

DỮ LIỆU (DATA)

I. Dữ liệu

Dữ liệu là các sự kiện và con số được thu thập, phân tích và tóm tắt nhằm phục vụ cho việc trình bày và diễn giải.

Toàn bộ dữ liệu được thu thập trong một nghiên cứu cụ thể được gọi là **tập dữ liệu (data set)** của nghiên cứu đó.

Company	Ticker	Sector	Volume Traded (shares)	Price (VND)	Market Capitalization (Billion VND)	Price Change (%)	Date
VinGroup	VIC	Real Estate	2,500,000	100,000	250,000	1.5	2024-03-01
Vietcombank	VCB	Banking	3,500,000	120,000	420,000	-0.5	2024-03-01
HoaPhat	HPG	Steel	4,200,000	55,000	300,000	0.3	2024-03-01
Masangroup	MSN	Food & Beverage	3,000,000	150,000	120,000	2.0	2024-03-01
BIDV	BID	Banking	2,800,000	42,000	220,000	-1.2	2024-03-01

Dưới đây là **Bảng 1.1** minh họa **dữ liệu giao dịch cổ phiếu** tại **Sở Giao dịch Chứng khoán Việt Nam (VNX)** cho tháng 3 năm 2024:

Giải thích:

- Volume Traded (shares):** Số lượng cổ phiếu giao dịch.
- Ticker:** Mã cổ phiếu
- Sector:** Ngành
- Price (VND):** Giá cổ phiếu tại thời điểm giao dịch.
- Market Capitalization (Billion VND):** Vốn hóa thị trường của công ty (theo tỷ lệ cổ phiếu phát hành và giá cổ phiếu).
- Price Change (%):** Tỷ lệ thay đổi giá cổ phiếu so với ngày hôm trước.
- Date:** Ngày giao dịch.

Theo báo cáo cập nhật từ **Sở Giao dịch Chứng khoán London (LSE)** tháng 3/2024, tổng khối lượng giao dịch cổ phiếu đạt **hơn 1.1 nghìn tỷ GBP**, với **các ngành năng lượng và AI tăng**

trưởng mạnh nhất.

Tập dữ liệu được sử dụng trong báo cáo này bao gồm:

- Tên công ty niêm yết
- Tổng giá trị giao dịch trong tháng
- Biến động giá cổ phiếu
- Tỷ lệ tăng trưởng so với tháng trước

“ Những tập dữ liệu như vậy đang ngày càng trở nên phổ biến nhờ sự phát triển của **dữ liệu lớn (big data)** và **AI phân tích dữ liệu (data analytics AI)**.

II. Các yếu tố (Elements), Biến số (Variables), Quan sát (Observations)

Các yếu tố (Elements): Là các thực thể mà dữ liệu được thu thập trên đó. Trong bảng dữ liệu này, mỗi công ty niêm yết là một yếu tố (VinGroup, Vietcombank, HoaPhat...). Có 5 công ty nên bộ dữ liệu chứa 5 yếu tố.

Biến số (Variables): Là các đặc tính quan tâm của các yếu tố. Bảng này gồm 7 biến:

- Công ty (Company)
- Mã chứng khoán (Ticker)
- Ngành (Sector)
- Khối lượng giao dịch (Volume Traded - shares)
- Giá (Price - VND)
- Vốn hóa thị trường (Market Capitalization - Billion VND)
- Thay đổi giá (%) (Price Change)
- Ngày (Date)

Quan sát (Observations): Là tập hợp các giá trị đo lường thu thập được cho một yếu tố cụ thể. Mỗi hàng trong bảng là một quan sát. Ví dụ quan sát đầu tiên (VinGroup) có các giá trị: VIC, Real Estate, 2,500,000, 100,000, 250,000, 1.5%, 2024-03-01.

Phân tích bảng dữ liệu

- Các yếu tố (Elements):** 5 công ty niêm yết trên sàn chứng khoán Việt Nam
 - VinGroup (VIC)
 - Vietcombank (VCB)
 - HoaPhat (HPG)
 - Masangroup (MSN)
 - BIDV (BID)
- Các biến số (Variables):**
 - Biến định tính (Qualitative):**

- Company: Tên công ty
- Ticker: Mã chứng khoán
- Sector: Ngành hoạt động
- Date: Ngày giao dịch
- **Biến định lượng (Quantitative):**
 - Volume Traded: Khối lượng cổ phiếu giao dịch (liên tục)
 - Price: Giá cổ phiếu (liên tục)
 - Market Capitalization: Vốn hóa thị trường (liên tục)
 - Price Change: % thay đổi giá (liên tục)
- 3. **Quan sát (Observations):** 5 quan sát tương ứng với 5 công ty
- 4. **Phân tích ngành:**
 - 2 công ty ngành ngân hàng (VCB, BID)
 - 1 công ty bất động sản (VIC)
 - 1 công ty thép (HPG)
 - 1 công ty thực phẩm & đồ uống (MSN)
- 5. **Giá trị nổi bật:**
 - Giá cao nhất: MSN (150,000 VND)
 - Giá thấp nhất: BID (42,000 VND)
 - Khối lượng giao dịch lớn nhất: HPG (4,200,000 shares)
 - Vốn hóa lớn nhất: VCB (420,000 tỷ VND)
 - Tăng giá mạnh nhất: MSN (+2.0%)
 - Giảm giá nhiều nhất: BID (-1.2%)

