóm còn lại

Nhân tố “Vùng cư trú”, được mã hóa thành bốn biến nhị phân MienNui, HaTay, HaNoi và BacTBo, với nhóm chứng là các sinh viên cư trú tại các tỉnh Đồng bằng sông Hồng. Nhân tố “Khóa học” cũng được mã hóa thành bốn biến nhị phân Khoa7, Khoa8, Khoa9 và Khoa10, lấy Khóa 6 làm nhóm chứng.

Các nhân tố “Khối thi đầu vào”, “Giới tính sinh viên”, “Nhập học đúng tuổi”, mối nhân tố chỉ có hai mức, được mã hóa thành các biến nhị phân ThiKhoiD1, NuSinhVien và HocMuon, với các nhóm chứng tương ứng lần lượt là nhóm sinh viên thi khối A, nhóm nam sinh viên và nhóm sinh viên nhập học đúng tuổi ( trước khi tròn 19 tuổi).

Trong bảng 3.13, mỗi mô hình hồi quy được trình bày trong hai cột, một cột chứa các hệ số hồi quy của các biến độc lập, cột còn lại chứa các xác suất ý nghĩa tương ứng của từng hệ số hồi quy. Hệ số hồi quy nào có xác suất ý nghĩa nhỏ hơn 5% được coi là khác 0 một cách có ý nghĩa thống kê, nhân tố tương ứng sẽ được đánh giá là có tác động đối với biến phụ thuộc của mô hình.

Điểm nổi bật trong cả bảy mô hình hồi quy bội nói trên là nhân tố “Giới tính sinh viên” có tác động thực sự đối với hiệu quả học tập, khi các xác suất ý nghĩa của biến NuSinhVien đều nhỏ hơn 5%, với hệ số hồi quy dương. Điều này cho thấy nữ sinh viên học tập tốt hơn nam sinh viên. Kết quả này hoàn toàn trùng hợp với kết quả của phân tích đơn biến đã trình bày trong mục trước.

Nhân tố “Nhập học đúng tuổi” không tác động đến kết quả học tập của hai học kỳ đầu mà có tác động đến kết quả học tập của các học kỳ 3, 4, 5, 6 và điểm trung bình chung, với các hệ số hồi quy âm cho biết sinh viên nhập học muộn có kết quả kém hơn sinh viên học đúng tuổi. Kết quả này khá thống nhất với phần phân tích đơn biến cho rằng yếu tố này ảnh hưởng đến kết quả học tập của tất cả các học kỳ và của toàn khóa học.

Nhân tố “Khối thi đầu vào” hầu như không ảnh hưởng đáng kể tới kết quả đào tạo, khi chỉ có xác suất ý nghĩa ở mô hình hồi quy của Học kỳ 6 là nhỏ hơn 5%. Trong khi đó theo mô hình phân tích đơn biến thì nhân tố này chỉ ảnh hưởng tới kết quả học tập Học kỳ 1.

Ảnh hưởng của nhân tố “Chuyên ngành” lên kết quả đào tạo đều được khẳng định trong phân tích đơn biến phía trên cũng như trong phân tích đa biến bằng mô hình hồi quy trình bày tại Bảng 3.13. Tuy nhiên, mô hình hồi quy cho biết rõ hơn về mức độ ảnh hưởng, chỉ ra rằng khi so với chuyên ngành CKT thì kết quả học tập của các chuyên ngành khác đều thấp hơn ở hầu hết các học kỳ, nhất là ở các học kỳ cuối và ở điểm trung bình chung

Qua mô hình hồi quy, nhân tố “Khóa học” cũng ảnh hưởng một cách có ý nghĩa thống kê đến kết quả học tập, giống với kết luận của phần phân tích đơn biến. Tuy nhiên các mô hình hồi quy cho biết nhiều thông tin hơn, chỉ ra rằng khi so sánh với Khóa 6, Khóa 7 có kết quả học tập tốt hơn ở Học kỳ 2, Học kỳ 5, Học kỳ 6 và điểm trung bình chung, nhưng lại kém hơn ở Học kỳ 4. Khóa 8 có kết quả tốt hơn Khóa 6 ở Học kỳ 1 và Học kỳ 6, song lại có kết quả kém hơn ở Học kỳ 3 và Học kỳ 4. Khóa 9 chỉ có kết quả tốt hơn so với Khóa 6 ở Học kỳ 6, mà lại có kết quả kém hơn ở Học kỳ 1, Học kỳ 3, Học kỳ 4 và điểm trung bình chung. Khóa 10 học khá hơn Khóa 6 ở Học kỳ 2 và Học kỳ 6 nhưng kém hơn ở Học kỳ 3.

