**Nền tảng Phoebe và Kết Quả Thử Nghiệm Ước Lượng Tuổi và Cân Nặng Thai**

Nguyễn Phước Lộc

Hội Nội Tiết Sinh Sản và Vô Sinh TPHCM, Việt Nam

Hồ Thị Thu Hằng

Bệnh Viện Đa Khoa Vĩnh Long, Việt Nam

Phan Trường Duyệt

Đại học Y Hà Nội, Việt Nam

**Tóm tắt**

Ước lượng tuổi và cân nặng thai có vai trò quan trọng trong điều trị sản khoa. Hiện có nhiều công thức ước lượng từ sự kết hợp giữa thống kê và ngành sản khoa nhưng mỗi công thức này chỉ tối ưu khi chúng được áp dụng trong một cộng đồng cụ thể. Nghiên cứu này giới thiệu một nền tảng Phoebe (hệ thống hỗ trợ nghiên cứu Phoebe) nhằm hỗ trợ các bác sĩ lâm sàng và nhà khoa học tìm ra những công thức chính xác nhất đối với cộng đồng nơi họ thực hiện nghiên cứu. Thuật toán lõi của Phoebe dùng mô hình thống kê hồi quy để ước lượng tuổi thai và cân nặng thai từ các số đo siêu âm như đường kính lưỡng đỉnh, chu vi đầu, chu vi bụng, chiều dài xương đùi, thể tích cánh tay và thể tích đùi. Thuật toán này dựa trên những giả định thông minh nhằm tìm ra công thức tối ưu nhanh nhất có thể. Kết quả thử nghiệm cho thấy Phoebe kết xuất được những công thức ước lượng có độ chính xác cao và thích hợp nhất với cộng đồng dân cư cụ thể nơi nghiên cứu được thực hiện. Hơn nữa, Phoebe còn cung cấp công cụ tiện ích cho các bác sĩ lâm sàng và nhà nghiên cứu để khai thác thông tin thống kê hữu ích từ dữ liệu thai kỳ. Nền tảng Phoebe là một phần mềm máy tính lưu trữ tại trang web http://phoebe.locnguyen.net.

**Từ khóa**: ước lượng tuổi thai, ước lượng cân nặng thai, số đo siêu âm, mô hình hồi quy, công thức ước lượng.

**1. Giới thiệu**

Ước lượng tuổi thai và cân nặng thai trước sinh rất cần thiết cho bác sĩ chẩn đoán bệnh lý hoặc các trường hợp bất thường nhằm có điều trị thích hợp. Để thuận tiện, thuật ngữ “ước lượng trước sinh” chỉ ước lượng tuổi thai và ước lượng cân nặng thai. Có hai phương pháp ước lượng trước sinh:

* Đo thể tích thai nhi bên trong tử cung mẹ để tính cân nặng thai dựa trên thể tích này và mật độ xương thịt. Tuổi và cân nặng thai còn có thể được ước lượng dựa trên kích thước tử cung mẹ.
* Áp dụng mô hình thống kê hồi quy: Các số đo siêu âm thai nhi như đường kính lưỡng đỉnh (*bpd*), chu vi đầu (*hc*), chu vi bụng (*ac*), chiều dài xương đùi (*fl*), thể tích cánh tay (*arm\_vol*) và thể tích đùi (*thigh\_vol*) trở thành dữ liệu đầu vào cho phân tích hồi quy, sẽ cho ra kết quả là *hàm hồi quy* tức chính là công thức để ước lượng tuổi và cân nặng thai theo các số đo siêu âm *bpd*, *hc*, *ac*, *fl, arm\_vol*, và *thigh\_vol*. Dữ liệu của các số đo siêu âm trên được gọi là mẫu thai kỳ. Hai thuật ngữ “*mẫu*” và “*dữ liệu*” có cùng ý nghĩa trong nghiên cứu này. Lưu ý mẫu là đại diện của quần thể dân cư nơi nghiên cứu thực hiện.

Phương pháp thứ hai thể hiện đặc trưng của quần thể dân cư từ dữ liệu thống kê nên chúng tôi áp dụng mô hình hồi quy cho việc ước lượng trước sinh. Các thuật ngữ sau đều có cùng ý nghĩa: *hàm*, *hàm hồi quy*, *hàm ước lượng*, *mô hình hồi quy*, *mô hình ước lượng*, *công thức*, *công thức hồi quy*, *công thức ước lượng*.

Hiện có nhiều công thức ước lượng từ các nghiên cứu (Hadlock, Harrist, Sharman, Deter, & Park, 1985), (Phan, 1985), (Phạm, 2000), (Ho T. T., Nghiên Cứu Phương Pháp Ước Lượng Trọng Lượng Thai, Tuổi Thai Bằng Siêu Âm Hai và Ba Chiều, 2011), (Shepard, Richards, Berkowitz, Warsof, & Hobbins, 1982), (Campbell & Wilkin, 1975), (Lee, et al., 2009), (Chang, et al., 1997), and (Varol, Saltik, Kaplan, Kilic, & Yardim, 2001), trong đó một số đạt độ chính xác cao nhưng chúng chỉ thích hợp cho những nhóm dân cư hay cộng động nơi nghiên cứu được thực hiện. Nếu các công thức này được áp dụng vào cộng đồng khác như ở Việt Nam, chúng không còn chính xác nữa. Hơn nữa, việc tìm ra công thức mới rất khó khăn và tốn nhiều chi phí về nguồn lực và thời gian.

Do đó *mục tiêu thứ nhất* của nghiên cứu là đề xuất một thuật toán hiệu quả để tìm ra những công thức chính xác phù hợp nhất với cộng đồng dân cư nơi nghiên cứu thực hiện. Tiến trình kết xuất công thức tối ưu này càng nhanh càng tốt. Ngoài ra, các bác sĩ lâm sàng và nhà nghiên cứu luôn có nhu cầu khai thác thông tin thống kê hữu ích từ dữ liệu siêu âm và từ mô hình hồi quy. Vì vậy *mục tiêu thứ hai* là cung cấp một ***nền tảng Phoebe*** (hệ thống Phoebe) cho các bác sĩ lâm sàng và nhà nghiên cứu như là một tiện ích hỗ trợ. Nền tảng Phoebe cài đặt thuật toán trong mục tiêu thứ nhất và hơn nữa cung cấp công cụ để khai thác thông tin thống kê hữu ích từ dữ liệu thai kỳ. Công cụ này là một phần mềm máy tính được lập trình. Hơn nữa, nền tảng Phoebe còn cho phép nhà phát triển phần mềm thay đổi chức năng của nó. Ví dụ, nhà phát triển phần mềm có thể cải tiến thuật toán tìm công thức tối ưu bằng cách thêm những ràng buộc thông minh.

Phần 2 là tổng quan kiến trúc của Phoebe. Phần 3 mô tả thuật toán đề xuất là lõi của Phoebe. Thuật toán này sẽ kết xuất công thức ước lượng sao cho tối ưu và phù hợp nhất với một cộng đồng dân cư cụ thể như Vietnam. Phần 4 thảo luận những trường hợp sử dụng Phoebe. Kết quả thử nghiệm được mô tả trong phần 5 với những công thức ước lượng tối ưu. Phần 6 là đề xuất ước lượng sớm cân năng thai. Phần 7 là kết luận. Chúng tôi đã sử dụng gói phần mềm thống kê của tác giả Michael Thomas Flanagan (Flanagan, 2004) và gói phần mềm xử lý biểu thức toán học của tác giả Jos de Jong (Jong, 2010) để phát triển Phoebe, trong đó gói phần mềm thống kê là quan trọng nhất. Phoebe được xây dựng bằng ngôn ngữ lập trình Java (Oracle, n.d.).

**2. Kiến trúc tổng quan của Phoebe**

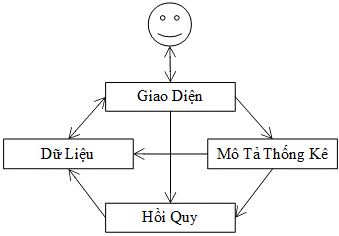
Từ số đo siêu âm như *bpd*, *hc*, *ac*, và *fl*, Phoebe sẽ kết xuất những công thức tối ưu để ước lượng tuổi và cân nặng thai với độ chính xác cao nhất. Hơn nữa thông tin thống kê về dữ liệu thai nhi và thai kỳ được mô tả bằng hai định dạng: số liệu và đồ thị. Nền tảng Phoebe được cấu thành từ bốn thành phần (component) sau:

* Thành phần *Dữ Liệu* (Dataset) quản lý thông tin về các số đo siêu âm thai nhi như *bpd*, *hc*, *ac*, *fl* và thông tin thai kỳ phụ trợ theo cách thức hợp lý và thông minh. Thành phần này cho phép các thành phần khác truy cập những thông tin này. Thông tin thai nhi – thai kỳ được tổ chức theo cấu trúc trừu tượng; ví dụ, cấu trúc ma trận với mỗi dòng là một mẫu của các số đo *bpd*, *hc*, *ac*, và *fl*. Bảng 1 là ví dụ của một cấu trúc trừu tượng.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *bpd* | *hc* |  | *fl* | *ac* | *tuổi thai*  (tuần) | *cân nặng thai*  (gram) |
| 74 | 262 |  | 51 | 255 | 28 | 900 |
| 72 | 260 |  | 51 | 232 | 28 | 900 |
| 68 | 260 |  | 50 | 229 | 28 | 900 |
| 72 | 275 |  | 52 | 240 | 28 | 900 |
| 72 | 274 |  | 52 | 240 | 28 | 950 |
| 74 | 253 |  | 50 | 235 | 28 | 950 |
| 71 | 257 |  | 52 | 239 | 28 | 950 |
| 71 | 255 |  | 53 | 236 | 28 | 950 |
| 70 | 264 |  | 52 | 246 | 28 | 950 |

