

PHƯƠNG PHÁP NGHIÊN CỨU

Lý thuyết & thực tiễn

Trương Sĩ Ánh

CRO, Kantar Media Vietnam

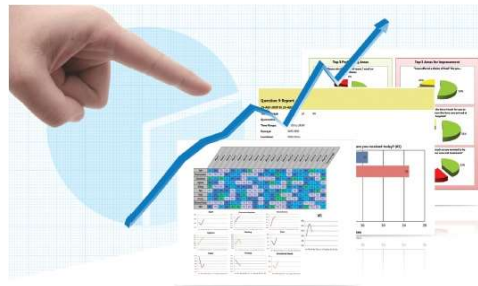
Tháng 5/2018

Bài 4

- Nhập liệu
Data entry/data punching
- Làm sạch dữ liệu & xử lý dữ liệu
Data cleaning & data processing
- Lập bảng phân tích kết quả
Tabulation
- Sử dụng phần mềm phân tích dữ liệu
- Báo cáo kết quả nghiên cứu

NHẬP LIỆU & PHÂN TÍCH KẾT QUẢ KHẢO SÁT

- ❑ Xử lý sơ bộ các bảng câu hỏi (BCH) đã hoàn tất
- ❑ Nhập thông tin trên BCH vào máy tính
- ❑ Áp dụng trọng số (*weight*) cho kết quả khảo sát
- ❑ Các đại lượng thống kê thường dùng trong phân tích
- ❑ Sử dụng chỉ số (*index*) để so sánh
- ❑ Diễn giải kết quả (*interpretation of results*)



XỬ LÝ SƠ BỘ CÁC BCH ĐÃ HOÀN TẤT (1)

- ❑ Tập kết, lưu giữ & bảo quản các bảng câu hỏi (BCH) đã hoàn tất
- ❑ Tổng kết và đúc rút kinh nghiệm
 - *Từ phòng vấn viên*
 - *Từ nhân viên kiểm soát chất lượng/giám sát viên*
- ❑ Kiểm tra & hiệu đính BCH trước hoặc trong quá trình nhập liệu
 - *Các thông tin hoặc chi tiết cần phải có trên BCH*
 - *Các câu hỏi chỉ cho phép 1 phương án trả lời duy nhất*
 - *Các câu hỏi bị bỏ trống (không ghi câu trả lời)*
 - *Các câu hỏi đúng ra phải bỏ trống nhưng lại được ghi câu trả lời*
 - *Các câu hỏi được điền vào mục "Khác (ghi rõ)"*



XỬ LÝ SƠ BỘ CÁC BCH ĐÃ HOÀN TẤT (2)

- Kiểm tra & hiệu đính BCH trước hoặc trong quá trình nhập liệu (*tiếp theo*)
 - ❖ Bổ sung mã số cho những câu trả lời trong mục "Khác (ghi rõ)" nếu tần suất xảy ra là có ý nghĩa về mặt thống kê. Ví dụ:

Anh/chị đang sử dụng ĐTDD nhãn hiệu nào?

Nokia 1

Samsung 2

LG 3

Iphone 4

Khác 9 (ghi rõ) Sky..... 

Nếu số người trả lời sử dụng ĐTDD nhãn hiệu **Sky** chiếm 1 tỷ lệ đáng kể trong tổng số các câu trả lời (ví dụ 1,5%), cần cung cấp riêng 1 mã số cho nhãn hiệu ĐTDD này, ví dụ mã số 5.

XỬ LÝ SƠ BỘ CÁC BCH ĐÃ HOÀN TẤT (3)

❑ Kiểm tra & hiệu đính BCH trước hoặc trong quá trình nhập liệu (*tiếp theo*)

❖ Gán mã số (*nhất quán*) cho câu trả lời của các câu hỏi mở. Ví dụ:

Anh/chị vui lòng kể tên 5 website Anh/chị truy cập thường xuyên nhất?

Website 1	<i>vnexpress.net</i>	5
Website 2	<i>youtube.com</i>	15
Website 3	<i>tuoitre.vn</i>	12
Website 4	<i>mp3.zing.vn</i>	37
Website 5	<i>không nhớ tên</i>	999

- ✓ Mỗi website khác nhau cần được thống nhất gán 1 mã số khác nhau.
- ✓ Những website có tần suất trả lời thấp (ví dụ dưới 5%), hoặc không thể nhận dạng được, có thể gán 1 mã số chung như 99 hay 999 (đại diện cho “Các website khác”).
- ✓ Công việc có thể thực hiện với sự hỗ trợ của máy tính (*sau khi nhập liệu hoàn tất*).

