



Tạp chí Khoa học Trường Đại học Cần Thơ
 Phần A: Khoa học Tự nhiên, Công nghệ và Môi trường

website: sj.ctu.edu.vn

DOI:10.22144/jvn.2017.063

PHÂN TÍCH THỐNG KÊ TRỌNG LƯỢNG TRẺ SƠ SINH

Võ Văn Tài¹, Lê Thị Mỹ Xuân¹, Nguyễn Thị Hồng Dân¹, Danh Ngọc Thắm¹ và Nguyễn Hữu Nghĩa²

¹Khoa Khoa học Tự nhiên, Trường Đại học Cần Thơ

²Khoa Cơ bản, Trường Đại học Xây dựng miền Tây

Thông tin chung:

Ngày nhận bài: 22/07/2016

Ngày nhận bài sửa: 30/08/2016

Ngày duyệt đăng: 27/06/2017

Title:

Statistical analysis for weight of newborn

Từ khóa:

Đa biến, đơn biến, phân loại, sai số, trọng lượng của trẻ sơ sinh

Keywords:

Classification, error, multivariate, multivariate, weight of newborn

ABSTRACT

From real data collected at the Maternity Hospital of Can Tho city, by methods of univariate and multivariate statistical analysis, the article is aimed to determine factors and group factors that have statistical significance to influence to weight of newborn. Based on the classification models, an optimal model in forecasting standard weight of newborns was built as well. The researched results could be useful information in caring health of pregnant women and applied to many other problems in reality.

TÓM TẮT

Từ số liệu thực tế được thu thập tại Bệnh viện Phụ sản Thành phố Cần Thơ, bằng các phương pháp phân tích thống kê đơn biến và đa biến, bài viết xác định các nhân tố và nhóm các nhân tố có ý nghĩa thống kê ảnh hưởng đến trọng lượng trẻ sơ sinh. Dựa trên các mô hình phân loại, bài viết cũng xây dựng mô hình tối ưu trong dự báo trọng lượng trẻ sơ sinh đủ cân. Kết quả nghiên cứu là thông tin hữu ích trong chăm sóc sức khỏe bà mẹ mang thai và có thể áp dụng cho nhiều vấn đề khác trong thực tế.

Trích dẫn: Võ Văn Tài, Lê Thị Mỹ Xuân, Nguyễn Thị Hồng Dân, Danh Ngọc Thắm và Nguyễn Hữu Nghĩa, 2017. Phân tích thống kê trọng lượng trẻ sơ sinh. Tạp chí Khoa học Trường Đại học Cần Thơ. 50a: 29-36.

1 GIỚI THIỆU

Trọng lượng trẻ sơ sinh (TLTSS) nói lên nhiều điều về sức khỏe và sự phát triển sau này của trẻ. Trẻ sơ sinh quá nặng hoặc quá nhẹ so với mức đạt chuẩn đều không tốt. Theo Bộ y tế Việt Nam, TLTSS đủ cân từ 2500 gam đến 3800 gam, cao hơn 3800 gam được xem là nặng cân và nhỏ hơn 2500 gam được xem là nhẹ cân. Hầu hết các nghiên cứu đều cho thấy, trẻ nhẹ cân có chỉ số thông minh kém hơn trẻ đủ cân. Những đứa trẻ nhẹ cân thường có tỉ lệ phát triển thần kinh bất thường, chậm phát triển cao hơn ở trẻ đủ cân. Trong lứa tuổi học đường, trẻ nhẹ cân có chỉ số thông minh, chỉ số phối hợp nhìn - vận động và khả năng đọc thấp hơn trẻ đủ cân. Các vấn đề về cư xử như kích động, kém phối hợp động tác, khó tập trung... cũng

thường gặp hơn ở trẻ nhẹ cân. Theo thống kê của Bộ y tế, tỉ lệ trẻ nhẹ cân ở nước ta là khoảng 10% và tập trung nhiều vào vùng nông thôn. Có rất nhiều nghiên cứu khác nhau về nguyên nhân dẫn đến trẻ nhẹ cân. Những nguyên nhân được tổng kết là di truyền, dinh dưỡng, môi trường, sức khỏe của mẹ, tuy nhiên có nhiều trường hợp vẫn chưa tìm được nguyên nhân. Nguyên nhân trẻ nhẹ cân còn phụ thuộc vào từng quốc gia, từng địa phương. Cho đến nay, các nghiên cứu để tìm nguyên nhân dẫn đến trẻ thiếu cân hầu hết đều dựa vào các phân tích thống kê. Việc tìm được các nguyên nhân chính là một thông tin hữu ích cho các bác sĩ, các bà mẹ trong chăm sóc sức khỏe mang thai để có được những đứa trẻ khỏe mạnh từ ban đầu, là tiền đề quan trọng để có một dân số khỏe mạnh.

