



CHƯƠNG 2

CÁC KỸ THUẬT TIỀN XỬ LÝ DỮ LIỆU



NỘI DUNG CHƯƠNG II

- Làm sạch dữ liệu
- Tích hợp dữ liệu
- Biến đổi dữ liệu
- Thu giảm dữ liệu
- Rời rạc hóa dữ liệu



Các kỹ thuật tiền xử lý dữ liệu

Tiền xử lý dữ liệu là gì?

- Là quá trình xử lý dữ liệu thô/gốc ban đầu nhằm cải thiện chất lượng dữ liệu → cải thiện chất lượng của kết quả khai thác.
- Dữ liệu thô/gốc:
 - Có cấu trúc, bán cấu trúc, phi cấu trúc;
 - Được đưa vào từ các nguồn khác nhau.



Các kỹ thuật tiền xử lý dữ liệu

1. Làm sạch dữ liệu – Data cleaning

-CSDL ít nhiều mang tính không nhất quán => KTDL không đảm bảo tính đúng đắn.



Các kỹ thuật tiền xử lý dữ liệu

1. Làm sạch dữ liệu – Data cleaning

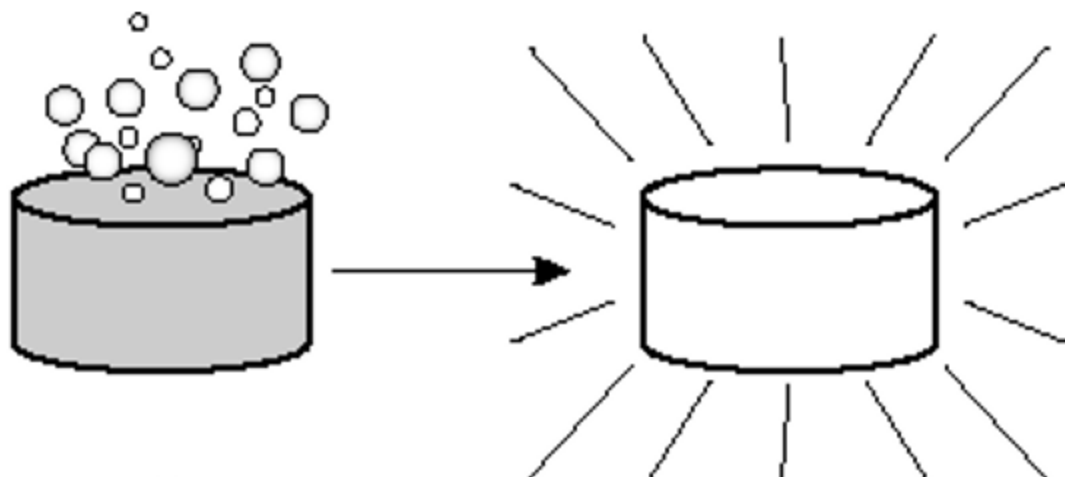
VD: Trong công ty bảo hiểm nhân thọ thì ngày sinh của khách hàng cần phải thật chính xác, trong khi đó có từ 30%-40% thông tin về tuổi khách hàng trong CSDL ngân hàng là để trống hoặc chứa dữ liệu sai => không thể khám phá ra mô hình quan hệ tuổi tác => xóa bỏ dữ liệu không cần thiết, dữ liệu sai càng nhiều càng tốt.

Các kỹ thuật tiền xử lý dữ liệu

1. Làm sạch dữ liệu – Data cleaning

- Là quá trình xử lý các giá trị thiếu, xác định và loại bỏ dữ liệu sai biệt, dữ liệu nhiễu và giải quyết dữ liệu mâu thuẫn.

Data cleaning





1. Làm sạch dữ liệu

Là quá trình xử lý các giá trị thiếu, loại bỏ dữ liệu sai biệt, dữ liệu nhiễu và giải quyết dữ liệu mâu thuẫn.

Dữ liệu thiếu:

- **Bỏ qua các mẫu tin có giá trị bị thiếu:** Phương pháp này dễ nhưng không hiệu quả đặc biệt khi tỉ lệ giá trị thiếu của trường nào đó cao.
- **Điền vào các giá trị thiếu bằng tay:** Phương pháp này thường tốn thời gian và có thể không khả thi cho một tập dữ liệu nguồn lớn với nhiều giá trị bị thiếu.
- **Điền giá trị thiếu tự động:** Thay thế bằng giá trị trung bình của thuộc tính, thay thế bằng giá trị có nhiều khả năng nhất.