NHẬP THÔNG TIN VÀO MÁY TÍNH

- ❑ Sử dụng các ứng dụng Spreadsheet như *MS Excel*.
- ❑ Sử dụng các phần mềm thống kê như *SPSS, STATA, SAS* v.v.
- ❑ Sử dụng các phần mềm chuyên dụng như *CSPRO, Epidata, SPSS Dataentry, SurveyCraft*, v.v.
- ❑ Phát triển chương trình nhập liệu riêng cho 1 dự án nghiên cứu
- ❑ Phần mềm nhập liệu có ý nghĩa quan trọng đối với việc kiểm tra và hiệu đính dữ liệu nếu đáp ứng tốt các điều kiện:
 - Kiểm soát logic của dữ liệu khảo sát
 - Kiểm soát logic của quá trình nhập liệu
 - Ước lượng thông tin cho những trường hợp có câu trả lời bị bỏ sót, hoặc câu trả lời bị bỏ trống do đáp viên từ chối trả lời (*data imputation*).
 - Kiểm tra & đánh giá chất lượng nhập liệu



Tải phần mềm miễn phí CSPRO tại đây: <https://www.census.gov/data/software/cspro.html>

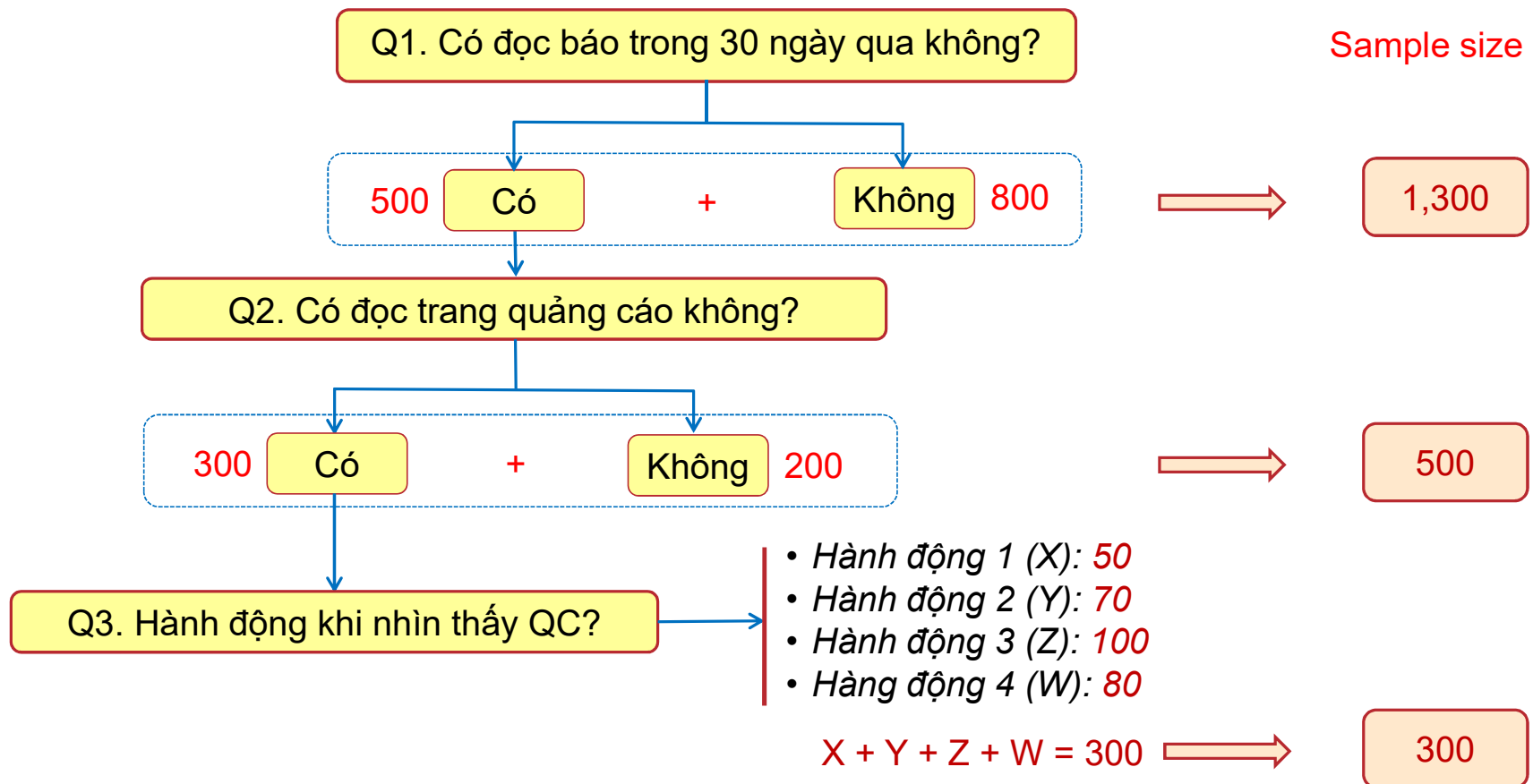
LÀM SẠCH DỮ LIỆU & XỬ LÝ DỮ LIỆU (1)

- ❑ Chỉnh sửa lỗi logic (*lỗi từ khâu phỏng vấn hoặc nhập liệu*)
 - Tuổi & trình độ học vấn: *15 tuổi đã tốt nghiệp đại học*
 - Nơi mua hàng hóa: *Mua đồ điện ở cửa hàng dược phẩm*
 - Tính chất nhất quán: *Trả lời có đọc báo hàng ngày nhưng trả lời không đọc báo trong 6 tháng qua.*
- ❑ Lập bảng tần suất (*frequency table*) cho tất cả các biến số để kiểm tra phân bố của các giá trị.
- ❑ Chạy các bảng Cross-tab quan trọng
- ❑ Chạy các tiêu chí thống kê đơn giản đối với các biến số *Numeric*
 - *Min, Max, Range, Average (Mean), Median, Standard deviation etc.*
- ❑ Phát hiện & xử lý các giá trị bị thiếu (*missing value*):
 - *Sai sót trong khâu phỏng vấn*
 - *Sai sót trong khâu nhập liệu*
 - *Đáp viên từ chối trả lời*



LÀM SẠCH DỮ LIỆU & XỬ LÝ DỮ LIỆU (2)

❑ Kiểm tra tính nhất quán của dữ liệu với các câu hỏi sàng lọc (*filter question*)



ÁP DỤNG TRỌNG SỐ CHO KẾT QUẢ KHẢO SÁT (1)

Weighting survey data

Trọng số là gì?