**Bảng 1**. An example of gestational sample matrix

* Thành phần *Hồi Quy* (Regression) mô hình hóa công thức ước lượng hoặc hàm hồi quy. Thành phần này đọc các số đo siêu âm từ thành phần *Dữ Liệu* và xây dựng công thức ước lượng tối ưu từ các số đo này. Phần 3 sẽ mô tả thuật toán lõi để tìm ra công thức tối ưu. Thành phần này quan trọng nhất vì thi công thuật toán lõi này.
* Thành phần *Mô Tả Thống Kê* (Statistical manifest) sẽ mô tả thông tin thống kê của các số đo siêu âm lẫn hàm hồi quy, ví dụ: trung bình và độ lệch chuẩn của mẫu *bpd*, tổng lỗi (sum of residuals), hệ số tương quan của hàm hồi quy, đồ thị bách phân vị của tuổi và cân nặng thai. Bản mô tả thống kê được tổ chức theo dạng số và dạng đồ thị.
* Thành phần *Giao Diện* (User Interface) hỗ trợ người dùng (bác sĩ lâm sàng, nhà nghiên cứu) tương tác với Phoebe. Một trường hợp sử dụng phổ biến là người dùng nhập các số đo siêu âm và yêu cầu hệ thống kết xuất công thức ước lượng tối ưu và kết xuất thông tin thống kê của các số đo siêu âm; hơn nữa người dùng còn có thể truy xuất thông tin khác được quản lý bởi thành phần Dữ Liệu. Thành phần Giao Diện kết nối tất cả các thành phần khác với nhau để cung cấp những tiện ích cho người dùng nhiều nhất có thể.



**Figure 1**. Kiến trúc tổng quan của Phoebe

Ba thành phần: *Dữ Liệu*, *Hồi Quy* và *Mô Tả Thống Kê* là những thành phần cơ bản. Thành phần thứ tư *Giao Diện* là cầu nối giữa ba thành phần cơ bản đó. Hình 1 mô tả kiến trúc tổng quan của Phoebe.

**3. Thuật toán lõi của Phoebe**

Nền tảng Phoebe dùng mô hình hồi quy để ước lượng. Giả sử hàm hồi quy tuyến tính *Y* = *α*0 + *α*1*X*1 + *α*2*X*2 + … + *αnXn* với *Y* là tuổi hoặc cân nặng thai trong khi đó *Xi* là các số đo siêu âm như *bpd*, *hc*, *ac*, and *fl*. Biến *Y* được gọi là *biến đáp ứng*, *biến tiên lượng* hay *biến phụ thuộc*. Mỗi biến *Xi* được gọi là *biến hồi quy* (regressor) hay *biến độc lập*. Mỗi *αi* được gọi là *hệ số hồi quy*. Với một tập các giá trị của những *Xi*, giá trị ước lượng của *Y* (ký hiệu *Y-estimated*) được tính từ hàm hồi quy sẽ được so sánh với giá trị thực của *Y* (ký hiệu *Y-real*) có trong mẫu dữ liệu. Trong nghiên cứu này, *Y* ngầm định *Y-estimated* nếu không có thêm giải thích. Độ lệch giữa *Y-estimated* và *Y-real* là tiêu chí đánh giá chất lượng hoặc độ chính xác của hàm hồi quy. Độ lệch này cũng được gọi là *lỗi ước lượng*. Độ lệch càng nhỏ, hàm hồi quy càng tốt. Mục đích của nghiên cứu này là tìm ra hàm hồi quy tối ưu (công thức ước lượng) với độ chính xác cao nhất.

Một hàm hồi quy tốt sẽ thỏa hai điều kiện sau:

* Sự tương quan giữa *Y-estimated* và *Y-real* thì lớn.
* Tổng lỗi, còn được gọi là tổng những phần dư (sum of residuals) thì nhỏ. Phần dư (residual) là bình phương độ lệch giữa *Y-estimated* và *Y-real*.

Hai điều kiện này được gọi là *cặp điều kiện tối ưu*. Hàm hồi quy tốt nhất sẽ thỏa cặp điều kiện tối ưu nhiều nhất, nghĩa là sự tương quan giữa *Y-estimated* và *Y-real* lớn nhất và tổng lỗi nhỏ nhất. Với tập biến hồi quy *Xi* (), một hàm hồi quy là một tổ hợp của *k* biến *Xi* với *k n* sao cho tổ hợp này thỏa cặp điều kiện tối ưu ưu. Giả sử với tập biến hồi quy *VAR =* {*X*1, *X*2,…, *Xn*}là các số đo siêu âm, thuật toán vét cạn sẽ tìm hàm hồi tối ưu theo ba bước sau:

1. Đặt *k* là chỉ số được khởi tạo bằng 1, tương ứng với tổ hợp gồm *k* biến hồi quy.
2. Tất cả tổ hợp chập *k* của *n* biến hồi quy được tạo ra. Với mỗi tổ hợp chập *k*, một hàm gồm *k* biến từ tổ hợp này được xây dựng. Hàm nào thỏa cặp điều kiện tối ưu nhiều nhất sẽ hiện là hàm tối ưu.
3. Chỉ số *k* được tăng lên 1. Nếu *k = n* thì thuật toán dừng; ngược lại, quay về bước 2.

Số tổ hợp mà thuật toán vét cạn duyệt qua là:

Với *n* là số biến hồi quy và “*k*!” là ký hiệu của *k* giai thừa. Khi *n* đủ lớn, sẽ có sự bùng nổ tổ hợp khiến thuật toán không dừng và hậu quả là không thể tìm ra hàm tốt nhất. Hơ nữa, có rất nhiều dạng hàm hồi quy như tuyến tính, bậc hai, bậc ba, lô-ga-rít, mũ và tích. Do đó chúng tôi đề xuất một thuật toán mới nhằm khắc phục nhược điểm này để luôn tìm ra hàm tối ưu. Điều kiện dừng của thuật toán đề xuất luôn xác định và thời gian tìm kiếm giảm đáng kể vì không gian tìm kiếm được thu hẹp nhỏ nhất có thể. Thuật toán đề xuất được gọi là thuật toán *Hạt Giống Nảy Mầm* (Seed Germination), viết tắt là ***SG***. SG là thuật toán nội tại (built-in), chính là phần lõi của Phoebe. SG là thuật toán thông minh (heuristic) dựa trên *cặp giả định thông minh* về hàm hồi quy như sau (Nguyen & Ho, A framework of fetal age and weight estimation, 2014, p. 22):

* Giả định thứ nhất: các biến hồi quy *Xi* có xu hướng độc lập lẫn nhau, nghĩa là mỗi cặp *Xi* và *Xj* với *i ≠ j* trong cùng một hàm tối ưu sẽ độc lập lẫn nhau. Điều kiện mở rộng của sự độc lập là “*hệ số tương quan giữa mỗi cặp Xi và Xj nhỏ hơn một ngưỡng δ*”. Đây là *giả định* *cực tiểu*.
* Giả định thứ hai: mỗi biến hồi quy *Xi* đều đóng góp vào chất lượng (độ chính xác) của hàm tối ưu và tỉ lệ đóng góp ấy được định nghĩa là hệ số tương quan giữa *Xi* và giá trị thực của biến đáp ứng *Y*. Tỉ lệ đóng góp càng cao, biến *Xi* tương ứng càng quan trọng. Những biến có tỉ lệ đóng góp cao được gọi là *biến cổ đông*. Vì thế một hàm tối ưu sẽ có nhiều biến cổ đông. Giả định thứ hai được phát biểu rằng “*hệ số tương quan giữa mỗi Xi và giá trị thực của Y lớn hơn một ngưỡng ε*”. Đây là *giả định* *cực đại*.

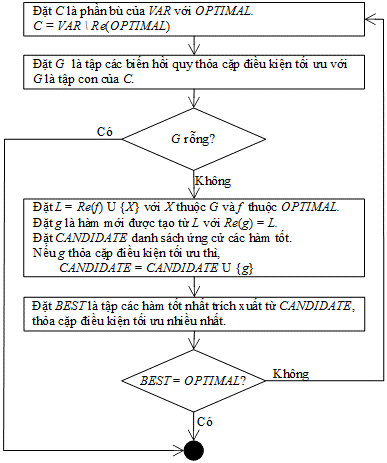
Thuật toán SG lặp và thử nhiều tổ hợp các biến *Xi* sao cho mỗi tổ hợp thỏa cặp giả định thông minh. Nói cách khác, kỳ vọng rằng mỗi tổ hợp có thể tạo thành một hàm hồi quy tối ưu thỏa *cặp điều kiện thông minh*, như sau (Nguyen & Ho, A framework of fetal age and weight estimation, 2014, p. 22):

* Hệ số tương quan giữa mỗi cặp *Xi* và *Xj* nhỏ hơn ngưỡng *δ* > 0. Đây là *điều kiện cực tiểu* hay *điều kiện độc lập*, tương ứng với giả định cực tiểu.
* Hệ số tương quan giữa mỗi cặp *Xi* và *Y* lớn hơn ngưỡng *ε* > 0. Đây là *điều kiện cực đại* hay *điều kiện đóng góp*, tương ứng với giả định cực đại.