Dựa vào số liệu thực tế được lấy tại Bệnh viện Phụ sản Thành phố Cần Thơ, bài viết nghiên cứu các vấn đề liên quan đến trọng lượng trẻ sơ sinh. Việc nghiên cứu được thực hiện theo hai hướng: (i) phân tích thống kê đơn biến và đa biến để xác định các nhân tố cũng như nhóm nhân tố ảnh hưởng đến trọng lượng trẻ sơ sinh, và (ii) đánh giá trọng lượng trẻ sơ sinh đủ cân qua các mô hình phân loại để tìm sự tối ưu. Mục đích của các nghiên cứu này là xác định các yếu tố thật sự ảnh hưởng đến TLSS, trẻ thiếu cân để từ đó có những khuyến cáo cho các bà mẹ mang thai, cho những người có trách nhiệm trong chăm sóc sức khỏe các bà mẹ này. Kết quả của bài viết cũng là thông tin hữu ích cho ngành y tế trong việc chăm sóc những bà mẹ mang thai ở khu vực Đồng bằng sông Cửu Long.

Cấu trúc của bài viết như sau: Phần 2 trình bày các phương pháp phân tích số liệu, các phương

pháp phân loại; đồng thời, trình bày về số liệu mẫu và các bước thực hiện. Phần 3 trình bày kết quả thực hiện theo hai hướng (i) và (ii). Phần cuối cùng là kết luận của bài viết.

2 TỔNG QUAN VIỆC THỰC HIỆN

2.1 Nguồn số liệu và cơ cấu mẫu

Được sự cho phép của Bệnh viện Phụ Sản Thành phố Cần Thơ, việc trích xuất các thông tin từ hồ sơ lưu của các bà mẹ đã sinh ở đây trong suốt 1 tháng cuối năm 2015 được tiến hành. Việc xác định các biến ban đầu có thể ảnh hưởng đến TLSS được sự tư vấn từ các bác sĩ tại đây. Mẫu nghiên cứu chỉ gồm hai đối tượng: Trẻ sơ sinh đủ cân và nhẹ cân. Số liệu gồm có 11 biến, trong đó có 2 biến định lượng và 9 biến định tính được khảo sát trên 513 sản phụ với cơ cấu mẫu theo từng nhóm đối tượng được trình bày ở Bảng 1.

Bảng 1: Cơ cấu mẫu các biến được khảo sát

| Chỉ tiêu | Ký hiệu | Thang đo | Tỉ lệ (%) |
|-------------------|------------|-------------|-----------|
| Tuổi | <i>T</i> | < 21 | 12,1 |
| | | 21 – 25 | 29,0 |
| | | 26 – 30 | 31,6 |
| | | 31 – 35 | 19,1 |
| | | > 35 | 8,2 |
| Nghề nghiệp | <i>N</i> | Nhân viên | 16,0 |
| | | Công nhân | 10,1 |
| | | Nội trợ | 55,6 |
| | | Nông dân | 7,8 |
| | | Khác | 10,5 |
| Dân tộc | <i>DT</i> | Kinh | 97,5 |
| | | Khác | 2,5 |
| Nơi sống | <i>NS</i> | Nông thôn | 52,4 |
| | | Thành thị | 47,6 |
| Bệnh | <i>B</i> | Không Bệnh | 95,5 |
| | | Có bệnh | 4,5 |
| Tuổi thai | <i>TT</i> | Non thai | 6,2 |
| | | Chuẩn | 67,8 |
| | | Già thai | 25,9 |
| Giới tính trẻ | <i>GT</i> | Bé gái | 51,1 |
| | | Bé trai | 48,9 |
| Số lần có thai | <i>ST</i> | Lần 1 | 52,4 |
| | | Lần 2 | 39,8 |
| | | Hơn 2 lần | 7,8 |
| Ti trọng cơ thể | <i>BMI</i> | < 18,5 | 0,6 |
| | | 18,5 – 25,0 | 45,8 |
| | | 25,0 – 30,0 | 42,3 |
| | | 30,0 – 40,0 | 10,7 |
| | | > 40,0 | 0,6 |
| Chiều cao tử cung | <i>CC</i> | Cm | 100 |
| Vòng bụng | <i>VB</i> | Cm | 100 |

Cơ cấu mẫu cho từng nhóm đối tượng của Bảng 1 đảm bảo được các phân tích thống kê thực hiện được trong phần 3.

2.2 Các phân tích thống kê thực hiện

a. Các phân tích số liệu

Để làm rõ các nhân tố và nhóm nhân tố có ý nghĩa thống kê ảnh hưởng đến TLSS, các phương pháp phân tích thống kê đơn biến và đa biến đã được sử dụng. Ngoài thống kê mô tả, bài toán ước lượng để đánh giá số liệu ban đầu, các phân tích thống kê sau đã được sử dụng:

Kiểm định chi bình phương: Kiểm định sự khác nhau của TLSS theo từng nhóm đối tượng đối với mỗi biến (Roxy et al., 2008).

So sánh hai trung bình và phân tích phương sai đơn biến: So sánh trọng lượng trung bình trẻ sơ sinh của hai và nhiều hơn hai nhóm đối tượng (Roxy et al., 2008).

Phân tích hiệp phương sai đơn biến: Kiểm chứng sự khác nhau về véc tơ trung bình giữa hai hay nhiều hơn hai nhóm độc lập, sự tương tác của các biến đối với TLSS (Andrew, 2011).

