HOÀNG TRỌNG - CHU NGUYỄN MỘNG NGỌC

PHÂN TÍCH DỮ LIỆU NGHIÊN CỨU VỚI

SPSS

TRƯỜNG ĐẠI HỌC KINH TẾ TP HỒ CHÍ MINH
NHÀ XUẤT BẢN HỒNG ĐỨC

ĐẠI HỌC KINH TẾ TP HỒ CHÍ MINH

HOÀNG TRỌNG – CHU NGUYỄN MỘNG NGỌC

PHÂN TÍCH DỮ LIỆU NGHIÊN CỨU VỚI SPSS

(dùng với SPSS các phiên bản 11.5, 13, 14, 15, 16)

Tập 1

TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI
TRUNG TÂM THÔNG TIN THƯ VIỆN
04 04
05619

NHÀ XUẤT BẢN HỒNG ĐỨC
NĂM 2008

Quyển sách dành cho những bạn đang làm
đề tài nghiên cứu khoa học, khóa luận hay luận văn tốt nghiệp

Hãy đọc kỹ lời nói đầu và xem mục lục
trước khi bạn đi vào nội dung của quyển sách

ĐỊA CHỈ TẢI FILE THỰC HÀNH

Để lấy các file dữ liệu thực hành cùng với sách Phân Tích Dữ Liệu Nghiên Cứu với SPSS, bạn vào một trong các trang web sau để tải file xuống:

Trang web của Khoa Toán – Thống Kê, ĐH Kinh Tế TPHCM (chọn mục Sách và Tài Liệu):
http://sems.ueh.edu.vn/danh-muc/tai-nguyen-113

Trang web cao học kinh tế:
http://caohockinhte.info/forum/showthread.php?t=3680

Trang web của công ty tư vấn:
http://www.thepathfinder.vn/index.php?option=thongtinnghiencuu&task=view&id=14

Nếu có trục trặc xin vui lòng email đến địa chỉ:

hoangtrong00@gmail.com

LỜI NÓI ĐẦU

Quyển sách Phân tích Dữ Liệu Nghiên Cứu Với SPSS (Nhà Xuất Bản Thống Kê, 2005) đã ra đời được ba năm. Tác giả đã nhận được nhiều ý kiến góp ý, yêu cầu bổ sung của nhiều bạn đọc. Ý kiến của bạn đọc xoay quanh các vấn đề chính.

Một là bổ sung nội dung, ví dụ như: Sử dụng Custom Tables (một số phiên bản SPSS sau này do bạn đọc cài đặt hay do đĩa nguồn cài đặt thiếu, các bạn chỉ có Custom Tables mà không có Basic Tables hay General Tables _); Vẽ đồ thị trong Excel (vẽ đồ thị trong SPSS không quen và ít tiện lợi như trong Excel); Phân tích phương sai hai yếu tố nguyên nhân; Tạo các biến giả và sử dụng biến giả trong hồi qui đối với các biến độc lập định tính, chẩn đoán và tuyến tính hóa các biến nguyên nhân; Lập các bản đồ nhận thức (bản đồ định vị); Gia trọng các quan sát, Ghép trộn dữ liệu... Chúng tôi đã cố gắng bổ sung theo những yêu cầu này. Còn một vài nội dung khác, hiện nay do số lượng bạn đọc có nhu cầu sử dụng còn ít chúng tôi sẽ bổ sung trong lần tái bản sau.

Hai là việc sử dụng quyển sách này với các phiên bản mới hơn của SPSS như 13.0, 15.0 và 16.0. Về phiên bản của SPSS, sau khi khảo sát và sử dụng thử các phiên bản SPSS 13, 15, 16, chúng tôi nhận thấy các giao diện, các lệnh thực hiện hoàn toàn tương tự nhau. Chúng tôi cũng vẫn sử dụng Phiên bản 11.5 và 13.0 vì sự gọn nhẹ, ít lỗi của hai phiên bản chuẩn này. Các phiên bản sau có bổ sung một vài tiện ích mới nhưng hiếm khi được sử dụng đối với người sử dụng thông thường. Bạn đọc yên tâm sử dụng quyển sách này với bất kỳ phiên bản của SPSS từ 11.5 đến 16.0.

Trong lần xuất bản này, chúng tôi tách thành hai tập. Tập 1 phục vụ cho nhu cầu xử lý và phân tích căn bản của các sinh viên bậc cử nhân đang học các môn học liên quan như Thống Kê, Kinh Tế Lượng, Phương Pháp Nghiên Cứu, Phân tích Dữ Liệu. Tập 2 dành cho sinh viên học chuyên ngành muốn đi sâu vào phân tích dữ liệu, học viên cao học , người phân tích dữ liệu chuyên nghiệp.

