

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Đồng Thị Thùy Linh

**PHÂN TÍCH THỐNG KÊ DỮ LIỆU TRONG
ĐIỀU TRA XÃ HỘI HỌC**

Chuyên ngành: Khoa học máy tính

Mã số: 60 48 01

LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

Hướng dẫn khoa học: PGS.TS Bùi Thế Hồng

Thái Nguyên - 2011

LỜI CẢM ƠN!

Em xin gửi lời cảm ơn chân thành nhất đến **PGS.TS Bùi Thế Hồng**, người đã tận tình hướng dẫn, giúp đỡ em trong suốt thời gian thực hiện luận văn này.

Con cảm ơn Cha, Mẹ và gia đình, những người đã dạy dỗ, khuyến khích, động viên con trong những lúc khó khăn, tạo mọi điều kiện cho chúng con nghiên cứu học tập.

Em cảm ơn các thầy, cô trong Viện Công Nghệ Thông Tin Hà Nội cùng các thầy cô trong trường Đại học Công nghệ thông tin và truyền thông – ĐH Thái Nguyên đã dìu dắt, giảng dạy em, giúp em có những kiến thức quý báu trong những năm học qua.

Cảm ơn các bạn đã tận tình động viên đóng góp ý kiến cho luận văn của tôi.

Mặc dù đã cố gắng hết sức cùng với sự tận tâm của thầy giáo hướng dẫn song do trình độ còn hạn chế, nội dung đề tài còn mới mẻ nên Luận văn khó tránh khỏi những thiếu sót. Em rất mong nhận được sự thông cảm và góp ý của thầy cô và các bạn.

Thái Nguyên, ngày 30 tháng 9 năm 2011.

Học viên

Đông Thị Thùy Linh

LỜI CAM ĐOAN

Em xin cam đoan toàn bộ nội dung bản luận văn này là do em tự sưu tầm, tra cứu và sắp xếp cho phù hợp với nội dung yêu cầu của đề tài.

Tất cả các thử nghiệm của luận văn đều do em tự thiết kế và xây dựng, trong đó có sử dụng một số ứng dụng chuẩn của phần mềm SPSS 18.0 và bộ số liệu cuộc điều tra về “*Mô tả thực trạng, kiến thức, hành vi phòng chống tai nạn thương tích của học sinh phổ thông trước và sau can thiệp. Xây dựng và đánh giá mô hình phòng chống TNTT tại trường phổ thông thành phố Thái Nguyên,*” .

Nếu sai em xin hoàn toàn chịu trách nhiệm.

Thái Nguyên, ngày 30 tháng 9 năm 2011

Người cam đoan

Đồng Thị Thùy Linh

MỤC LỤC

Trang phụ bìa	
LỜI CAM ĐOAN.....	i
MỤC LỤC	ii
Danh mục các bảng	v
Lời mở đầu.....	1
Chương 1: Tổng quan về thống kê dữ liệu và điều tra xã hội học.....	3
1.1. Khái quát về dữ liệu và thống kê	3
1.1.1. Dữ liệu là gì?	3
1.1.2. Thống kê là gì?	3
1.1.3. Một số khái niệm thường dùng của thống kê.....	4
1.1.4. Khái quát quá trình nghiên cứu thống kê.....	6
1.2. Giới thiệu các phương pháp tiến hành điều tra xã hội học.....	7
1.2.1. Đối tượng nghiên cứu của xã hội học.....	7
1.2.2. Những phương pháp nghiên cứu xã hội học	8
1.3. Các phương pháp và kỹ thuật đo phân tích thống kê dữ liệu.	11
1.3.1. Ước lượng	11
1.3.2. Tương quan và hồi quy	11
1.3.3. Kiểm định giả thuyết	12
1.4. Giới thiệu về các phần mềm ứng dụng phân tích thống kê dữ liệu	13
1.4.1. Phần mềm SPSS	13
1.4.2. Phần mềm Epidata	13
Chương 2: Một số vấn đề trong phân tích dữ liệu thống kê	14
2.1. Thống kê mô tả	14
2.1.1 Mục tiêu của các nghiên cứu mô tả	14
2.1.2. Những nội dung chính của các nghiên cứu mô tả	14
2.1.3. Thiết kế nghiên cứu mô tả	15
2.1.4. Mối quan hệ nhân quả.....	17

