

CHƯƠNG 4

XỬ LÝ VÀ PHÂN TÍCH DỮ LIỆU

NGUYỄN THỊ MINH THU

NỘI DUNG

4.1. Xử lý dữ liệu

4.2. Một số phương pháp phân tích dữ liệu

4.3. Trình bày kết quả xử lý và phân tích



4.1. Xử lý dữ liệu

4.1.1. Hiệu chỉnh và mã hóa dữ liệu

4.1.2. Đánh giá chất lượng dữ liệu

4.1.3. Xác định các mối liên hệ

4.1.1. Hiệu chỉnh và mã hóa dữ liệu

a) Khái niệm và lợi ích hiệu chỉnh và mã hóa dữ liệu

- Hiệu chỉnh là kiểm tra dữ liệu và thông tin theo yêu cầu
- Mã hóa là thay đổi số liệu 'nguyên thủy' về một dạng ngắn gọn hơn để có thể lưu trong máy tính và xử lý được dễ dàng bằng các mã số hoặc ký hiệu thích hợp (Xây dựng bộ mã hóa).
- **Các lợi ích của mã hóa dữ liệu**
 - Giảm công suất, không gian lưu trữ
 - Nếu mã hóa số lượng vừa phải có thể giúp nâng cao giá trị của số liệu
 - Giúp cho các phương pháp phân tích định lượng

b) Kỹ thuật mã hóa dữ liệu

Lựa chọn mã hóa, thang đo gắn cho từng dữ liệu

- Các loại thang đo:
 - Định danh/danh nghĩa (Nominal Scale)
 - Thứ bậc (Ordinal Scale)
 - Khoảng (Interval Scale)
 - Tỷ lệ (Ratio Scale)
- Các mã số: Đánh dấu, ký hiệu, cho điểm
- Lựa chọn số lượng và giới hạn của từng hành vi của thông tin.
- Gắn thang đo, mã số cho từng hành vi của thông tin

c) Những chú ý khi mã hóa dữ liệu

- Người sử dụng cần phải biết mã của dữ liệu
 - Nếu người sử dụng không biết mã của số liệu thì không thể phân tích được
 - Thí dụ Mã hóa thông tin về giới: 1 là nam; 2 là nữ
- Mức độ chính xác của dữ liệu mã hóa
 - Ví dụ: Mã hóa thông tin về mức độ kinh tế
Hộ giàu: thu nhập/1 người > 2,5 triệu/tháng; Hộ nghèo: < 0,5 triệu/tháng
- Mã hóa thường thể hiện bằng số
 - Ví dụ: Nghèo 1, Cận nghèo 2, Khá 3, Giàu 4

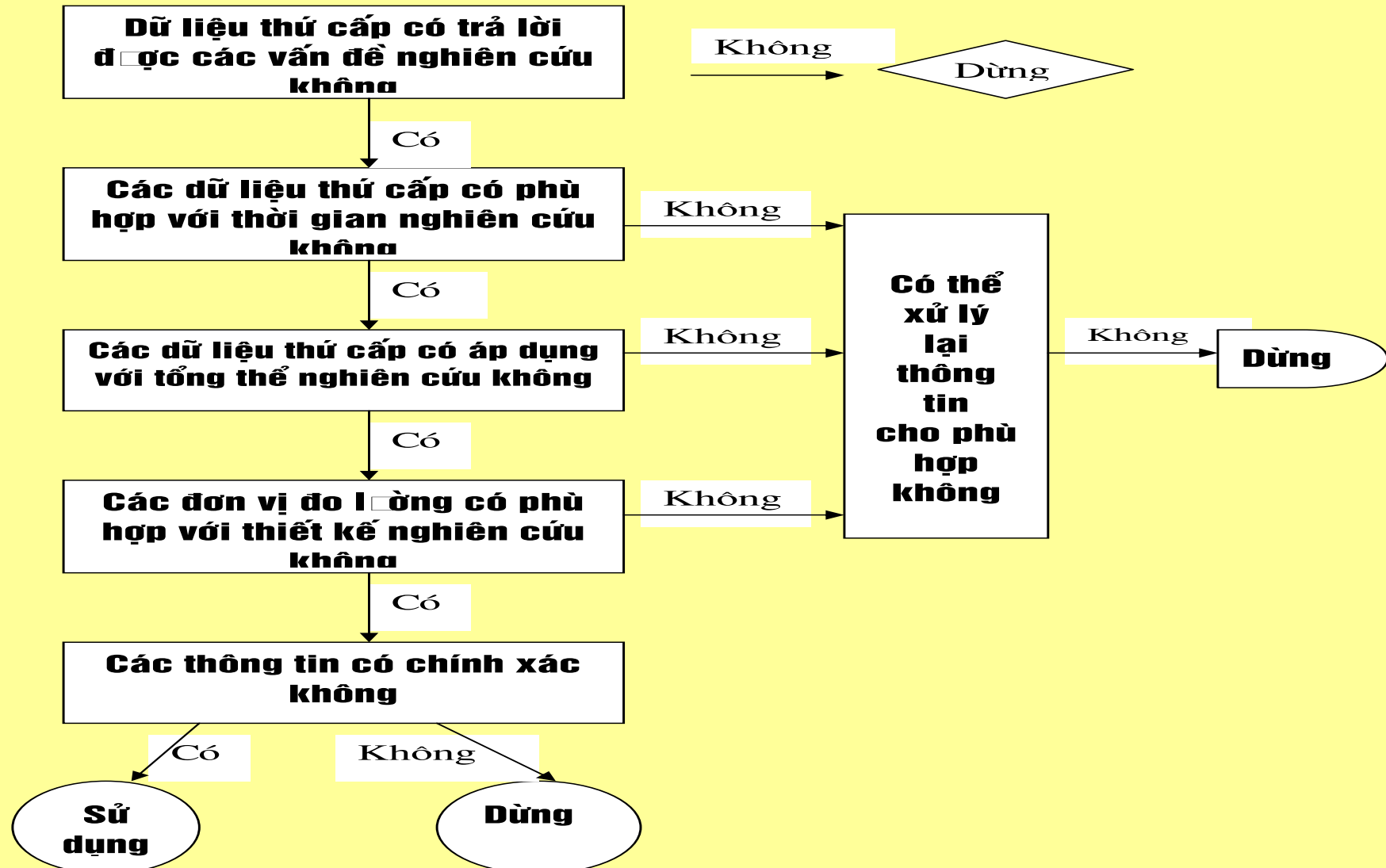
4.1.2. Đánh giá chất lượng dữ liệu



a) Thế nào là đánh giá chất lượng số liệu

- Tại sao số liệu phải ĐÁNG TIN CẬY và số liệu phải THỰC?
- Có thể biết được thông qua kiểm định số liệu
- Đánh giá chất lượng số liệu giúp người sử dụng và nhà quản lý chắc chắn rằng số liệu “tốt” có thể sử dụng cho NC
- **Tiêu chí đánh giá chất lượng số liệu**
 - Tính chính xác
 - Hợp lí
 - Thời gian (trước, sau, mới?)
 - Đầy đủ
 - Mức hiện diện (có sẵn)
 - Mức độ chi tiết