III. Thang đo trong thống kê và phân tích dữ liệu

Việc thu thập dữ liệu đòi hỏi phải xác định thang đo phù hợp, bao gồm: định danh (nominal), thứ bậc (ordinal), khoảng cách (interval) hoặc tỷ lệ (ratio). Thang đo quyết định lượng thông tin chứa trong dữ liệu và giúp lựa chọn phương pháp tổng hợp, phân tích thống kê phù hợp.

1. Thang đo định danh (Nominal Scale)

- **Định nghĩa:** Dùng để phân loại dữ liệu dựa trên **nhãn hoặc tên gọi**, không có thứ tự hay giá trị số.
- **Đặc điểm:**
 - Chỉ phân biệt các nhóm, không so sánh hơn/kém.
 - Có thể dùng mã số thay cho nhãn (ví dụ: 1 = VinGroup, 2 = Vietcombank), nhưng con số không mang ý nghĩa toán học.
- **Ví dụ trong bảng 1.1:**
 - *Company* (tên công ty), *Ticker* (mã cổ phiếu), *Sector* (ngành) – đều là biến **định danh**.
 - Ví dụ: *Sector* gồm "Real Estate", "Banking", "Steel"... chỉ phân loại, không xếp hạng.

2. Thang đo thứ bậc (Ordinal Scale)

- **Định nghĩa:** Dữ liệu có thể **sắp xếp theo thứ tự**, nhưng khoảng cách giữa các hạng mục không đồng nhất.

- **Đặc điểm:**

- Có thể dùng số để mã hóa (ví dụ: 1 = Rất tốt, 2 = Tốt, 3 = Kém), nhưng phép trừ/chia giữa các số vô nghĩa.
- Đánh giá dịch vụ: "Xuất sắc", "Khá", "Trung bình".
- Nếu có biến như *Xếp hạng rủi ro* (AAA, AA, B), đó là thang **thứ bậc**.

- **Ví dụ**

Rank	Company	...
1	VinGroup	
2	Vietcombank	
3	HoaPhat	

Rank – là thứ hạng của công ty theo tổng giá trị giao dịch (Total Trading Value).

Mặc dù **VinGroup** xếp hạng 1 và **Vietcombank** xếp hạng 2, chúng ta không biết **VinGroup** vượt bao nhiêu phần trăm so với Vietcombank – chỉ biết là đúng **cao hơn**.

3. Thang đo khoảng cách (Interval Scale)

- **Định nghĩa:** Dữ liệu có thứ tự và khoảng cách giữa các giá trị có ý nghĩa, nhưng không có điểm "0 tuyệt đối".

- **Đặc điểm:**

- Phép cộng/trừ có nghĩa, nhưng phép nhân/chia không hợp lý.
- Ví dụ: Nhiệt độ (°C), điểm GMAT.

- **Ví dụ trong bảng:**

- *Price Change (%)* (thay đổi giá): Có thể tính chênh lệch (ví dụ: MSN tăng 2.0%, BID giảm 1.2% → khoảng cách MSN và BID là 3.2%), nhưng không thể nói "tăng gấp đôi".

4. Thang đo tỷ lệ (Ratio Scale)

- **Định nghĩa:** Có đủ tính chất của thang khoảng cách và có điểm 0 tuyệt đối, cho phép so sánh tỷ lệ.

- **Đặc điểm:**

- Có thể dùng mọi phép toán (+, -, ×, ÷).
- Ví dụ: Chiều cao, cân nặng, doanh thu.

- **Ví dụ trong bảng:**

- *Price* (giá cổ phiếu): 100,000 VND gấp đôi 50,000 VND.
- *Volume Traded* (khối lượng giao dịch): 2,500,000 cổ phiếu có thể so sánh tỷ lệ với 3,500,000 cổ phiếu.
- *Market Cap* (vốn hóa): 250,000 tỷ VND = 2.5 lần 100,000 tỷ VND.

Lưu ý quan trọng

- Thang đo **tỷ lệ (ratio)** là mạnh nhất, cho phép sử dụng mọi phương pháp thống kê.
- Biến **định danh** và **thứ bậc** thường dùng cho phân loại hoặc kiểm định phi tham số (ví dụ: Kiểm định Chi-square).

- Hiểu đúng thang đo giúp chọn **đồ thị** phù hợp (ví dụ: biểu đồ cột cho nominal, biểu đồ xếp hạng cho ordinal).

Tác giả: Đỗ Ngọc Tú
Công Ty Phần Mềm VHTSoft

Phiên bản #3

Được tạo 24 tháng 4 2025 01:54:27 bởi Đỗ Ngọc Tú

Được cập nhật 24 tháng 4 2025 10:15:13 bởi Đỗ Ngọc Tú