Theo phân tích đơn biến (Bảng 3.8), nhân tố “Vùng cư trú” không ảnh hưởng tới kết quả học tập của hai kỳ đầu nhưng ảnh hưởng tới kết quả học tập của các học kỳ còn lại và của điểm trung bình chung. Nhưng kết quả phân tích đa biến bằng mô hình hồi quy lại chỉ ra rằng nhân tố này chỉ ảnh hưởng tới kết quả học tập của các Học kỳ 2,3 và 4, chỉ rõ ở Học kỳ 2 sinh viên cư trú ở khu vực Bắc Trung bộ có kết quả học tập tốt hơn sinh viên vùng Đồng bằng sông Hồng. Trong khi đó, sinh viên khu vực Hà Tây cũ học tốt hơn sinh viên các tỉnh Đồng bằng sông Hồng tại Học kỳ 3 và Học kỳ 4.

Như đã trình bày trong Mục 1.4, các phép kiểm định trong mô hình hồi quy bội phân tích ảnh hưởng của các nhân tố lên điểm trung bình của các học kỳ trên đây chỉ có hiệu lực nếu năm giả thiết của mô hình được thỏa mãn. Trong số năm giả thiết đó cần chú ý nhất đến giả thiết về tính độc lập và giả thiết về phương sai không đổi của các sai số. Do vậy ta cần kiểm tra các giả thiết đó. Qua các đồ thị trong Hình 3.14 ta thấy rõ ràng giữa sai số hồi quy và biến phụ thuộc của mô hình có mối ràng buộc khá chặt chẽ, khi giá trị của biến phụ thuộc tăng lên thì giá trị của các sai số cũng tăng theo. Như vậy, rõ ràng tính độc lập của các sai số đã bị vi phạm, các phép kiểm định của mô hình hồi quy bội không thực sự có hiệu lực và ta cần xây dựng mô hình hồi quy nhiều mức để khắc phục các hạn chế đó.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

**Hình 3.14. Đồ thị sai số phần dư và biến phụ thuộc trong các mô hình hồi quy cổ điển**

**3.4. Phân tích hiệu quả theo mô hình hồi quy tuyến tính nhiều mức**

Kết quả phân tích mô hình hồi quy nhiều mức bẳng phần mềm STATA để đánh giá tác động của các nhân tố lên hiệu quả đào tạo được trình bày trong Bảng 3.14. Trong các mô hình này, đơn vị ở mức 1 là sinh viên, đơn vị ở mức 2 là các khóa (K6 đến K10). Việc phân tích ảnh hưởng của các yếu tố ngẫu nhiên ở mức 1 và mức 2 lên phần dư và hệ số chặn của mô hình được tổng hợp lại thành Bảng 3.15.