Đặt *VAR =* {*X*1, *X*2,…, *Xn*} là tập tất cả các biến số đo siêu âm. Đặt *f* = *α*0 *+ α*1*X*1 *+ α*2*X*2 *+ … + αkXk* (*k* *n*) là hàm ước lượng và đặt *Re*(*f*) = {*X*1, *X*2,…, *Xk*} là tập các biến của *f*. Giá trị của *f* là tuổi hoặc cân nặng thai và có thể xem *Re*(*f*) đại diện cho chính *f*. Đặt *OPTIMAL* là kết quả cuối cùng của thuật toán SG gồm tất cả các hàm tối ưu. *OPTIMAL* được khởi tạo là tập rỗng. Đặt *Re*(*OPTIMAL*) là tập các biến hồi quy có trong tất cả các hàm tối ưu *f OPTIMAL*. Thuật toán SG gồm 4 bước sau (Nguyen & Ho, A framework of fetal age and weight estimation, 2014, p. 22):

1. Đặt *C* là phần bù của *VAR* với *OPTIMAL*, ta có *C* = *VAR* \ *Re*(*OPTIMAL*). Dấu vạch chéo ngược “\” ký hiệu toán tử bù trong lý thuyết tập hợp. Điều này có nghĩa là *C* thuộc *VAR* nhưng không thuộc *Re*(*OPTIMAL*).
2. Đặt *G* *C* là tập các biến hồi quy thỏa cặp điều kiện tối ưu. *G* là tập con của *C*. Nếu *G* rỗng, thuật toán dừng; ngược lại tiếp bước 3.
3. Duyệt qua *G* để xác định danh sách ứng cử các hàm tốt. Với mỗi biến hồi quy *X* *G*, đặt *L* là tập hợp của những biến tối ưu và *X* nênta có *L* = *Re*(*f*) {*X*} với *f* *OPTIMAL*. Đặt *CANDIDATE* là danh sách ứng cử các hàm tốt, được khởi tạo rỗng. Đặt *g* là hàm mới được tạo từ *L*; như vậy tập các biến của *g* là *L*, ta có *Re*(*g*) = *L*. Nếu *g* thỏa cặp điều kiện tối ưu, nó được thêm vào *CANDIDATE*, nên ta có *CANDIDATE* = *CANDIDATE* {*g*}.
4. Đặt *BEST* là tập các hàm tốt nhất trích xuất từ *CANDIDATE*. Những hàm này vừa thuộc *CANDIDATE* vừa thỏa cặp điều kiện tối ưu nhiều nhất với hệ số tương quan lớn nhất và tổng lỗi nhỏ nhất. Nếu *BEST* bằng *OPTIMAL*, thuật toán dừng, ngược lại gán *OPTIMAL* = *BEST* và trở lại bước 1.

Hình 2 mô tả lược đồ của thuật toán SG.



**Hình 2**. Lược đồ của thuật toán SG

Thuật toán SG đã được mô tả chi tiết trong bài báo “A framework of fetal age and weight estimation” (Nguyen & Ho, A framework of fetal age and weight estimation, 2014, pp. 21-23). Ý tưởng chính của SG là giảm không gian tìm kiếm bằng cách chọn các biến hồi quy thỏa cặp điều kiện tối ưu như là những “hạt giống” và các hàm tối ưu được tạo thành từ những “hạt giống” này. SG luôn tìm ra những hàm tốt nhất nhưng có thể bỏ qua những hàm khá tốt khác. Độ dài của một hàm là số biến hồi quy của nó. Thuật toán SG dừng khi không thể tìm thêm hàm tối ưu hoặc các biến hồi quy được duyệt hết. Vì thế hàm được tìm ra cuối cùng sẽ dài nhất và tốt nhất nhưng những hàm ngắn hơn khác có thể tốt đáng kể.

Phiên bản thi công hiện tại của SG thiết lập ngưỡng cực tiểu *δ* dương bất kỳ. SG cũng hỗ trợ các mô hình hồi quy phi tuyến trong bảng 2 như sau:

|  |  |
| --- | --- |
| Đa thức |  |
| Lô-ga-rít |  |
|  |
| Hàm mũ |  |
|  |
| Hàm tích |  |

**Table 3.** Các mô hình hồi quy phi tuyến

Ký hiệu “exp” và “log” biểu diễn hàm mũ và hàm lô-ga-rít tự nhiên. Hầu hết mô hình hồi quy phi tuyến được biến đổi thành mô hình tuyến tính. Ví dụ, mô hình hàm tích được biến đổi tuyến tính như sau:

Đặt,

Mô hình hàm tích trở thành mô hình tuyến tính quen thuộc với các biến *U*, *Zi* và những hệ số *βi* như sau:

Bảng 3 liệt kê cách biến đổi mô hình phi tuyến thành tuyến tính.

|  |  |
| --- | --- |
| Biến đổi  đa thức |  |
| Where |
| Biến đổi  lô-ga-rít |  |
| Where *Zi* = log(*Xi*) |
| Biến đổi  lô-ga-rít |  |
| Where |
| Biến đổi  hàm mũ |  |
| Where *U* = log(*Y*) |
| Biến đổi  hàm mũ |  |
| Where *U* = log(*Y*) and |
| Biến đổi  hàm tích |  |
| Where |

**Bảng 3.** Biến đổi mô hình phi tuyến thành tuyến tính

Tóm lại, với thuật toán thông minh SG, nền tảng Phoebe có thể được sử dụng cho bất kỳ ứng dụng hồi quy khác bên cạnh ứng dụng ước lượng tuổi và cân nặng thai – vấn đề chính của nghiên cứu này.

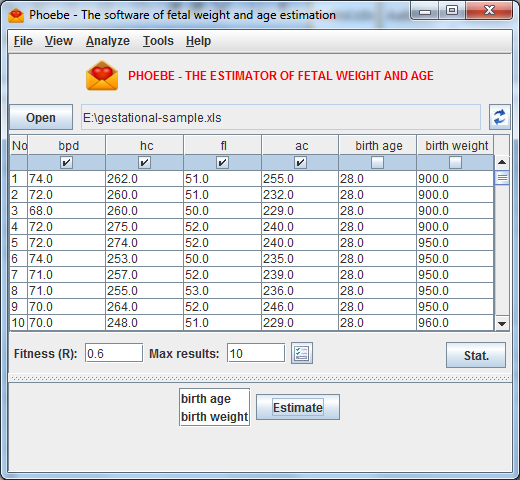
**4. Các trường hợp sử dụng Phoebe**

Ba trường hợp sử dụng Phoebe phổ biến được hỗ trợ bởi ba thành phần cơ bản (Dữ Liệu, Hồi Quy, Bản Mô Tả Thống Kê) của kiến trúc tổng quan của Phoebe được đề cập trong phần 2. Ba trường hợp này bao gồm:

1. Tìm những công thức tối ưu với độ chính xác cao là kết quả của thuật toán SG được mô tả trong phần 3.
2. Cung cấp thông tin thống kê của dữ liệu thai kỳ dưới dạng số và đồ thị.
3. So sánh các công thức với nhau.

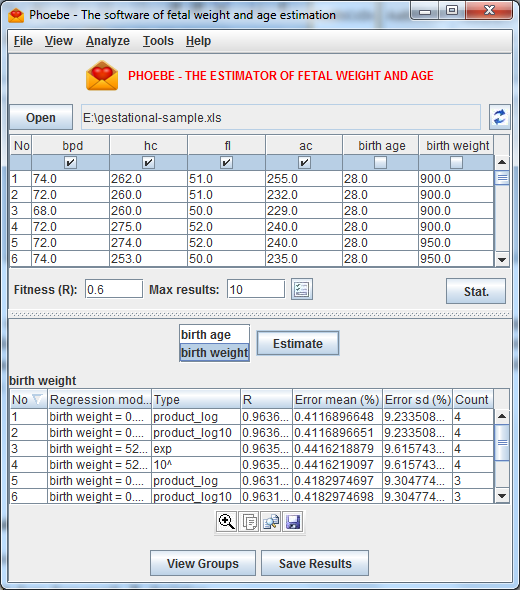
*Trường hợp sử dụng 1: Tìm công thức tối ưu*

Với dữ liệu thai kỳ (Ho & Phan, Ước lượng cân nặng của thai từ 37 – 42 tuần bằng siêu âm 2 chiều, 2011) gồm các số đo siêu âm hai chiều của những sản phụ, được đo tại Bệnh Viện Đa Khoa Vĩnh Long – Việt Nam, gồm đường kính lưỡng đỉnh (*bpd*), chu vi đầu (*hc*), chu vi bụng (*ac*), chiều dài xương đùi (*fl*). Tuổi thai từ 28 đến 42 tuần. Đơn vị cân nặng thai là gram. Hình 3 là một giao diện của Phoebe hiển thị dữ liệu thai kỳ này.



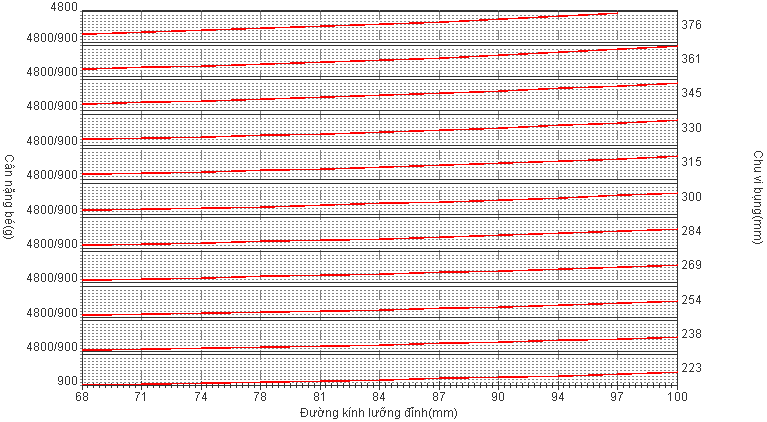
**Figure 3**. Gestational sample

Sau khi đặc tả ngưỡng cực đại *ε* (giá trị fitness), biến hồi quy và biến đáp ứng, người dùng nhấn nút “Estimate” để nhận những công thức tối ưu là kết quả của thuật toán SG. Hình 4 hiển thị những công thức tối ưu này. Lưu ý, trong hình 4, biến hồi quy là *bpd*, *hc*, *ac*, *fl* và biến đáp ứng là cân nặng thai. Ngưỡng cực đại *ε* là 0.6.



