

Tóm tắt dữ liệu định tính

Dữ liệu định tính (qualitative data) là loại dữ liệu mô tả các đặc điểm hoặc thuộc tính mà không thể đo lường bằng số. Ví dụ trong thực tế bao gồm loại phương tiện giao thông ưa thích, thương hiệu điện thoại di động, hoặc sự hài lòng của khách hàng qua khảo sát.

Mặc dù **dữ liệu định tính không thể đo lường bằng số**, nhưng **chúng ta vẫn có thể đếm số lần một giá trị xuất hiện** – và đó chính là “tần suất” (frequency).

Định tính (Qualitative): Những yếu tố không thể đo đếm trực tiếp mà mô tả đặc điểm, tính chất hoặc phân loại. Ví dụ, các thương hiệu xe như Toyota, Kia, Hyundai, Honda, và VinFast thuộc về nhóm **định tính** vì chúng đại diện cho các phân loại hoặc nhóm sản phẩm, không thể đo đếm số lượng trực tiếp từ chính bản thân tên thương hiệu mà phải thông qua các chỉ số như doanh thu, số lượng xe bán ra của từng hãng

Để hiểu và phân tích dữ liệu định tính một cách hiệu quả, chúng ta thường sử dụng các phương pháp tóm tắt như **bảng phân phối tần suất**, **biểu đồ cột** (bar chart), hoặc **biểu đồ tròn** (pie chart). Những phương pháp này giúp thể hiện số lượng hoặc tỷ lệ xuất hiện của từng hạng mục một cách trực quan, dễ hiểu.

1. Phân phối tần suất (Frequency Distribution)

Định nghĩa

Phân phối tần suất là một bảng tóm tắt dữ liệu, hiển thị số lượng (tần suất) các mục rơi vào từng nhóm (lớp) không trùng lặp.

Hiểu đơn giản: Nó giúp bạn trả lời câu hỏi “**Có bao nhiêu lần mỗi mục xuất hiện?**” trong một danh sách.

Tại sao dùng Frequency Distribution?

- **Dễ đọc - dễ hiểu** hơn danh sách dữ liệu thô.
- **So sánh nhanh** giữa các nhóm hoặc loại dữ liệu.
- Là **nền tảng** để xây dựng biểu đồ như bar chart, pie chart, histogram...

Ví dụ

Phân phối tần suất theo thương hiệu xe bán ra trong 1 tháng (tháng 3/2025) tại TP.HCM

Bối cảnh

Dữ liệu được thu thập từ một số đại lý ô tô lớn tại TP.HCM, ghi nhận **50 xe ô tô mới được bán ra trong tháng 3 năm 2025**, thuộc 5 thương hiệu phổ biến:

- Toyota
- Kia
- Hyundai
- Honda
- VinFast

2. Tần suất tương đối

Định nghĩa: Là **tỷ lệ** (hoặc phần của tổng thể) mà một giá trị xuất hiện trong dữ liệu.

Công thức:

$$\text{Tần suất tương đối} = \frac{\text{Tần suất}}{\text{Tổng số quan sát}}$$

3. Tần suất phần trăm

Định nghĩa: Là tần suất tương đối nhưng được chuyển thành phần trăm.

$$\text{Tần suất phần trăm} = \text{Tần suất tương đối} \times 100\%$$

Giả sử bạn có dữ liệu bán xe ô tô tại TP.HCM trong **tháng 3/2025**, gồm 50 chiếc xe bán ra:

Hãng xe	Tần suất	Tần suất tương đối	Tần suất phần trăm
Toyota	16	$16/50 = 0.32$	32%
Kia	11	$11/50 = 0.22$	22%
Hyundai	10	$10/50 = 0.20$	20%
Honda	7	$7/50 = 0.14$	14%
VinFast	6	$6/50 = 0.12$	12%

4. Phân biệt Tần suất phân phối, Tần suất tương đối, Tần suất phần trăm

Loại	Nội dung
Phân phối	Số lần xuất hiện (vd: Toyota bán được 16 chiếc)

Loại	Nội dung
Tương đối	Tỷ lệ trong tổng số (vd: Toyota chiếm 32% trong 50 xe bán ra)
Phần trăm	Tần suất tương đối chuyển thành phần trăm (vd: 32%)

Biểu đồ cột (Bar chart) và biểu đồ tròn (Pie chart)

