

Phân tích số liệu và tạo biểu đồ bằng



hướng dẫn thực hành

Mục lục

1	Lời nói đầu
2	Giới thiệu ngôn ngữ R
2.1	R là gì ?
2.2	Tải và cài đặt R vào máy tính
2.3	Package cho các phân tích đặc biệt
2.4	Khởi động và ngưng chạy R
2.5	“Văn phạm” ngôn ngữ R
2.6	Cách đặt tên trong R
2.7	Hỗ trợ trong R
2.8	Môi trường vận hành
3	Nhập dữ liệu
3.1	Nhập số liệu trực tiếp: <code>c()</code>
3.2	Nhập số liệu trực tiếp: <code>edit(data.frame())</code>
3.3	Nhập số liệu từ một textfile: <code>read.table()</code>
3.4	Nhập số liệu từ Excel: <code>read.csv</code>
3.5	Nhập số liệu từ SPSS: <code>read.spss</code>
3.6	Tìm thông tin cơ bản về dữ liệu
4	Biên tập dữ liệu
4.1	Kiểm tra số liệu trống không: <code>na.omit()</code>
4.2	Tách rời dữ liệu: <code>subset</code>
4.3	Chiết số liệu từ một data.frame
4.4	Nhập hai data.frame thành một: <code>merge</code>
4.5	Mã hóa số liệu (data coding)
4.5.1	Mã hoá bằng hàm <code>replace</code>
4.5.2	Đổi một biến liên tục thành biến rời rạc
4.6	Chia một biến liên tục thành nhóm: <code>cut</code>
4.7	Tập hợp số liệu bằng <code>cut2</code> (Hmisc)

5	Sử R cho các phép tính đơn giản và ma trận
5.1	Tính toán đơn giản
5.2	Số liệu về ngày tháng
5.3	Tạo dãy số bằng seq, rep và gl
5.4	Sử dụng R cho các phép tính ma trận
5.4.1	Chiết phần tử từ ma trận
5.4.2	Tính toán với ma trận
6	Tính toán xác suất và mô phỏng (simulation)
6.1	Tính toán đơn giản
6.1.1	Phép hoán vị (permutation)
6.1.2	Tổ hợp (combination)
6.2	Biến số ngẫu nhiên và hàm phân phối
6.3	Các hàm phân phối xác suất (probability distribution function)
6.3.1	Hàm phân phối nhị phân (Binomial distribution)
6.3.2	Hàm phân phối Poisson (Poisson distribution)
6.3.3	Hàm phân phối chuẩn (Normal distribution)
6.3.4	Hàm phân phối chuẩn chuẩn hóa (Standardized Normal distribution)
6.3.5	Hàm phân phối t, F và χ^2
6.4	Mô phỏng (simulation)
6.4.1	Mô phỏng phân phối nhị phân
6.4.2	Mô phỏng phân phối Poisson
6.4.3	Mô phỏng phân phối χ^2 , t, F, gamma, beta, Weibull, Cauchy
6.5	Chọn mẫu ngẫu nhiên (random sampling)
7	Kiểm định giả thiết thống kê và ý nghĩa trị số P
7.1	Trị số P
7.2	Giả thiết khoa học và phản nghiệm
7.3	Ý nghĩa của trị số P qua mô phỏng
7.4	Vấn đề logic của trị số P
7.5	Vấn đề kiểm định nhiều giả thiết (multiple tests of hypothesis)
8	Phân tích số liệu bằng biểu đồ
8.1	Môi trường và thiết kế biểu đồ
8.1.1	Nhiều biểu đồ cho một cửa sổ (windows)
8.1.2	Đặt tên cho trục tung và trục hoành
8.1.3	Cho giới hạn của trục tung và trục hoành
8.1.4	Thẻ loại và đường biểu diễn
8.1.5	Màu sắc, khung, và kí hiệu
8.1.6	Ghi chú (legend)
8.1.7	Viết chữ trong biểu đồ