**Figure 4**. Optimal weight estimation formulas

Công thức ước lượng với một hoặc 2 biến hồi quy (số đo siêu âm) có thể được biểu diễn bằng đồ thị. Trong hình minh họa 5, trục hoành biểu diễn số đo *bpd* và trục tung bên phải biểu diễn số đo *ac*. Trục tung bên trái là cân nặng ước lượng.



**Hình 5**. Đồ thị ước lượng cân nặng thai

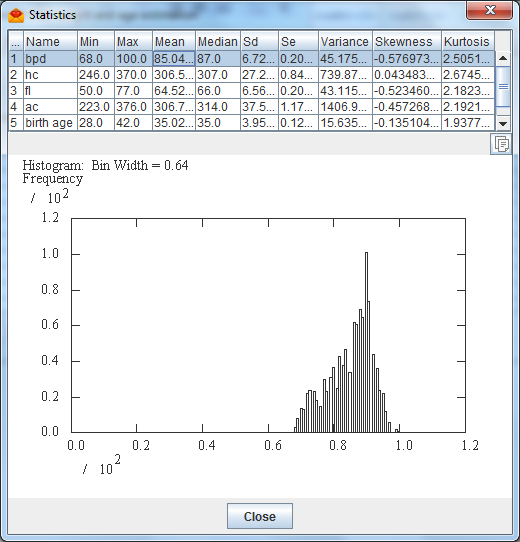
Đồ thị trong hình 5 có 11 đường ước lượng được vẽ như những đường (màu đỏ) bên trong. Mỗi đường ước lượng ứng với một khoảng nhỏ của *ac*. Cân nặng thai trên mỗi đường ước lượng dao động từ 900 gram đến 4800 gram. Đồ thị này là một cách để biểu diễn hàm ba chiều thành đồ thị hai chiều. Ví dụ, chúng ta cần ước lượng cân nặng thai với *bpd* = 90 and *ac* = 300. Trước tiên chúng ta nhìn vào đường ước lượng thứ 6 từ dưới đếm lên. Giao điểm giữa *bpd* = 90 và đường ước lượng thứ 6 được chiếu lên trục tung bên trái, sẽ cho ra là kết quả cân nặng thai xấp xỉ (4800 – 900) / 2 + 900 ≈ 2850 grams vì giao điểm này gần điểm giữa trong khoảng dao động của giá trị cân nặng trên đường ước lượng thứ 6 từ 900 gram đến 4800 gram.

*Trường hợp sử dụng 2: Cung cấp thông tin thống kê*

Thông tin thống kê được chia thành hai nhóm: thông tin thai kỳ và thông tin ước lượng.

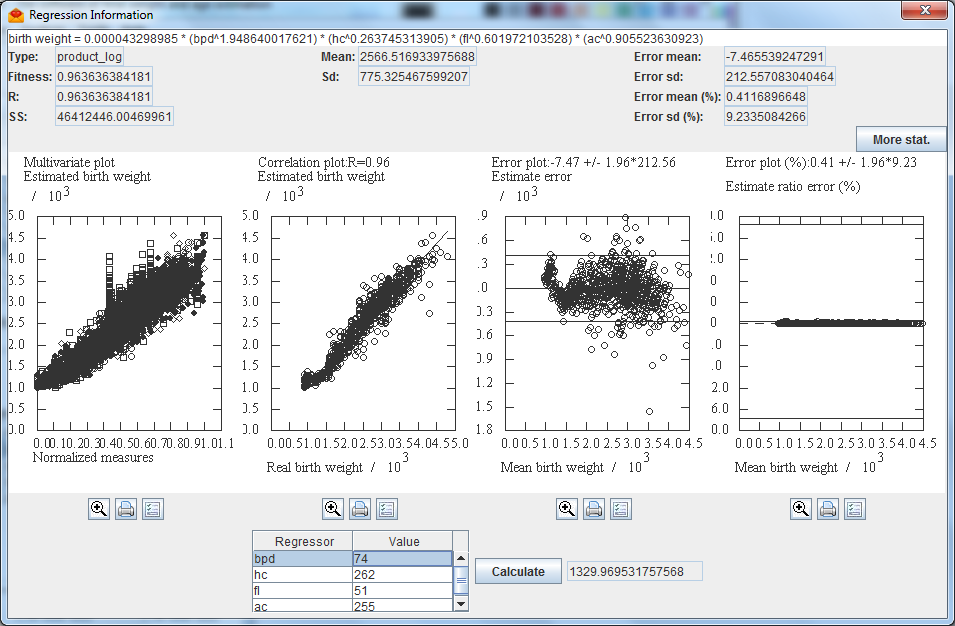
* Thông tin thai kỳ chứa các thuộc tính thống kê về số đo siêu âm, ví dụ như trung bình và độ lệch chuẩn của số đo *bpd*.
* Thông tin ước lượng chứa các thuộc tính thống kê của công thức ước lượng, ví dụ như hệ số tương quan, tổng phần dư (sum of residuals), lỗi ước lượng của công thức ước lượng

Thông tin thống kê được biểu diễn dưới dạng số và dạng đồ thị. Hình 6 mô tả các thuộc tính thống kê (trung bình, trung vị, độ lệch chuẩn, histogram,…) của các số đo siêu âm và tuổi thai.



**Figure 6**. Gestational statistical information

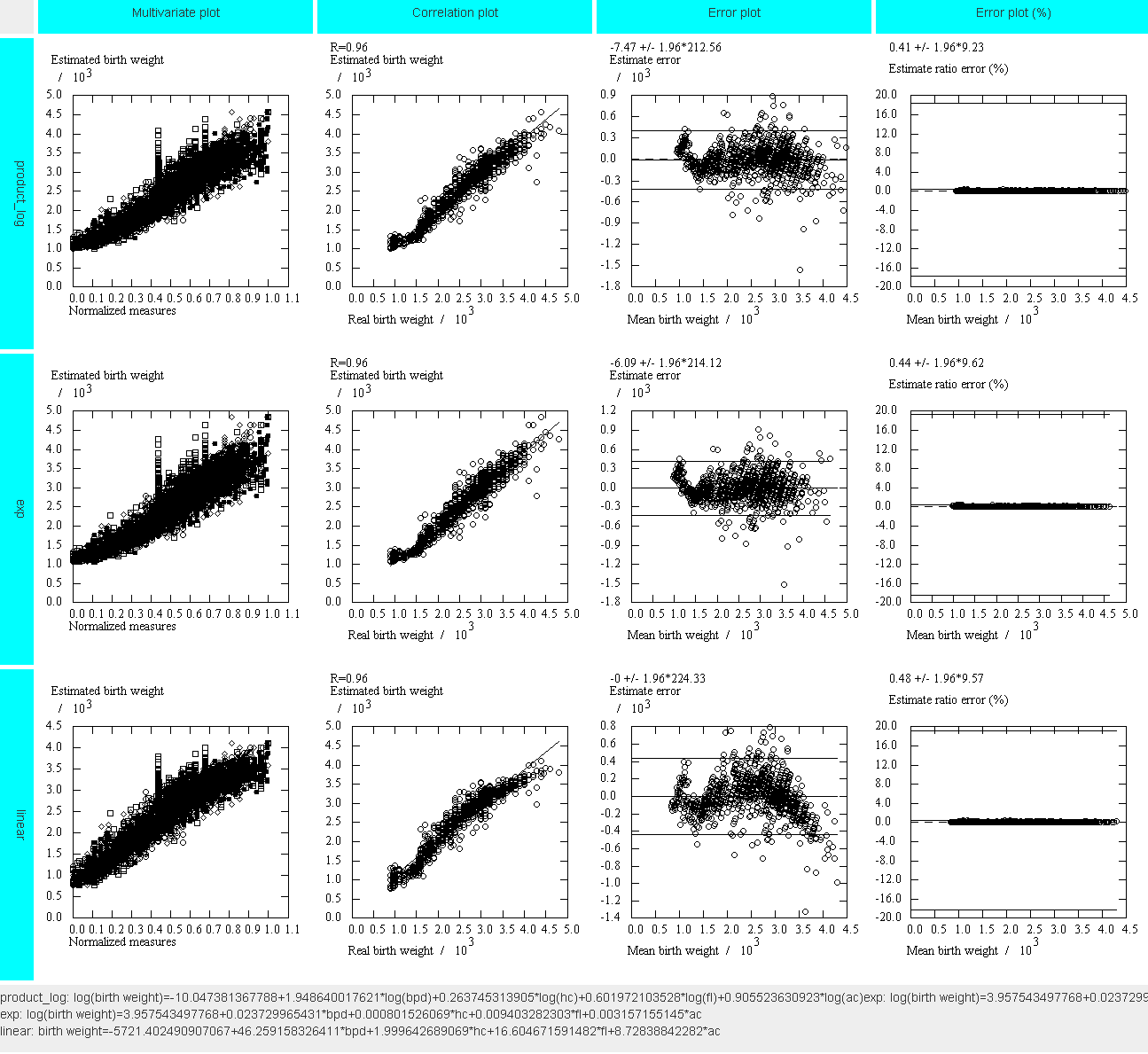
Hình 7 mô tả đầy đủ một công thức ước lượng cân nặng thai: *weight* = 0.000043 \* (*bpd*^1.948640) \* (*hc*^0.263745) \* (*fl*^0.601972) \* (*ac*^0.905524). Ví dụ, tổng phần dư (SS) là 46412446.0047 và khoảng lỗi ước lượng là -7.4655 ± 212.5571. Lưu ý, ký hiệu “^” biểu diễn hàm mũ, như 2^3 = 8.