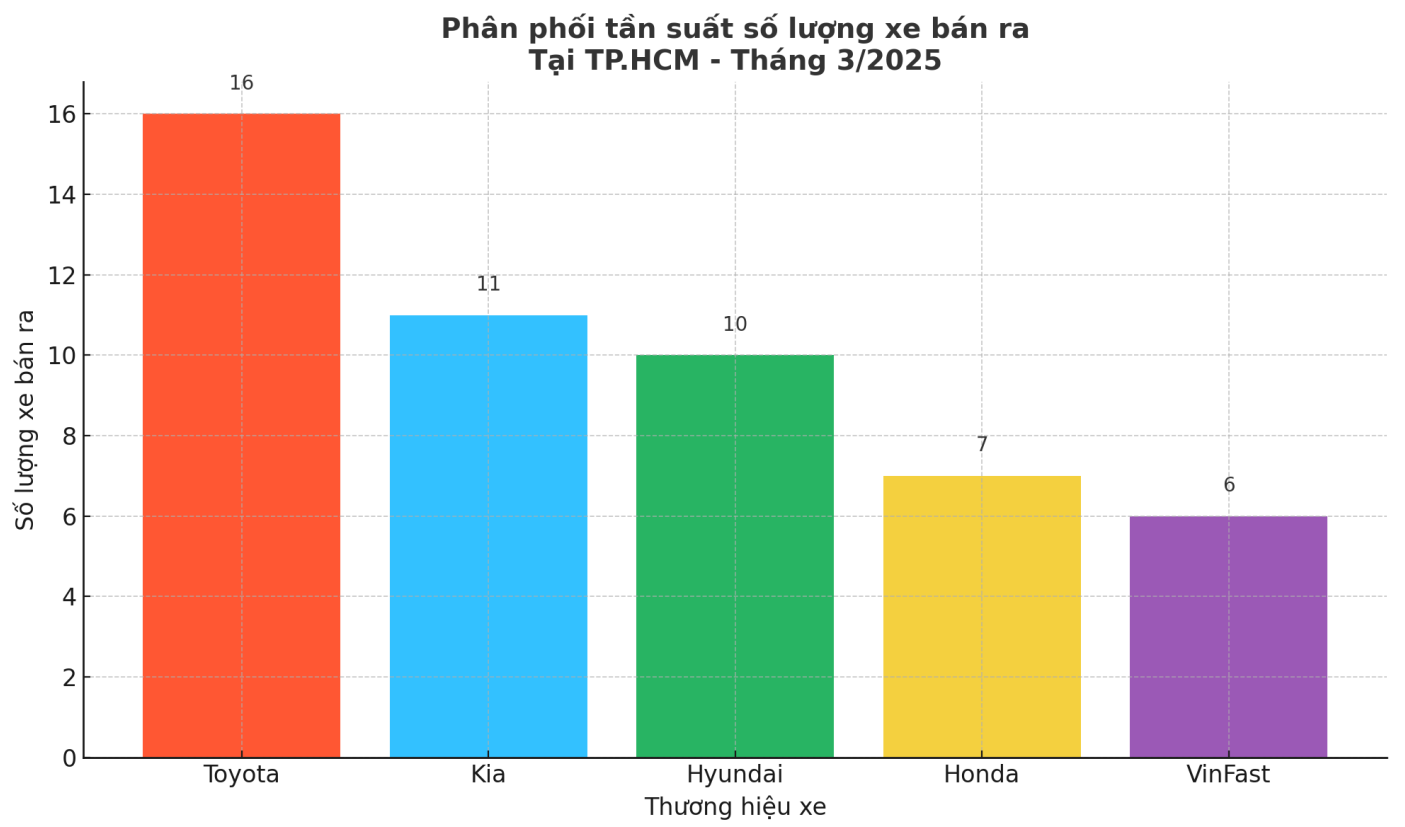
Biểu đồ cột (Bar chart)

Biểu đồ cột, hay còn gọi là **biểu đồ thanh**, là một cách thể hiện trực quan của bảng phân phối tần suất, tần suất tương đối, hoặc tần suất phần trăm.

Trên một trục của biểu đồ (thường là **trục ngang**), ta ghi nhãn cho các **nhóm (loại)** dữ liệu. Trục còn lại (thường là **trục dọc**) sẽ thể hiện giá trị **tần suất**, **tần suất tương đối** hoặc **tần suất phần trăm**.

Sau đó, ta vẽ các **thanh dọc có độ rộng bằng nhau** trên từng nhãn, chiều cao của mỗi thanh sẽ tương ứng với giá trị tần suất của nhóm đó.

Với dữ liệu định tính (ví dụ: thương hiệu xe, ngành nghề...), các cột nên được **cách nhau** để nhấn mạnh rằng các nhóm là **riêng biệt**.