1. Làm sạch dữ liệu

Dữ liệu nhiễu

- Nhiễu dữ liệu là một lỗi ngẫu nhiên hoặc sự ghi chép nhầm lẫn không được kiểm soát.
- Dùng các phương pháp đo lường sự tập trung của dữ liệu → xử lý nhiễu như: chia giỏ, hồi qui, gom cụm.



1. Làm sạch dữ liệu

Chống trùng lặp: xóa bỏ thông tin trùng của các bản ghi

Nguyên nhân:

- Sai sót trong phần nhập dữ liệu
- Thông tin không được cập nhật kịp thời
- Thông tin được cung cấp bị sai

...



1. Làm sạch dữ liệu

Ví dụ:

Bảng 1.3: Bảng kết quả thu thập dữ liệu

Makh	SoDonHang	MaSanPham	TenKH	DiaChiKH
23003	10021	90008	Toàn	số 1 Đường A
23003	10021	50034	Toàn	số 1 Đường A
<i>23003</i>	<i>10021</i>	<i>73006</i>	<i>Toàn</i>	<i>số 1 Đường A</i>
23009	22038	90008	Phương	số 4 Đường B
23013	23001	50034	Chính	số 9 Đường D
<i>23019</i>	<i>34062</i>	<i>73006</i>	<i>Tàn</i>	<i>số 1 Đường A</i>

Khách hàng Tàn và Toàn với hai mã khách hàng khác nhau nhưng lại có cùng giá trị địa chỉ, điều đó xác định họ chính là một người => có 1 bảng ghi sai chính tả tên khách hàng => loại bỏ trùng lặp => 2 bảng ghi này của cùng 1 khách hàng.



1. Làm sạch dữ liệu

Giới hạn vùng giá trị: Giá trị nằm ngoài miền giá trị cho phép (thông tin chứa các giá trị không hợp lệ theo một quy tắc nào đó) => mâu thuẫn trong các vùng dữ liệu (khó phát hiện)

Để sửa chữa mâu thuẫn, ta dùng giá trị NULL để thay thế vào những phần thông tin chưa biết trong các trường của bảng và cứ tiến hành sửa chữa các mâu thuẫn trong các vùng dữ liệu khác.



1. Làm sạch dữ liệu

Ví dụ:

Bảng 1.4: Bảng dữ liệu ví dụ

MaKh	SoDon Hang	MasanPham	TenKH	ĐịaChỉKh	Ngày DHang
23003	10021	90008	Toàn	số 1 Đường A	15/4/94
23003	10021	50034	Toàn	số 1 Đường A	21/6/93
23003	10021	73006	Toàn	số 1 Đường A	01/01/01
23009	22038	90008	Phượng	số 4 Đường B	30/5/92
23013	23001	50034	Chính	số 9 Đường D	31/12/95
23003	34062	73006	Toàn	số 1 Đường A	21/01/94

Trường NgayDHang có giá trị 01/01/01 => giá trị không hợp lý vì tại thời điểm đó siêu thị chưa thành lập => kiểm tra tính nhất quán miền giá trị => thay bằng giá trị NULL



1. Làm sạch dữ liệu

Bảng 1.5: Bảng dữ liệu ví dụ

MaKH	So DonHang	MaSanPham	TenKH	DiaChiKH	NgayDHng
23003	10021	90008	Toàn	số 1 Đường A	15/4/94
23003	10021	50034	Toàn	số 1 Đường A	21/6/93
23003	10021	73006	Toàn	số 1 Đường A	NULL
23009	22038	90008	Phương	số 4 Đường B	30/5/92
23013	23001	50034	Chính	số 9 Đường D	31/12/95



1. Làm sạch dữ liệu

Làm giàu dữ liệu: bổ sung thêm nhiều loại thông tin có liên quan vào CSDL gốc.

Cần có các CSDL khác bên ngoài có liên quan đến CSDL gốc ban đầu => bổ sung những thông tin cần thiết, làm tăng khả năng khám phá nhiều tri thức từ CSDL.



1. Làm sạch dữ liệu

Ví dụ:

Bảng 1.5: Bảng dữ liệu ví dụ

MaKH	So DonHang	MaSanPham	TenKH	DiaChiKH	NgayDHng
23003	10021	90008	Toàn	số 1 Đường A	15/4/94
23003	10021	50034	Toàn	số 1 Đường A	21/6/93
23003	10021	73006	Toàn	số 1 Đường A	NULL
23009	22038	90008	Phương	số 4 Đường B	30/5/92
23013	23001	50034	Chính	số 9 Đường D	31/12/95

Bảng 1.6: Bảng dữ liệu

MaKh	NgaySinh	ThuNhap	TinDung	SoHuu
23003	13/4/76	\$ 1,200,000.00	\$ 12,000,000.00	có
23009	20/10/71	\$ 3,600,000.00	\$ 50,000,000.00	có
23013	30/8/69	\$ 200,000.00	\$ 0.00	không
23021	4/9/67	\$ 500,000.00	\$ 200,000	không



1. Làm sạch dữ liệu

Vấn đề đặt ra là làm thế nào để kết hợp thông tin giữa dữ liệu gốc và dữ liệu được bổ sung. Bên cạnh đó cần lưu ý đến vấn đề khôi phục các quan hệ trong CSDL sau khi đã được làm giàu thông tin.