Phân tích nhân tố: Xác định các nhân tố và nhóm nhân tố chính ảnh hưởng đến TLSS (Alvin, 2002).

b. Mô hình xác định trẻ sơ sinh đủ cân

Theo tiêu chuẩn của Bộ y tế Việt Nam, trẻ sơ sinh được chia thành 2 nhóm: trẻ đủ cân (ĐC) và không đủ cân (KĐC). Để xây dựng mô hình đánh giá trẻ ĐC, các mô hình sau được sử dụng:

i) *Mô hình hồi quy logistic* (Donald, 1997): Gọi p là xác suất để một đứa trẻ sinh ra đủ cân, $X_i, i = 1, 2, \dots, k$ là các biến có thể ảnh hưởng đến TLSS, khi đó mô hình hồi quy logistic có dạng:

$$\ln[p / (1 - p)] = \beta_0 + \sum_{i=1}^k \beta_i X_i, (i = 1, k), \tag{1}$$

trong đó $\beta_i, i=1,2,\dots,k$ là các hệ số của mô hình. Từ mô hình (1), xác suất p sẽ được xác định, khi đó nếu $p > 0.5$ trẻ được xếp vào nhóm ĐC, ngược lại trẻ bị xếp vào nhóm KĐC.

ii) *Mô hình Fisher* (Webb, 2000): Xét tập hợp các bà mẹ có biến quan sát x trong không gian n chiều, trong đó đối tượng thứ j có số liệu quan sát là véc tơ cột x_j . Gọi $\bar{x}_i, i=1,2$ lần lượt là trung bình mẫu của nhóm ĐC và KĐC; S_1 và S_2 lần lượt là ma trận hiệp phương sai của nhóm ĐC và KĐC. Ta có hàm phân biệt Fisher của hai nhóm như sau:

$$d_i(x) = \bar{x}_i^T S^{-1} x - \frac{1}{2} \bar{x}_i^T S^{-1} \bar{x}_i, i=1,2, \tag{2}$$

trong đó $S = \sum_{i=1}^k (n_i - 1) S_i / (\sum_{i=1}^k n_i - k)$ là ma trận hiệp phương sai gộp của hai nhóm.

Lúc này nếu đứa trẻ có biến quan sát x_0 sẽ được xếp vào nhóm ĐC nếu $d_1(x_0) > d_2(x_0)$ ngược lại ta xếp vào nhóm KĐC.

iii) *Mô hình Bayes* (Pham-Gia et al., 2008): Giả sử nhóm ĐC và KĐC với biến quan sát có hàm mật độ xác suất lần lượt là $f_1(x)$ và $f_2(x)$ với xác suất tiên nghiệm lần lượt là q_1 và q_2 ($q_1 + q_2 = 1$). Khi đó, một bà mẹ có biến quan sát x_0 sẽ sinh ra đứa trẻ đủ cân nếu

$$q_1 f_1(x_0) > q_2 f_2(x_0). \tag{3}$$

Sai lầm trong phân loại này được xác định bởi công thức sau:

$$P_{e,1,2}^{(q)} = 1 - \int_{R^n} \max_{1 \leq l \leq k} \{q_l f_l(x)\} dx, i = 1, 2. \tag{4}$$

trong đó n là số chiều của biến quan sát.

2.3 Vấn đề tính toán

i) Phần mềm SPSS phiên bản 22 được sử dụng để thực hiện các phân tích thống kê đơn biến và đa biến. Phần mềm này cũng được sử dụng trong xây dựng mô hình hồi quy logistic và Fisher để đánh giá TLSS đủ cân.

ii) Để áp dụng thực tế của phương pháp Bayes, từ dữ liệu rời rạc hàm mật độ xác suất phải được ước lượng. Có nhiều phương pháp tham số và phi tham số để thực hiện vấn đề này. Trong bài viết này, phương pháp hàm hạt nhân, một phương pháp được đánh giá có nhiều ưu điểm nhất hiện nay được sử dụng. Trong phương pháp này, nghiên cứu chọn hàm hạt nhân dạng chuẩn và tham số tron theo Scott (1992).

iii) Vấn đề xác định xác suất tiên nghiệm trong mô hình Bayes rất quan trọng. Trong bài viết này, tất cả các xác suất tiên nghiệm truyền thống và cả những phương pháp mới được công bố gần đây đều được sử dụng. Đó là phương pháp tiên nghiệm đều, phương pháp tỉ lệ mẫu, phương pháp Laplace và phương pháp dựa vào phân tích chùm mờ (Thao và Tai, 2016).

iv) Hiện tại, mô hình Bayes chưa có phần mềm thực hiện, do đó chương trình liên quan đã được viết ra để thực hiện. Đó là chương trình ước lượng hàm mật độ xác suất n chiều, phân loại phần tử mới và tính sai số Bayes với tích phân được tính theo phương pháp Monte-Carlo (Tai, 2016). Các

chương trình này đã hỗ trợ hiệu quả các tính toán phức tạp của phần ứng dụng.