- ❑ Giá trị gán cho từng đối tượng được khảo sát (*đáp viên*).
- ❑ Sử dụng để cân chỉnh tỷ lệ đại diện của các nhóm đối tượng khác nhau trong mẫu khảo sát (*ví dụ tỷ lệ nam & nữ*) cho phù hợp với tỷ lệ dân số trong thực tế.
- ❑ Sử dụng như những hệ số phóng đại số mẫu (**projection factor**) áp dụng cho từng nhóm đối tượng khảo sát khác nhau. Ví dụ tỷ lệ chọn mẫu trong nhóm dân số 15-19 tuổi là 1/10.000, tức là 1 mẫu khảo sát đại diện cho 10.000 người trong thực tế. Như vậy, kết quả khảo sát của mỗi người trong nhóm 15-19 tuổi sẽ được nhân với 10.000 để đại diện cho 10.000 người.



ÁP DỤNG TRỌNG SỐ CHO KẾT QUẢ KHẢO SÁT (2)

Weighting survey data

Trọng số dùng để cân chỉnh tỷ lệ đại diện

Mẫu khảo sát & tổng thể nghiên cứu

Giới tính	Mẫu khảo sát		Dân số thực	Trọng số
	Số lượng	Tỷ lệ		
Nam	140	35%	51%	$51/35 = 1.4571$
Nữ	260	65%	49%	$49/65 = 0.7538$
Tổng số	400	100%	100%	-

Kết quả khảo sát về tỷ lệ đọc báo trong 7 ngày qua (*Bạn có đọc báo trong 7 ngày qua không?*)

	Tổng số mẫu	Số mẫu có đọc	% mẫu có đọc	Weight	% có đọc sau khi áp dụng trọng số
Nam	140	30	21.4%	1.4571	21.4%
Nữ	260	40	15.4%	0.7538	15.4%
Tổng số	400	70	17.5%	-	19.4%

ÁP DỤNG TRỌNG SỐ CHO KẾT QUẢ KHẢO SÁT (3)

Weighting survey data

Trọng số dùng để cân chỉnh tỷ lệ đại diện

❑ Kết quả khảo sát về tỷ lệ đọc báo trong 7 ngày qua (*không áp dụng trọng số*)

❖ Tỷ lệ đọc báo của dân số NAM = $30/140 = 21.4\%$

❖ Tỷ lệ đọc báo của dân số NỮ = $40/260 = 15.4\%$

❖ Tỷ lệ đọc báo của dân số CHUNG = $(30+40)/(140+260) = 17.5\%$

❑ Kết quả khảo sát về tỷ lệ đọc báo trong 7 ngày qua (*có áp dụng trọng số*)

❖ Tỷ lệ đọc báo của dân số NAM = $(30*1.4571)/(140*1.4571) = 21.4\%$

❖ Tỷ lệ đọc báo của dân số NỮ = $(40*0.7538)/(260*0.7538) = 15.4\%$

❖ Tỷ lệ đọc báo của dân số CHUNG = $(30*1.4571+40*0.7538)/(140*1.4571+260*0.7538) = 19.4\%$

ÁP DỤNG TRỌNG SỐ CHO KẾT QUẢ KHẢO SÁT (4)

Weighting survey data

Trọng số dùng làm hệ số phóng đại mẫu (*projection factor*)

Mẫu khảo sát & tổng thể nghiên cứu

Giới tính	Mẫu khảo sát	Dân số thực (Universe)	Trọng số
Nam	140	3.57 mil	$3.57 \text{ mil}/140 = 25,500.0$
Nữ	260	3.43 mil	$3.43 \text{ mil}/260 = 13,192.3$
Tổng số	400	7.00 mil	-

Kết quả khảo sát về tỷ lệ đọc báo trong 7 ngày qua

	Số mẫu	Có đọc	Weight	Ước lượng số người đọc trong thực tế
Nam	140	30	25,500.0	765,000
Nữ	260	40	13,192.3	527,692
Tổng số	400	70	-	1,292,692

PHƯƠNG PHÁP TÍNH TRỌNG SỐ (1)

Kết hợp nhiều thuộc tính kiểm soát mẫu cùng 1 lúc (dùng bảng cross-tab)