Phương pháp đánh giá chất lượng số liệu



Đánh giá chất lượng thông tin thứ cấp

Phương pháp đánh giá chất lượng số liệu sơ cấp

- Tình trạng bình thường
 - Kiểm tra số liệu có thể sử dụng trong điều kiện bình thường với những số liệu “bình thường”
- Tình trạng “cực đoan”
 - Kiểm tra mức độ chính xác của số liệu nhưng ở mức thấp hơn hoặc cao hơn trong khoảng số liệu cần
- Tình trạng “sai”
 - Kiểm tra với số liệu sai
- Kiểm tra số liệu trong mọi tình trạng

Mọi người và với các phương pháp khác nhau, kiểm tra thường xuyên theo các tiêu chí để hạn chế tới mức thấp nhất các sai số

4.1.3. Tổng hợp dữ liệu

Thế nào là tổng hợp dữ liệu?

- Tập trung
- Hệ thống hoá (phân loại)
- Sắp xếp
- Trình bày một cách khoa học các dữ liệu thu thập được

Mục đích: Khái quát đặc trưng của từng nhóm, toàn bộ tổng thể

Mã hoá các biến trên EXCEL

1	C1	Nhân khẩu và lao động
2	C11	Tên chủ hộ
3	C12	Thôn
4	C13	Xã
5	<i>C111</i>	<i>Giới tính</i>
6	<i>C112</i>	<i>Tuổi</i>
7	<i>C113</i>	<i>Trình độ văn hoá</i>
8	<i>C114</i>	<i>Trình độ chuyên môn</i>
9	C14	Số nhân khẩu của hộ
10	<i>C141</i>	<i>Số nhân khẩu dưới 16 tuổi</i>
11	<i>C142</i>	<i>Số nhân khẩu từ 16 đến 45 tuổi</i>
12	<i>C143</i>	<i>Số nhân khẩu từ trên 45 tuổi</i>
13	C15	Trình độ văn hoá của nhân khẩu trong hộ
14	<i>C151</i>	<i>Số nhân khẩu có trình độ văn hoá tiểu học cơ sở</i>
15	<i>C152</i>	<i>Số nhân khẩu có trình độ văn hoá trung học cơ sở</i>
16	<i>C153</i>	<i>Số nhân khẩu có trình độ văn hoá trung học phổ thông</i>
17	C16	Trình độ chuyên môn nghiệp vụ của nhân khẩu trong hộ
18	<i>C161</i>	<i>Số nhân khẩu có trình độ sơ cấp</i>
19	<i>C162</i>	<i>Số nhân khẩu có trình độ trung cấp</i>
20	<i>C163</i>	<i>Số nhân khẩu có trình độ đại học</i>
21	C2	Đất đai
22		

4.2. Phân tích thông tin

4.2.1. Phân tích thông tin định tính

4.2.2. Phân tích thông tin định lượng

4.2.3. Phân tích thông tin thứ cấp

4.2.1. Phân tích thông tin định tính

Một số phương pháp chính

- a) Nghiên cứu tình huống
- b) Tổng quan lịch sử
- c) Phân tích điểm mạnh yếu (SWOT)
- d) Phân tích thông tin thứ cấp
- e) Xếp hạng theo thứ tự ưu tiên
- f) Cây vấn đề, cây giải pháp

a) Nghiên cứu tình huống

* Thế nào là nghiên cứu tình huống? (Murray, 1938)

Nghiên cứu tình huống là nghiên cứu kỹ một người như **một “chủ thể” thống nhất** – chứ không phải là một phần trong dân số

Nghiên cứu tình huống là **nghiên cứu sâu hay rất kỹ** về một đơn vị nhằm làm rõ hơn những vấn đề nghiên cứu

* Điểm mạnh

- Sâu và chi tiết
- Bao quát cả những sự việc phức tạp

* Điểm yếu

- Vấn đề khái quát: Có ý kiến chủ quan
- Kết quả có thể bị chệch và có ấn tượng chủ quan
- Làm cho mối quan hệ phức tạp giữa các biến số

Kỹ thuật thể hiện: Viết, hộp

b) Tổng quan lịch sử

Tổng quan lịch sử hay tổng quan tài liệu nghiên cứu

- Là tóm tắt những hiểu biết về những vấn đề, những lĩnh vực có liên quan đến đề tài hay nội dung NC
- Nêu mẫu thuẫn, ưu nhược điểm của các quan điểm NC
- Xem lại chương 3 phần phương pháp thu thập thông tin bằng phương pháp nghiên cứu tài liệu theo trình tự:
 1. Lựa chọn câu hỏi NC
 2. Lựa chọn các cơ sở dữ liệu, bài báo, tài liệu để tìm
 3. Lựa chọn nội dung, từ khóa để tìm
 4. Nghiên cứu tài liệu và Thực hiện tổng quan
 5. Khái quát hóa kết quả

c) Phân tích điểm mạnh, yếu, cơ hội, thách thức

- Strengths (mạnh), Weaknesses (yếu), Opportunities (cơ hội), Threats (thách thức) - SWOT - Phân tích SWOT
- Phân tích SWOT là phương pháp xác định các **điểm mạnh** (ưu điểm), các **điểm yếu** (nhược điểm) và đồng thời tìm ra các **cơ hội** và **thách thức** mà chúng ta có thể có được hoặc phải đối mặt (với vấn đề nghiên cứu)
- Là phân tích một hiện tượng dưới quan điểm hệ thống từ bên trong (S, W) ra bên ngoài (O, T) hay đồng thời kết hợp cả trong và ngoài
- Đây là công cụ sử dụng nhiều trong phân tích các hiện tượng dưới dạng định tính – xã hội, chính sách
- Có thể sử dụng cả trong thảo luận hoặc cá nhân tự nghiên cứu

Phân tích SWOT

- Mạnh (S)

- Khả năng bên trong

Chúng ta có cái gì (hiện tại)?