8.2	Số liệu cho phân tích biểu đồ
8.3	Biểu đồ cho một biến số rời rạc (discrete variable): <code>barplot</code>
8.4.	Biểu đồ cho hai biến số rời rạc (discrete variable): <code>barplot</code>
8.5	Biểu đồ hình tròn
8.6	Biểu đồ cho một biến số liên tục: <code>stripchart</code> và <code>hist</code>
8.6.1	<code>Stripchart</code>
8.6.2	Histogram
8.6.3	Biểu đồ hộp (<code>boxplot</code>)
8.6.4	Biểu đồ thanh (<code>barchart</code>)
8.6.5	Biểu đồ điểm (<code>dotchart</code>)
8.7	Phân tích biểu đồ cho hai biến liên tục
8.7.1	Biểu đồ tán xạ (<code>scatter plot</code>)
8.8	Phân tích Biểu đồ cho nhiều biến: <code>pairs</code>
8.9	Một số biểu đồ “đa năng”
8.9.1	Biểu đồ tán xạ và hình hộp
8.9.2	Biểu đồ tán xạ với kích thước biến thứ ba
8.9.3	Biểu đồ thanh và xác suất tích lũy
8.9.4	Biểu đồ hình đồng hồ (<code>clock plot</code>)
8.9.5	Biểu đồ với sai số chuẩn (standard error)
8.9.6	Biểu đồ vòng (<code>contour plot</code>)
8.9.10	Biểu đồ với kí hiệu toán

9 Phân tích thống kê mô tả

9.0	Khái niệm về tổng thể (population) và mẫu (sample)
9.1	Thống kê mô tả: <code>summary</code>
9.2	Kiểm định xem một biến có phải phân phối chuẩn
9.3	Thống kê mô tả theo từng nhóm
9.4	Kiểm định t (<code>t.test</code>)
9.4.1	Kiểm định t một mẫu
9.4.2	Kiểm định t hai mẫu
9.5	So sánh phương sai (<code>var.test</code>)
9.6	Kiểm định Wilcoxon cho hai mẫu (<code>wilcox.test</code>)
9.7	Kiểm định t cho các biến số theo cặp (paired t-test, <code>t.test</code>)
9.8	Kiểm định Wilcoxon cho các biến số theo cặp (<code>wilcox.test</code>)
9.9	Tần số (frequency)
9.10	Kiểm định tỉ lệ (proportion test, <code>prop.test</code> , <code>binom.test</code>)
9.11	So sánh hai tỉ lệ (<code>prop.test</code> , <code>binom.test</code>)
9.12	So sánh nhiều tỉ lệ (<code>prop.test</code> , <code>chisq.test</code>)
9.12.1	Kiểm định Chi bình phương
9.12.2	Kiểm định Fisher