Biểu đồ tròn (Pie chart)

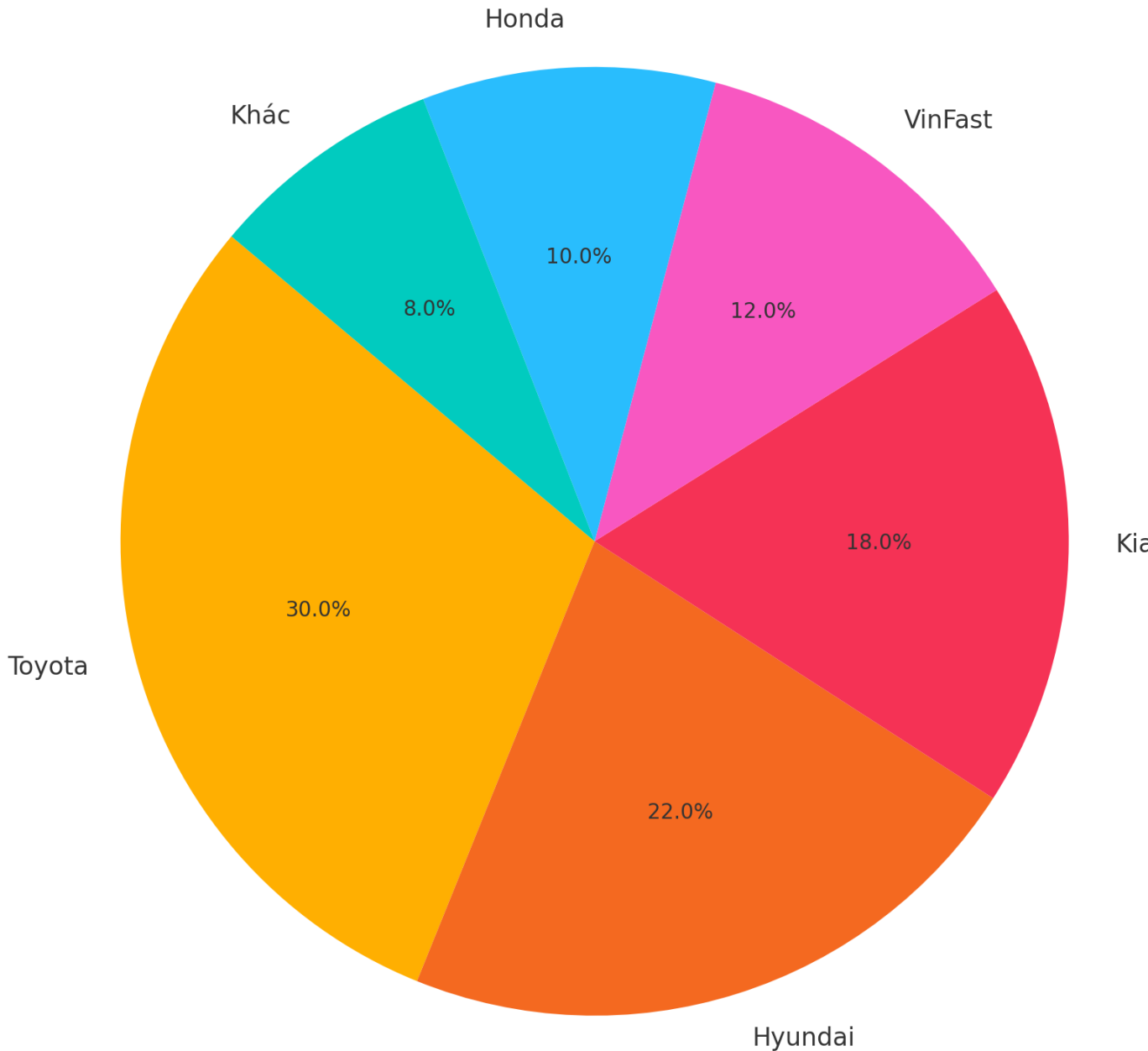
Biểu đồ tròn là một cách khác để thể hiện **tần suất tương đối** hoặc **tần suất phần trăm**.

Cách vẽ:

1. Vẽ một **hình tròn** đại diện cho toàn bộ dữ liệu (100%).
2. Chia hình tròn thành các **hình quạt** tương ứng với tần suất của từng nhóm.
3. Vì một hình tròn có **360 độ**, nên:
 - Nếu một nhóm chiếm 38% (tần suất tương đối 0.38), thì góc hình quạt là:
 $0.38 \times 360 = 136.8$ độ
 - Tương tự, nhóm Mercedes chiếm 26% thì có góc:
 $0.26 \times 360 = 93.6$ độ

Những con số hiển thị trên biểu đồ tròn có thể là **số lượng (frequency)**, **tần suất tương đối**, hoặc **phần trăm**.

Thị phần các hãng xe tại Việt Nam (Tương đối)



Ứng dụng trong kiểm soát chất lượng:

Biểu đồ cột còn được dùng trong **quản lý chất lượng** để thể hiện **nguyên nhân chính của lỗi**. Khi các cột được **sắp xếp theo thứ tự giảm dần từ trái qua phải**, với nguyên nhân phổ biến nhất đứng đầu,

Tác giả: **Đỗ Ngọc Tú**
Công Ty Phần Mềm **VHTSoft**

Phiên bản #3

Được tạo 24 tháng 4 2025 08:47:47 bởi Đỗ Ngọc Tú

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