**19**

**Hướng dẫn sử dụng WinBUGS**

Trong khoảng một thập niên gần đây, WinBUGS đã trở thành một chương trình máy tính phân tích thống kê quan trọng. WinBUGS do một nhóm chuyên gia thống kê thuộc Hội đồng Nghiên cứu Y khoa Anh (Medical Research Council - MRC) phát triển, với sự tài trợ trực tiếp của chính phủ Anh. Sự ra đời của WinBUGS và phiên bản trước đó là BUGS (Bayesian Inference Using Gibbs Sampling) là một trong những phát triển quan trọng nhất trong thế kỉ 21 trong khoa học thống kê học, vì chương trình này giải được rất nhiều vấn đề mà trước đây người ta chỉ bàn luận trên lí thuyết. WinBUGS dùng phương pháp tính toán (Markov chain Monte Carlo – MCMC) để thực hiện những tính toán rất phức tạp mà trước đây không ai giải được. Có thể nói không ngoa rằng nếu không có WinBUGS thì các mô hình phân tích theo phương pháp Bayes không có diện mạo như ngày nay.

Do đó, chúng ta cần phải học WinBUGS. Bài này chỉ dẫn cách cài đặt và sử dụng WinBUGS cho phân tích thống kê theo mô hình Bayes. Tôi cũng minh họa hai phân tích đơn giản qua WinBUGS. Trong bất cứ vấn đề kĩ thuật nào, chỉ có thể thành thạo khi thực hành chứ không chỉ đọc. Do đó, bạn đọc có thể làm theo những qui trình từng bước được mô tả trong bài để có thêm kinh nghiệm.

Vài năm gần đây, đã có khá nhiều tài liệu chỉ dẫn cách sử dụng WinBUGS. Tuy nhiên, kinh nghiệm của tôi cho thấy những tài liệu này không dễ đọc chút nào! Thậm chí, có tài liệu viết khá cụ thể, nhưng vẫn làm cho những người mới làm quen với WinBUGS cảm thấy lúng túng và rối rắm. Do đó, tôi soạn bài này để giúp cho những người mới sử dụng WinBUGS lần đầu, và tôi sẽ cố gắng mô tả các qui trình sử dụng một cách chi tiết và cụ thể.

**19.1 Cài đặt WinBUGS**

WinBUGS là chương trình được chạy từ máy tính cá nhân. Do đó, cần phải tải các nhu liệu từ website của winBUGS để cài đặt trong máy.

* Truy cập website của nhóm thống kê Hội đồng nghiên cứu y khoa Anh:

<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>

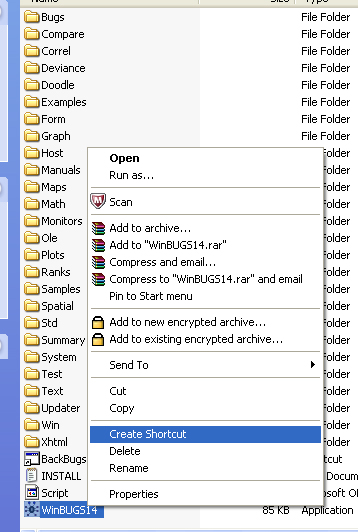
* Tải về máy các nhu liệu sau đây: [WinBUGS14.exe](http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/WinBUGS14.exe) và [keys](http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/WinBUGS14_immortality_key.txt) (WinBUGS14\_immortality\_key.txt).
* Khởi động chương trình WinBUGS14.exe, và nhu liệu này sẽ tự động cài đặt WinBUGS trong máy. Sau khi cài đặt xong, máy tính sẽ có một folder mới:

C:\Program Files\WinBUGS14

và trong đó có nhiều folder nhỏ khác, nhưng có lẽ folder liên quan nhất là

C:\Program Files\WinBUGS14\Bugs

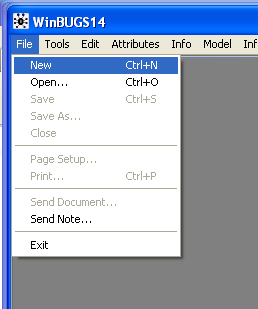
* Nên tạo ra một icon trong desktop để sử dụng WinBUGS về lâu về dài. Cách tạo icon có thể làm theo qui trình như sau:
  + Vào folder C:\Program Files\WinBUGS14
  + Dùng chuột bấm nút phải nhấn vào file có WinBUgS icon và chọn “**Create Shortcut**”