Cell weighting

- ❑ Phải có sẵn số liệu về universe (*total population*) được trình bày dưới dạng cross-tab
 - *Dân số chia theo nhóm tuổi & giới tính*
 - *Dân số chia theo nhóm tuổi, giới tính & thành phần kinh tế*
 - *Dân số chia theo nhóm tuổi, giới tính & trình độ học vấn*
- ❑ Không có ô dữ liệu (*cell*) nào của bảng cross-tab có số mẫu bằng 0
 - ❖ *Nếu có 1 nhóm người nào đó không có trong mẫu khảo sát thì sao?*
- ❑ Số mẫu trong mỗi ô của bảng cross-tab không quá nhỏ
 - ❖ *Chúng ta không muốn trọng số áp dụng cho các nhóm mẫu chênh lệch nhau quá lớn*
- ❑ Khó có thể thực hiện trong điều kiện:
 - *Thiếu thông tin cần thiết về tổng thể (universe/total population)*
 - *Phải kết hợp nhiều hơn 2 thuộc tính kiểm soát số mẫu khảo sát*
 - *Tổng số mẫu khảo sát tương đối nhỏ (nên không có đủ đại diện trong mẫu khảo sát)*

PHƯƠNG PHÁP TÍNH TRỌNG SỐ (2)

Kết hợp nhiều thuộc tính kiểm soát mẫu cùng 1 lúc (dùng bảng cross-tab)

Cell weighting

Ví dụ về cách tính trọng số theo giới tính & nhóm tuổi

Nhóm tuổi	TAM Panel		Universe (000)		Trọng số (weight)	
	Nam	Nữ	Nam	Nữ	Nam	Nữ
4-9	68	61	255	232	3,750.0	3,803.3
10-14	66	50	184	171	2,787.9	3,420.0
15-19	64	71	278	302	4,343.8	4,253.5
20-24	80	85	367	422	4,587.5	4,964.7
25-29	82	102	335	380	4,085.4	3,725.5
30-34	85	78	286	303	3,364.7	3,884.6
35-39	55	66	268	275	4,872.7	4,166.7
40-44	57	72	228	237	4,000.0	3,291.7
45-49	63	77	188	214	2,984.1	2,779.2
50-54	87	71	153	180	1,758.6	2,535.2
55-59	55	79	91	118	1,654.5	1,493.7
60+	71	110	167	257	2,352.1	2,336.4
Tổng số	833	922	2800	3091	-	-

PHƯƠNG PHÁP TÍNH TRỌNG SỐ (3)

Tách biệt các thuộc tính kiểm soát mẫu thành nhiều vòng lặp tính toán

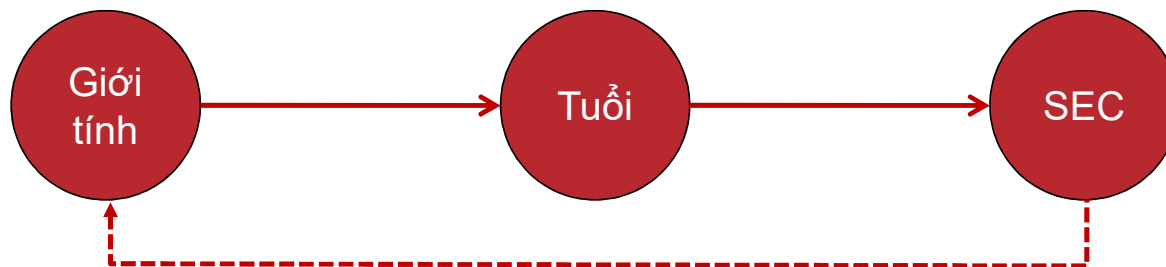
Rim weighting

Giới tính	Nam	Nữ	Tổng số
Số mẫu	833	922	1755
Universe	$g1$	$g2$	$g3$

Nhóm tuổi	4-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60+
Số mẫu	61	50	71	85	102	78	66	72	77	71	79	110
Universe	$t1$	$t2$	$t3$	$t4$	$t5$	$t6$	$t7$	$t8$	$t9$	$t10$	$t11$	$t12$

Socio-economic class (SEC)	Class A	Class B	Class C	Class D	Class E/F
Sample	176	351	527	527	176
Universe	$c1$	$c2$	$c3$	$c4$	$c5$

Các bước tính weight theo vòng lặp



PHƯƠNG PHÁP TÍNH TRỌNG SỐ (3)

Tách biệt các thuộc tính kiểm soát mẫu thành nhiều vòng lặp tính toán

Rim weighting

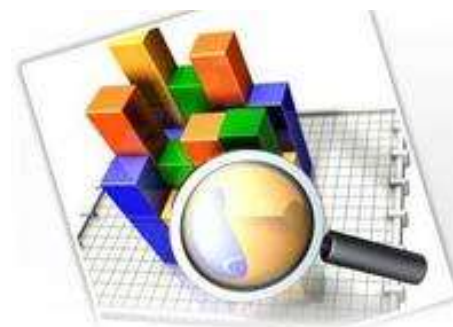
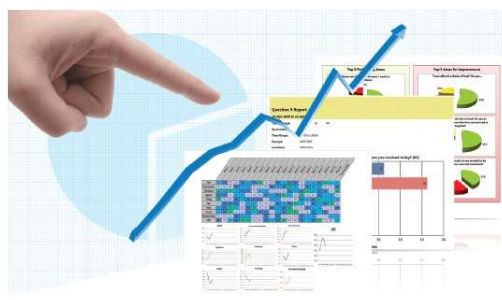
- ❑ Tính Rim weight bằng phương pháp thủ công (*có/không có sự hỗ trợ của máy tính*)
- ❑ Tính Rim weight tự động bằng phần mềm chuyên dùng
 - ❖ SAS
 - ❖ STATA
 - ❖ Python plug-in for SPSS