2.2. Cơ sở dữ liệu.....	17
2.2.1. Cơ sở dữ liệu là gì?	17
2.2.2. Các dạng cơ sở dữ liệu.....	18
2.2.3. Biểu diễn thông tin thống kê trong cơ sở dữ liệu.....	19
2.2.4. Mã hóa các thông tin trong cơ sở dữ liệu	20
2.2.5. Xác định và xử lý các giá trị bị thiếu và vượt trội trong cơ sở dữ liệu ...	22
2.3. Ước lượng.....	23
2.3.1. Khái niệm.....	23
2.3.2. Ước lượng một trung bình quần thể	24
2.3.3. Ước lượng tỷ lệ của một quần thể	26
2.3.4. Ước lượng sự khác nhau giữa hai trung bình quần thể	27
2.3.5. Ước lượng sự khác nhau giữa hai tỷ lệ quần thể.....	29
2.4. Kiểm định giả thuyết thống kê	30
2.4.1. Khái niệm.....	30
2.4.2. Hình thành các giả thuyết	30
2.4.3. Các kết luận và kết quả có được từ việc kiểm định giả thuyết	31
2.4.4. Các bước của việc kiểm định giả thuyết thống kê	32
2.4.5. Các thống kê kiểm định và miền bác bỏ.....	33
2.4.6. Ứng dụng lý thuyết kiểm định	34
Chương 3: Ứng dụng thử nghiệm trong điều tra xã hội học	40
3.1. Giới thiệu về cuộc điều tra	40
3.1.1. Đối tượng nghiên cứu	40
3.1.2. Địa điểm và thời gian nghiên cứu	40
3.1.3. Phương pháp nghiên cứu	41
3.2. Phương pháp thu thập và xử lý số liệu.....	43
3.2.1. Phương pháp thu thập số liệu.....	43
3.2.2. Phương pháp xử lý nhập số liệu.....	43
3.2.3. Kết quả của cuộc nghiên cứu	44
3.3. Sử dụng phương pháp ước lượng các tham số rút ra từ quần thể.....	45

3.3.1. Ước lượng về sự hiểu biết trung bình của học sinh phổ thông về phòng tránh TNTT do bỏng.....	46
3.3.2. Ước lượng sự khác nhau về sự hiểu biết trung bình của học sinh trước và sau can thiệp.....	49
3.3.3. Ước lượng tỷ lệ học sinh bị TNTT trên tổng số học sinh.....	50
3.4. Kiểm định giả thuyết.....	51
3.4.1. Kiểm định giả thuyết về sự khác nhau về mức độ hiểu biết trung bình của học sinh về phòng tránh TNTT do bỏng trước và sau can thiệp.	52
3.4.2. Kiểm định giả thuyết về tỷ lệ học sinh bị TNTT trên tổng số học sinh ..	54
3.4.3. Kiểm định giả thuyết về tỷ lệ học sinh bị TNTT trước và sau can thiệp và theo giới	55
Kết luận.....	60
Kiến nghị những nghiên cứu tiếp theo	61
Tài liệu tham khảo.....	62
PHỤ LỤC.....	63

Danh mục các bảng

Bảng 2.1. Các kết luận và kết quả khi kiểm định một giả thuyết

Bảng 3.1. Tỷ lệ học sinh có hành vi đúng về TNTT do bỏng trước can thiệp

Bảng 3.2. Bảng kết quả khoảng tin cậy cho trung bình quần thể chung cho toàn thể mẫu.

Bảng 3.3. Bảng mô tả về trung bình quần thể về kiến thức phòng tránh TNTT trước can thiệp và sau can thiệp của trường Quang Trung

Bảng 3.4. Kết quả tìm khoảng tin cậy cho sự khác nhau giữa hai trung bình mẫu với độ tin cậy 95%

Bảng 3.5. Kết quả tìm khoảng tin cậy cho sự khác nhau giữa hai trung bình mẫu với độ tin cậy 98%

Bảng 3.6. Tỷ lệ học sinh bị TNTT

Bảng 3.7. Kết quả các nhóm thống kê (Group Statistics)

Bảng 3.8. Kết quả kiểm định về mức độ hiểu biết trung bình về phòng tránh TNTT do bỏng của học sinh trước và sau can thiệp

Bảng 3.9. Tỷ lệ học sinh bị TNTT

Bảng 3.10. Kết quả phân tích kiểm định (Test Statistics)

Bảng 3.11: Tỷ lệ TNTT về giới

Bảng 3.12. Kiểm định Chi bình phương (Chi-Square Tests)

Bảng 3.13: Tỷ lệ TNTT về giới

Bảng 3.14. Kiểm định Chi bình phương (Chi-Square Tests)

Bảng 3.15: Tỷ lệ TNTT về giới

Bảng 3.16. Kiểm định Chi bình phương (Chi-Square Tests)

Lời mở đầu

1. Tính cấp thiết của đề tài

Thống kê là một bộ môn khoa học bao gồm các phương pháp thu thập, tổ chức và phân tích các dữ liệu. Ngày nay thống kê đã đi vào hầu hết các lĩnh vực của xã hội như kinh tế thị trường, chỉ số giá cả, tỷ lệ tăng giảm dân số, khí hậu, thời tiết, y tế ... Những cuộc khảo sát thăm dò dư luận, các cuộc điều tra chọn mẫu về dân số, sức khỏe, giáo dục và các dự báo dân số đã và đang được tiến hành, tất cả đều chứng thực cho tầm quan trọng của các phương pháp thống kê trong các sự kiện quan trọng hằng ngày.

Phân tích thống kê là một bộ phận khoa học đã hình thành từ lâu, đã chứng minh được sự cần thiết và quan trọng của nó trong đời sống và phát triển kinh tế xã hội hiện đại.