**Hình 1. Tạo ra shortcut cho WinBUGS**

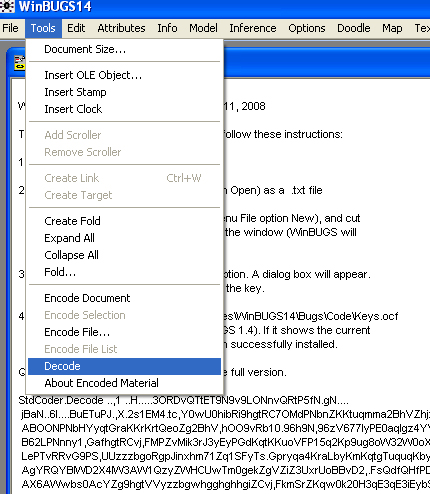
Nhưng trước khi sử dụng WinBUGS, chúng ta cần phải báo cho WinBUGS biết các mã số trong một file gọi là chìa khóa: WinBUGS14\_immortality\_key.txt. Cách cài đặt khóa trong WinBUGS như sau:

* Khởi động WinBUGS (từ desktop)
* Click vào tab **File** và chọn **New** để mở một window mới hoàn toàn



**Hình 2. Khởi động WinBUGS**

* Mở file WinBUGS14\_immortality\_key.txt (dùng notepad), và copy toàn bộ (Control-A)
* Chuyển sang WinBUGS và dán vào window trống. Từ tab **Tools**, chọn **Decode**:



**Hình 3. Cài mã khóa (key)**

Một box mới sẽ xuất hiện. Click vào **Decode All**. Đến đây thì chúng ta đã xong công đoạn cài đặt WinBUGS, và có thể sử dụng chương trình này một cách … lâu dài.

**19.2 Phân tích dữ liệu với WinBUGS**

So với các chương trình máy tính về phân tích thống kê khác (như SAS, R, SPSS, v.v.), cách phân tích trên WinBUGS tương đối khó hơn. Tuy nhiên, các mô hình phân tích Bayes chỉ có thể thực hiện một cách hữu hiệu qua WinBUGS, chứ không thể qua một chương trình khác. Ngay cả các chương trình như SAS và R có vài phần để phân tích Bayes nhưng vẫn còn rất “nhiêu khê” và rất chậm. Ngược lại WinBUGS có thể ứng dụng cho tất cả mô hình Bayes và rất hữu hiệu. Qui trình phân tích bằng WinBUGS có thể tóm lược qua 10 bước sau đây:

**Tóm lược qui trình chạy một mô hình trên WinBUGS**

|  |  |  |
| --- | --- | --- |
| **Bước** | **Mô tả** | **WinBUGS** |
| 1 | Mô tả mô hình (model specification) | **File 🡪 New** |
| 2 | Kiểm tra lệnh và mô hình (model checking) | **File 🡪 Specification = Specification Tool 🡪**  **(check model)** |
| 3 | Nhập dữ liệu (data loading) | **Specification Tool 🡪 load data** |
| 4 | Chạy thử (compile) | **Specification Tool 🡪 compile** |
| 5 | Giá trị khởi đầu (initial values) | **Specification Tool 🡪 gen inits** |
| 6 | Thông số (specification of parameters) | **Inference 🡪 Samples = Sample Monitor Tool** |
| 7 | Mô phỏng (simulation) | **Model 🡪 Update = Update Tool**  **(updates: số lần mô phỏng**  **update)** |
| 8 | Kết quả (results) | **Sample Monitor Tool 🡪 node 🡪**  **(stats, density)** |
| 9 | Chẩn đoán (diagnosis) | **Sample Monitor Tool 🡪 history** |

Chúng ta có thể bắt đầu bằng một ví dụ đơn giản và cụ thể. Một nghiên cứu về hiệu quả của thuốc trên 25 bệnh nhân cho thấy có 11 bệnh nhân đáp ứng (tức thuốc có hiệu quả). Chúng ta muốn biết khoảng tin cậy 95% của tỉ lệ hiệu quả là bao nhiêu. Chúng ta sẽ ứng dụng phương pháp Bayes để phân tích. Phương pháp Bayes có 3 phần: thông tin tiền định (prior information), dữ liệu thực tế (likelihood), và thông tin hậu định (posterior information).

Dữ liệu thực tế là số liệu chúng ta có được: *n* = 25 và *x* = 11. Có thể chọn luật phân bố nhị phân (binomial distribution) để mô tả kết quả này. Luật phân phối nhị phân cũng có 2 thông số: số cỡ mẫu *n* và tỉ lệ *p.* Nói cách khác:

*x ~* Bin(*p*, *n*)

Thông tin tiền định: chúng ta không có một “thiên kiến” nào về hiệu quả của thuốc, nên sẽ chọn luật phân bố beta để mô tả thông tin tiền định. Phân bố beta có 2 thông số, tạm gọi là *a* và *b*:

*p ~* Beta(*a*, *b*)

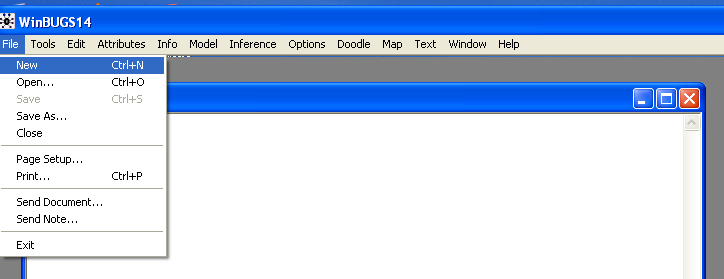
Chúng ta cho thông số *a = b =* 1.