1. Làm sạch dữ liệu

- Phương pháp chia giỏ (Binning)
- **Chia theo độ rộng:**
 - Chia vùng giá trị thành N khoảng cùng kích thước, độ rộng của từng khoảng bằng (giá trị lớn nhất - giá trị nhỏ nhất)/N
 - **Ví dụ:** Có dãy 64 65 67 69 70 71 72 72 75 75 80 81 83 85

với $N = 7$ khoảng, Biên trái \leq giá trị $<$ biên phải

64 67	67 70	70 73	73 76	76 79	79 82	82 85
-------	-------	-------	-------	-------	-------	-------

Dãy mới 1 1 2 2 3 3 3 3 4 4 6 6 7 7 được chuẩn hóa



1. Làm sạch dữ liệu

- Phương pháp chia giỏ (Binning)
- **Chia theo độ rộng:**
 - Chia vùng giá trị thành N khoảng cùng kích thước, độ rộng của từng khoảng bằng (giá trị lớn nhất - giá trị nhỏ nhất)/N

Bài tập : Có dãy 64 65 67 69 70 71 72 72 75 75 80 81 83 84

với $N = 5$ khoảng, hãy sử dụng phương pháp chia giỏ theo độ rộng cho dãy trên, và cho biết dãy mới đã được chuẩn hóa.



1. Làm sạch dữ liệu

- **Chia theo độ rộng:**
 - Khi dữ liệu lệch nhiều thì sao?
 - Ví dụ: Mức lương trong một công ty

170 200 210 40040000 50000. Trong đó lương mức 170 200 chiếm hơn 80% → chia theo độ rộng không phù hợp.



1. Làm sạch dữ liệu

- Chia theo độ sâu:
 - Chia vùng giá trị thành N khoảng, mỗi khoảng có cùng số lượng mẫu (trừ khoảng cuối cùng)
 - Ví dụ: Dãy 4 8 15 21 21 24 25 28 34

4 8 15	21 21 24	25 28 34
--------	----------	----------

Dãy mới ???



1. Làm sạch dữ liệu

- **Tiến hành khử nhiễu (làm trơn dữ liệu):**
 - Khử nhiễu bằng giá trị trung vị: Biến đổi giá trị trong giỏ theo giá trị trung vị.
 - Khử nhiễu bằng giá trị trung bình: Biến đổi giá trị trong giỏ theo giá trị trung bình của các phần tử trong giỏ.
 - Khử nhiễu bằng giá trị biên: Xét các giá trị trong giỏ gần với biên nào nhất thì quy về biên đó.



1. Làm sạch dữ liệu

◦ **Ví dụ:** Dữ liệu về giá cả như sau (đơn vị \$)

◦ 4 8 15 21 21 24 25 28 34

◦ Chia giỏ theo độ sâu với $N = 3$

◦ Bin 1: 4 8 15

◦ Bin 2: 21 21 24

◦ Bin 3: 25 28 34

• **Khử nhiễu:** theo giá trị trung vị

• Bin 1: 8 8 8

• Bin 2: 21 21 21

• Bin 3: 28 28 28



1. Làm sạch dữ liệu

- **Ví dụ:** Dữ liệu về giá cả như sau (đơn vị \$)

- 4 8 15 21 21 24 25 28 34

- Chia giỏ theo độ sâu với $N = 3$

- Bin 1: 4 8 15

- Bin 2: 21 21 24

- Bin 3: 25 28 34

- **Khử nhiễu:** theo giá trị trung bình

- Bin 1: 9 9 9

- Bin 2: 22 22 22

- Bin 3: 29 29 29



1. Làm sạch dữ liệu

◦ **Ví dụ:** Dữ liệu về giá cả như sau (đơn vị \$)

◦ 4 8 15 21 21 24 25 28 34

◦ Chia giỏ theo độ sâu với $N = 3$

◦ Bin 1: 4 8 15

◦ Bin 2: 21 21 24

◦ Bin 3: 25 28 34

• **Khử nhiễu:** theo giá trị biên

• Bin 1: 4 4 15

• Bin 2: 21 21 24

• Bin 3: 25 25 34



1. Làm sạch dữ liệu

Bài tập: Cho dữ liệu giá cả (\$) 15, 17, 19, 25, 29, 31, 33, 41, 42, 45, 45, 47, 52, 52, 64

a/ Dùng phương pháp chia giỏ theo độ rộng và độ sâu với $N = 4$.