2. Mục tiêu nghiên cứu của đề tài

- Tìm hiểu về điều tra xã hội học và phân tích dữ liệu thống kê.
- Ứng dụng thực tế trong điều tra xã hội học.

3. Đối tượng và phạm vi nghiên cứu

- Các khảo sát điều tra xã hội học.
- Phân tích thống kê số liệu điều tra xã hội học.
- Các công cụ phân tích thống kê dữ liệu.
- Các phương pháp đánh giá ước lượng hỗ trợ ra quyết định về các xu hướng hay các chính sách xã hội.

4. Phương pháp nghiên cứu

- Tìm hiểu các phương pháp tiến hành điều tra xã hội học.
- Nghiên cứu các phương pháp phân tích thống kê dữ liệu.
- Chọn lựa các phương pháp phân tích tổng hợp và các công cụ đánh giá dữ liệu.

5. Kết cấu luận văn

Bố cục của Luận văn gồm ba chương chính như sau:

Chương 1: Tổng quan về thống kê dữ liệu và điều tra xã hội học.

- Khái quát về dữ liệu thống kê.
- Giới thiệu các phương pháp tiến hành điều tra xã hội học.
- Các phương pháp và kỹ thuật phân tích thống kê dữ liệu.

Chương 2: Một số vấn đề trong phân tích dữ liệu thống kê.

- Thống kê mô tả.
- Cơ sở dữ liệu.
- Ước lượng.
- Kiểm định giả thuyết.

Chương 3: Ứng dụng thử nghiệm trong điều tra xã hội học.

Dựa trên số liệu của cuộc điều tra, sử dụng thống kê suy luận để:

- Tìm khoảng tin cậy.
- Kiểm định các giả thuyết.

6. Kết quả đạt được

Qua quá trình thực hiện luận văn em đã thực hiện được một số công việc sau:

- Hiểu được về điều tra xã hội học, quá trình thực hiện một cuộc điều tra xã hội học, và ý nghĩa của nó trong xã hội hiện đại.
- Hiểu được về thống kê dữ liệu và phân tích thống kê dữ liệu.
- Áp dụng được ước lượng và kiểm định giả thuyết vào phân tích thống kê dữ liệu.

7. Ý nghĩa khoa học của đề tài

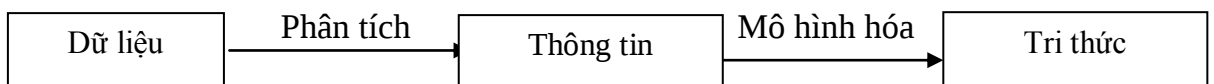
- Đề tài có ý nghĩa khoa học, thời sự và mang đậm tính thực tiễn.
- Thông qua quá trình phân tích dữ liệu thống kê trong điều tra xã hội học đưa ra các đánh giá hỗ trợ cho việc ra quyết định về các xu hướng hay các chính sách xã hội.

Chương 1: Tổng quan về thống kê dữ liệu và điều tra xã hội học

1.1. Khái quát về dữ liệu và thống kê

1.1.1. Dữ liệu là gì?

Dữ liệu được định nghĩa là những điều được biết đến, hoặc giả định sự kiện và con số, mà từ đó kết luận có thể được suy ra. Nói chung, dữ liệu là thông tin ở dạng thô về chất lượng cũng như số lượng. Các nguồn có thể được bất cứ điều gì từ tin đồn kết quả của nghiên cứu và điều tra. Các điều khoản của báo cáo có thể được mô tả, số, hoặc kết hợp khác nhau của cả hai. Việc chuyển đổi từ dữ liệu kiến thức có thể được xem xét bao gồm các trình tự phân cấp [9]:



Dữ liệu thường được xem là cấp thấp nhất của trừu tượng mà từ đó thông tin và sau đó kiến thức có nguồn gốc. Dữ liệu thô, tức là dữ liệu chưa qua chế biến, đề cập đến một tập hợp các số, ký tự, hình ảnh, kết quả đầu ra khác từ các thiết bị thu thập thông tin để chuyển đổi số lượng vật lý vào biểu tượng [9].

1.1.2. Thống kê là gì?

a. Khái niệm thống kê

Thống kê là hệ thống các phương pháp dùng để thu thập, xử lý và phân tích các con số (mặt lượng) của những hiện tượng số lớn để tìm hiểu bản chất và qui luật vốn có của chúng (mặt chất) trong điều kiện không gian và thời gian cụ thể [7].

Mọi sự vật, hiện tượng đều có hai mặt chất và lượng không tách rời nhau, và khi chúng ta nghiên cứu hiện tượng, điều chúng ta muốn biết đó là bản chất của hiện tượng. Nhưng mặt chất thường ẩn bên trong, còn mặt lượng biểu hiện ra bên ngoài dưới dạng các đại lượng ngẫu nhiên. Do đó phải thông qua các phương pháp xử lý thích hợp trên mặt lượng của số lớn đơn vị cấu thành hiện tượng, tác động của