Để phân tích mô hình trên với WinBUGS, chúng ta sẽ làm theo qui trình từng bước sau đây:

**Bước 1**. **Mô tả mô hình phân tích**

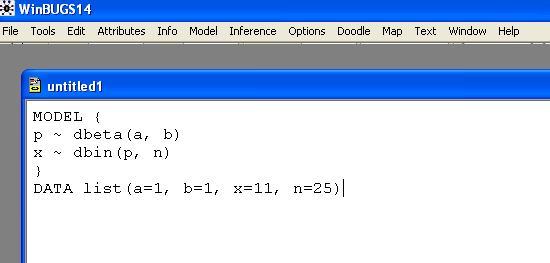
Bước đầu tiên trong WinBUGS là mô tả mô hình bằng cách viết ra lệnh, và lưu trữ một một file. Bước này có thể thực hiện theo qui trình sau đây:

* Khởi động WinBUGS, nhấn chuột vào tab **File** (trên WinBUGS menu) và chọn **New**:



**Hình 4. Khởi động và tạo ra file mới**

* Trong bước này, chúng ta phải viết ra mô hình trong window trống như dưới đây:

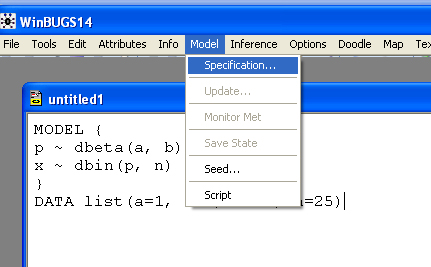


**Hình 5. Mô tả mô hình bằng các lệnh**

Bất cứ một chương trình nào cũng có 2 phần: mô hình (MODEL) và dữ liệu (DATA). Trong phần mô hình, chúng ta mô tả mô hình phân tích. Cách viết mô hình trong WinBUGS rất giống như cách viết thống kê, nên cũng tiện cho việc giảng dạy và bàn luận. Trong phần dữ liệu, chúng ta cung cấp dữ liệu để mô hình có thể sử dụng.

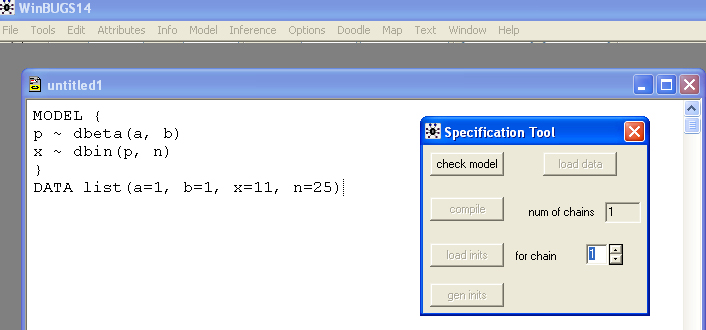
**Bước 2. Kiểm tra “ngữ pháp” (check model)**

Ngữ pháp rất quan trọng. Nếu ngữ pháp sai, WinBUGS sẽ không chạy. Do đó, trước khi chạy WinBUGS chúng ta lúc nào cũng phải kiểm tra các lệnh và hàm cẩn thận. Để kiểm tra, nhấn chuột vào tab **Model** (trên menu của **WinBUGS)** và chọn **Specification**:



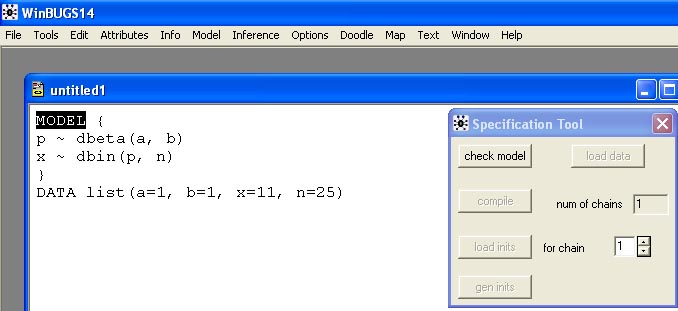
**Hình 6. Chọn specification từ Model**

Sau đó, một dialogue box có tên là **Specification Tool** sẽ xuất hiện như sau:



**Hình 7. Specification Tool**

Nhấn chuột và tô đậm (highlight) **MODEL,** di dời chuột sang **Specification Tool** và nhấn **check model**;