2.4 Các bước thực hiện

Từ số liệu, các phân tích lần lượt được thực hiện, bao gồm:

i) Mã hóa số liệu, nhập số liệu vào phần mềm SPSS 22;

ii) Thực hiện các thống kê mô tả để đánh giá số liệu ban đầu về TLSS theo từng nhóm đối tượng khảo sát;

iii) Kiểm tra sự khác biệt về TLSS theo từng nhóm đối tượng, sự tương tác của các nhân tố và nhóm nhân tố ảnh hưởng đến TLSS;

iv) Tìm các nhân tố có ý nghĩa thống kê ảnh hưởng đến TLSS đủ cân và không đủ cân qua mô hình hồi quy logistic;

v) Đánh giá TLSS trẻ đủ cân qua tất cả các mô hình phân loại. Tìm phương pháp phân loại tối ưu nhất.

Từ mỗi phân tích thống kê sẽ rút ra những nhận xét và đánh giá. Trước khi tiến hành những phân tích, các điều kiện để thực hiện đều được kiểm tra. Các phân tích thống kê được thực hiện với độ tin cậy 90% hoặc 95%.

3 KẾT QUẢ THỰC HIỆN

3.1 Phân tích thống kê các vấn đề liên quan đến trọng lượng trẻ sơ sinh

a. Phân tích TLSS theo từng biến và từng nhóm đối tượng

Để xét sự ảnh hưởng của từng biến đến TLSS, nghiên cứu sử dụng phương pháp phân tích phương sai một yếu tố. Một số tham số thống kê và kết quả phân tích phương sai được tổng kết trong Bảng 2.

Bảng 2: Ước lượng và kiểm định TLSS theo từng biến và từng nhóm đối tượng

| Biến | Đối tượng | Trọng lượng TB | Sig. | Kết luận |
|-----------|--------------------|--|-------|-----------------|
| T | < 21 | (2855,36; 3083,35) | 0,001 | Có ảnh hưởng |
| | 21 – 25 | (3013,72; 3144,67) | | |
| | 26 – 30 | (3110,72; 3256,56) | | |
| | 31 – 35 | (3137,67; 3325,60) | | |
| | > 35 | (2912,86; 3215,71) | | |
| N | Nhân viên | (3114,67; 3334,11) | 0,033 | Có ảnh hưởng |
| | Công nhân | (3135,24; 3357,06) | | |
| | Nội trợ | (3047,59; 3149,95) | | |
| | Nông dân Khác | (2897,29; 3192,71) (2936,09; 3208,35) | | |
| DT | Kinh | (3096,20; 3176,20) | 0,004 | Có ảnh hưởng |
| Khác | (2549,12; 2981,65) | | | |
| NS | Nông thôn | (3047,65; 3160,16) | 0,233 | Không ảnh hưởng |
| Thành thị | (3096,27; 3207,83) | | | |
| B | Không bệnh | (3089,75; 3170,46) | 0,450 | Không ảnh hưởng |
| Có bệnh | (2840,37; 3272,67) | | | |
| TT | Non thai | (2393,67; 2721,96) | 0,000 | Có ảnh hưởng |
| Chuẩn | (3080,48; 3170,38) | | | |
| Già thai | (3193,84; 3340,75) | | | |
| GT | Bé gái | (3019,40; 3121,06) | 0,004 | Có ảnh hưởng |
| Bé trai | (3125,25; 3246,46) | | | |
| ST | Lần 1 | (3021,59; 3123,76) | 0,001 | Có ảnh hưởng |
| | Lần 2 | (3152,05; 3285,70) | | |
| | Hơn 2 lần | (2879,84; 3162,66) | | |
| BMI | < 18,5 | (2169,73; 3030,27) | 0,000 | Có ảnh hưởng |
| | 18,5 – 25,0 | (2942,67; 3046,27) | | |
| | 25,0 – 30,0 | (3135,77; 3259,17) | | |
| | 30,0 – 40,0 | (3312,69; 3560,04) | | |
| | > 40,0 | (2853,88; 3612,79) | | |

Kết quả trên cho ta thấy hầu hết các biến đều ảnh hưởng đến TLSS. Đó là các biến: tuổi, nghề, dân tộc của thai phụ, tuổi thai, giới tính trẻ, số lần có thai, tỉ trọng cơ thể, chiều cao tử cung và vòng bụng. Hai biến tiền sử bệnh (B) và nơi sống (NS) không có ảnh hưởng đến TLSS.

Đối với hai biến định lượng CC và VB, sử dụng phương pháp hồi quy tuyến tính để tìm mối quan hệ giữa từng biến này với TLSS. Kết quả cho thấy có tương quan thuận với kết quả được cho bởi Bảng 3.

Bảng 3: Mô hình hồi quy tuyến tính giữa CC và VB với TLSS

| R^2 | R | Mô hình hồi quy | Sig. |
|-------|-------|-----------------------|-------|
| 0,209 | 0,208 | $1061,327 + 67,909CC$ | 0,000 |
| 0,130 | 0,129 | $1486,898 + 17,156VB$ | 0,000 |

b. Phân tích nhân tố

Phân tích các nhân tố ảnh hưởng đến TLSS theo ma trận hệ số tương quan và sử dụng phép quay Varimax, rút ra được 5 nhóm nhân tố chính (Bảng 4).