**Figure 7**. Statistical estimation information

*Trường hợp sử dụng 3: So sánh các công thức với nhau*

Có nhiều tiêu chuẩn đánh giá sự hiệu quả và chính xác của công thức ước lượng như: hệ số tương quan, tổng phần dư, lỗi ước lượng. Mỗi công thức đều có ưu điểm và nhược điểm. Công thức này tốt hơn công thức kia ở một vài tiêu chuẩn nhưng có thể kém hơn ở tiêu chuẩn khác. Một công thức tối ưu sẽ có nhiều ưu điểm hơn nhược điểm trong hầu hết các tiêu chuẩn. Do đó, Phoebe hỗ trợ việc so sánh giữa các công thức qua *ma trận đánh giá* như trong hình 8. Mỗi dòng đại diện một công thức và mỗi cột đại diện một tiêu chuẩn. Ví dụ, dòng 1, dòng 2 và dòng 3 đại diện ba công thức tương ứng ở ba dạng: hàm lô-ga-rít, hàm mũ, hàm tuyến tính. Bốn tiêu chuẩn: tương quan đa biến, tương quan ước lượng (R), khoảng lỗi và khoảng tỉ lệ lỗi được xếp theo cột tương ứng.



**Figure 8**. Comparison among different formulas

Bảng 4, 5, 6, 7, 8 trong phần kết quả thử nghiệm là biểu diễn ở dạng số của ma trận đánh giá trong hình 8.

**5. Kết Quả Thử Nghiệm**

Nghiên cứu này tập trung vào thử nghiệm Phoebe với mẫu dữ liệu của chúng tôi để tìm công thức tối ưu dùng ước lượng tuổi và cân nặng thai trước sinh, phù hợp nhất với cộng đồng dân cư nơi chúng tôi thực hiện nghiên cứu. Chúng tôi dùng hai mẫu dữ liệu, trong đó mẫu số đo siêu âm 2 chiều (2D) gồm 1027 trường hợp và mẫu số đo siêu âm 3 chiều (3D) gồm 506 trường hợp. Giáo sư Hồ Thị Thu Hằng và giáo sư Phan Trường Duyệt (Ho & Phan, Ước lượng cân nặng của thai từ 37 – 42 tuần bằng siêu âm 2 chiều, 2011), (Ho & Phan, 2011) đã xây dựng hai mẫu này từ những sản phụ tại bệnh viện Đa Khoa Vĩnh Long - Việt Nam với sự tuân thủ nghiêm ngặt mọi tiêu chuẩn y đức. Các sản phụ và chồng của họ đều là người Việt Nam. Mỗi sản phụ chỉ có một thai còn sống. Tuổi thai từ 28 đến 42 tuần. Thời điểm sinh không quá 48 giờ sau lần siêu âm cuối cùng. Các số đo siêu âm 2D là *bpd*, *hc*, *ac* và *fl*. Các số đo siêu âm 3D là *bpd*, *hc*, *ac*, *fl*, *arm\_vol* và *thigh\_vol*. Đơn vị của *bpd*, *hc*, *ac*, *fl* is mi-li-mét. Đơn vị của *arm\_vol* và *thigh\_vol* là cm3. Đơn vị của tuổi thai là tuần và của cân nặng thai là gram.

Nền tảng Phoebe luôn tìm ra công thức tối ưu khi so sánh với các công thức khác theo hai tiêu chuẩn là hệ số tương quan (*R*) và khoảng lỗi. Chúng tôi dùng hai mẫu dữ liệu 2D và 3D này. Đặt *Y* = {*y*1, *y*2, *yn*} là tập tuổi/cân nặng thai từ mẫu dữ liệu và *Z* = {*z*1, *z*2, *zn*} là tập tuổi/cân nặng thai từ ước lượng. Theo phương trình 1, hệ số tương quan ước lượng *R* là hệ số tương quan giữa *Y* và *Z*, thể hiện độ tương thích với cộng đồng dân cư của công thức ước lượng. *R* càng lớn thì công thức càng tốt.

|  |  |
| --- | --- |
|  | (1) |

Lỗi ước lượng *di* là độ lệch giữa *zi* and *yi*. Trung bình lỗi *µ* thể hiện sự chính xác của công thức ước lượng. Công thức càng chính xác khi *µ* càng nhỏ. Nếu *µ* dương, công thức dễ cho ra kết quả ước lượng vượt mức. Nếu *µ* âm, công thức có xu hướng ước lượng thấp. Độ lệch chuẩn lỗi *σ* thể hiện sự ổn định của công thức. Công thức càng ổn định khi *σ* càng nhỏ. *Khoảng lỗi* là sự kết hợp giữa trung bình lỗi *µ* và độ lệch chuẩn lỗi *σ*. Phương trình 2 liệt kê cách tính *μ*, *σ* và khoảng lỗi.

|  |  |
| --- | --- |
|  | (2) |

Ví dụ, nếu *µ* = -0.0292 và *σ* = 1.45 thì khoảng lỗi là -0.0292±1.45, nghĩa là sai số của giá trị ước lượng tuổi/cân nặng thai dao động từ -1.4792 = -0.0292-1.45 to 1.4208 = -0.0292+1.45. Khoảng lỗi thể hiện toàn diện nhất độ chính xác của công thức.

Bảng 4 so sánh công thức của chúng tôi về ước lượng tuổi thai với các công thức khác theo mẫu 2D. Quy ước rằng tên của công thức chính là tên của tác giả tương ứng sẽ được liệt kê trong phần tham khảo. Ví dụ, công thức “Ho 1” là công thức đầu tiên của tác giả Ho (Ho T. T., Nghiên Cứu Phương Pháp Ước Lượng Trọng Lượng Thai, Tuổi Thai Bằng Siêu Âm Hai và Ba Chiều, 2011). Trong bảng 4, công thức chúng tôi tốt nhất với *R*=0.9303 và khoảng lỗi -0.0292±1.4500 tuần. Theo quy ước, tên của công thức chúng tôi có tiền tố “NH”.

|  |  |  |  |
| --- | --- | --- | --- |
| *Công thức* | *Biểu thức* | *R* | *Khoảng lỗi* |
| **NH 1** | log(*age*) = 2.419638 + 0.002012 \* *bpd* + 0.000934 \* *hc* + 0.00547 \* *fl* + 0.001042 \* *ac* | 0.9303 | -0.0292±1.4500 |
| **NH 2** | *age* = -3.364759 + 0.056285 \* *bpd* + 0.034697 \* *hc* + 0.188156 \* *fl* + 0.035304 \* *ac* | 0.9285 | 0±1.4682 |
| Ho 1 | *age* = 331.022308 - 1.611774 \* (*hc* + *ac*) + 0.00278 \* ((*hc* + *ac*)^2) - 0.000002 \* ((*hc* + *ac*)^3) | 0.9212 | 0±1.5384 |
| Varol 6 | *age* = 11.769 + 1.275 \* *fl*/10 + 0.449 \* ((*fl*/10)^2) - 0.02 \* ((*fl*/10)^3) | 0.8949 | -1.6807±1.8525 |
| Varol 1 | *age* = 5.596 + 0.941 \* *ac*/10 | 0.8941 | -0.5683±1.7711 |
| Varol 5 | *age* = 1.863 + 6.280 \* *fl*/10 - 0.211 \* ((*fl*/10)^2) | 0.8934 | -1.5182±2.1150 |

**Bảng 4.** So sánh ước lượng tuổi thai với mẫu 2D.

Kí hiệu “^” biểu diễn hàm mũ. Các biểu thức được ghi chú rất mềm dẻo để có thể trở thành đầu vào của bất kỳ công cụ tính toán. Bảng 5 so sánh công thức của chúng tôi về ước lượng cân nặng thai với các công thức khác theo mẫu 2D. Trong bảng 5, công thức chúng tôi tốt nhất với *R*=0.9636 và khoảng lỗi -7.4656±212.5573 gram.

|  |  |  |  |
| --- | --- | --- | --- |
| *Công thức* | *Biểu thức* | *R* | *Khoảng lỗi* |
| **NH 3** | log(*weight*) = -10.047381 + 1.94864 \* log(*bpd*) + 0.263745 \* log(*hc*) + 0.601972 \* log(*fl*) + 0.905524 \* log(*ac*) | 0.9636 | -7.4656±212.5573 |
| **NH 4** | log(*weight*) = 3.957543 + 0.02373 \* *bpd* + 0.000802 \* *hc* + 0.009403 \* *fl* + 0.003157 \* *ac* | 0.9635 | -6.0901±214.1153 |
| Sherpard | *weight* = 10^(1.2508 + 0.166 \* *bpd*/10 + 0.046 \* *ac*/10 - 0.002646 \* *ac* \* *bpd*/100) | 0.9619 | -65.8121±219.0392 |
| Ho 2 | *weight* = 10^(1.746 + 0.0124 \* *bpd* + 0.001906 \* *ac*) | 0.9602 | -11.5576±223.5124 |
| Hadlock | *weight* = 10^(1.304 + 0.05281 \* *ac*/10 + 0.1938 \* *fl*/10 - 0.004 \* *ac* \* *fl*/100) | 0.9395 | -76.4960±272.9474 |
| Campbell & Wilkin | *weight* = 1000 \* exp(-4.564 + 0.282 \* *ac*/10 - 0.00331 \* *ac* \* *ac*/100) | 0.9215 | 68.1261±308.5728 |

**Bảng 5.** So sánh ước lượng cân nặng thai với mẫu 2D.

Bảng 6 so sánh công thức của chúng tôi về ước lượng tuổi thai với các công thức khác theo mẫu 3D. Trong bảng 6, công thức chúng tôi tốt nhất với *R*=0.9970 và khoảng lỗi ±0.2696 tuần.

|  |  |  |  |
| --- | --- | --- | --- |
| *Công thức* | *Biểu thức* | *R* | *Khoảng lỗi* |
| **NH 5** | *age* = 20.759763 + 0.170859 \* (*thigh\_vol* + *arm\_vol*) - 0.000545 \* ((*thigh\_vol* + *arm\_vol*)^2) + 0.000001 \* ((*thigh\_vol* + *arm\_vol*)^3) | 0.9970 | 0±0.2696 |
| **NH 6** | *age* = 21.816252 + 0.137531 \* (*thigh\_vol* + *arm\_vol*) - 0.000228 \* ((*thigh\_vol* + *arm\_vol*)^2) | 0.9969 | 0±0.2752 |
| Ho 3 | *age* = 21.1148 + 0.2381 \* *thigh\_vol* - 0.001 \* (*thigh\_vol*^2) + 0.000002 \* (*thigh\_vol*^3) | 0.9960 | -0.0150±0.3173 |
| Ho 4 | *age* = 167.079079 - 1.553705 \* *ac* + 0.005559 \* (*ac*^2) - 0.000006 \* (*ac*^3) | 0.8482 | 0.3723±1.8985 |