Khác với lần xuất bản trước, lần xuất bản này chúng tôi không kèm đĩa chứa dữ liệu mẫu với sách vì dung lượng các file này khá nhỏ. Mặt khác nhiều bạn đọc đã yêu cầu chúng tôi gửi file thực hành vì đôi khi sách không có đĩa, hay đĩa bị hỏng, hay khi có nhu cầu sử dụng mà đĩa đã thất lạc đâu mất, nhất là các bạn ở xa. Chúng tôi để các file này trong 1 file nén và để trên mạng để bạn đọc ở tất cả mọi nơi đều có thể download xuống. Trong trường hợp các bạn gặp trục trặc về việc tải file xuống hoặc có thắc mắc về việc sử dụng các file này, bạn hãy liên lạc với chúng tôi qua hộp thư điện tử sau đây:

hoangtrong00@gmail.com

Việc biên soạn một quyển sách nào cũng khó tránh khỏi các sai sót. Mọi sai sót, nếu có, trong quyển sách này hoàn toàn là do người biên soạn. Chúng tôi mong được nhiều ý kiến đóng góp của tất cả các sinh viên, giảng viên và những người làm công tác nghiên cứu để lần tái bản tiếp theo, quyển sách được hoàn chỉnh hơn. Thư góp ý về nội dung quyển sách xin gửi về:

Hoàng Trọng
Khoa Toán – Thống Kê, Đại Học Kinh Tế TPHCM
Số 91 đường 3/2, quận 10, TP Hồ Chí Minh
Email: htrong@ueh.edu.vn

Chu Nguyễn Mộng Ngọc
Email: chunguyenmongngoc@ueh.edu.vn

Xin chân thành cảm ơn và chúc các bạn thành công!

TP Hồ Chí Minh, tháng 09 năm 2008
Tác giả

Hoàng Trọng
Chu Nguyễn Mộng Ngọc

MỤC LỤC

CHƯƠNG MỞ ĐẦU: GIỚI THIỆU PHÂN TÍCH DỮ LIỆU

1. NGHIÊN CỨU VÀ PHÂN TÍCH DỮ LIỆU 1
2. BẢN CHẤT CỦA PHÂN TÍCH DỮ LIỆU 4
3. THỐNG KÊ VÀ PHÂN TÍCH DỮ LIỆU 5

CHƯƠNG I: PHÂN LOẠI DỮ LIỆU, MÃ HÓA, NHẬP LIỆU VÀ MỘT SỐ XỬ LÝ TRÊN BIẾN

1. PHÂN LOẠI DỮ LIỆU 7
2. CÁC LOẠI THANG ĐO 8
3. NGUYÊN TẮC MÃ HOÁ VÀ NHẬP LIỆU 11
4. CỬA SỔ LÀM VIỆC CỦA SPSS 14
5. TẠO TẬP TIN DỮ LIỆU TRONG SPSS FOR WINDOWS 15
   5.1. Khai báo biến 15
   5.2. Lưu tập tin dữ liệu 19
6. MỘT SỐ XỬ LÝ TRÊN BIẾN 19
   6.1. Mã hoá lại biến (Recode) 19
   6.2. Chuyển một biến dạng Category thành dạng Dichotomy 26
   6.3. Thủ tục Compute để tính toán giá trị biến mới từ biến có sẵn 28
7. THAY ĐỔI MỘT SỐ MẶC ĐỊNH CỦA CHƯƠNG TRÌNH 29
8. THỂ HIỆN TIẾNG VIỆT TRONG SPSS 30

CHƯƠNG II: LÀM SẠCH DỮ LIỆU

1. SỰ CẦN THIẾT 35
2. CÁC BIỆN PHÁP NGĂN NGỪA 35
3. CÁC PHƯƠNG PHÁP LÀM SẠCH DỮ LIỆU 36
   3.1. Dùng bảng tần số 36
   3.2. Dùng bảng phối hợp hai biến hay ba biến 37
   3.3. Cách tìm lỗi đơn giản ngay trên cửa sổ dữ liệu (Data View) 41