Bảng 5: Các nhân tố ảnh hưởng đến TLSS

| Biến quan sát | Y1 | Y2 | Y3 | Y4 | Y5 |
|--|-------|--------------------|--------|---------|-------|
| Vòng bụng | 0,791 | | | | |
| Tỉ trọng cơ thể | 0,777 | | | | |
| Chiều cao tử cung | 0,595 | | | | |
| Số lần có thai | | 0,874 | | | |
| Tuổi | | 0,790 | | | |
| Tuổi thai | | | 0,779 | | |
| Giới tính trẻ sơ sinh | | | -0,617 | | |
| Nghề nghiệp | | | | 0,790 | |
| Nơi sống | | | | -0,609 | |
| Dân tộc | | | | | 0,726 |
| Tiền sử bệnh mẹ | | | | | 0,671 |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | | | | |
| Bartlett's Test of Sphericity | | | | | |
| | | Approx. Chi-Square | | 435,273 | |
| | | Df | | 55 | |
| | | Sig. | | 0,000 | |

Vì giá trị KMO = 0,559 nên ta kết luận dữ liệu đủ điều kiện để phân tích nhân tố khám phá. Mặt khác, giá trị Sig. = 0,000 nên các biến có tương quan với nhau trong tổng thể. Vì vậy, phân tích nhân tố khám phá với dữ liệu này là thích hợp.

Kết quả qua nhiều lần xoay trong phân tích nhân tố, ta được 5 nhóm nhân tố chính có ảnh hưởng đến cân nặng trẻ sơ sinh là:

Nhóm Y1: gồm biến vòng bụng, tỉ trọng cơ thể và chiều cao tử cung.

Nhóm Y2: gồm biến số lần có thai và tuổi của thai phụ.

Nhóm Y3: chứa biến tuổi thai và giới tính trẻ sơ sinh.

Nhóm Y4: gồm 2 biến nghề nghiệp và nơi sinh sống của thai phụ.

Nhóm Y5: biến dân tộc và biến tiền sử bệnh.

Mô hình thể hiện qua 5 nhân tố giải thích được 59,63% biến thiên của dữ liệu.

Qua kết quả phân tích ở Bảng 5, năm nhóm nhân tố tác động đến TLSS đượ xác định, do đó, xây dựng mô hình hồi quy với 5 biến: Y1, Y2, Y3, Y4 và Y5.

Bảng 4: Các thành phần chính được giữ lại

| Nhân tố | Giá trị riêng | Tỉ lệ giải thích được (%) | Tỉ lệ tích lũy (%) |
|---------|---------------|---------------------------|--------------------|
| 1 | 1,74 | 15,78 | 15,78 |
| 2 | 1,50 | 13,61 | 29,39 |
| 3 | 1,16 | 10,56 | 39,96 |
| 4 | 1,13 | 10,27 | 50,22 |
| 5 | 1,04 | 9,41 | 59,63 |

Năm nhân tố có ảnh hưởng cụ thể lên các biến được cho bởi Bảng 5:

Vì ba nhóm nhân tố Y1, Y3 và Y5 có giá trị Sig. = 0,000 nên chúng có ý nghĩa thống kê trong mô hình đang xét. Như vậy, một lần nữa cho thấy các biến VB, BMI, CC, TT, GT, DT và B có ảnh hưởng đến TLSS. Ngoài ra, nhìn vào các chỉ số phương sai VIF tương ứng với các nhóm nhân tố độc lập đều nhỏ hơn 10 (tất cả đều bằng 1). Điều này cho thấy các nhóm nhân tố này không xảy ra hiện tượng đa cộng tuyến.

Bảng 6: Hệ số hồi quy trong mô hình

| | B | Sig. | VIF |
|----------|----------|--------------|-------|
| Constant | 3126,803 | 0,000 | |
| Y1 | 234,992 | 0,000 | 1,000 |
| Y2 | 18,452 | 0,258 | 1,000 |
| Y3 | 115,316 | 0,000 | 1,000 |
| Y4 | -31,713 | 0,052 | 1,000 |
| Y5 | -61,302 | 0,000 | 1,000 |

Lập được phương trình hồi quy tuyến tính bộ như sau:

$$Y = 3126,803 + 234,992Y1 + 115,316Y3 - 61,302Y5.$$

Dấu của các hệ số trước các nhóm nhân tố Y1 và Y3 đều dương nên tương quan giữa chúng với biến TLSS là tương quan thuận. Trong điều kiện các

yếu tố khác không đổi, khi nhân tố Y1 tăng/giảm 1 đơn vị thì TLSS sẽ tăng/giảm 234,992 đơn vị.

Trong mô hình hồi quy, dấu hệ số trước nhóm nhân tố Y5 âm nên tương quan giữa TLSS và nhóm nhân tố Y5 là tương quan nghịch. Cụ thể, trong điều kiện các yếu tố khác không thay đổi, khi nhân tố Y5 tăng một đơn vị thì TLSS sẽ giảm 61,302 đơn vị.

c. *Kết quả phân tích phương sai đơn biến nhiều nhân tố*

Kiểm tra sự ảnh hưởng đến TLSS của từng biến và sự tương tác của các biến bằng phương pháp phân tích phương sai đơn biến nhiều nhân tố theo từng nhóm biến ta lần lượt có các kết quả sau:

Bảng 7: Các nhân tố nhóm 1, nhóm 2 và nhóm 3 ảnh hưởng đến TLSS

| Nhóm 1 | Sig. | Nhóm 2 | Sig. | Nhóm 3 | Sig. |
|------------|--------------|---------------|--------------|--------------|--------------|
| T | 0,623 | CC | 0,000 | ST | 0,038 |
| N | 0,142 | VB | 0,001 | TT | 0,000 |
| DT | 0,033 | B | 0,463 | GT | 0,071 |
| NS | 0,634 | BMI | 0,005 | ST * TT | 0,457 |
| T * N | 0,819 | CC * VB | 0,080 | ST * GT | 0,286 |
| T * DT | 0,917 | CC * B | 0,627 | TT * GT | 0,645 |
| T * NS | 0,550 | CC * BMI | 0,085 | ST * TT * GT | 0,173 |
| N * DT | 0,373 | VB * B | 0,119 | | |
| N * NS | 0,546 | VB * BMI | 0,004 | | |
| DT * NS | 0,906 | B * BMI | 0,157 | | |
| T * N * NS | 0,439 | CC * VB * BMI | 0,394 | | |

Từ Bảng 7 cho thấy, các biến DT, CC, VB, BMI, ST, TT và sự tương tác của VB*BMI có ảnh hưởng đến TLSS. Các trường hợp còn lại không ảnh hưởng đến TLSS.

d. *Kết quả phân tích hiệp phương sai*

Ngoài yếu tố nghề, tuổi, dân tộc, tỉ trọng cơ thể, vòng bụng có thể ảnh hưởng đến cân nặng sơ sinh thì tuổi thai cũng là một biến ảnh hưởng lớn đến TLSS. Tuy nhiên, ở những ngày tuổi thai khác nhau thì tốc độ phát triển và khả năng tăng cân nặng cũng khác nhau. Để đảm bảo tính khách quan của nghiên cứu, chúng ta sử dụng phương pháp phân tích hiệp phương sai (ANCOVA) để kiểm soát ảnh hưởng của yếu tố ngoại sinh đến kết quả nghiên cứu theo từng nhóm biến.

Bảng 8 cho thấy sự tương tác của các nhân tố N*TT và N*ST*TT có ảnh hưởng tương tác với

nhau. Các nhân tố NS*TT, T*TT, NS*T*TT và ST*TT không có ảnh hưởng tương tác nhau đến TLSS.

Bảng 8: Kết quả trong thực hiện ANCOVA

| Nhóm 1 | Sig. | Nhóm 2 | Sig. |
|-------------|--------------|-------------|--------------|
| NS | 0,144 | N | 0,021 |
| TT | 0,000 | ST | 0,185 |
| T | 0,452 | TT | 0,000 |
| NS * TT | 0,146 | N * ST | 0,015 |
| NS * T | 0,859 | N * TT | 0,019 |
| T * TT | 0,519 | ST * TT | 0,207 |
| NS * T * TT | 0,856 | N * ST * TT | 0,014 |

3.2 Đánh giá trọng lượng trẻ sơ sinh đủ cân qua mô hình bài toán phân loại

Từ số liệu, tính hệ số tương quan giữa các biến ta có kết quả:

Bảng 9: Hệ số tương quan cặp của biến đưa vào mô hình

| | T | N | DT | NS | B | ST | TT | CC | VB | GT | BMI |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|------|------|-----|
| T | 1 | | | | | | | | | | |
| N | -0,06 | 1 | | | | | | | | | |
| DT | -0,05 | -0,01 | 1 | | | | | | | | |
| NS | 0,12 | -0,01 | -0,03 | 1 | | | | | | | |
| B | 0,05 | 0,04 | 0,03 | -0,04 | 1 | | | | | | |
| ST | 0,47 | 0,12 | 0,06 | 0,07 | -0,04 | 1 | | | | | |
| TT | 0,04 | -0,01 | 0,01 | -0,04 | -0,01 | -0,02 | 1 | | | | |
| CC | 0,02 | -0,01 | -0,03 | -0,02 | -0,01 | -0,03 | 0,09 | 1 | | | |
| VB | 0,18 | -0,05 | -0,06 | 0,09 | -0,01 | 0,13 | 0,06 | 0,05 | 1 | | |
| GT | 0,08 | -0,03 | -0,03 | -0,02 | 0,01 | 0,04 | -0,11 | -0,04 | 0,02 | 1 | |
| BMI | 0,11 | -0,01 | -0,08 | 0,09 | -0,01 | 0,05 | 0,02 | 0,01 | 0,46 | 0,04 | 1 |

Bảng 9 cho thấy các biến không phụ thuộc tuyến tính với nhau hay có 11 biến độc lập. Vì vậy, có thể đưa các biến này vào mô hình để giải thích

cho tình trạng trẻ sơ sinh đủ cân.