**Bảng 6.** So sánh ước lượng tuổi thai với mẫu 3D.

Bảng 7 so sánh công thức của chúng tôi về ước lượng cân nặng thai với các công thức khác theo mẫu 3D. Trong bảng 7, công thức chúng tôi tốt nhất với *R*=0.9708 và khoảng lỗi -0.0001 ± 180.9803 gram.

|  |  |  |  |
| --- | --- | --- | --- |
| *Công thức* | *Biểu thức* | *R* | *Khoảng lỗi* |
| **NH 7** | *weight* = -3617.936175 + 0.513171 \* *hc* + 1.960176 \* *ac* + 39.804645 \* *bpd* + 17.016936 \* *fl* + 8.366404 \* *thigh\_vol* + 5.828808 \* *arm\_vol* | 0.9708 | -0.0001±180.9803 |
| **NH 8** | *weight* = -3626.314419 + 43.426744 \* *bpd* + 23.645338 \* *fl* + 11.414273 \* *thigh\_vol* | 0.9698 | 0±184.0439 |
| Ho 5 | *weight* = -3306 + 55.477 \* *bpd* + 13.483 \* *thigh\_vol* | 0.9663 | -0.0072±194.0956 |
| Lee 3 | *weight* = exp(0.5046 + 1.9665 \* log(*bpd*/10) - 0.3040 \* (log(*bpd*/10)^2) + 0.9675 \* log(*ac*/10) + 0.3557 \* log(*arm\_vol*)) | 0.9620 | 247.8761±206.1607 |
| Lee 5 | *weight* = exp(2.1264 + 1.1461 \* log(*ac*/10) + 0.4314 \* log(*thigh\_vol*)) | 0.9514 | 289.2660±234.0763 |
| Lee 2 | *weight* = exp(-3.6138 + 4.6761 \* log(*ac*/10) - 0.4959 \* (log(*ac*/10)^2) + 0.3795 \* log(*arm\_vol*)) | 0.9472 | 316.4974±242.7964 |
| Ho 6 | *weight* = -882.7049 + 73.9955 \* *thigh\_vol* - 0.497 \* (*thigh\_vol*^2) + 0.0014 \* (*thigh\_vol*^3) | 0.9385 | -7.5001±260.4596 |
| Lee 4 | *weight* = exp(4.7806 + 0.7596 \* log(*thigh\_vol*)) | 0.9298 | 737.4932±344.1904 |
| Lee 1 | *weight* = exp(4.9588 + 1.0721 \* log(*arm\_vol*) - 0.0526 \* (log(*arm\_vol*)^2)) | 0.9281 | 867.0836±309.5779 |
| Chang | *weight* = 1080.8735 + 22.44701 \* *thigh\_vol* | 0.9229 | 456.5168±298.2517 |

**Bảng 7.** So sánh ước lượng cân nặng thai với mẫu 3D.

Khi nghiên cứu phần siêu âm 3D trong luận văn tiến sĩ của giáo sư Hồ Thị Thu Hằng (Ho T. T., Nghiên Cứu Phương Pháp Ước Lượng Trọng Lượng Thai, Tuổi Thai Bằng Siêu Âm Hai và Ba Chiều, 2011), tôi nhận thấy cân nặng thai và tuổi thai có sự phụ thuộc lẫn nhau. Cụ thể, cân nặng thai tăng khi tuổi thai tăng. Kết quả là ước lượng cân nặng thai sẽ cải thiện đáng kể nếu biết trước tuổi thai. Nếu tuổi thai trở thành một biến trong hàm hồi quy của cân nặng thai thì ta có *công thức đối ngẫu* để ước lượng cân nặng thai còn chính xác hơn những công thức tốt nhất trong bảng 5 và 7. Công thức đối ngẫu không những chính xác mà còn dễ áp dụng bởi vì nhiều sản phụ biết tuổi thai của họ trước khi khám siêu âm. Với các mẫu 2D và 3D, bảng 8 so sánh công thức đối ngẫu của chúng tôi về ước lượng cân nặng thai với các công thức tốt nhất vốn đã được liệt kê trong bảng 5 và 7. Tiêu chuẩn so sánh vẫn là *R* và khoảng lỗi. Quy ước rằng tên công thức đối ngẫu có tiền tố “NHD”. Ký hiệu “log10” biểu diễn hàm lô-ga-rít cơ số 10.

|  |  |  |  |
| --- | --- | --- | --- |
| *Công thức* | *Biểu thức* | *R* | *Khoảng lỗi* |
| **NHD 1**  (mẫu 2D) | log10(*weight*) = -3.715073 + 1.873457 \* log10(*bpd*) + 0.363783 \* log10(*fl*) + 0.691683 \* log10(*ac*) + 0.722245 \* log10(*age*) | 0.9674 | -5.6422 ± 202.0395 |
| **NHD 2**  (mẫu 2D) | log10(*weight*) = -3.761798 + 2.001731 \* log10(*bpd*) + 0.811078 \* log10(*ac*) + 0.826279 \* log10(*age*) | 0.9667 | -5.6111 ± 204.1477 |
| **NHD 3**  (mẫu 3D) | *weight* = -4988.000528 + 66.374156 \* *age* + 0.370084 \* *hc* + 1.943247 \* *ac* + 39.464816 \* *bpd* + 13.215505 \* *fl* + 3.658463 \* *thigh\_vol* | 0.9715 | 0 ± 178.8091 |
| **NHD 4**  (mẫu 3D) | *weight* = -4982.099978 + 68.089354 \* *age* + 2.001675 \* *ac* + 39.85375 \* *bpd* + 13.229377 \* *fl* + 3.619405 \* *thigh\_vol* | 0.9714 | 0 ± 178.9114 |
| NH 3  (mẫu 2D) | log(*weight*) = -10.047381 + 1.94864 \* log(*bpd*) + 0.263745 \* log(*hc*) + 0.601972 \* log(*fl*) + 0.905524 \* log(*ac*) | 0.9636 | -7.4656±212.5573 |
| NH 4  (mẫu 2D) | log(*weight*) = 3.957543 + 0.02373 \* *bpd* + 0.000802 \* *hc* + 0.009403 \* *fl* + 0.003157 \* *ac* | 0.9635 | -6.0901±214.1153 |
| NH 7  (mẫu 3D) | *weight* = -3617.936175 + 0.513171 \* *hc* + 1.960176 \* *ac* + 39.804645 \* *bpd* + 17.016936 \* *fl* + 8.366404 \* *thigh\_vol* + 5.828808 \* *arm\_vol* | 0.9708 | -0.0001±180.9803 |
| NH 8  (mẫu 3D) | *weight* = -3626.314419 + 43.426744 \* *bpd* + 23.645338 \* *fl* + 11.414273 \* *thigh\_vol* | 0.9698 | 0±184.0439 |

**Bảng 8.** Ước lượng cân nặng với công thức đối ngẫu.

Trong bảng 8, tất cả công thức đối ngẫu NHD \* đều tốt hơn những công thức thông thường NH \* theo tiêu chí *R* và khoảng lỗi. Hơn nữa, NHD \* không cần quá nhiều biến hồi quy. Với mẫu 2D, NHD 1 và NHD 2 lần lượt dùng 4 và 3 biến hồi quy kể luôn biến tuổi thai trong khi đó NH 3 và NH 4 đều dùng 4 biến hồi quy. Với mẫu 3D, NHD 3 và NHD 4 lần lượt dùng 6 và 5 biến hồi quy kể luôn biến tuổi thai trong khi đó NH 7 và NH 8 lần lượt dùng 5 và 3 biến hồi quy.

Mặc dù công thức chúng tôi tốt nhất với độ tương thích cao (*R* lớn) và độ chính xác cao (khoảng lỗi nhỏ) nhưng các nghiên cứu của những nhà khoa học khác vẫn luôn rất quan trọng vì công thức của họ đơn giản và dễ dùng. Hơn nữa, công thức chúng tôi không mang tính toàn cục vì nếu áp dụng chúng vào cộng đồng dân cư khác thì độ chính xác có thể giảm, có thể kém những công thức truyền thống như của Sherpard và Hadlock. Tuy nhiên có thể kết luận từ kết quả thử nghiệm rằng nếu dùng Phoebe với mẫu dữ liệu khác thì luôn được công thức ưu việt cho chính mẫu đó. Để đạt sự tối ưu toàn cục với nền tảng Phoebe, chúng tôi có hai đề xuất then chốt sau:

* Thử nghiệm Phoebe trên nhiều mẫu dữ liệu.
* Thêm nhiều hơn tri thức về sản khoa và kỹ thuật siêu âm vào nền tảng Phoebe. Những tri thức bổ sung này sẽ được mô tả thành các ràng buộc trong thuật toán SG.

Hai đề xuất trên vượt ra ngoài nghiên cứu này. Theo tôi, chúng ta không thể đạt tối ưu toàn cục một cách tuyệt đối vì Phoebe tập trung vào tối ưu cục bộ với cộng đồng dân cư cụ thể. Chính xác hơn, hai đề xuất chỉ giảm nhẹ nhược điểm của thuật toán SG.