Tiến hành khử nhiễu theo trung vị, trung bình, biên.

a/ Dùng phương pháp chia giỏ theo độ rộng và độ sâu với $N = 3$.

Tiến hành khử nhiễu theo trung vị, trung bình, biên.



1. Làm sạch dữ liệu

Dữ liệu nhiễu

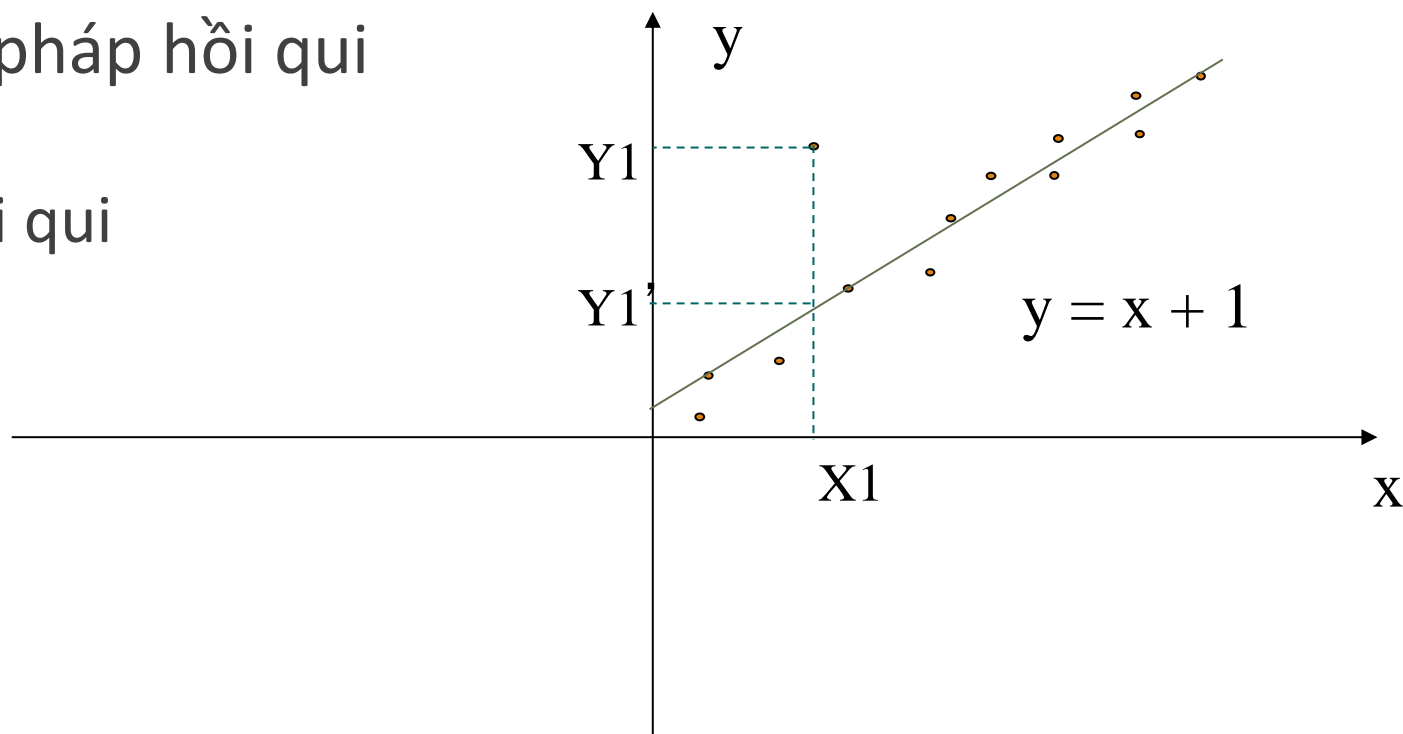
- Phương pháp hồi qui
 - Là phương pháp ghép một phương trình hồi qui vào một tập hợp các điểm số liệu.
 - Dùng để tìm ra được một mối quan hệ tốt nhất giữa hai thuộc tính (hoặc biến), từ đó một thuộc tính có thể dùng để dự đoán thuộc tính khác.