**Hình 8. Kiểm tra ngữ pháp của mô hình**

và chú ý đến thông điệp phía dưới (bên trái) của window:



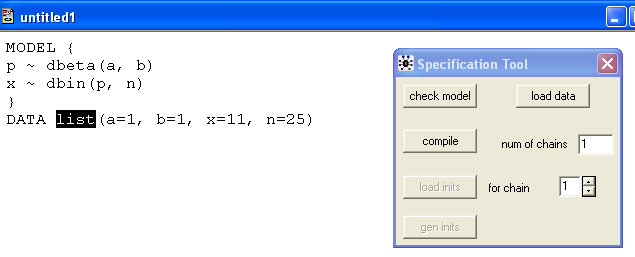
**Hình 9. Thông điệp sau khi kiểm tra ngữ pháp của mô hình**

Trong trường hợp này, chúng ta có thông điệp cho biết ngữ pháp của mô hình là đúng (model is syntactically correct). Nếu có vấn đề, chúng ta cần phải xem xét cẩn thận các lệnh một lần nữa và kiểm tra cho đến khi nào có thông điệp này mới có thể tiến đến bước kế tiếp.

**Bước 3. Nhập dữ liệu (load data)**

Trong bước này, chúng ta vẫn dùng **Specification Tool** để cho WinBUGS biết số liệu thực tế. Chúng ta làm theo qui trình:

* Nhấn chuột vào và bôi đậm **list** (trong phần DATA) như dưới đây:



**Hình 10. Nhập dữ liệu**

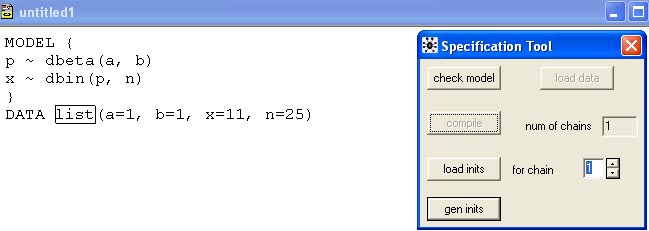
* Di chuyển chuột sang **Specification Tool**, nhấn vào nút **load data**. Một thông điệp sẽ hiện phía dưới (phần trái) cho biết dữ liệu đã được nhập (data loaded):



**Hình 11. Thông điệp sau khi nhập dữ liệu**

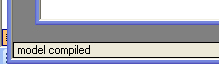
**Bước 4. Chạy thử mô hình (compile)**

Trong bước này, chúng ta báo cho WinBUGS biết có thể sử dụng mô hình cho dữ liệu mới nhập. Trong **Specification Tool**, chọn **compile**.



**Hình 12. Chạy thử (compile)**

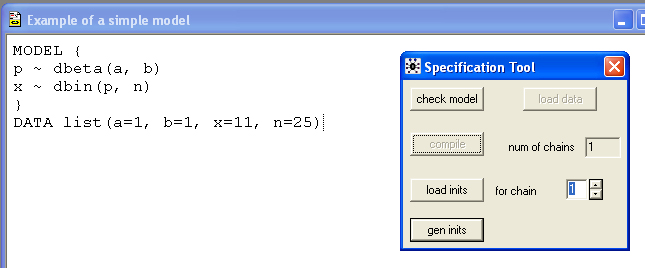
Một thông điệp sẽ hiện phía dưới (phần trái) cho biết dữ liệu đã được thử (model compiled):



**Hình 13. Thông điệp sau khi chạy thử**

**Bước 5. Cho giá trị khởi đầu (generating initials – gen inits)**

Vì yêu cầu của phương pháp mô phỏng Monte Carlo, nên các mô hình Bayes đều cần có một giá trị khởi đầu, gọi là initial values. Cách tốt nhất là để cho WinBUGS tự động cho các giá trị này, bằng cách dùng **Specification Tool**, và chọn **gen inits:**



**Hình 14. Dữ liệu khởi đầu**

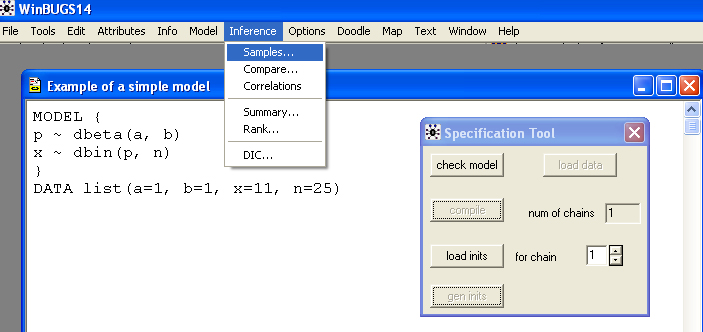
Sau khi chọn **gen inits**, một thông điệp phía dưới (bên trái) sẽ xuất hiện:

initial values generated, model initialized

**Bước 6. Mô tả thông số (parameter)**

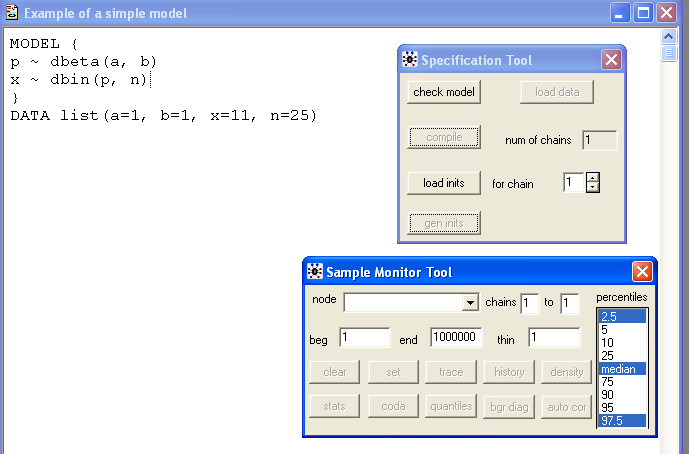
Một mô hình có ít nhất là một thông số (parameter). Chúng ta cần phải cho WinBUGS biết thông số cần ước tính là gì. Trong trường hợp trên, thông số chúng ta cần biết là *p* – tỉ lệ bệnh nhân đáp ứng thuốc.

Để thực hiện bước này, nhất chuột vào tab **Inference** trên menu của WinBUGS và chọn chọn **Sample**:



**Hình 15. Chọn Sample từ Inference**

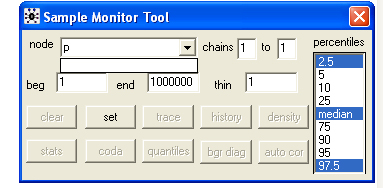
Một dialogue mới có tên là **Sample Monotor Tool** sẽ xuất hiện như hình dưới đây:



**Hình 16. Sample Monitor Tool**

Trong **Sample Monotor Tool** chúng ta có thể cho WinBUGS biết những thông số chúng ta cần phân tích. Như đề cập trên, chúng ta muốn biết khoảng tin cậy 95% của *p* (là tên thông số trong MODEL), nên chọn bách phân vị (**percentiles**) 2.5, trung vị (median) và 97.5 bách phân vị.

Đưa chuột vào box có tên **node** trong **Sample Monotor Tool** và gõ **p**:

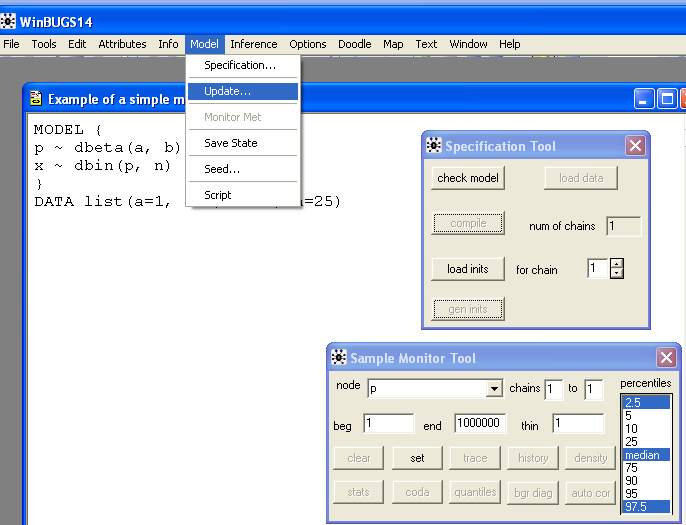


**Hình 17. Chọn thông số cho phân tích**

Sau đó, nút **set** sẽ được kích hoạt (activited). Nhấn chuột vào **set** cho đến khi tất cả các thông số đã được liệt kê. Cũng có thể thấy các thông số (nếu cần) bằng cách nhấn chuột vào nút **node**.

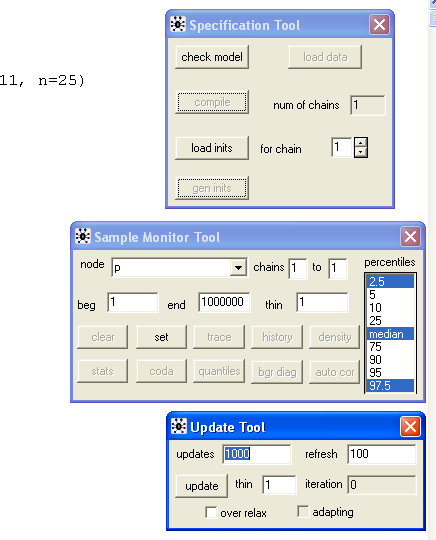
**Bước 7. Mô phỏng (generating simulated values)**