- Yếu (W)

- Sự thiếu khả năng bên trong

Chúng ta thiếu cái gì (hiện tại)?

- Cơ hội (O)

- Những điểm tích cực từ hoàn cảnh bên ngoài

Chúng ta có thể nhận được cái gì (tương lai)?

- Thách thức (T)

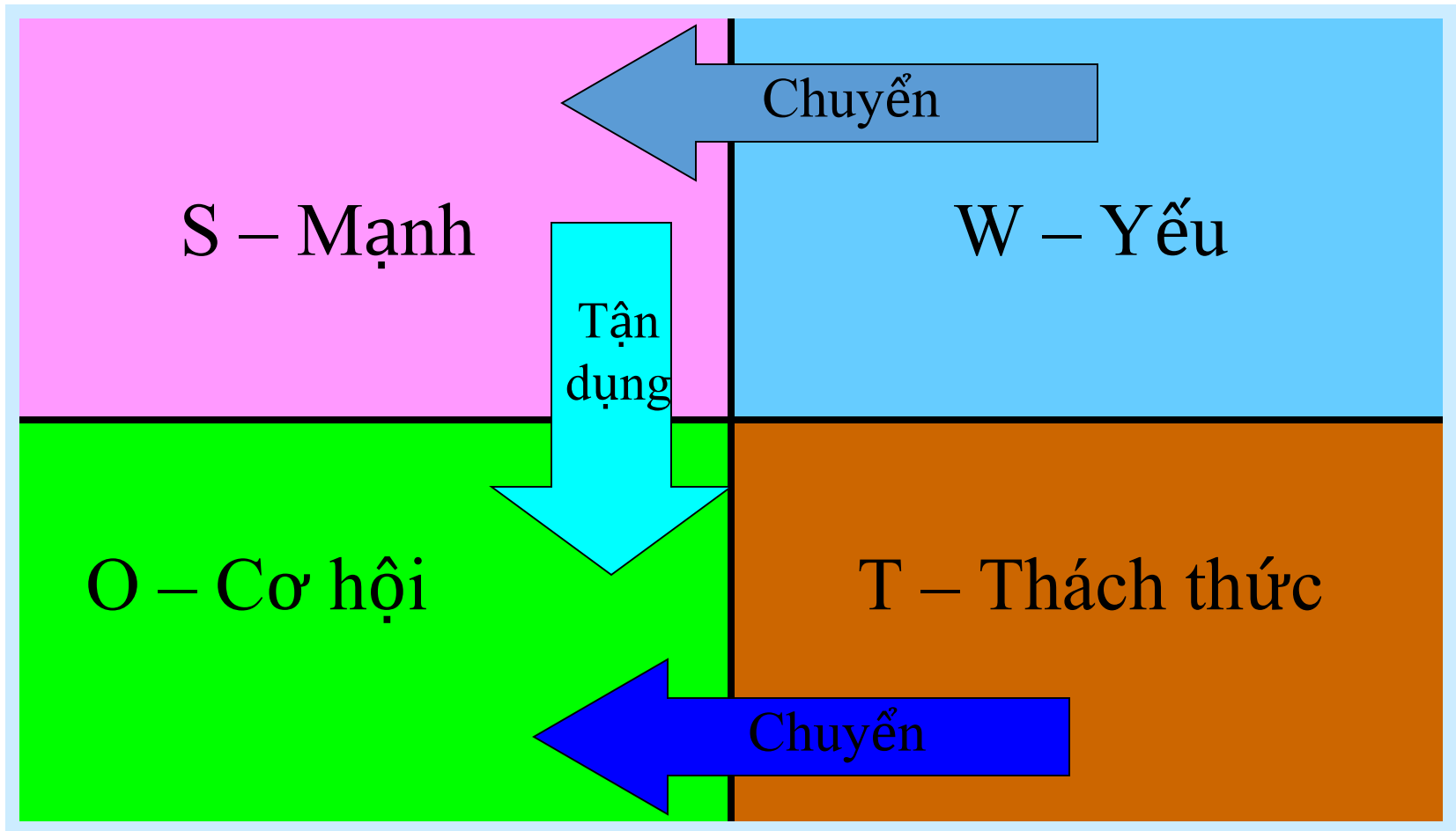
- Những điểm tiêu cực từ hoàn cảnh bên ngoài

Chúng ta có thể mất cái gì (tương lai)?

Phân tích SWOT

<p>Các đặc tính</p>	<p>Tích cực</p>	<p>Tiêu cực</p>
<p>Bên trong Có thể kiểm soát</p>	<p>S — <i>Mạnh</i> Tạo ra lợi thế cạnh tranh</p>	<p>W- <i>Yếu</i> Khả năng xuất hiện vấn đề</p>
<p>Bên ngoài Ngoài tầm kiểm soát</p>	<p>O – <i>Cơ hội</i> Tiềm năng cho tăng trưởng</p>	<p>T- <i>Thách thức</i> Có thể cản trở sự tiến bộ</p>

4 ô của ma trận SWOT



Phân tích SWOT

SWOT	Cơ hội thực hiện (O)	Thách thức/ Nguy cơ (T)
Mặt mạnh (S)	Tận dụng cơ hội để phát huy thế mạnh (O/S)	Tận dụng mặt mạnh để giảm thiểu nguy cơ (S/T)
Mặt yếu (W)	Nắm bắt cơ hội để khắc phục mặt yếu (O/W)	Giảm thiểu mặt yếu để ngăn chặn nguy cơ (W/T)