1. Làm sạch dữ liệu

Dữ liệu nhiễu

- Phương pháp hồi qui
- Hàm hồi qui





1. Làm sạch dữ liệu

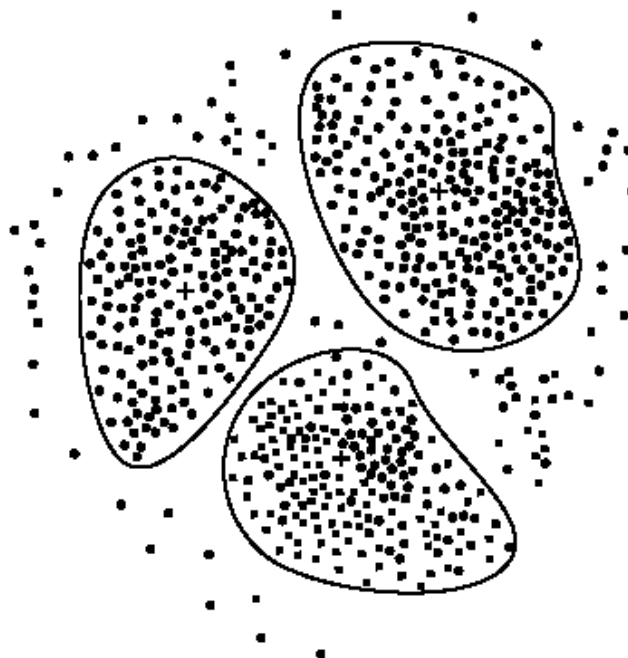
Dữ liệu nhiễu

- Phương pháp gom cụm
- Các giá trị tương tự nhau được tổ chức thành các nhóm hay “cụm”. Các giá trị rơi ra bên ngoài các nhóm này sẽ được xem xét để làm mịn, sau đó đưa chúng vào cụm phù hợp.

1. Làm sạch dữ liệu

Dữ liệu nhiễu

- Phương pháp gom cụm

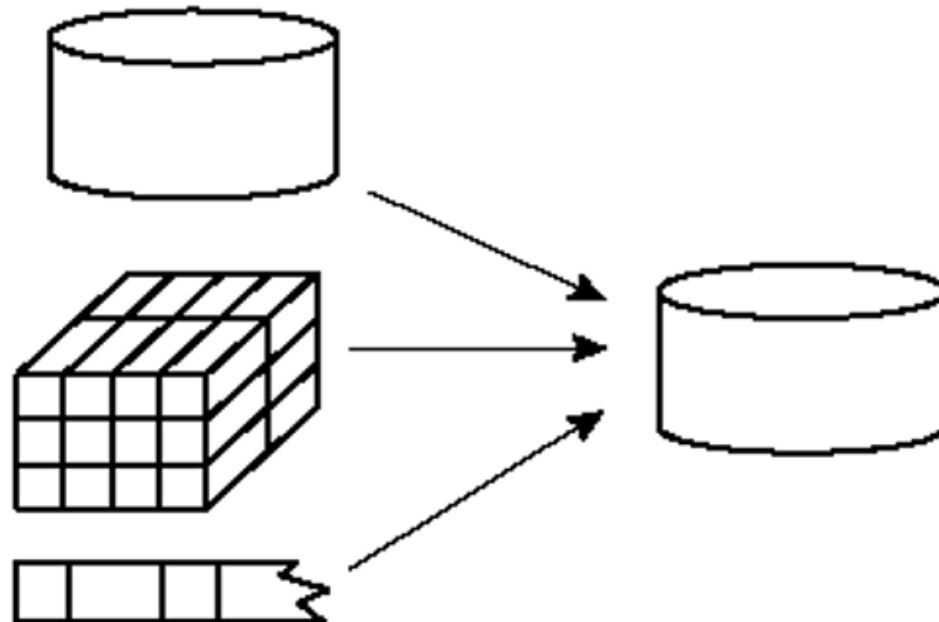


2. Tích hợp dữ liệu

2. Tích hợp dữ liệu – Data integration

- Tổng hợp, tích hợp dữ liệu từ nhiều CSDL, tập tin dữ liệu khác nhau.

Data integration





2. Tích hợp dữ liệu

Là quá trình trộn dữ liệu từ các nguồn khác nhau vào một kho dữ liệu sẵn sàng cho quá trình khai thác.

Để đảm bảo tính tương đương của thông tin giữa các nguồn khi tích hợp cần xem xét các vấn đề sau:

- Vấn đề nhận dạng thực thể
- Vấn đề dư thừa
- Vấn đề mâu thuẫn giá trị dữ liệu



2. Tích hợp dữ liệu

- **Vấn đề nhận dạng thực thể**
 - Cùng một thực thể nhưng đôi khi có sự **biểu diễn khác nhau** trong nhiều nguồn dữ liệu.
 - **Ví dụ:**
 - Ở mức lược đồ thuộc tính id của khách hàng trong một cơ sở dữ liệu A và số hiệu custom trong một cơ sở dữ liệu B chỉ một thuộc tính.
 - Ở mức thể hiện giá trị $\{male, female\}$ của trường giới tính trong cơ sở dữ liệu A và $\{Nam, Nữ\}$ trong cơ sở dữ liệu B.