MỘT SỐ ĐẠI LƯỢNG THỐNG KÊ THƯỜNG DÙNG TRONG PHÂN TÍCH ĐỊNH LƯỢNG

- ❑ Giá trị tối thiểu, tối đa và trung bình (*Min, max and mean or average*)
- ❑ Giá trị Mod (*Mode*)
- ❑ Giá trị trung vị (*Median*)
- ❑ Phương sai (*Variance*)
- ❑ Độ lệch chuẩn (*Standard Deviation*)
- ❑ Sai số chọn mẫu (*Sampling Error or Margin of Error*)
- ❑ Tần suất (*Frequency*)
- ❑ Tỷ lệ và phần trăm (*Proportion, Percentage and N-tile*)
- ❑ Tỷ số và tỷ suất (*Ratio and Rate*)
- ❑ Chỉ số (*Index*)

PHÂN TÍCH KẾT QUẢ KHẢO SÁT VỚI SỰ HỖ TRỢ CỦA MÁY TÍNH

- ❑ 10 tiền đề cho xử lý & phân tích dữ liệu khảo sát
- ❑ Phác thảo/thiết kế các bảng biểu phân tích dữ liệu (*Tabulation plan*)
- ❑ Sử dụng phần mềm SPSS trong phân tích dữ liệu khảo sát



10 TIỀN ĐỀ CHO PHÂN TÍCH & XỬ LÝ KẾT QUẢ

- ❑ Bạn có phải là người tham gia vào các giai đoạn khác nhau của nghiên cứu/khảo sát mà bạn muốn phân tích kết quả?
 - Thiết kế mẫu
 - Thiết kế Bảng câu hỏi
 - Mã hóa thông tin thu thập từ cuộc khảo sát và nhập liệu vào máy tính
 - Chuẩn bị dữ liệu để sẵn sàng cung cấp cho người dùng
- ❑ Với sự hỗ trợ của các phần mềm máy tính mang tính thân thiện & dễ sử dụng, người phân tích có xu hướng “nhảy” ngay vào việc phân tích mà bỏ qua khâu nghiên cứu, tìm hiểu kỹ lưỡng về cuộc khảo sát nghiên cứu.
- ❑ Không hiểu biết về cuộc khảo sát/nghiên cứu sẽ dễ dàng dẫn đến những sai lầm không đáng có.
- ❑ Kiến thức & kinh nghiệm của người phân tích là yếu tố then chốt cho sự thành công.

10 TIỀN ĐỀ CHO PHÂN TÍCH & XỬ LÝ KẾT QUẢ

1

Tìm hiểu & nắm vững nội dung Bảng câu hỏi khảo sát

- ❖ Ý nghĩa đích thực của câu hỏi bạn muốn phân tích là gì?
 - ✓ Ví dụ: Biến số cần phân tích “*Những hoạt động chính khi sử dụng Internet*”
- ❖ Câu hỏi này dành cho những đối tượng khảo sát nào?
 - *Giới hạn phạm vi không gian: NỘI THÀNH TP. HCM*
 - *Giới hạn phạm vi thời gian: TRONG 30 NGÀY QUA, QUÝ 1/20XX*
 - *Những giới hạn khác có liên quan: CÓ SỬ DỤNG INTERNET*

Bạn không thể có những phân tích có ý nghĩa
nếu bạn không nắm vững nội dung của bảng câu hỏi

10 TIỀN ĐỀ CHO PHÂN TÍCH & XỬ LÝ KẾT QUẢ

2

Xây dựng kế hoạch làm báo cáo kết quả khảo sát

- ❖ Mục đích của báo cáo
 - *Mô tả*
 - *Chứng minh*
 - *Xây dựng mô hình, lý thuyết*
- ❖ Đối tượng mà báo cáo sẽ nhắm tới là ai
- ❖ Các giả thiết ban đầu
- ❖ Các kết quả nghiên cứu khác có thể sử dụng
- ❖ Thiết kế các bảng dữ liệu & đồ thị cần đưa vào báo cáo

10 TIỀN ĐỀ CHO PHÂN TÍCH & XỬ LÝ KẾT QUẢ

3

Nắm vững phương pháp luận của nghiên cứu/khảo sát

- ❖ Đối tượng khảo sát
- ❖ Thiết kế mẫu & phương pháp chọn mẫu
- ❖ Số lượng mẫu & tính đại diện của mẫu khảo sát
- ❖ Phương pháp thu thập dữ liệu
 - ✓ Phỏng vấn trực diện (*face-to-face interviewing*)
 - ✓ Phỏng vấn qua điện thoại
 - ✓ Khảo sát bằng FF tự điền BCH (phát BCH qua bưu điện, phát trực tiếp, on-line)
- ❖ Thời gian tiến hành khảo sát
- ❖ Khả năng suy rộng kết quả khảo sát

10 TIỀN ĐỀ CHO PHÂN TÍCH & XỬ LÝ KẾT QUẢ

4

Tìm hiểu xem dữ liệu khảo sát đã được gán trọng số chưa

- ❖ Dữ liệu đã được gán sẵn trọng số hay chưa?
- ❖ Có mấy trọng số có thể sử dụng?
 - *Trọng số áp dụng cho cá thể (repondent)*
 - *Trọng số áp dụng cho HGD*
- ❖ (Các) trọng số được tính toán như thế nào?
 - *Cở sở tính trọng số: Nhóm tuổi, Giới tính, SEC, TV platform, v.v.*
 - *Phương pháp tính trọng số: Cell weighting vs. Rim weighting*
- ❖ Kiểm tra mức độ ảnh hưởng của trọng số
 - *So sánh giữa kết quả chưa gán trọng số và kết quả đã gán trọng số: Variance, Standard Dev, Sampling error, Frequencies, v.v.*