**6. Đề xuất ước lượng sớm cân nặng thai**

Mẫu siêu âm được lấy với tuổi thai từ 28 tuần đến 42 tuần vì thời điểm sinh không quá 48 giờ từ lần siêu âm cuối cùng. Vì vậy ước lượng cân nặng chỉ chính xác khi siêu âm được thực hiện sau 28 tuần tuổi thai. Trong phần này chúng tôi đề xuất một phương pháp ước lượng sớm cân nặng thai với những số đo siêu âm được lấy trước 28 tuần tuổi thai. Chúng tôi chưa khẳng định được độ chính xác của phương pháp đề xuất vì chúng tôi chưa thử nghiệm nhưng việc dữ liệu siêu âm có thể được thu thập tại bất cứ thời điểm nào trong thai kỳ sẽ mang lại tiện lợi cho các bác sĩ thực hành / nhà nghiên cứu bởi vì họ không cần quan tâm đến cân nặng thai khi thực hiện siêu âm. Tóm lại, với phương pháp đề xuất, mẫu dữ liệu có thể thiếu trọng lượng thai và kết quả là cân nặng thai có thể được ước lượng sớm trước 28 tuần tuổi thai. Trong phần này, quy ước các vector là vector cột nếu không có giải thích thêm.

Không mất tính tổng quát, giả sử mô hình hồi quy là tuyến tính *Y = α*0 *+ α*1*X*1 *+ α*2*X*2 *+ … + αnXn* and *Z = β*0 *+ β*1*X*1 *+ β*2*X*2 *+ … + βnXn* với *Y* là tuổi thai và *Z* là cân nặng thai trong khi *Xi* là những số đo siêu âm như *bpd*, *hc*, *ac*, *fl*. Giả sử *Y* và *Z* tuân theo phân bố chuẩn với phương trình 3 (Lindsten, Schön, Svensson, & Wahlström, 2017, pp. 8-9).

|  |  |
| --- | --- |
|  | (3) |

Theo đó, *α* = (*α*1, *α*2,…, *αn*)*T* và *β* = (*β*1, *β*2,…, *βn*)*T* là hai vector tham số và *X* = (1, *X*1, *X*2,…, *Xn*)*T* là vector dữ liệu. Trung bình mẫu của *Y* và *Z* là *αTX* và *βTX*, trong khi phương sai của chúng là và . Ghi chú, chữ cái “*T*” ký hiệu toán tử chuyển vị của vector và ma trận. Đặt ***D*** = (***X***, ***y***, ***z***) là mẫu thu thập, trong đó ***X*** là tập các số đo siêu âm, ***y*** là tập các tuổi thai và ***z*** là tập các cân nặng thai với lưu ý rằng ***z*** rỗng hoặc không đầy đủ. Nếu ***z*** rỗng thì không có bất kỳ giá trị cân nặng thai nào. Nếu ***z*** không đầy đủ thì ***z*** có một số giá trị nhưng một số giá trị khác bị thiếu. Tuy nhiên, ***y*** phải đầy đủ là một ràng buộc, nghĩa là tất cả sản phụ tham gia nghiên cứu đều biết tuổi thai của họ. Chúng ta tập trung vào ước lượng *α* và *β* dựa trên ***D***. Quy ước *α\** và *β\** là những giá trị ước lượng tối ưu của *α* và *β* (Lindsten, Schön, Svensson, & Wahlström, 2017, p. 8).

Cho *X*, xác suất hợp của *Y* và *Z* là tích của xác suất của *Y* trên *X* và xác suất của *Z* trên *X* bởi vì *Y* và *Z* đều phụ thuộc điều kiện vào *X*, theo phương trình 4.

|  |  |
| --- | --- |
|  | (4) |

Phương trình 5 đặc tả kỳ vọng điều kiện của thống kê đủ *Z* trên *X* với *P*(*Z* | *X*, *β*).

|  |  |
| --- | --- |
|  | (5) |

Khi *Z* là biến ẩn thì có mối liên hệ tiềm ẩn giữa *Y* và *Z* được đặc tả bởi xác suất hợp của *Y* và *Z*.

Hai biến *Y* và *Z* khác độ đo, cụ thể *Y* có đơn vị là tuần trong khi *Z* có đơn vị là gram. Giả sử *Y* là biến rời rạc có giá trị trong khoảng từ 1 đến *K*, ví dụ *K* có thể bằng 42 (tuần). Xác suất *P*(*Y*) trở thành tham số *θY*, là xác suất để *Y* có giá trị từ 1 đến *K*.

Với mỗi *Z*, giả sử xác suất điều kiện *P*(*Z* | *Y*) tuân theo phân bố chuẩn với trung bình *μY* và phương sai *σY*2. Phương trình 6 đặc tả xác suất hợp *P*(*Y*, *Z*).

|  |  |
| --- | --- |
|  | (6) |

Phương trình 7 đặc tả kỳ vọng điều kiện của thống kê đủ *Z* trên *Y* với *P*(*Z* | *Y*, *μY*,*σY*2).

|  |  |
| --- | --- |
|  | (7) |

Vui lòng lưu ý phương trình 7 vì *Z* sẽ được ước lượng bởi kỳ vọng này. Phương trình 8 đặc tả kỳ vọng của thống kê đủ *Z* với *P*(*Y*, *Z* | *θY*, *μY*, *σY*2).

|  |  |
| --- | --- |
|  | (8) |

Do

Phương trình 9 đặc tả xác suất hợp đầy đủ của *Y* và *Z* cho *X* và các tham số *α*, *β*, *θY*, *μY*, and *σY*2.

|  |  |
| --- | --- |
|  | (9) |

Với *P*(*Y*, *Z* | *X*, *α*, *β*) và *P*(*Y*, *Z* | *θY*, *μY*, *σY*2) đặc tả bởi các phương trình 4 and 6. Phương trình 9 thể hiện sự phụ thuộc hiển hiện qua xác suất *P*(*Y*, *Z* | *X*, *α*, *β*) và sự phụ thuộc ngầm định qua xác suất *P*(*Y*, *Z* | *θY*, *μY*, *σY*2) giữa *Y* và *Z*. Sự phụ thuộc hiển hiện và sự phụ thuộc ngầm định có cùng mức ảnh hưởng trên *Z* nếu *E*(*Z* | *X*) được tính bởi phương trình 5 bằng với *E*(*Z*) được tính bởi phương trình 8, như sau:

|  |  |
| --- | --- |
|  | (10) |

Cho mẫu ***D***, tất cả *θY* trở thành hằng số và được xác định bởi phương trình 11.

|  |  |
| --- | --- |
|  | (11) |

Để thuận tiện, đặt Θ = (*α*, *β*, *μY*)*T* là tập các tham số cần thiết. Xác suất hợp đầy đủ từ phương trình 9 được xác định lại như sau:

(Do tất cả dữ liệu thu thập đều độc lập và cùng phân bố)

Với,

Theo quy ước, nếu *δ*(*yi*, *Y*) = 0 thì xác suất *P*(*yi*, *zi* | *μY*, *σY*2) tương ứng bị loại khỏi tích trên. Hàm khả năng log là lô-ga-rít của xác suất hợp đầy đủ:

Khi log(2*π*) và *θY* là hằng số, hàm khả năng log giản lược được dẫn xuất từ hàm khả năng log theo phương trình 12 như sau:

|  |  |
| --- | --- |
|  | (12) |

Ước lượng tối ưu Θ*\** là điểm cực đại của *l*(Θ), theo phương trình 13 (Lindsten, Schön, Svensson, & Wahlström, 2017, p. 9).

|  |  |
| --- | --- |
|  | (13) |

Những đạo hàm riêng bậc nhất của *l*(Θ) là (Nguyen, Matrix Analysis and Calculus, 2015, p. 34):

Khi mọi đạo hàm riêng bậc nhất của *l*(Θ) bằng 0 thì *l*(Θ) đạt cực đại địa phương. Do đó Θ*\** là nghiệm của hệ phương trình 14. Lưu ý, hệ phương trình 14 được lập từ việc đặt một số đạo hàm riêng bậc nhất của *l*(Θ) bằng 0 và đặt *E*(*Z* | *X*) = *E*(*Z*).

|  |  |
| --- | --- |
|  | (14) |

Với,

Kí hiệu **0** = (0, 0,…, 0)*T* chỉ vector không. Tất cả phương trình trong hệ 14 đều tuyến tính với ẩn là Θ = (*α*, *β*, *μY*)*T*. Phương trình cuối cùng trong hệ 14 là phương trình 10 mô hình hóa giả định thông minh rằng sự phụ thuộc hiển hiện và sự phụ thuộc ngầm định có cùng mức ảnh hưởng trên *Z*. Phương trình này chỉ được dùng để điều chỉnh *μY* nếu giả định thông minh trên được quan tâm; ngược lại phương trình này sẽ bị bỏ qua.

Chúng tôi áp dụng thuật toán cực đại kỳ vọng (expectation maximization, viết tắt EM) để ước lượng Θ = (*α*, *β*, *μY*)*T* khi dữ liệu mẫu thiếu cân nặng thai. Lưu ý, xác suất hợp đầy đủ *P*(*Y*, *Z* | *X*, *α*, *β*, *μY*) được đặc tả bởi phương trình 9 là tích của những phân bố hàm mũ chính quy. Thuật toán EM gồm nhiều lần lặp, mỗi lần lặp có hai bước, bước E (expectation step) xác định kỳ vọng dựa trên tham số của lần lặp trước và bước M (maximization step) để ước lượng tham số mới cho lần lặp sau. Giả sử tham số hiện hành tại lần lặp thứ *t* là Θ*t* = (*αt*, *βt*, *μYt*)*T*, bảng 9 mô tả bước E và bước M tại lần lặp thứ *t* (Dempster, Laird, & Rubin, 1977, p. 4).

|  |
| --- |
| 1. Bước E: Chỉ những cân nặng bị thiếu *zi* được ước lượng bởi kỳ vọng của chính chúng với trung bình mẫu hiện tại , theo phương trình 6. Lưu ý, mỗi *zi* luôn có một *yi* kèm theo vì mọi sản phụ tham gia nghiên cứu đều biết tuổi thai. 2. Bước M: Tham số tiếp theo Θ*t*+1 là điểm cực đại của *l*(Θ), chính là nghiệm của hệ phương trình 14. Lưu ý, Θ*t*+1 sẽ trở thành tham số hiện tại của lần lặp tiếp theo. |

**Table 9.** Bước E and bước M của thuật toán EM

Hệ phương trình 14 luôn giải được vì những giá trị bị thiếu *zi* đã được ước lượng trong bước E. Thuật toán EM dừng nếu ta có Θ*t* = Θ*t*+1 = Θ*\** tại lần lặp thứ *t* nào đó; lúc đó Θ*\** = (*α\**, *β\**, *μY\**)*T* là giá trị ước lượng tối ưu của thuật toán EM và do đó những hàm hồi quy (công thức hồi quy) của *Y* và *X* được xác định với *α\** và *β\**.