2. Tích hợp dữ liệu

- **Vấn đề dư thừa**

- Ví dụ thuộc tính doanh thu hàng năm và thuộc tính doanh thu quý trong một CSDL là dư thừa.
- **Phân tích sự tương quan:** chỉ ra mức độ một thuộc tính phụ thuộc vào thuộc tính kia → **Sự tương quan giữa hai thuộc tính cao có thể bỏ bớt 1 thuộc tính.**
- Với thuộc tính số: Đánh giá sự tương quan giữa hai thuộc tính A và B bằng cách tính hệ số tương quan theo công thức Karl Pearson.



2. Tích hợp dữ liệu

- Phân tích sự tương quan:
- Với thuộc tính số:

Công thức Karl Pearson:

$$r_{A,B} = \frac{n \sum AB - (\sum A)(\sum B)}{\sqrt{(n(\sum A^2) - (\sum A)^2)(n(\sum B^2) - (\sum B)^2)}}$$

Ta có $-1 \leq r_{A,B} \leq 1$



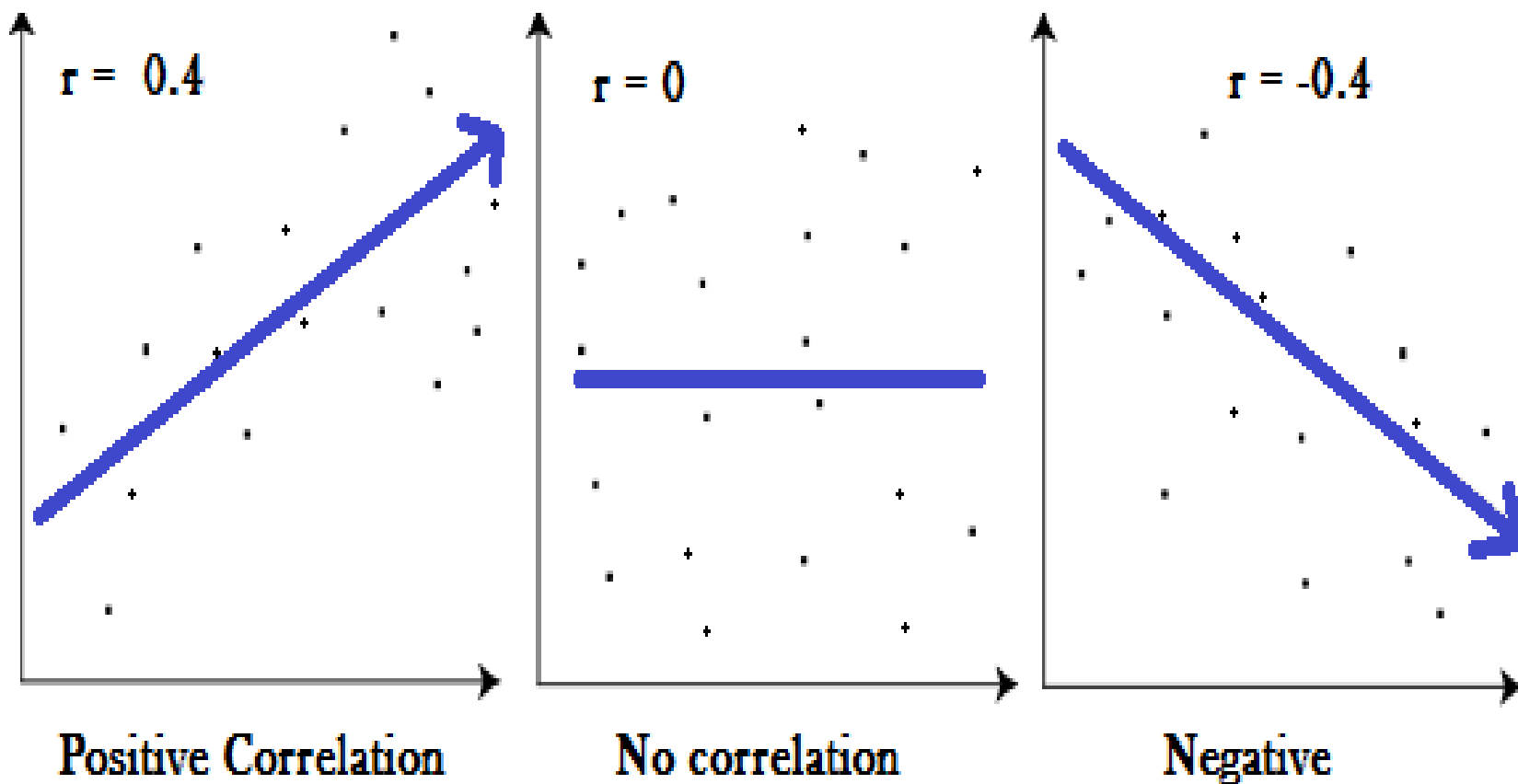
2. Tích hợp dữ liệu

Nếu $r_{A,B}$ lớn hơn 0, thì A và B có khả năng có mối liên hệ tương quan với nhau.

Nếu $r_{A,B}$ bằng 0 thì A và B là độc lập với nhau và giữa chúng không có mối quan hệ nào.

Nếu $r_{A,B}$ nhỏ hơn 0 thì A và B có mối quan hệ tương quan nghịch.

2. Tích hợp dữ liệu





2. Tích hợp dữ liệu

- Phân tích sự tương quan:
 - Với thuộc tính số:
 - Ví dụ: Xét độ tương quan của hai thuộc tính tuổi tác (X) và mức đường (Y) trong bảng

ĐỐI TƯỢNG	X	Y
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81



2. Tích hợp dữ liệu

- Tính các giá trị XY , X^2 , Y^2 , và giá trị tổng các cột

ĐỐI TƯỢNG	X	Y	XY	X^2	Y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
Tổng	247	486	20485	11409	40022



2. Tích hợp dữ liệu

$$r_{A,B} = \frac{n \sum AB - (\sum A)(\sum B)}{\sqrt{(n(\sum A^2) - (\sum A)^2)(n(\sum B^2) - (\sum B)^2)}}$$

$$\begin{aligned} r_{A,B} &= \frac{6 \times 20485 - 247 \times 486}{\sqrt{(6 \times 11409 - 61009)(6 \times 40022 - 236196)}} \\ &= 0,209 \end{aligned}$$