10 TIỀN ĐỀ CHO PHÂN TÍCH & XỬ LÝ KẾT QUẢ

5

ĐỐI CHIẾU CÁC BIẾN SỐ TRONG CSDL VỚI CÁC CÂU HỎI TRONG BCH

- ❖ CSDL có tất cả các biến số tương ứng với các câu hỏi trong BCH?
- ❖ Các biến số trong CSDL được tổ chức & sắp xếp như thế nào
- ❖ Có câu hỏi nào trong BCH không có trong CSDL?
 - *Nếu có, nó có ảnh hưởng gì đến việc phân tích & làm báo cáo của bạn?*
 - *Các câu hỏi nhận dạng đáp viên, địa chỉ, số điện thoại v.v.*
- ❖ Biến số nào có mặt trong CSDL nhưng không có trong BCH?
 - *Nếu có, ý nghĩa hay công dụng của nó là gì?*
 - *Weight, SEC, SEL, Media Quintiles, Claimed/Regular Reader v.v.*

10 TIỀN ĐỀ CHO PHÂN TÍCH & XỬ LÝ KẾT QUẢ

6

Luôn thận trọng với cấp độ hay phạm vi phân tích dữ liệu khảo sát mẫu

- ❖ Đừng giả thiết mình có thể phân tích bất cứ thứ gì được thu thập trong cuộc khảo sát.
- ❖ Luôn chú ý đến điều kiện tiên quyết thứ 3
 - *Thông tin về Tỉnh/TP, Địa bàn khảo sát, v.v., được thu thập từ mỗi đáp viên trong khảo sát VHLSS 2012, nhưng bạn có thể phân tích kết quả theo Tỉnh/TP hoặc Địa bàn khảo sát không?*
 - *Bạn có thông tin demographics của tất cả đáp viên, nhưng liệu bạn có thể phân tích kết quả khảo sát cho tất cả các nhóm đáp viên mà bạn quan tâm?*

Hãy kiểm tra số lượng mẫu khảo sát của nhóm đối tượng mà bạn quan tâm trước khi bắt tay phân tích dữ liệu.

10 TIỀN ĐỀ CHO PHÂN TÍCH & XỬ LÝ KẾT QUẢ

7

Hãy tìm ra những câu hỏi (*biến số*) quan trọng cần cho phân tích của bạn

❖ Những câu hỏi/thông tin nào sẽ giúp bạn trả lời những vấn đề cần quan tâm nghiên cứu.

✓ **Bài tập:** Thông tin nào trong bảng câu hỏi VHLSS giúp bạn trả lời các câu hỏi sau:

- *Tỷ lệ hộ gia đình có 1, 2 hoặc 3 tivi trở lên*
- *Chỉ tiêu cho ăn uống hàng ngày chiếm bao nhiêu % thu nhập khả dụng của HGD, và có sự khác biệt nào giữa các nhóm hộ có mức sống khác nhau, hoặc giữa các vùng trong cả nước hay không??*
- *Có sự khác biệt nào giữa nam & nữ về tỷ lệ tốt nghiệp đại học hay không?*

10 TIỀN ĐỀ CHO PHÂN TÍCH & XỬ LÝ KẾT QUẢ

8

Hãy tìm ra những câu hỏi (*biến số*) giúp xác định các phân khúc (segments) khác nhau về đối tượng nghiên cứu

- ❖ Các câu hỏi xác định phạm vi & đối tượng nghiên cứu của bạn
 - *Thông tin demographics, lifestyle, segmentation, v.v.*
 - *Thông tin về hành vi, mức độ sử dụng hay mức độ xảy ra 1 hiện tượng nào đó cần quan tâm.*
 - ✓ **Bài tập:** Thông tin nào trong BCH MHS giúp bạn có thể phân tích các đối tượng sau:
 - *Người sử dụng Internet nhiều (heavy user)*
 - *Người ít xem TV (light viewer)*
 - *Người thích sử dụng các sản phẩm mới*
- ❖ Tạo lập các nhóm đối tượng cần nghiên cứu từ những thông tin có sẵn
 - *Phụ nữ có con 0-4 tuổi; Nam giới 30-45 tuổi thích tin tức, thời sự.*

10 TIỀN ĐỀ CHO PHÂN TÍCH & XỬ LÝ KẾT QUẢ

9

Hãy quan tâm đến những trường hợp đáng chú ý

- ❖ Hiện tượng xảy ra phổ biến nhất
- ❖ Hiện tượng được cho là mới mẻ, thú vị, đáng chú ý v.v.
- ❖ Hiện tượng được cho là không bình thường theo hiểu biết thông thường
 - Cần tìm hiểu kỹ lưỡng hơn
 - ✓ Kiến thức & kinh nghiệm
 - ✓ Các kết quả nghiên cứu (tương đương khác)
 - ✓ Tiến hành thêm nghiên cứu khảo sát khác
 - Nếu có bằng chứng là số liệu không hợp lý
 - ✓ Quyết định loại bỏ những trường hợp có dữ liệu bất thường ra khỏi phân tích
 - ✓ Phản ánh cho bộ phận xử lý dữ liệu để tìm hiểu & khắc phục