Sau khi mô hình đã được mô tả, thông số đã được chọn, bước kế tiếp là mô phỏng. Để thực hiện bước này, nhấn chuột vào tab **Model** trên menu của WinBUGS và chọn **Update**:



**Hình 18. Chọn Model và Update**

Sau đó một dialogue box có tên là **Update Tool** sẽ xuất hiện:



**Hình 19. Update Tool**

Trong **Update Tool** có phần update, nơi chúng ta cần cho WinBUGS biết số lần mô phỏng là bao nhiêu. Chẳng hạn như tôi cho số lần mô phỏng là 5000:

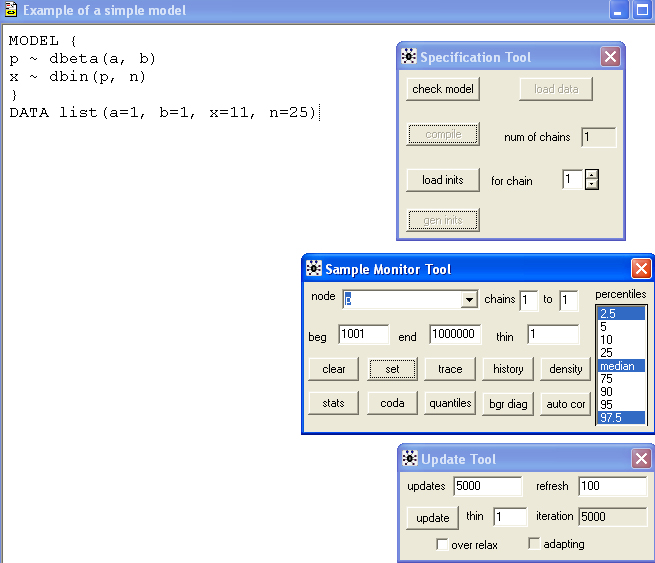


**Hình 20. Update**

Nhấn vào nút **update**, WinBUGS sẽ chạy (tùy theo mô hình) vài giây hay vài phút. Sau khi bước này xonng, chúng ta đã có kết quả, và bước kế tiếp là xem xét kết quả.

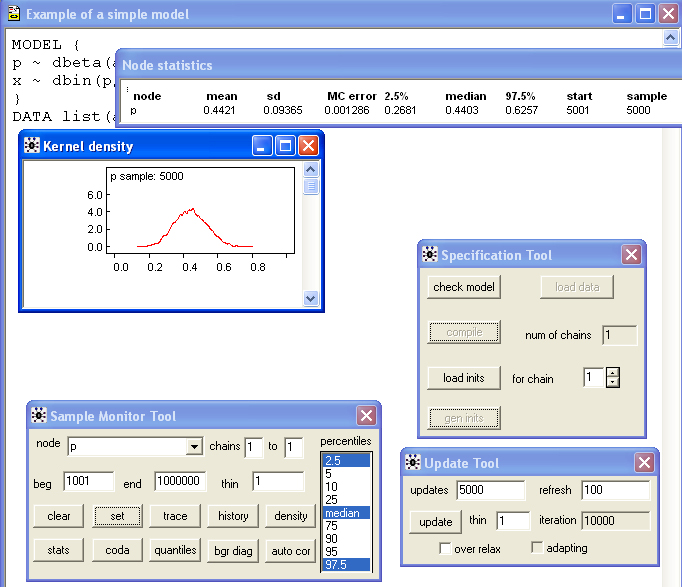
**Bước 8. Xem xét kết quả thông tin hậu định (posterior information)**

Trong bước này, chúng ta sẽ xem kết quả phân tích. Để thực hiện bước này, chúng ta quay lại **Sample Monotor Tool** vàcung cấp 1001 vào field **beg** (có nghĩa là yêu cầu WinBUGS bỏ 1000 mô phỏng đầu vốn chỉ để lấy các giá trị initials). Cần nhắc lại rằng nhớ chọn thông số (trường hợp này là p) từ field **node**:



**Hình 21. Cung cấp tên thông số cho phân tích**

Sau khi chọn p, các nút trong **Sample Monitor Tool** sẽ kích hoạt, và chúng ta có thể xem qua kết quả. Nhấn chuột vào **stats** (để có các thông số khoảng tin cậy 95%) và **density** để có biểu đồ phân bố của *p*. Kết quả là:

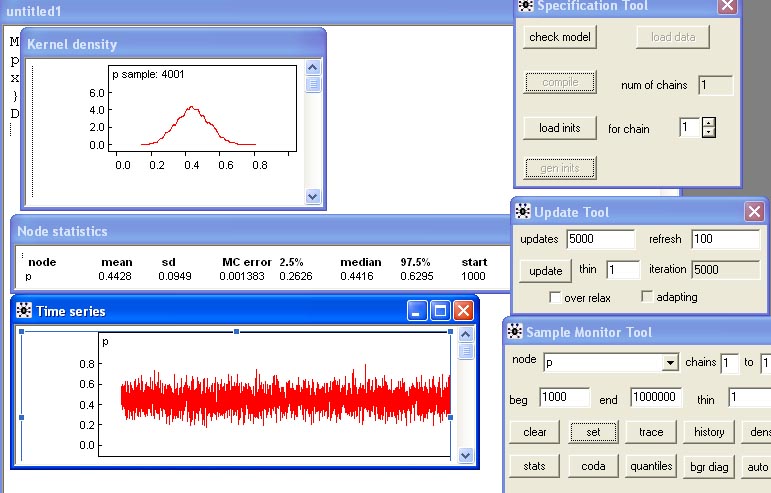


**Hình 22. Kết quả phân tích**

Trong phần stats, chúng ta thấy tỉ lệ trung bình là 44.2%, với khoảng tin cậy 95% dao động trong khoảng 26.8% đến 62.6%. Biểu đồ phân bố thể hiện mô phỏng 5000 lần cũng phù hợp với các kết quả trên.

**Bước 9. Chẩn đoán (diagnosis)**

Trước khi chấp nhận kết quả trên, chúng ta cần phải kiểm tra phần kĩ thuật tính toán có đạt yêu cầu hay chưa. Đạt yêu cầu khi các bước tính toán theo chu kì đã “converged” (tức ổn định). Để kiểm tra kĩ thuật, chúng ta sử dụng **Sample Monitor Tool** và nhấn chuột vào nút **history**. Kết quả là:



**Hình 23. Chẩn đoán**

Kết quả trình bày trong biểu đồ có tên **Time Series** cho thấy mô hình và tính toán ổn định.

**19.3 Phân tích 2 nhóm**

Số liệu về tử vong trong 2 nhóm phẫu thuật như sau: Trong số 18 bệnh nhân được phẫu thuật bằng phương pháp 1, có 3 người tử vong; trong nhóm 23 người được phẫu thuật bằng phương pháp mới có 2 người tử vong.

|  |  |  |
| --- | --- | --- |
| **Nhóm** | **Số bệnh nhân (*n*)** | **Số tử vong (*x*)** |
| Nhóm 1 | 18 | 3 |
| Nhóm 2 | 23 | 2 |

Câu hỏi nghiên cứu là có khác biệt gì giữa hai nhóm. Nêu gọi *P*1 và *P*2 là tỉ lệ tử vong của nhóm 1 và 2, câu hỏi nghiên cứu có thể đặt ra theo ba cách diễn đạt:

* Gọi *D* = *P*1 – *P*2, câu hỏi là *D* > 0?
* Gọi *RR* = *P*1 / *P*2 (tỉ số nguy cơ), câu hỏi cũng có thể là *RR* > 1?
* Nếu đặt *OR* = *P*1/(1–*P*1) / *P*2/(1–*P*2) (tỉ số odds), câu hỏi cũng có thể là *OR* > 1?

Chúng ta có thể phân tích 3 câu hỏi trên bằng WinBUGS như sau:

MODEL {

# prior information

p1 ~ dbeta(0.5, 0.5)

p2 ~ dbeta(0.5, 0.5)

# data or likelihood

x1 ~ dbin(p1, n1)

x2 ~ dbin(p2, n2)

# effect size

diff = p1-p2

rr = p1/p2

or = (p1/(1-p1)) / (p2/(1-p2))

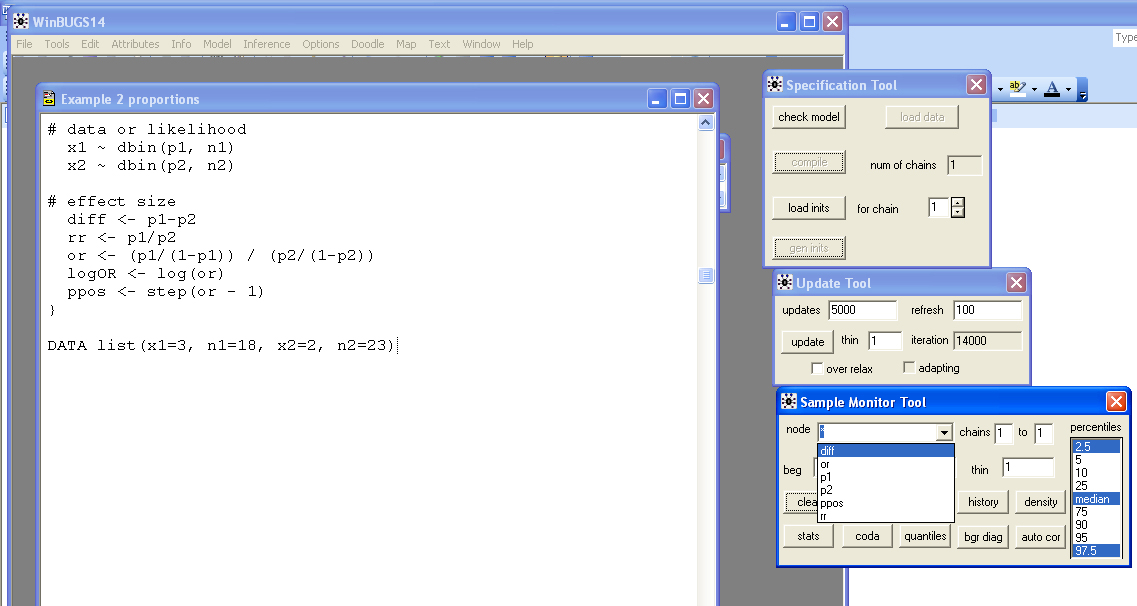
logOR = log(or)