Phân tích SWOT

SWOT	Cơ hội thực hiện (O)	Thách thức/ Nguy cơ (T)
Mặt mạnh (S)	Tận dụng cơ hội để phát huy thế mạnh (O/S)	Tận dụng mặt mạnh để giảm thiểu nguy cơ (S/T)
Mặt yếu (W)	Nhắm bắt cơ hội để khắc phục mặt yếu (O/W)	Giảm thiểu mặt yếu để ngăn chặn nguy cơ (W/T)

Phân tích SWOT

SWOT	Cơ hội thực hiện (O)	Thách thức/ Nguy cơ (T)
Mặt mạnh (S)	Tận dụng cơ hội để phát huy thế mạnh (O/S)	Tận dụng mặt mạnh để giảm thiểu nguy cơ (S/T)
Mặt yếu (W)	Nắm bắt cơ hội để khắc phục mặt yếu (O/W)	Giảm thiểu mặt yếu để ngăn chặn nguy cơ (W/T)

8 bước xây dựng ma trận SWOT (1)

1. Liệt kê các điểm mạnh chủ yếu
2. Liệt kê những điểm yếu cơ bản (tiêu biểu) bên trong
3. Liệt kê các cơ hội chính
4. Liệt kê các mối đe dọa chủ yếu bên ngoài
5. Kết hợp S-O và đề xuất phương án chiến lược phát huy điểm mạnh để nắm bắt cơ hội.

8 bước xây dựng ma trận SWOT (2)

6. Kết hợp W-O để đề ra phương án chiến lược khắc phục điểm yếu bằng cách tận dụng cơ hội.
7. Kết hợp W-T và đề xuất phương án chiến lược nhằm tối thiểu hoá tác dụng của điểm yếu và phòng thủ trước các mối đe dọa từ bên ngoài.
8. Kết hợp S-T và đề xuất phương án chiến lược lợi dụng thế mạnh của mình để đối phó với nguy cơ đe dọa từ bên ngoài.

Nguồn: <http://www.nguoikinhdoanh.com>

4.2.2. Phân tích thông tin định lượng

1. Thống kê mô tả

2. Thống kê phân tích

- Kiểm định thống kê
- Phân tích phương sai
- Phân tích tương quan
- Phân tích hồi qui tuyến tính
- Hồi qui phi tuyến (KTL)

3. Mô hình hóa

- * Mô hình KTL và Toán kinh tế
 - Mô phỏng
 - Mô hình cân đối
 - Mô hình khác
- 4. Các phương pháp phân tích khác

a) Thống kê mô tả



Các chỉ tiêu phân tích thống kê mô tả

- Mode
- Trung vị
- Trung bình
- Tổng số
- Tần số
- Tần suất
- Khoảng biến động
- Độ lệch bình quân
- Phương sai
- Độ lệch chuẩn
- Hệ số biến động
- Độ nghiêng
- Độ tù

Thống kê mô tả

- Thống kê mô tả là các chỉ tiêu thống kê mức độ cho 1 biến định lượng
- Mode, trung vị, trung bình (có 3 phương pháp tính số bình quân), tổng số, tần số, tần suất
- Biến động về lượng của 1 biến được đo bằng các chỉ tiêu
 - Khoảng biến động
 - Độ lệch chuẩn (mô-men bậc 1)
 - Phương sai (mô-men bậc 2)
 - Độ nghiêng (skewness, mô-men bậc 3)
 - Độ tù (Kurtosis, mô-men bậc 4)

Tóm tắt

- Trung bình số học
- Phương sai
 - Đo độ phân tán của số bình quân
- Độ lệch chuẩn
 - Căn bậc 2 của phương sai

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\delta^2 \cong s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\delta \cong s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

Phân phối của trung bình mẫu

- Mẫu được chọn từ tổng thể
 - Sử dụng phân phối của trung bình mẫu
- Phân phối của trung bình mẫu có

- Bình quân

$$\mu_{\bar{x}} = \mu$$

- P sai

$$\delta^2_{\bar{x}} = \frac{\delta^2}{n}$$

- Độ lệch chuẩn

$$\delta_{\bar{x}} = \frac{\delta}{\sqrt{n}}$$

- Chứng minh

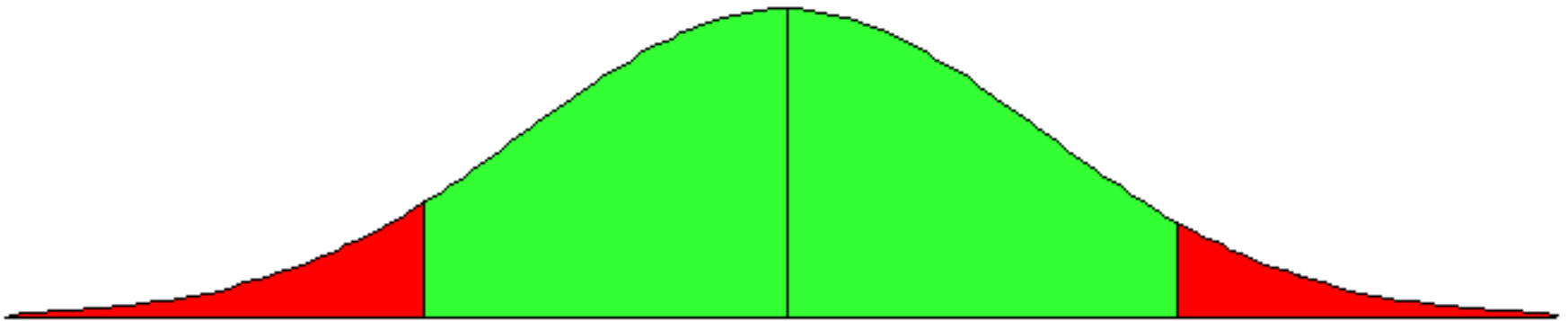
$$\bar{X} = \frac{1}{n} X_1 + \frac{1}{n} X_2 + \dots + \frac{1}{n} X_n$$

$$\delta^2_{\bar{X}} = \frac{1}{n^2} \delta^2_{X_1} + \frac{1}{n^2} \delta^2_{X_2} + \dots + \frac{1}{n^2} \delta^2_{X_n}$$

$$\delta^2_{\bar{X}} = \frac{1}{n^2} \delta^2 + \frac{1}{n^2} \delta^2 + \dots + \frac{1}{n^2} \delta^2$$

$$\delta^2_{\bar{X}} = n \left(\frac{1}{n^2} \delta^2 \right) = \frac{1}{n} (\delta^2)$$

b) Thống kê so sánh (Kiểm định giả thuyết)