10 TIỀN ĐỀ CHO PHÂN TÍCH & XỬ LÝ KẾT QUẢ

10

Luôn kiểm tra kết quả phân tích của bạn

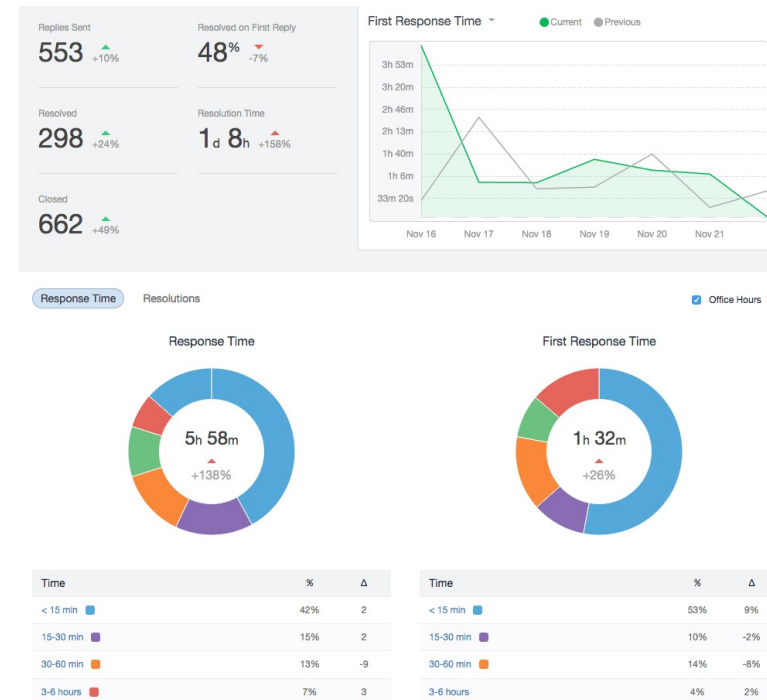
- ❖ Vận dụng kiến thức và kinh nghiệm của bản thân.
- ❖ So sánh với các kết quả nghiên cứu (*tương đương*) khác, nếu có thể.
- ❖ Nếu kết quả phân tích có những vấn đề được cho là mới mẻ, thú vị, đáng chú ý, hoặc khác biệt với hiểu biết thông thường (*common sense*)
 - *Kiểm tra lại toàn bộ quá trình phân tích.*
 - ✓ *Cách thức/kỹ năng sử dụng phân mềm.*
 - ✓ *Xác định đối tượng nghiên cứu hoặc các phân khúc đối tượng nghiên cứu khác nhau.*
 - ✓ *Sử dụng các kỹ thuật sàng lọc dữ liệu hoặc tạo lập các biến số mới từ các biến số sẵn có.*
 - *Tham vấn các chuyên gia trong lĩnh vực nghiên cứu.*
 - *Kiến nghị phương châm hành động tiếp theo.*

PHÁC THẢO/THIẾT KẾ BẢNG BIỂU PHÂN TÍCH DỮ LIỆU (1)

Tabulation plan

❑ Bạn sẽ thể hiện kết quả nghiên cứu như thế nào.

- ❖ Văn xuôi (*Text*)
- ❖ Biểu/bảng dữ liệu (*Tables*)
- ❖ Đồ thị (*Graphs*)
- ❖ Hình ảnh, video



PHÁC THẢO/THIẾT KẾ BẢNG BIỂU PHÂN TÍCH DỮ LIỆU (2)

Tabulation plan

- ❑ Bạn cần những bảng dữ liệu nào để phục vụ cho việc phân tích & làm báo cáo kết quả nghiên cứu?
 - ❖ Bảng tần suất đơn giản (frequency tables)
 - ❖ Bảng chéo (crosstab)
 - Nhiều hơn 1 biến số được thể hiện trong bảng
 - So sánh, đối chiếu (ví dụ để tìm hiểu sự giống nhau, khác nhau giữa các đối tượng nghiên cứu)
 - Tìm kiếm mối liên hệ (tương quan) giữa các đối tượng nghiên cứu, giữa hiện tượng này với hiện tượng khác.



Bảng tần suất

Frequency table

Bảng x. Phân bố hộ gia đình theo mức thu nhập bình quân/người/tháng

Thu nhập BQ/người/tháng	Số hộ	% hộ
Dưới 500.000		
500.000-1.000.000		
>1.000.000-2.000.000		
>2.000.000-3.000.000		
>3.000.000-5.000.000		
>5.000.000		
Tổng số		

Bảng chéo

Cross table

Bảng xx. Phân bố người lao động theo mức thu nhập bình quân/tháng

Thu nhập BQ/người/tháng	Tổng số	Nam	Nữ
Dưới 3.000.000			
3.000.000-5.000.000			
>5.000.000-7.000.000			
>7.000.000-10.000.000			
>10.000.000-15.000.000			
>15.000.000			
Tổng số			