  **Bảng 3.14. Mô hình hồi quy nhiều mức đánh giá tác động của các nhân tố lên kết quả học tập**

|  | Điểm Học kỳ 1 | Điểm Học kỳ 2 | Điểm Học kỳ 3 | Điểm Học kỳ 4 | Điểm Học kỳ 5 | Điểm Học kỳ 6 | ĐiểmTBC |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Biến độc lập | Hệ số hồi qui | XS ý nghĩa | Hệ số hồi qui | XS ý nghĩa | Hệ số hồi qui | XS ý nghĩa | Hệ số hồi qui | XS ý nghĩa | Hệ số hồi qui | XS ý nghĩa | Hệ số hồi qui | XS ý nghĩa | Hệ số hồi qui | XS ý nghĩa |
| CKT | 0 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| CKD | -0.106 | **0.000** | -0.090 | **0.001** | -0.123 | **0.000** | -0.734 | **0.000** | -0.383 | **0.000** | -0.573 | **0.000** | -0.429 | **0.000** |
| CKN | -0.023 | 0.623 | 0.074 | 0.098 | -0.035 | 0.480 | -0.452 | **0.000** | -0.347 | **0.000** | -0.716 | **0.000** | -0.364 | **0.000** |
| CKS | -0.060 | **0.047** | -0.239 | **0.000** | -0.153 | **0.000** | -0.693 | **0.000** | -0.457 | **0.000** | -0.975 | **0.000** | -0.582 | **0.000** |
| CTD | -0.211 | **0.004** | -0.052 | 0.466 | 0.120 | 0.131 | -0.420 | **0.000** | -0.375 | **0.000** | -0.137 | 0.126 | -0.265 | **0.000** |
| CTK | -0.143 | **0.048** | 0.018 | 0.799 | -0.064 | 0.409 | -0.520 | **0.000** | -0.633 | **0.000** | -0.357 | **0.000** | -0.432 | **0.000** |
| Thi khoi A | 0 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| Thi khoi D1 | 0.0155 | 0.627 | 0.014 | 0.642 | 0.035 | 0.307 | 0.045 | 0.232 | 0.079 | 0.056 | 0.094 | **0.015** | 0.048 | 0.067 |
| Nam SV | 0 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| Nu sinh vien | 0.204 | **0.000** | 0.165 | **0.000** | 0.365 | **0.000** | 0.369 | **0.000** | 0.463 | **0.000** | 0.471 | **0.000** | 0.345 | **0.000** |
| Hoc dung t | 0 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| Hoc muon | -0.039 | 0.072 | -0.021 | 0.315 | -0.097 | **0.000** | -0.114 | **0.000** | -0.078 | **0.006** | -0.057 | **0.031** | -0.072 | **0.000** |
| DbsHong | 0 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| MienNui | 0.006 | 0.829 | 0.022 | 0.412 | 0.023 | 0.450 | 0.001 | 0.981 | -0.033 | 0.373 | -0.011 | 0.749 | 0.008 | 0.731 |
| HaTay | 0.011 | 0.714 | 0.038 | 0.187 | 0.077 | **0.016** | 0.076 | **0.032** | 0.022 | 0.569 | 0.047 | 0.202 | 0.045 | 0.069 |
| HaNoi | 0.025 | 0.444 | 0.052 | 0.102 | -0.009 | 0.805 | -0.076 | **0.049** | -0.073 | 0.091 | -0.022 | 0.592 | -0.019 | 0.493 |
| BacTBo | 0.006 | 0.853 | 0.058 | **0.050** | -0.000 | 0.999 | 0.017 | 0.644 | 0.041 | 0.307 | 0.027 | 0.481 | 0.027 | 0.285 |
| (Constant) | 5.859 | 0.000 | 5.692 | 0.000 | 6.110 | 0.000 | 6.757 | 0.000 | 6.834 | 0.000 | 7.580 | 0.000 | 6.680 | 0.000 |

So sánh hai Bảng 3.13 và 3.14 ta thấy các tác động có ý nghĩa thống kê của các nhân tố lên kết quả học tập đã bộc lộ trong Bảng 3.13 đều giữ nguyên vai trò của mình trong bảng 3.14. Ngoài ra, mô hình hồi quy nhiều mức còn chỉ rõ thêm vai trò của một số nhân tố khác nữa. Cụ thể trong mô hình hồi quy đối với điểm trung bình Học kỳ 1, Học kỳ 5 và trung bình chung, sinh viên chuyên ngành CTK có kết quả học tập thấp hơn một cách có ý nghĩa so với chuyên ngành CKT (hệ số hồi quy bằng -0.143, -0.633 và -0.432 với xác suất ý nghĩa tương ứng bằng 0.048, 0.000 và 0.000). Đối với điểm trung bình học kỳ 4 sinh viên cư trú tại Hà Nội có điểm trung bình thấp hơn một cách có ý nghĩa so với sinh viên các tỉnh Đồng bằng sông Hồng (hệ số hồi quy bằng -0.076 với xác suất ý nghĩa bằng 0.049).