Thông thường mọi tham số đều thay đổi sau mỗi lần lặp của thuật toán EM nhưng may mắn là *α\** được xác định ngay tại lần lặp đầu tiên như là nghiệm của phương trình thứ nhất của hệ 14 bởi vì dữ liệu ***X*** và ***y*** đều đầy đủ. Nói cách khác, *β* và *μY* thay đổi trong khi *α\** cố định trong suốt tiến trình EM. Phương trình 15 (Montgomery & Runger, 2003, p. 417) được dùng để tính *α\**.

|  |  |
| --- | --- |
|  | (15) |

Với ký hiệu “–1” chỉ ma trận nghịch đảo.

Tại lần lặp đầu tiên, thông thường Θ1 được khởi tạo bất kỳ nhưng chúng ta có thể tăng tốc độ hội tụ của thuật toán EM bằng cách khởi tạo bằng trung bình mẫu. Không mất tính tổng quát, giả sử bác sĩ thực hành đo được *n* < *N* giá trị cân nặng thai *z*1, *z*2,…, *zn* từ *n* lần siêu âm. Hơn nữa, những sản phụ trong *n* lần siêu âm đó có cùng tuổi thai là *Y*. Khi đó, được khởi tạo bởi phương trình 16.

|  |  |
| --- | --- |
|  | (16) |

Tham số *β*1 tại lần lặp đầu tiên được khởi tạo theo những nghiên cứu trước trong y văn.

**7. Kết luận**

Từ kết quả thử nghiệm, Phoebe luôn kết xuất công thức ước lượng tối ưu với độ tương thích và độ chính xác cao; vui lòng xem các bảng 4, 5, 6, 7 và 8 để xem lại chi tiết những công thức này. Tuy nhiên nhược điểm trong nghiên cứu của chúng tôi là thuật toán SG có thể bỏ qua một số công thức tốt do cặp điều kiện tối ưu. Giải pháp được đề nghị là thêm ràng buộc vào cặp điều kiện này; vui lòng đọc bài báo “A framework of fetal age and weight estimation” (Nguyen & Ho, A framework of fetal age and weight estimation, 2014, pp. 24-25) để biết thêm chi tiết. Phương pháp đề xuất ước lượng sớm cân nặng thai thật sự đã dùng một ràng buộc, chính là sự liên hệ ngầm định giữa tuổi thai và cân nặng thai. Mối liên hệ ngầm định này được mô hình hóa bằng xác suất hợp giữa tuổi thai và trọng lượng thai, là một điểm tri thức của ngành sản khoa. Trong những nghiên cứu tiếp theo, chúng tôi sẽ thử nghiệm phương pháp đề xuất này cùng với nỗ lực tìm tòi những tri thức khác.

Các công thức của chúng tôi phức tạp để đạt độ chính xác cao nên thật sự khó tính nhanh chúng. Trong tương lai chúng tôi sẽ nhúng các công thức này vào phần mềm hay phần cứng của máy siêu âm sao cho người dùng dễ đọc kết quả ước lượng từ máy.

**Lời cảm ơn**

Chúng tôi bày tỏ sự cảm ơn sâu sắc đối với tác giả Michael Thomas Flanagan – University College London và tác giả Jos de Jong đã cung cấp cho chúng tôi những gói phần mềm máy tính quý giá, giúp chúng tôi xây dựng và phát triển nền tảng Phoebe.

**Tham khảo**

Campbell, S., & Wilkin, D. (1975, September). Ultrasonic measurement of fetal abdomen circumference in the estimation of fetal weight. *BJOG: An International Journal of Obstetrics & Gynaecology, 82*(9), 689-697. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1101942

Chang, F.-M., Liang, R.-I., Ko, H.-C., Yao, B.-L., Chang, C.-H., & Yu, C.-H. (1997, September). Three-dimensional ultrasound-assessed fetal thigh volumetry in predicting birth weight. *Obstetrics & Gynecology, 90*(3), 331-339. doi:10.1016/S0029-7844(97)00280-9

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. (M. Stone, Ed.) *Journal of the Royal Statistical Society, Series B (Methodological), 39*(1), 1-38.

Flanagan, M. T. (2004, May 2). Java Scientific Library. *Michael Thomas Flanagan's Java Scientific Library*. (M. T. Flanagan, Ed.) London, Enland, UK: University College London. Retrieved from http://www.ee.ucl.ac.uk/~mflanaga/java

Hadlock, F. P., Harrist, R. B., Sharman, R. S., Deter, R. L., & Park, S. K. (1985, February 1). Estimation of fetal weight with use of head, body and femur measurements: A prospective study. *American Journal of Obstetrics and Gynecology, 151*(3), 333-337. doi:10.1016/0002-9378(85)90298-4

Ho, T. T. (2011). *Nghiên Cứu Phương Pháp Ước Lượng Trọng Lượng Thai, Tuổi Thai Bằng Siêu Âm Hai và Ba Chiều.* Hanoi Univerisy of Medicine. Hanoi: Hanoi Univerisy of Medicine. Retrieved 2011

Ho, T. T., & Phan, D. T. (2011, December). Ước lượng cân nặng của thai từ 37 – 42 tuần bằng siêu âm 2 chiều. (D. Thai, Ed.) *Journal of Practical Medicine, 12*(797), 8-9.

Ho, T.-H. T., & Phan, D. T. (2011, December). Ước lượng tuổi thai qua các số đo thể tích cánh tay bằng siêu âm 3 chiều và các số đo bằng siêu âm 2 chiều. (D. Thai, Ed.) *Journal of Practical Medicine, 12*(798), 12-15.

Jong, J. d. (2010, January 31). A Java expression parser. *A Java expression parser*. Rotterdam, Netherlands: SpeQ Mathematics. Retrieved 2011, from http://www.speqmath.com/tutorials/expression\_parser\_java

Lee, W., Balasubramaniam, M., Deter, R. L., Yeo, L., Hassan, S. S., Gotsch, F., . . . Romero, R. (2009, November 1). New fetal weight estimation models using fractional limb volume. (M. A. Zoppi, Ed.) *Ultrasound in Obstetrics & Gynecology, 34*(5), 556-565. doi:10.1002/uog.7327

Lindsten, F., Schön, T. B., Svensson, A., & Wahlström, N. (2017). *Probabilistic modeling – linear regression & Gaussian processes.* Uppsala University, Department of Information Technology. Uppsala: Uppsala University. Retrieved January 24, 2018, from http://www.it.uu.se/edu/course/homepage/sml/literature/probabilistic\_modeling\_compendium.pdf

Montgomery, D. C., & Runger, G. C. (2003). *Applied Statistics and Probability for Engineers* (3rd Edition ed.). New York, NY, USA: John Wiley & Sons, Inc.

Nguyen, L. (2015). *Matrix Analysis and Calculus* (1st ed.). (C. Evans, Ed.) Hanoi, Vietnam: Lambert Academic Publishing. Retrieved from https://www.shuyuan.sg/store/gb/book/matrix-analysis-and-calculus/isbn/978-3-659-69400-4

Nguyen, L., & Ho, H. (2014, March 30). A framework of fetal age and weight estimation. (B. S. Shetty, J. Morales, a. badawy, C. Mowa, K. K. Shukla, T. Chen, . . . G. Androutsopoulos, Eds.) *Journal of Gynecology and Obstetrics (JGO), 2*(2), 20-25. doi:10.11648/j.jgo.20140202.13

Nguyen, L., & Ho, T.-H. T. (2017, March 13). Experimental Results of Phoebe Framework: Optimal Formulas for Estimating Fetus Weight and Age. (H. J. Shaji, M. C. Portillo, & M. M. Zdanowicz , Eds.) *Journal of Community & Public Health Nursing, 3*(2), 1-5. doi:10.4172/2471-9846.1000163

Oracle. (n.d.). *Java language*. (Oracle Corporation) Retrieved December 25, 2014, from Java website: https://www.oracle.com/java

Phạm, T. T. (2000). *Ước lượng cân nặng thai nhi qua các số đo của thai trên siêu âm.* Ho Chi Minh University of Medicine and Pharmacy. Ho Chi Minh: Ho Chi Minh University of Medicine and Pharmacy.

Phan, D. T. (1985). *Ứng dụng siêu âm để chẩn đoán tuổi thai và cân nặng thai trong tử cung.* Hanoi University of Medicine. Hanoi: Hanoi University of Medicine.

Shepard, J. M., Richards, A. V., Berkowitz, L. R., Warsof, L. S., & Hobbins, C. J. (1982, January 1). An evaluation of two equations for predicting fetal weight by ultrasound. *American Journal of Obstetrics and Gynecology, 142*(1), 47-54. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/7055171

Varol, F., Saltik, A., Kaplan, P. B., Kilic, T., & Yardim, T. (2001, June). Evaluation of Gestational Age Based on Ultrasound Fetal Growth Measurements. (J.-W. Park, Ed.) *Yonsei Medical Journal, 42*(3), 299-303. doi:10.3349/ymj.2001.42.3.299