CHƯƠNG III: TÓM TẮT VÀ TRÌNH BÀY DỮ LIỆU

1. PHƯƠNG PHÁP VÀ CÔNG CỤ 43
2. BẢNG TẦN SỐ ĐƠN GIẢN 43
3. CÁC ĐẠI LƯỢNG THỐNG KÊ MÔ TẢ 45
4. LẬP BẢNG TẦN SỐ ĐỒNG THỜI TÍNH TOÁN CÁC ĐẠI LƯỢNG THỐNG KÊ MÔ TẢ 50
5. THỐNG KÊ MÔ TẢ VỚI THỦ TỤC EXPLORE 53
6. LẬP BẢNG TỔNG HỢP NHIỀU BIẾN 60

6.1. Bảng kết hợp các biến định tính 60
6.1.1 Bảng kết hợp 2 biến định tính 60
6.1.2 Bảng kết hợp 3 biến định tính 67
6.2. Bảng kết hợp biến định tính với biến định lượng 71
6.2.1 Bảng kết hợp 1 biến định tính và 1 biến định lượng 71
6.2.2 Bảng kết hợp 2 biến định tính và 1 biến định lượng 73
6.3. Bảng tần số phức tạp với Tables of Frequencies 75
7. XỬ LÝ CÂU HỎI CÓ THỂ CHỌN NHIỀU TRẢ LỜI – Multiple Answer (MA) 77
8. TRÌNH BÀY KẾT QUẢ BẰNG ĐỒ THỊ 83
8.1. Các loại đồ thị cơ bản của SPSS 83
8.1.1 Đồ thị hình thanh (Bar) 84
8.1.2 Đồ thị dạng đường và diện tích (Line and Area Chart) 88
8.1.3 Đồ thị hình tròn (Pie) 88
8.2. Hiệu chỉnh đồ thị trên SPSS 89
8.2.1 Hiệu chỉnh các điểm phân tán trên đồ thị (Markers) 89
8.2.2 Thể hiện đường kẻ ngang trên đồ thị 90
8.2.3 Thể hiện trị số tuyệt đối của từng trường hợp 91
8.2.4 Hiệu chỉnh các vấn đề liên quan đến hai trục đồ thị 91
8.2.5 Chuyển đổi giữa các loại đồ thị 92
8.3. Lưu đồ thị 93
8.4. Vẽ đồ thị bằng Excel 93
Các bước thực hiện một đồ thị/ biểu đồ trên Excel 94
8.4.1. Biểu đồ thanh ngang 94
8.4.2. Biểu đồ thanh đứng (cột dọc) 102
8.4.3. Biểu đồ stack 104
8.4.4. Đồ thị hình tròn 105
9. BẢNG TÙY BIẾN (Custom tables) 107

CHƯƠNG IV: KIỂM ĐỊNH MỐI LIÊN HỆ GIỮA HAI BIẾN ĐỊNH TÍNH

1. KIỂM ĐỊNH MỐI LIÊN HỆ GIỮA HAI BIẾN ĐỊNH DANH -ĐỊNH DANH HOẶC ĐỊNH DANH - THỨ BẬC 116
1.1 Tóm tắt lý thuyết Kiểm định Chi-bình phương 117
1.2 Sử dụng SPSS thực hiện kiểm định Chi-bình phương 118
1.3 Một số đại lượng thống kê khác về mối liên hệ giữa 2 biến định danh 123
1.3.1 Cramer V 123
1.3.2 Hệ số liên hợp (Coefficient of contigency) 124
1.3.3. Lambda 124
2. KIỂM ĐỊNH MỐI LIÊN HỆ GIỮA 2 BIẾN THỨ BẬC 125
2.1. Gamma của Goodman và Kruskal 126
2.2. tau-b của Kendall (τ_b) 126

CHƯƠNG V: PHÂN TÍCH LIÊN HỆ GIỮA BIẾN NGUYÊN NHÂN ĐỊNH TÍNH VÀ BIẾN KẾT QUẢ ĐỊNH LƯỢNG: KIỂM ĐỊNH TRUNG BÌNH TỔNG THỂ

1. KIỂM ĐỊNH GIẢ THUYẾT VỀ TRỊ TRUNG BÌNH CỦA MỘT TỔNG THỂ 132
2. KIỂM ĐỊNH GIẢ THUYẾT VỀ SỰ BẰNG NHAU GIỮA HAI TRUNG BÌNH TỔNG THỂ 134
   2.1 Kiểm định giả thuyết về trị trung bình của hai tổng thể – trường hợp mẫu độc lập (Independent-samples T-test) 134
   2.2 Kiểm định trị trung bình của hai mẫu phụ thuộc hay mẫu phối hợp từng cặp. (Paired-samples T-test) 139