Kết quả trong Bảng 3.15 chỉ rõ trong tất cả các mô hình hồi quy, độ lệch tiêu chuẩn của hệ số chặn và độ lệch tiêu chuẩn của phần dư đều thực sự khác 0 một cách có ý nghĩa. Điều đó thể hiện ở chỗ các khoảng tin cậy của chúng đều tách biệt rõ ràng khỏi điểm 0. Độ lệch tiêu chuẩn của hệ số chặn biểu diễn độ biến động của hệ số chặn đó giữa các đơn vị ở mức 2, tức là giữa các khóa học. Điều này cũng đã được phản ánh trong Bảng 3.13 khi nhiều hệ số hồi quy của các biến nhị phân Khoa7 - Khoa10 ứng với các mức của nhân tố “Khóa học” (khác mức Khoa10) khác 0 một cách có ý nghĩa thống kê. Độ lệch tiêu chuẩn của phần dư thể hiện độ biến động của phần dư đó giữa các đơn vị mức 2, giữa các khóa học. Độ lệch tiêu chuẩn này khác 0, khẳng định phương sai của phần dư biến động một cách có ý nghĩa giữa các đơn vị mức 2. Điều đó cho thấy giả thiết phương sai phần dư không đổi của mô hình hồi quy tuyến tính cổ điển bị vi phạm, càng làm tăng ý nghĩa của việc sử dụng mô hình hồi quy nhiều mức thay cho mô hình hồi quy tuyến tính cổ điển để đánh giá được đúng đắn vai trò của các nhân tố tác động lên hiệu quả đào tạo.

**Bảng 3.15. Ước lượng tham số ngẫu nhiên trong mô hình hồi quy nhiều mức**

|  |  |  |
| --- | --- | --- |
| **Biến phụ thuộc** | **Độ lệch chuẩn của hệ số chặn** | **Độ lệch chuẩn của phần dư hồi qui** |
|  | Ước lượng | Sai số chuẩn | Khoảng tin cậy 95% | Ước lượng | Sai số chuẩn | Khoảng tin cậy 95% |
| Điểm Học kỳ 1 | 0.15387 | 0.05685 | (0.07459;0.31741) | 0.49404 | 0.00711 | (0.48029;0.50817) |
| Điểm Học kỳ 2 | 0.06500 | 0.02834 | (0.02844;0.15314) | 0.47172 | 0.00679 | (0.45859;0.48522) |
| Điểm Học kỳ 3 | 0.14378 | 0.05260 | (0.07019;0.29449) | 0.52580 | 0.00757 | (0.51117;0.54080) |
| Điểm Học kỳ 4 | 0.06151 | 0.02767 | (0.02547;0.14853) | 0.58081 | 0.00836 | (0.56465;0.59743) |
| Điểm Học kỳ 5 | 0.12180 | 0.04557 | (0.05850;0.25358) | 0.63923 | 0.00920 | (0.62145;0.65752) |
| Điểm Học kỳ 6 | 0.16976 | 0.06260 | (0.08240;0.34970) | 0.59789 | 0.00861 | (0.58125;0.61450) |
| ĐiểmTBC | 0.05015 | 0.02015 | (0.02282;0.11021) | 0.40546 | 0.00586 | (0.39410;0.41712) |

**3.5. Kết luận**

Mô hình hồi quy nhiều mức giúp ta có được những kết luận đúng đắn về ảnh hưởng của các nhân tố lên kết quả học tập của sinh viên từng học kỳ và điểm Trung bình chung. Trước tiên, nữ sinh viên có ý thức học tập tốt hơn nam sinh viên, thể hiện ở chỗ với cùng điều kiện như nhau thì điểm trung bình của nữ sinh viên luôn cao hơn nam sinh viên. Nhân tố “Nhập học đúng tuổi” ảnh hưởng đến đáng kể đến kết quả học tập của tất cả các học kỳ và của toàn khóa học. Sinh viên nhập học vào trường muộn so với năm tốt nghiệp phổ thông của mình thường sẽ có kết quả học tập kém hơn. Điều này có thể do những sinh viên này hoặc không trúng tuyển ngay ở kỳ thi Cao đẳng – Đại học đầu tiên sau khi tốt nghiệp phổ thông (phần nhiều là do sức học yếu hơn), hoặc là những người đã đi làm một vài năm rồi mới tham gia thi tuyển vào trường (kiến thức cơ bản có phần bị rơi rụng theo thời gian).

Nhân tố “Khối thi đầu vào” rất ít ảnh hưởng tới kết quả đào tạo. Do vậy, trong quá trình tuyển sinh không cần phân biệt khối thi của học sinh dự tuyển, nên đặt ra mức điểm sàn như nhau đối với học sinh thi các khối thi khác nhau. Trong khi đó nhân tố “Chuyên ngành” lại có vai trò quan trọng đối với điểm trung bình các học kỳ cũng như điểm trung bình chung. Vì thế, có thể dựa vào các kết quả phân tích trên đây để điều chỉnh mức điểm sàn thích hợp cho học sinh thi tuyển vào từng chuyên ngành cụ thể. Chẳng hạn, điểm sàn đầu vào của chuyên ngành CKT (Kế toán) có thể cao hơn điểm sàn các chuyên ngành khác 0.5 điểm.