Với $r_{A,B} = 0,209$ nghĩa là thuộc tính tuổi tác có liên quan đều thuộc tính đường



2. Tích hợp dữ liệu

- **Bài tập 1:** Độ tuổi của 10 cặp vợ chồng được cho trong bảng sau. Hãy phân tích sự tương quan theo công thức Karl Pearson:

Husband (x)	36	72	37	36	51	50	47	50	37	41
Wife (y)	35	67	33	35	50	46	47	42	36	41



2. Tích hợp dữ liệu

- Phân tích sự tương quan:

- Với nguồn dữ liệu rời rạc:

Sự tương quan giữa hai thuộc tính A và B có thể được khám phá qua phép kiểm chi – bình phương (χ^2).

Phép kiểm chi – bình phương kiểm tra **giả thuyết liệu A và B có độc lập với nhau không**.

Nếu giả thuyết bị loại bỏ thì A và B có sự liên hệ với nhau dựa trên thống kê.



2. Tích hợp dữ liệu

Giả sử A có c giá trị rời rạc được ký hiệu là a_1, a_2, \dots, a_c . B có r giá trị rời rạc ký hiệu b_1, b_2, \dots, b_r .

Bảng biểu diễn mối quan hệ A và B có thể được xây dựng như sau:

- c giá trị của A tạo thành cột
- r giá trị của B tạo thành hàng.



2. Tích hợp dữ liệu

Giá trị χ^2 được tính như sau

$$\chi^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Trong đó:

o_{ij} là tần xuất quan sát được (đếm) các trường hợp (A_i, B_j)

e_{ij} là tần xuất mong đợi được tính toán dựa trên phân bố dữ liệu của hai thuộc tính A_i, B_j theo công thức



2. Tích hợp dữ liệu

$$e_{ij} = \frac{\text{count}(A = a_i) \times \text{count}(B = b_j)}{N}$$

Với N là tổng số bộ, $\text{count}(A = a_i)$ là tổng số bộ có chứa giá trị a_i cho thuộc tính A , $\text{count}(B = b_j)$ là tổng số bộ có chứa trị b_j cho thuộc tính B .



2. Tích hợp dữ liệu

- Độ tự do (degree of freedom): $(r-1)*(c-1)$
- Tra bảng phân bố chi-square để xác định giá trị χ^2 .
- Nếu giá trị tính toán được lớn hơn hay bằng giá trị tra bảng thì giả thuyết hai thuộc tính A và B độc lập nhau bị bác bỏ.