1. Kiểm định giả thuyết về sự khác nhau giữa 2 số trung bình của 2 tổng thể

a. Lấy mẫu từng cặp

+ Bài toán

n quan sát sẽ được lấy mẫu theo từng cặp phối hợp từ 2 tổng thể số liệu **X** và **Y** nh sau:

Quan sát	X	Y	X-Y
1	X1	Y1	X1-Y1
2	X2	Y2	X2-Y2
3	X3	Y3	X3-Y3
.	.	.	.
.	.	.	.
.	.	.	.
n	Xn	Yn	Xn-Yn
Trung bình	μ_x	μ_y	Đ
Phương sai	δ_x^2	δ_y^2	S_d^2
Độ lệch chuẩn	δ_x	δ_y	Sd

μ_x Trung bình của tổng thể X

μ_y Trung bình của tổng thể Y

Đ là trung bình của tổng thể sai lệch X - Y

Sd là độ lệch chuẩn của tổng thể X-Y

Giả sử tổng thể các sai lệch giữa X và Y

(X-Y) có phân phối chuẩn. Ta cần kiểm

định giả thuyết sau:

Ho: $\mu_x - \mu_y = Do$ (Do là giá trị cho trước
Do=0)

H1: $\mu_x - \mu_y \neq Do$

Hay:

Ho: $\mu_x - \mu_y = 0$; H1: $\mu_x - \mu_y \neq 0$

Nguyên tắc kiểm định

Trong đó: D_0 : giá trị cụ thể cho tróc

$$T = \frac{\check{D} - D_0}{\frac{Sd}{\sqrt{n}}}$$

\check{D} : là trung bình của tổng thể sai lệch X - Y

n: số đơn vị mẫu quan sát

T: Tiêu chuẩn kiểm định (T thực nghiệm)

- T×m T lý thuyết với bậc tự do là n-1; $\alpha/2$. Ta có thể tra bảng phân phối Student với n-1 và $\alpha/2$; hoặc t×m hàm **TINV(n-1, α)**.

Gia thuyết	Bác bỏ Ho khi
$H_0 : \mu_x - \mu_y = D_0$ $H_1 : \mu_x - \mu_y \neq D_0$	$T > T_{n-1, \alpha/2}$ hoặc $T < -T_{n-1, \alpha/2}$ Hay $ T > T_{n-1, \alpha/2}$

- So sánh T thực nghiệm với T lý thuyết”

Nếu $|T| \leq T(n-1, \alpha/2)$ ta chấp nhận giả thuyết H_0 ,

Nếu $|T| > T(n-1, \alpha/2)$ ta bác bỏ giả thuyết H_0 và khi đó:

- Nếu $\check{D} > D_0$ thì $\mu_x - \mu_y > 0$

- Nếu $\check{D} < D_0$ thì $\mu_x - \mu_y < 0$

+ Thí dụ: Công ty VINAMILK áp dụng công nghệ mới trong chế biến sữa chua. Hãy kiểm định xem năng suất lao động của công nhân sau khi sử dụng công nghệ mới với công nghệ cũ có khác nhau không với mức ý nghĩa là 5% ?

Thứ tự công nhân quan sát	NSLD (kg/ngày)		X-Y
	Trước khi X	Sau khi Y	
1	50	52	-2
2	48	46	2
3	45	50	-5
4	60	65	-5
5	70	78	-8
6	62	61	1
7	55	58	-3
8	62	70	-8
9	58	67	-9
10	53	65	-12
Trung bình	56.30	61.20	-4.90
Phương sai	57.57	97.07	20.10
§ộ lệch chuẩn	7.59	9.85	4.4833

μ_x Trung bình của 10 công nhân theo công nghệ cũ = 56,30

μ_y Trung bình của 10 công nhân theo công nghệ mới = 61,20

Đ là trung bình của tổng thể sai lệch $X - Y = - 4,9$

Sd là độ lệch chuẩn của tổng thể $X - Y = 4,4833$

Ta cần kiểm định giả thuyết sau:

$H_0: \mu_x - \mu_y = D_0 = 0$

$H_1: \mu_x - \mu_y \neq D_0 \neq 0$

Tính T kiểm định

$$T = \frac{\bar{D} - D_0}{\frac{Sd}{\sqrt{n}}} = \frac{4,9 - 0}{\frac{4,4833}{\sqrt{10}}} = \frac{4,9}{1,4177} = 3,456$$

Tim T lý thuyết với bậc tự do là 9; $\alpha = 0,025$: Ta tìm hàm $T_{INV}(9, 0,05) = 2,262$;

Nh vậy, $|T|$ kiểm định = 3,456 > T lý thuyết = 2,262 ta bác bỏ H_0 , nghĩa là năng suất lao động của công nhân sau khi áp dụng công nghệ mới khác với công nghệ cũ.

Vì $\bar{D} = 4,9 > D_0$ nên $\mu_x - \mu_y > 0$, nghĩa là ở mức ý nghĩa 5% áp dụng công nghệ mới đã làm tăng năng suất so với công nghệ cũ.

b. Trường hợp lấy mẫu độc lập

+ Bài toán:

Gia sử ta có n_x và n_y là số đơn vị mẫu được chọn ngẫu nhiên, độc lập từ hai tổng thể X và Y có phân phối chuẩn, thể hiện ở bảng sau:

Quan sát	X	Y
1	X1	Y1
2	X2	Y2
3	X3	Y3
.	.	.
.	.	.
n	Xn	Yn
Số quan sát	n_x	n_y
Trung bình mẫu	\bar{x}	\hat{y}
Trung bình	μ_x	μ_y
Phương sai	σ_x^2	σ_y^2
§ lệch chuẩn	σ_x	σ_y

μ_x Trung bình của tổng thể X
 μ_y Trung bình của tổng thể Y
 \bar{x} , \hat{y} là trung bình của 2 mẫu chọn ngẫu nhiên từ 2 tổng thể X ; Y
 σ_x^2 và σ_y^2 là phương sai của tổng thể X và Y
Với mức ý nghĩa α , ta cần kiểm định giả thuyết sau:
 $H_0: \mu_x - \mu_y = D_0$ (D0 là giá trị cho trước D0=0)
 $H_1: \mu_x - \mu_y \neq D_0$
Hay:
 $H_0: \mu_x - \mu_y = 0$; $H_1: \mu_x - \mu_y \neq 0$