ppos = step(OR - 1)

}

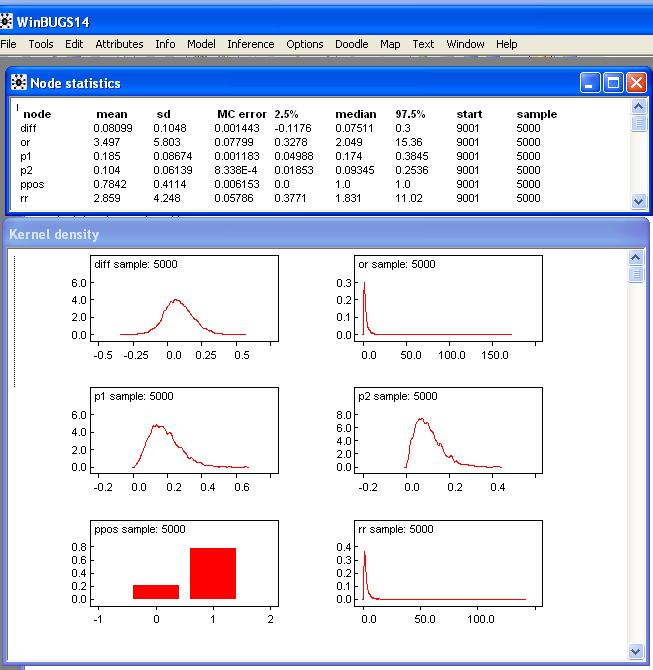
DATA list(x1=3, n1=18, x2=2, n2=23)

Chúng ta vẫn tiến hành những bước kiểm tra “ngữ pháp” (check model), cung cấp dữ liệu (load data), chạy thử mô hình (compile), cho giá trị khởi đầu (generating initials – gen inits), và tính toàn thông tin hậu định (posterior information). Vì mô hình này có nhiều thông số, nên trong field **node** của **Sample Monitor Tool** chúng ta phải cung cấp những thông số như p1, p2, or, rr, và ppos (xác suất OR > 1). Chú ý dấu “\*” trong node có nghĩa là tất cả các thông số:



**Hình 24. Sample Monitor Tool: mô tả thông số cho phân tích**

Kết quả sẽ là:



**Hình 25. Kết quả phân tích 2 tỉ lệ**

\*\*\*

Trên đây là chỉ dẫn đơn giản từng bước một để sử dụng WinBUGS cho các phân tích theo mô hình Bayes. Hi vọng rằng các chỉ dẫn trên đây đã và sẽ giúp cho các bạn thực hiện được những phân tích đơn giản và trung bình.

Để sử dụng WinBUGS một cách hữu hiệu, người sử dụng cần phải làm quen với “ngữ pháp” của chương trình máy tính. Bài này không có mục đích mô tả các lệnh đó. Bạn đọc có thể truy cập trang web www.mrc-bsu.cam.ac.uk để tải về các tài liệu chỉ dẫn và cẩm nang để tham khảo.

Phương pháp phân tích dữ liệu theo trường phái Bayes được đánh giá là phương pháp chính trong thế kỉ 21. Vì phương pháp tần số (frequentist methods) có nhiều khiếm khuyết, và càng ngày càng có nhiều người nhận ra sự hữu ích của phương pháp Bayes, nên hiện nay WinBUGS càng ngày càng được nhắc đến thường xuyên. Ở Việt Nam phương pháp Bayes chưa được quan tâm vì thiếu tài liệu và các đại học cũng chưa có những chương trình giảng dạy có hệ thống. Chương này mang tính giới thiệu này hi vọng sẽ góp một phần nhỏ để giới thiệu phương pháp Bayes đến các bạn quan tâm.