2. Tích hợp dữ liệu

Ví dụ:

Giả sử có một nhóm 1500 người được khảo sát. Giới tính của họ được ghi nhận sau đó họ sẽ được hỏi về thể loại sách yêu thích thuộc hai dạng hư cấu và viễn tượng.

Như vậy ở đây có hai thuộc tính “giới tính” và “sở thích đọc”



2. Tích hợp dữ liệu

Số lần xuất hiện các trường hợp cho trong bảng sau

	Nam	Nữ	Tổng
Hư cấu	250	200	450
Viễn tưởng	50	1000	1050
Tổng	300	1200	1500



2. Tích hợp dữ liệu

Tính:

$$e_{11} = \frac{\text{count}(\text{Nam}) \times \text{count}(\text{Hư cầu})}{N} = \frac{300 \times 450}{1500} = 90$$

$$e_{12} = \frac{\text{count}(\text{Nữ}) \times \text{count}(\text{Hư cầu})}{N} = \frac{1200 \times 450}{1500} = 360$$



2. Tích hợp dữ liệu

	Nam	Nữ	Tổng
Hư cấu	250 (90)	200 (360)	450
Viễn tưởng	50 (210)	1000 (840)	1050
Tổng	300	1200	1500

$$\begin{aligned}\chi^2 &= \frac{(250-90)^2}{90} + \frac{(50-210)^2}{210} + \frac{(200-360)^2}{360} + \frac{(1000-840)^2}{840} \\ &= 507.93\end{aligned}$$



2. Tích hợp dữ liệu

Với bậc tự do $df = (r-1)(c-1) = (2-1)(2-1) = 1$

Giá trị χ^2 cần để bác bỏ giả thiết này ở mức α được chọn là 1% (nghĩa là độ tin cậy là 99%) tra bảng phân phối chi – square được giá trị là 6.63. Giá trị tính được $>$ giá trị tra bảng \rightarrow bác bỏ giả thuyết.

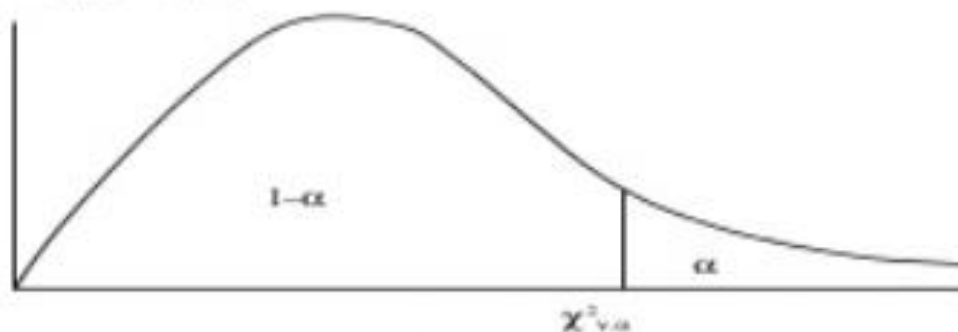
\rightarrow Hai thuộc tính này có một quan hệ tương quan khá mạnh trong nhóm người được khảo sát (nghĩa là sở thích đọc liên quan chặt chẽ với giới tính).

BẢNG TRA PHÂN PHỐI CHI-SQUARE

(Cho $\alpha=0.1$, bậc tự do $=6$, giá trị Chi-square α sẽ là 10.64.

Ý nghĩa: $P(\text{Chi-square} > \text{Chi-square}_{\alpha}) = \alpha$

Mật độ XS



Bậc tự do	Chi-Square Alpha									
v	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	3.93E-05	1.57E-04	9.82E-04	3.93E-03	0.01579	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.1026	0.2107	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.610	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81	18.55
7	0.989	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80



2. Tích hợp dữ liệu

- Phân tích sự tương quan
- **Bài tập:** phân tích sự tương quan của bảng số liệu sau với giả thuyết thuộc tính A độc lập thuộc tính B là $\alpha = 5\%$ bằng chi – bình phương.

$= 0.05$

	A	B
T/c 1	6	6
T/c 2	14	15
T/c 3	4	3



2. Tích hợp dữ liệu

- **Vấn đề mâu thuẫn giá trị dữ liệu**
 - Cùng một thực thể, các giá trị thuộc tính đến từ các nguồn khác nhau có thể khác nhau về cách biểu diễn hay mã hóa.
- **Ví dụ:**
 - Dữ liệu ngày tháng 31/12/2016 hay 12/31/2016
 - Dữ liệu logic là {1, 0} hay {yes, no}.
 - Thuộc tính giá trong các hệ thống tiền tệ khác nhau.