+ Nguyên tắc kiểm định: có 2 trường hợp xảy ra

1). Nếu $n_x, n_y \geq 30$, với X, Y tuân theo phân phối chuẩn và $\delta^2_x \neq \delta^2_y$

. Tính tiêu chuẩn kiểm định Z (Z thực nghiệm)

Trong đó: D_0 : giá trị cụ thể cho trước ($D_0 = 0$)

$$Z = \frac{x - \hat{y} - D_0}{\sqrt{\frac{\delta^2_x}{n_x} + \frac{\delta^2_y}{n_y}}}$$

\bar{x}, \hat{y} là trung bình của 2 mẫu
 δ^2_x và δ^2_y là phương sai của tổng thể
 X và Y

n_x, n_y : số đơn vị mẫu quan sát của tổng thể X và Y

Z : Tiêu chuẩn kiểm định (Z thực nghiệm)

- T×m Z lý thuyết:

T×m $Z_{\alpha/2}$ bằng cách tra bảng hoặc dùng hàm NORMSINV với $\alpha/2$ trong EXCEL

+ Nếu $|Z| \leq Z_{\alpha/2}$ ta chấp nhận giả thuyết H_0 , coi $\mu_x - \mu_y = D_0$

+ Nếu $|Z| > Z_{\alpha/2}$ ta bác bỏ giả thuyết H_0 , coi $\mu_x - \mu_y \neq D_0$ và khi đó :

Nếu $x > \hat{y}$ ta xem $\mu_x > \mu_y$

Nếu $x < \hat{y}$ ta xem $\mu_x < \mu_y$

**Thí dụ: Một trại chăn nuôi gà tiến hành thí nghiệm sử dụng 2 loại thức ăn A và B trên cùng một giống. Sau một thời gian thử nghiệm cho ăn, người ta điều tra 50 con nuôi bằng thức ăn A và 40 con nuôi bằng thức ăn B thu được các số liệu sau:
Bảng : Một số chỉ tiêu của 2 mẫu thí nghiệm cho ăn 2 loại thức ăn A và B**

Diễn giải	ĐVT	Thức ăn A	Thức ăn B
1. Số đơn vị mẫu quan sát	con	50	40
2. Khối lượng trung bình 1 con	Kg/con	2,2	1,2
3. Độ lệch chuẩn	Kg/con	1,25	1,02

Yêu cầu: Anh chi hãy cho biết khối lượng trung bình 1 con sử dụng ở 2 loại thức ăn sau thời gian nuôi có khác nhau không với mức ý nghĩa là 5%?

Giai:

- Gọi μ_x và μ_y là khối lượng trung bình 1 con sau khi nuôi sử dụng thức ăn A và B;

- Đặt giả thuyết: $H_0 : \mu_x - \mu_y = 0$

$H_1 : \mu_x - \mu_y \neq 0$

- Tính tiêu chuẩn kiểm định Z

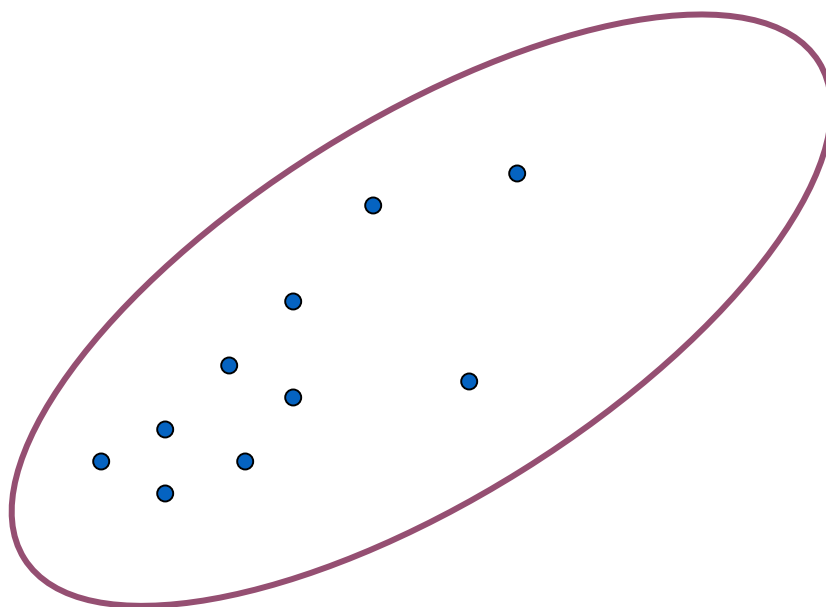
$$Z = \frac{\bar{x} - \bar{y} - D_0}{\sqrt{\frac{\delta_x^2}{n_x} + \frac{\delta_y^2}{n_y}}} = \frac{2,2 - 1,2 - 0}{\sqrt{\frac{1,25^2}{50} + \frac{1,02^2}{40}}} = \frac{1}{0,2392} = 4,179$$

- Tìm Z lý thuyết qua hàm NORMSINV với $\alpha = 0,025$ trong EXCEL ta đọc Z lý thuyết = 1,96.

- $|Z| = 4,179 > Z_{\alpha/2} = 1,96$ ta bác bỏ giả thuyết H_0 , coi $\mu_x - \mu_y \neq 0$.