Nhân tố “Vùng cư trú” ảnh hưởng phần nào tới kết quả học tập. So với sinh viên các tỉnh thuộc khu vực Đồng bằng sông Hồng, sinh viên cư trú ở khu vực Bắc Trung bộ và sinh viên khu vực Hà Tây cũ có kết quả học tập tốt hơn, còn sinh viên cư trú tại Hà Nội lại có kết quả kém hơn ở một số học kỳ. Do nhà trường thuộc địa bàn Hà Tây cũ, các sinh viên của khu vực này sẽ có điều kiện sinh hoạt, đi lại thuận lợi hơn sinh viên các khu vực khác, dẫn đến một số ưu thế nào đó trong quá trình học tập. Trong khi đó, sinh viên cư trú tại Hà Nội, là địa bàn có nhiều lựa chọn để thi tuyển vào hơn, nên thường chỉ theo học trường này khi không có đủ điều kiện về lực học để được tuyển vào các trường khác trong nội thành. Kết quả học tập có phần trội hơn của sinh viên khu vực Bắc Trung bộ có thể phản ánh tinh thần học tập cao hơn của các sinh viên này. Đặc điểm đó có thể thấy được ở cả các trường Cao đẳng – Đại học khác.

Kết quả phân tích trên đây có thể coi là những bằng chứng khoa học xác đáng mà Nhà trường nên tham khảo để đề ra các biện pháp thích hợp điều hành các hoạt động tuyển sinh và đào tạo nói chung một cách có hiệu quả nhất.

Phương pháp phân tích thống kê như đã sử dụng ở trong nghiên cứu này có thể được ứng dụng rộng rãi trong nghiên cứu giáo dục đào tạo, hỗ trợ cho việc hoạch định chính sách của các nhà quản lý, cũng như việc xây dựng chiến lược phát triển của ngành Giáo dục - Đào tạo trong cả nước.

Nghiên cứu này mới chỉ thực hiện trong phạm vi số liệu thu thập của Trường Cao đẳng Kinh tế - Kỹ thuật Thương mại trong một số năm và trong một số chuyên ngành đào tạo chính của Trường, do đó có thể mới phản ánh các quy luật trong phạm vi hẹp về cả không gian và thời gian. Muốn có được các kết luận chính xác hơn và có tính phổ biến hơn, cần phải tiến hành các nghiên cứu trên diện rộng ở nhiều trường Cao đẳng – Đại học và của nhiều năm. Hơn nữa, Trình độ nghiên cứu cũng như chuyên môn của tác giả luận văn còn có nhiều hạn chế, không tránh được những thiếu sót. Tác giả luận văn xin tiếp thu mọi sự đóng góp phê bình của các thầy cô, các đồng nghiệp và các bạn để bản luận văn được hoàn thiện hơn. Xin chân thành cảm ơn.

**TÀI LIỆU THAM KHẢO**

[1] Trần Mạnh Tuấn, *Xác suất & Thống kê lý thuyết và thực hành tính toán*, Nhà xuất bản ĐHQGHN, 2004.

[2] Đào Hữu Hồ, *Xác suất thống kê*, Nhà xuất bản ĐHQGHN, 2001

[3] Phạm Văn Kiều, *Xác suất thống kê*, Nhà xuất bản ĐHSP, 2006.

[4] Lê Sĩ Đồng, *Xác suất thống kê và ứng dụng*, Nhà xuất bản Giáo Dục, 2006.

[5] Hoàng Trọng – Chu Nguyễn Mộng Ngọc, *Phân Tích dữ liệu nghiên cứu với SPSS*, Nhà xuất bản Hồng Đức, 2008

[6] Harvey Goldstein, *Multilevel Statistical Models*, London: Institute of Education, Multilevel Models, Project, 1999.

  Quelle: http://www.arnoldpublishers.com /support/ Goldstein

[7] Http://www.ats.ucla.edu/stat/stata/modules/