CHƯƠNG VI: PHÂN TÍCH LIÊN HỆ GIỮA BIẾN NGUYÊN NHÂN ĐỊNH TÍNH VÀ BIẾN KẾT QUẢ ĐỊNH LƯỢNG: PHÂN TÍCH PHƯƠNG SAI

1. PHÂN TÍCH PHƯƠNG SAI MỘT YẾU TỐ (ANOVA) 145
   1.1 Khái niệm và vận dụng 145
   1.2 Tóm tắt lý thuyết phân tích phương sai một yếu tố (One-Way ANOVA) ... 146
   1.3 Thực hiện phân tích phương sai một yếu tố với SPSS 148
   1.4 Đọc kết quả phân tích phương sai của SPSS 150
   1.5. Xác định chỗ khác biệt (phân tích sâu ANOVA) 151
2. PHÂN TÍCH PHƯƠNG SAI HAI YẾU TỐ (Two -way anova) 154

CHƯƠNG VII: KIỂM ĐỊNH PHI THAM SỐ

1. KIỂM ĐỊNH DẤU (SIGN TEST) VÀ KIỂM ĐỊNH MCNEMAR 165
   1.1 Trình tự tiến hành kiểm định dấu 166
   1.2 Thực hiện kiểm định dấu bằng SPSS 167
2. KIỂM ĐỊNH DẤU VÀ HẠNG WILCOXON (WILCOXON SIGNED-RANK TEST) 169
   2.1 Trình tự tiến hành kiểm định dấu và hạng Wilcoxon 170
   2.2 Thực hiện kiểm định dấu và hạng Wilcoxon trên SPSS 171
3. KIỂM ĐỊNH MANN-WHITNEY 2 MẪU ĐỘC LẬP 172
   3.1 Trình tự thực hiện kiểm định Mann-Whitney 2 mẫu độc lập 173
   3.2 Thực hiện kiểm định Mann-Whitney hai mẫu độc lập trên SPSS 174
4. KIỂM ĐỊNH KRUSKAL-WALLIS 176
5. KIỂM ĐỊNH CHI-BÌNH PHƯƠNG MỘT MẪU 179
6. KIỂM ĐỊNH KOLMOGOROV-SMIRNOV MỘT MẪU 184

CHƯƠNG VIII: KIỂM ĐỊNH TỶ LỆ 189

CHƯƠNG IX: TƯƠNG QUAN VÀ HỒI QUI TUYẾN TÍNH

1. TƯƠNG QUAN TUYẾN TÍNH 195
1.1 Hệ số tương quan đơn r (Pearson Correlation Coefficient) 197
1.1.1 Tính toán r 197
1.1.2 Một số đặc điểm của r 198
1.1.3 Kiểm định giả thuyết về hệ số tương quan tuyến tính r 199
Cách thực hiện tính r bằng SPSS 200
1.2. Hệ số tương quan hạng (Rank correlation coefficient) 204
2. HỒI QUI TUYẾN TÍNH 205
2.1. Hồi qui đơn tuyến tính 207
2.1.1 Xây dựng phương trình của mô hình hồi qui đơn tuyến tính từ dữ liệu mẫu 207
Cách thức xây dựng mô hình hồi qui đơn tuyến tính bằng SPSS 209
2.1.2 Các giả định đối với phân tích hồi qui tuyến tính. 211
2.1.3 Độ chính xác khi ước lượng các tham số của tổng thể từ các hệ số hồi qui mẫu 213
2.1.4 Đánh giá độ phù hợp của mô hình 214
2.1.5 Kiểm định các giả thuyết 218
2.1.6 Dự đoán bằng mô hình hồi qui 221
2.1.7 Dò tìm sự vi phạm các giả định cần thiết trong hồi qui tuyến tính 224
2.2 Mô hình hồi qui tuyến tính bội 236
2.2.1 Mô hình hồi qui tuyến tính bội, kí hiệu và các giả định 236
2.2.2 Xây dựng mô hình 237
2.2.3 Xác định tầm quan trọng của các biến trong mô hình 242
2.2.4 Lựa chọn biến cho mô hình 245
2.2.5. Dò tìm các vi phạm giả định cần thiết 251
2.2.6 Các thủ tục chọn biến 253
2.3. Các lựa chọn trong hộp thoại Linear Regression của SPSS 259
3. HỒI QUI VỚI QUAN HỆ PHI TUYẾN 267
4. HỒI QUI VỚI BIẾN ĐỘC LẬP ĐỊNH TÍNH (BIẾN GIẢ) 277

TÀI LIỆU THAM KHẢO 285

PHỤ LỤC: BẢN CÂU HỎI 287