$\mu_x = 2,2 \text{ kg/con} > \mu_y = 1,2 \text{ kg/con}$ nên ta xem $\mu_x > \mu_y$, chứng tỏ khối lượng trung bình 1 con nuôi bằng thức ăn A lớn hơn nuôi bằng thức ăn B.

c) Phân tích tương quan

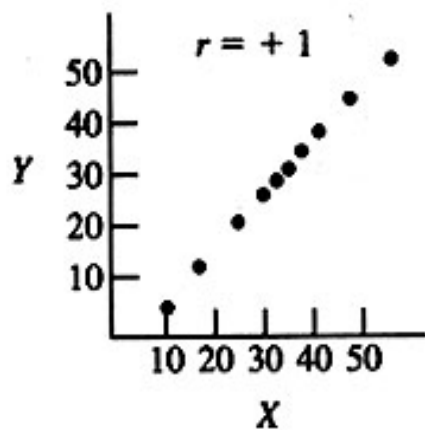


Quan hệ tương quan

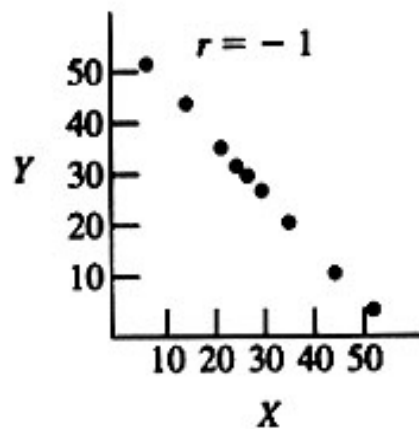
- Là quan hệ tuyến tính giữa 2 biến độc lập
- Có thể dự báo giá trị 1 biến khi biết giá trị của biến kia
- Hệ số tương quan nằm trong khoảng -1 đến 1
- Giá trị -1 và 1 là tương quan hoàn hảo, rất chặt, tuyến tính giữa 2 biến
- Giá trị 0 là không có tương quan, không có mối quan hệ

Tương quan – mức độ và xu hướng

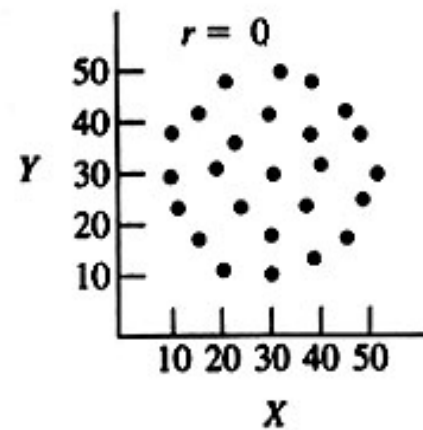
a.



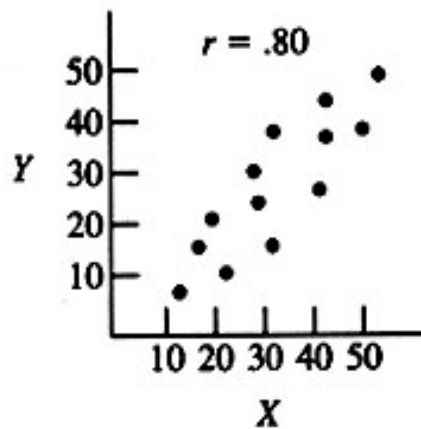
b.



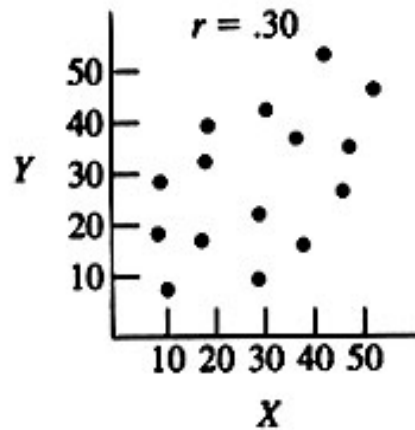
c.



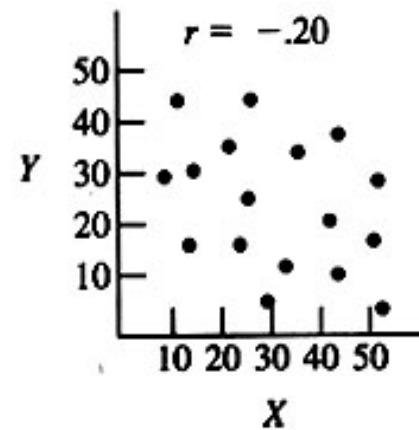
d.



e.



f.