Tiến hành phân tích hồi quy logistic, ta có kết quả xử lý bởi bảng tổng hợp sau:

Bảng 10: Kết quả phân tích hồi quy logistic cho 11 biến

| | B | S.E. | Wald | Df | Sig. | Exp(B) |
|----------|---------|-------|--------|----|--------------|--------|
| T | 0,010 | 0,164 | 0,004 | 1 | 0,952 | 1,010 |
| N | -0,151 | 0,163 | 0,864 | 1 | 0,353 | 0,860 |
| DT | -1,335 | 0,813 | 2,700 | 1 | 0,100 | 0,263 |
| NS | 0,201 | 0,338 | 0,355 | 1 | 0,551 | 1,223 |
| B | -0,774 | 0,633 | 1,497 | 1 | 0,221 | 0,461 |
| ST | 0,249 | 0,294 | 0,720 | 1 | 0,396 | 1,283 |
| TT | 0,086 | 0,015 | 31,337 | 1 | 0,000 | 1,090 |
| CC | 0,213 | 0,056 | 14,654 | 1 | 0,000 | 1,237 |
| VB | 0,018 | 0,023 | 0,634 | 1 | 0,426 | 1,019 |
| GT | 0,866 | 0,349 | 6,157 | 1 | 0,013 | 2,378 |
| BMI | 0,207 | 0,071 | 8,625 | 1 | 0,003 | 1,231 |
| Constant | -33,183 | 4,936 | 45,196 | 1 | 0,000 | 0,000 |

Với mức ý nghĩa chung 10% cho thấy có năm biến DT, TT, CC, GT và BMI có ý nghĩa thống kê, các biến còn lại không có ý nghĩa thống kê. Điều này có nghĩa là các biến này đóng vai trò quan

trọng đối với khả năng thai phụ sinh con đủ cân được khảo sát. Chính vì vậy, bài toán phân loại với năm biến này và lần lượt gọi chúng là X1, X2, X3, X4 và X5 được thực hiện.

Bảng 11: Bảng tổng hợp khả năng phân loại đúng (%) của ba phương pháp

| Biến | Khả năng phân loại đúng (%) | | | | | |
|--------------------|-----------------------------|--------|--------|-------------|--------|-------------|
| | Logistic | Fisher | BayesU | BayesP | BayesL | BayesC |
| X1 | 88,3 | 87,7 | 81,2 | 94,7 | 94,6 | 95,8 |
| X2 | 89,5 | 75,8 | 84,0 | 94,7 | 94,6 | 95,2 |
| X3 | 87,1 | 69,4 | 84,7 | 94,6 | 94,5 | 95,6 |
| X4 | 88,3 | 51,7 | 83,3 | 95,4 | 95,4 | 96,4 |
| X5 | 88,7 | 61,8 | 82,9 | 94,4 | 94,4 | 95,9 |
| X1, X2 | 89,3 | 75,6 | 93,7 | 97,7 | 97,8 | 94,5 |
| X1, X3 | 87,7 | 76,8 | 93,8 | 97,6 | 97,6 | 94,8 |
| X1, X4 | 88,3 | 51,5 | 93,4 | 98,0 | 97,9 | 97,4 |
| X1, X5 | 89,1 | 70,4 | 92,7 | 97,3 | 97,3 | 97,9 |
| X2, X3 | 88,7 | 78,8 | 94,5 | 97,8 | 97,8 | 97,4 |
| X2, X4 | 89,1 | 76,6 | 94,3 | 98,1 | 98,0 | 98,4 |
| X2, X5 | 90,6 | 77,8 | 93,3 | 97,5 | 97,5 | 96,3 |
| X3, X4 | 87,1 | 66,5 | 94,6 | 97,9 | 97,9 | 98,5 |
| X3, X5 | 88,7 | 71,3 | 93,7 | 97,4 | 97,4 | 97,4 |
| X4, X5 | 88,7 | 65,5 | 93,8 | 97,7 | 97,6 | 98,3 |
| X1, X2, X3 | 89,1 | 80,7 | 98,2 | 98,7 | 98,7 | 98,5 |
| X1, X2, X4 | 88,9 | 78,8 | 98,9 | 99,3 | 99,3 | 99,5 |
| X1, X2, X5 | 90,6 | 80,3 | 97,3 | 98,8 | 98,8 | 98,3 |
| X1, X3, X4 | 87,7 | 72,5 | 96,2 | 98,8 | 98,8 | 98,6 |
| X1, X3, X5 | 88,7 | 74,5 | 92,9 | 98,2 | 98,1 | 92,1 |
| X1, X4, X5 | 88,7 | 70,6 | 95,9 | 98,9 | 98,9 | 98,0 |
| X2, X3, X4 | 88,3 | 81,1 | 98,6 | 98,9 | 98,9 | 99,5 |
| X2, X3, X5 | 90,3 | 78,9 | 96,6 | 98,2 | 98,2 | 98,0 |
| X2, X4, X5 | 90,8 | 78,0 | 98,0 | 98,9 | 98,9 | 99,3 |
| X3, X4, X5 | 87,9 | 72,1 | 93,8 | 98,5 | 98,5 | 98,1 |
| X1, X2, X3, X4 | 88,9 | 81,5 | 99,4 | 99,6 | 99,6 | 99,1 |
| X1, X2, X3, X5 | 90,6 | 80,5 | 98,5 | 99,2 | 99,2 | 98,7 |
| X1, X2, X4, X5 | 91,0 | 79,9 | 99,1 | 99,6 | 99,6 | 99,6 |
| X1, X3, X4, X5 | 97,9 | 75,0 | 97,5 | 99,3 | 99,3 | 98,8 |
| X2, X3, X4, X5 | 90,3 | 81,3 | 98,8 | 99,4 | 99,4 | 99,6 |
| X1, X2, X3, X4, X5 | 90,8 | 81,5 | 99,4 | 99,8 | 99,8 | 99,4 |

Đối với phân loại bằng phương pháp Bayes khi biết hàm mật độ xác suất, bài viết phân loại dựa vào các xác suất tiên nghiệm khác nhau. Cụ thể, xác suất tiên nghiệm được xác định dựa trên phân phối đều, tỉ lệ mẫu, ước lượng Laplace, thuật toán phân tích chùm mờ lần lượt được gọi là BayesU, BayesP, BayesL và BayesC.