10	Phân tích hồi qui tuyến tính (regression analysis)
10.1	Hệ số tương quan
10.1.1	Hệ số tương quan Pearson
10.1.2	Hệ số tương quan Spearman
10.1.3	Hệ số tương quan Kendall
10.2	Mô hình của hồi qui tuyến tính đơn giản
10.2.1	Vài dòng lí thuyết
10.2.2	Phân tích hồi qui tuyến tính đơn giản bằng R
10.2.3	Giả định của phân tích hồi qui tuyến tính
10.2.4	Mô hình tiên đoán
10.3	Mô hình hồi qui tuyến tính đa biến (multiple linear regression)
10.4	Phân tích hồi qui đa thức (Polynomial regression analysis)
10.5	Xây dựng mô hình tuyến tính từ nhiều biến
10.6	Xây dựng mô hình tuyến tính bằng Bayesian Model Average (BMA)
11	Phân tích phương sai (analysis of variance)
11.1	Phân tích phương sai đơn giản (one-way analysis of variance - ANOVA)
11.1.1	Mô hình phân tích phương sai
11.1.2	Phân tích phương sai đơn giản với R
11.2	So sánh nhiều nhóm (multiple comparisons) và điều chỉnh trị số p
11.2.1	So sánh nhiều nhóm bằng phương pháp Tukey
11.2.2	Phân tích bằng biểu đồ
11.3	Phân tích bằng phương pháp phi tham số
11.4	Phân tích phương sai hai chiều (two-way analysis of variance - ANOVA)
11.4.1	Phân tích phương sai hai chiều với R
11.5	Phân tích hiệp biến (analysis of covariance - ANCOVA)
11.5.1	Mô hình phân tích hiệp biến
11.5.2	Phân tích bằng R
11.6	Phân tích phương sai cho thí nghiệm giai thừa (factorial experiment)
11.7	Phân tích phương sai cho thí nghiệm hình vuông Latin (Latin square experiment)
11.8	Phân tích phương sai cho thí nghiệm giao chéo (cross-over experiment)
11.9	Phân tích phương sai cho thí nghiệm tái đo lường (repeated measure experiment)
12	Phân tích hồi qui logistic (logistic regression analysis)
12.1	Mô hình hồi qui logistic

12.2	Phân tích hồi qui logistic bằng R
12.3	Ước tính xác suất bằng R
12.4	Phân tích hồi qui logistic từ số liệu giản lược bằng R
12.5	Phân tích hồi qui logistic đa biến và chọn mô hình
12.6	Chọn mô hình hồi qui logistic bằng Bayesian Model Average
12.7	Số liệu dùng cho phân tích
13	Phân tích biến cố (survival analysis)
13.1	Mô hình phân tích số liệu mang tính thời gian
13.2	Ước tính Kaplan-Meier bằng R
13.3	So sánh hai hàm xác suất tích lũy: kiểm định log-rank (log-rank test)
13.4	Kiểm định log-rank bằng R
13.5	Mô hình Cox (hay Cox's proportional hazards model)
13.6	Xây dựng mô hình Cox bằng Bayesian Model Average (BMA)
14	Phân tích tổng hợp (meta-analysis)
14.1	Nhu cầu cho phân tích tổng hợp
14.2	Ảnh hưởng ngẫu nhiên và ảnh hưởng bất biến (Fixed-effects và Random-effects)
14.3	Quy trình của một phân tích tổng hợp
14.4	Phân tích tổng hợp ảnh hưởng bất biến cho một tiêu chí liên tục (Fixed-effects meta-analysis for a continuous outcome)
14.4.1	Phân tích tổng hợp bằng tính toán “thủ công”
14.4.2	Phân tích tổng hợp bằng R
14.5	Phân tích tổng hợp ảnh hưởng bất biến cho một tiêu chí nhị phân (Fixed-effects meta-analysis for a dichotomous outcome)
14.5.1	Mô hình phân tích
14.5.2	Phân tích bằng R
15	Ước tính cỡ mẫu (estimation of sample size)
15.1	Khái niệm về “power”
15.2	Thử nghiệm giả thiết thống kê và chẩn đoán bệnh
15.3	Số liệu để ước tính cỡ mẫu
15.4	Ước tính cỡ mẫu
15.4.1	Ước tính cỡ mẫu cho một chỉ số trung bình
15.4.2	Ước tính cỡ mẫu cho so sánh hai số trung bình
15.4.3	Ước tính cỡ mẫu cho phân tích phương sai
15.4.4	Ước tính cỡ mẫu cho ước tính một tỉ lệ
15.4.5	Ước tính cỡ mẫu cho so sánh hai tỉ lệ
16	Phụ lục 1: Lập trình và viết hàm bằng ngôn ngữ R

17	Phụ lục 2: Một số lệnh thông dụng trong R
18	Phụ lục 3: Thuật ngữ dùng trong sách
19	Lời bạt (tài liệu tham khảo và đọc thêm)