Bảng chéo

Cross table

Bảng xx. Phân bố người lao động theo mức thu nhập bình quân/tháng

Thu nhập BQ/người/tháng	Tổng số		Nam		Nữ	
	Số lượng	%	Số lượng	%	Số lượng	%
Dưới 3.000.000						
3.000.000-5.000.000						
>5.000.000-7.000.000						
>7.000.000-10.000.000						
>10.000.000-15.000.000						
>15.000.000						
Tổng số		100%		100%		100%

Bảng chéo

Cross table

Bảng xxx. Phân bố dân số theo vùng & khu vực thành thị/nông thôn

Vùng KTXH	Tổng số	Thành thị	Nông thôn	Tổng số	Thành thị	Nông thôn
Miền Núi Trung Du Phía Bắc	100%					
Đồng Bằng Sông Hồng	100%					
Bắc Trung Bộ & DHMT	100%					
Tây Nguyên	100%					
Đông Nam Bộ	100%					
Đồng Bằng Sông Cửu Long	100%					
Cả nước	100%			100%	100%	100%

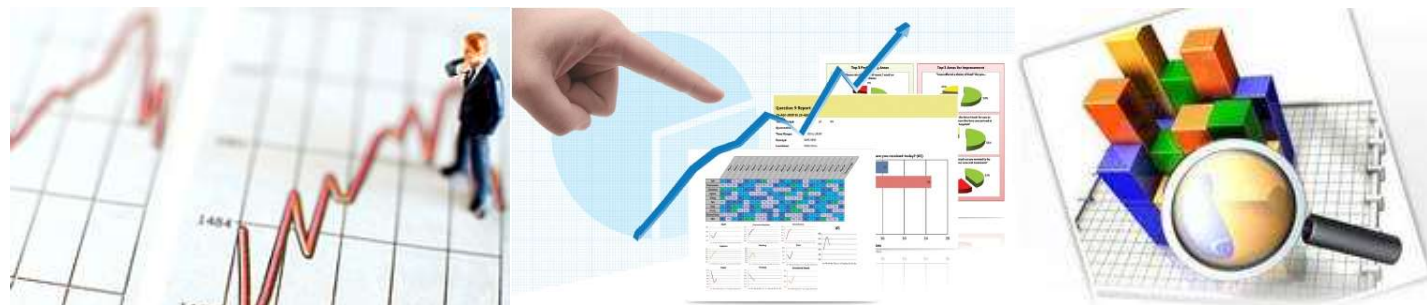
PHÂN TÍCH KẾT QUẢ KHẢO SÁT VỚI SỰ HỖ TRỢ CỦA MÁY TÍNH

❑ Phần mềm phân tích ĐỊNH LƯỢNG

- ✓ SPSS
- ✓ SAS
- ✓ STATA
- ✓ EVIEW

❑ Phần mềm phân tích ĐỊNH TÍNH

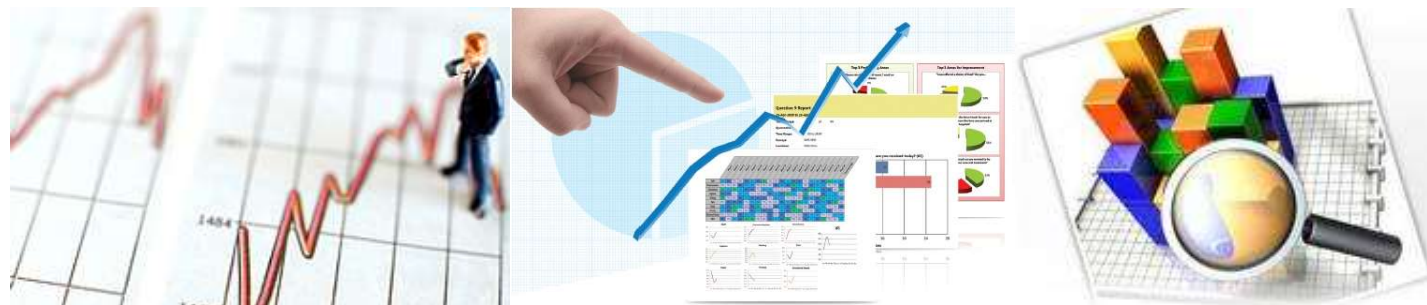
- ✓ NVIVO
- ✓ SPSS MODELER



PHÂN TÍCH KẾT QUẢ KHẢO SÁT VỚI SỰ HỖ TRỢ CỦA MÁY TÍNH

□ Nhiệm vụ chính

- ✓ Thao tác với các file dữ liệu có định dạng khác nhau
- ✓ Khảo sát dữ liệu
- ✓ Biến đổi dữ liệu
- ✓ Phân tích dữ liệu



BÁO CÁO KẾT QUẢ NGHIÊN CỨU

- ❑ Giới thiệu dự án nghiên cứu
 - *Đặt vấn đề hoặc bối cảnh ra đời*
 - *Mục đích/ý nghĩa*
- ❑ Phương pháp nghiên cứu
 - *Đối tượng nghiên cứu, phạm vi nghiên cứu, thiết kế nghiên cứu, chọn mẫu, thu thập thông tin, xử lý phân tích*
- ❑ Kết quả nghiên cứu
- ❑ Kết luận/khuyến nghị đúc kết từ kết quả nghiên cứu
- ❑ Tóm tắt kết quả nghiên cứu (*có thể đặt lên đầu*)
- ❑ Phụ lục (*nếu có*)
- ❑ Danh mục tài liệu trích dẫn