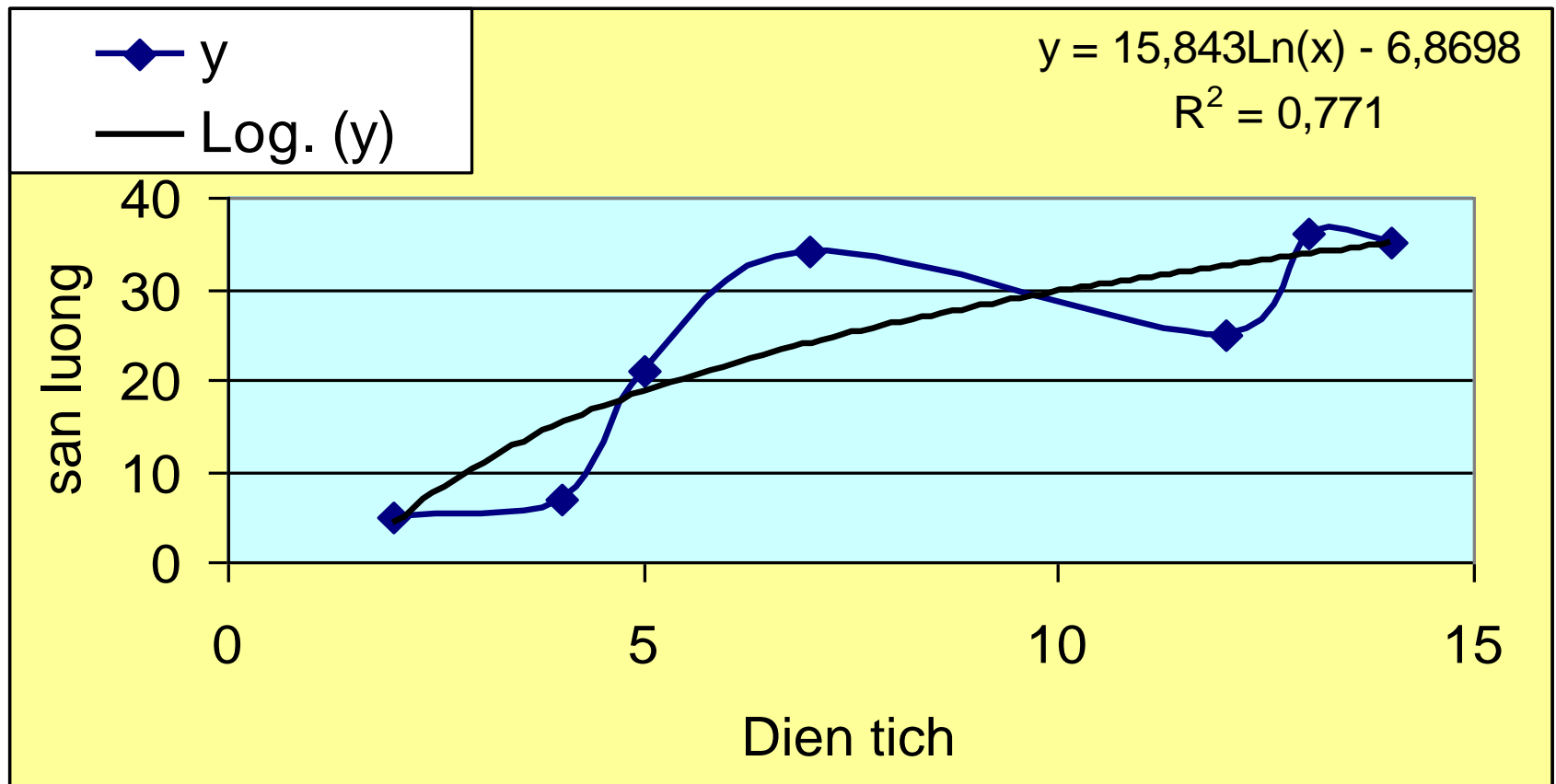
Phân tích tương quan

	<i>Xóm</i>	<i>Giới tính</i>	<i>Trình độ</i>	<i>Tổng diện</i>	<i>Tuổi</i>	<i>Số khẩu</i>	<i>lao động</i>	<i>tổng vốn</i>	<i>thu nhập 1 năm</i>
<i>Xóm</i>	1								
<i>Giới tính</i>	0.391	1.000							
<i>Trình độ</i>	-0.253	-0.218	1.000						
<i>Tổng diện</i>	-0.372	-0.377	0.087	1.000					
<i>Tuổi</i>	0.511	0.359	-0.312	-0.420	1.000				
<i>Số khẩu</i>	0.209	0.069	-0.389	-0.331	0.370	1.000			
<i>Số lao động</i>	0.444	0.215	-0.343	-0.423	0.435	0.797	1.000		
<i>tổng vốn</i>	-0.097	-0.309	0.261	0.347	-0.328	-0.010	0.048	1.000	
<i>thu nhập 1 năm</i>	0.120	0.001	0.064	-0.110	-0.183	0.208	0.481	0.749	

Ưu và nhược điểm của tương quan

- Có thể đo được mối quan hệ giữa các biến mà rất khó làm thực nghiệm (chỉ số IQ và tình trạng việc làm/ngành nghề nghiệp)
- Không xác định được mối quan hệ nhân – quả
- Có thể biến thứ 3 có ảnh hưởng
- Hướng có thể chưa rõ ràng
- R mang tính định tính

d) Phân tích hồi qui



<i>Regression Statistics</i>						
Multiple R	0.871779					
R Square	0.759998					
Adjusted R Square	0.729998					
Standard Error	10.03797					
Observations	19					
ANOVA						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	2	5105.162	2552.581	25.33305	1.1E-05	
Residual	16	1612.175	100.7609			
Total	18	6717.337				
Coefficients						
	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	
Intercept	-14.32992	6.906661	-2.074798	0.054496	-28.97139	0.311542
Số lao động	7.218543	1.982913	3.640373	0.002203	3.014955	11.42213
tổng vốn	0.512103	0.086259	5.936809	2.09E-05	0.329242	0.694964

Các phương pháp phân tích định lượng khác

1. Mô hình hóa (kết hợp KTL + Toán + các môn khác)
2. Mô hình tối ưu
3. Phân tích ngành hàng,
4. Phân tích ma trận chính sách (PAM)
5. Phân tích lợi thế so sánh
6. Phân tích tài chính

Xem lại các môn học liên quan và sẽ vận dụng ở chương sau

4.2.3. Phân tích thông tin thứ cấp

- Sử dụng thông tin thứ cấp cho nghiên cứu?

Số liệu thứ cấp

- Số liệu được thu thập không phải cho mục đích của NC hiện tại
- Phân tích số liệu/ thông tin đã có
- Số liệu thu thập cho sử dụng chung
- Nhà NC có ảnh hưởng nhỏ đến dạng số liệu (làm thế nào điều tra,...)
- Không nên nhầm với : ***Phân tích “thứ cấp”*** – là phân tích một phân tích khác – như kiểm tra tính chính xác

Số liệu thứ cấp

- Lợi ích
 - Chi phí và thời gian
 - Sự sẵn có
 - Đỡ tốn kém
 - Không đòi hỏi thời gian nhiều

Số liệu thứ cấp

- **Bất lợi:** Có thể không đáp ứng nhu cầu của NC hiện tại về
 - Đơn vị tính
 - Các khái niệm, định nghĩa khác
 - Thời gian
- **Hạn chế:** Tính chính xác
 - Ai thu thập?
 - Tại sao lại thu thập?
 - Thu thập thế nào?

4.3. Trình bày kết quả xử lý & phân tích thông tin

- (Xem lại Bài giảng Nguyên lý thống kê kinh tế)

Trình bày số liệu, thông tin

4 kiểu trình bày kết quả xử lý và phân tích

- Số liệu độc lập
- Bảng số liệu
- Sơ đồ, hình
- Đồ thị

Mục đích

- Tất cả các kiểu trình bày đều nhằm mục đích giúp người đọc dễ hiểu kết quả phân tích
- Đây chỉ là cách khác nhau – mục đích như nhau

Yêu cầu khi trình bày kết quả xử lý và phân tích thông tin

Mỗi hình thức trình bày đều có:

- Bao gồm tiêu đề mô tả
- Tên của tất cả các biến và đơn vị đo
- Các biến độc lập ở trục hoành
- Biến phụ thuộc ở trục tung
- Nguồn số liệu trong ghi chú
- Chỉ rõ kiểm định thống kê ở cuối bảng
- Luôn có số quan sát, số mẫu
- Chỉ rõ nếu dùng phần trăm.