Thực hiện phương pháp Fisher, logistic và Bayes để tìm mô hình có xác suất phân loại tốt nhất, ta được Bảng 11.

Bảng 11 cho thấy phân loại bằng phương pháp Bayes với xác suất tiên nghiệm dựa vào thuật toán phân tích chùm mờ cho kết quả phân loại tốt nhất, khả năng phân loại đúng bằng phương pháp Fisher là thấp nhất. Tuy nhiên, trong trường hợp năm biến thì khả năng phân loại đúng của Bayes P là cao nhất.

Nhìn chung, trường hợp năm biến cho kết quả phân loại đúng cao nhất là 99,8%, nghĩa là năm biến này ảnh hưởng lớn nhất đến TLSS. Khi ứng dụng vào thực tế, để có được kết quả phân loại tốt nhất chúng ta nên sử dụng mô hình phân loại với năm biến DT, TT, CC, GT và BMI. Điều này có nghĩa là thai phụ muốn sinh trẻ có trọng lượng sơ sinh đủ cân thì không những phụ thuộc yếu tố khách quan tuổi thai, chiều cao tử cung, giới tính trẻ, mà còn phụ thuộc nhiều vào kiến thức hiểu biết và sức khỏe của thai phụ thông qua biến dân tộc và biến tỉ trọng cơ thể. Trong nghiên cứu này, các thai phụ thuộc nhóm dân tộc thiểu số có tỉ lệ TLSS đủ cân thấp. Nguyên nhân của vấn đề này có thể do hoàn cảnh kinh tế, ý thức và kiến thức về sức khỏe sinh sản còn hạn chế của nhóm đối tượng này. Việc cung cấp thông tin, hỗ trợ về giáo dục sức khỏe sinh sản tới những thai phụ này là điều quan trọng.

4 KẾT LUẬN

Sử dụng các phương pháp thống kê khác nhau, dựa trên số liệu thực tế, bài viết đã xác định các nhân tố, sự tương tác của các nhóm nhân tố ảnh hưởng đến TLSS. Những nhân tố chính và mô hình hồi quy cho TLSS từ các nhân tố này đã được rút ra. Thông qua các mô hình phân loại, bài

viết cũng rút ra mô hình tối ưu trong đánh giá trọng lượng trẻ sơ sinh đủ cân từ các thông tin có được của sản phụ. Để có những kết quả đầy đủ và đáng tin cậy hơn, nghiên cứu phải bổ sung số liệu, phạm vi chọn mẫu và phải lấy mẫu nhiều lần. Tuy nhiên, kết quả nghiên cứu ban đầu này là thông tin rất hữu ích cho ngành y tế, cho các sản phụ và những người chăm sóc họ. Các phương pháp thực hiện cho vấn đề này có thể ứng dụng tương tự cho nhiều vấn đề thực tế khác.

LỜI CẢM ƠN

Nhóm tác giả xin gửi lời cảm ơn đến Bệnh viện Phụ sản Thành phố Cần Thơ đã hỗ trợ xác định biến và số liệu trong nghiên cứu này. Cảm ơn các phản biện đã cho những đóng góp có giá trị để bài viết được tốt hơn.

TÀI LIỆU THAM KHẢO

- Andrew, R., 2011. *Introducing ANOVA and ANCOVA*. SAGE. London, 192 pages.
- Alvin, C. R., 2002. *Methods of Multivariate Analysis*. John Wiley & Sons. New York, 727 pages.
- Donald, C., 1997. *Log-linear Models and logistic regression*. Springer. New York, 507 pages.
- Pham-Gia, T., Turkkan, N. and Tai, Vovan., 2008. The maximum function in statistical discrimination analysis. *Commun. in Stat-Simulation computation*. 37(2): 320-336.
- Roxy, P., Chris, O., Jay, D., 2008. *Statistics and data analysis*. Thomson. New York, 619 pages.
- Scott, D. W., 1992. *Multivariate density estimation: Theory, practice and visualization*. Wiley & Son, New York, 345 pages.
- Tai, V.V., 2016. L1-distance and classification problem by Bayesian. *J. Appl. Stat* (online first: <http://dx.doi.org/10.1080/02664763.2016.1174194>).
- Thao, N.T., Tai, V.V., 2016. A new approach for determining the prior probabilities in the classification problem by Bayesian method, *Adv Data Anal Classif* (online first: <http://link.springer.com/article/10.1007/s11634-016-0253>).
- Webb, A., 2000. *Statistical pattern recognition*. Wiley & Sons, New York, 645 pages.