

THỐNG KÊ ỨNG DỤNG

ĐỖ LÂN

dolan@tlu.edu.vn
Đại học Thủy Lợi

Ngày 13 tháng 11 năm 2018

Nội dung môn học

- ➊ Tổng quan về Thống kê
- ➋ Thu thập dữ liệu
- ➌ **Tóm tắt và trình bày dữ liệu bằng bảng và đồ thị**
- ➍ Tóm tắt dữ liệu bằng các đại lượng thống kê mô tả
- ➎ Xác suất căn bản và biến ngẫu nhiên
- ➏ Phân phối của tham số mẫu và ước lượng tham số tổng thể
- ➐ Kiểm định giả thuyết về tham số một tổng thể
- ➑ Kiểm định giả thuyết về tham số hai tổng thể
- ➒ Phân tích phương sai
- ➓ Kiểm định phi tham số
- ➔ Kiểm định chi - bình phương
- ➕ Hồi quy đơn biến
- ➖ Hồi quy đa biến

Phần III

Tóm tắt dữ liệu bằng bảng và biểu đồ

1 Tóm tắt và trình bày dữ liệu bằng bảng tần số

- Lập bảng tần số cho dữ liệu định tính
- Lập bảng tần số cho dữ liệu định lượng
- Bảng tần số chéo

2 Tóm tắt và trình bày dữ liệu bằng biểu đồ và đồ thị

- Một số biểu đồ và đồ thị minh họa cho tập dữ liệu định lượng
 - Biểu đồ phân phối tần số
 - Đa giác tần số
 - Biểu đồ thân và lá
- Biểu đồ cho dữ liệu định tính
 - Biểu đồ hình tròn
 - Biểu đồ thanh
- Biểu đồ mô tả mối quan hệ giữa hai biến
 - Biểu đồ tán xạ
 - Biểu đồ chéo mô tả phân phối của biến định tính này theo biến định tính khác

Tình huống 1

Tình huống 1

Bạn có trong tay dữ liệu điểm thi môn Toán kỳ thi THPT quốc gia, bạn muốn công bố lên các phương tiện thông tin đại chúng, bạn làm thế nào để người đọc dễ dàng nhìn ra các điểm quan trọng nhất của phổ điểm năm nay?

Tình huống 2

Tình huống 2

Trong một cuộc họp về chất lượng học tập của sinh viên K58 khoa CNTT so với toàn trường, giáo viên phụ trách khoa làm thế nào để so sánh tập dữ liệu điểm của sinh viên trong khoa với sinh viên toàn trường một cách rõ ràng và dễ nhìn nhất?

Nội dung chính của chương

Chương 3 giới thiệu một số cách tóm tắt một dữ liệu bằng bảng và bằng biểu đồ, đồ thị.

- 1 Tóm tắt tập dữ liệu định tính bằng bảng tần số.

Nội dung chính của chương

Chương 3 giới thiệu một số cách tóm tắt một dữ liệu bằng bảng và bằng biểu đồ, đồ thị.

- 1 Tóm tắt tập dữ liệu định tính bằng bảng tần số.
- 2 Tóm tắt tập dữ liệu định lượng bằng bảng tần số.

Nội dung chính của chương

Chương 3 giới thiệu một số cách tóm tắt một dữ liệu bằng bảng và bằng biểu đồ, đồ thị.

- 1 Tóm tắt tập dữ liệu định tính bằng bảng tần số.
- 2 Tóm tắt tập dữ liệu định lượng bằng bảng tần số.
- 3 Phân một tập dữ liệu nhiều giá trị biểu hiện bằng các tổ.

Nội dung chính của chương

Chương 3 giới thiệu một số cách tóm tắt một dữ liệu bằng bảng và bằng biểu đồ, đồ thị.

- 1 Tóm tắt tập dữ liệu định tính bằng bảng tần số.
- 2 Tóm tắt tập dữ liệu định lượng bằng bảng tần số.
- 3 Phân một tập dữ liệu nhiều giá trị biểu hiện bằng các tổ.
- 4 Tóm tắt, mô tả một tập dữ liệu định lượng bằng các kiểu biểu đồ: *biểu đồ phân phối tần số, đa giác tần số, biểu đồ thân và lá.*

Nội dung chính của chương

Chương 3 giới thiệu một số cách tóm tắt một dữ liệu bằng bảng và bằng biểu đồ, đồ thị.

- 1 Tóm tắt tập dữ liệu định tính bằng bảng tần số.
- 2 Tóm tắt tập dữ liệu định lượng bằng bảng tần số.
- 3 Phân một tập dữ liệu nhiều giá trị biểu hiện bằng các tổ.
- 4 Tóm tắt, mô tả một tập dữ liệu định lượng bằng các kiểu biểu đồ: *biểu đồ phân phối tần số, đa giác tần số, biểu đồ thân và lá.*
- 5 Tóm tắt, mô tả một tập dữ liệu định tính bằng các kiểu biểu đồ: *biểu đồ hình tròn, biểu đồ thanh, biểu đồ pareto.*

Nội dung chính của chương

Chương 3 giới thiệu một số cách tóm tắt một dữ liệu bằng bảng và bằng biểu đồ, đồ thị.

- 1 Tóm tắt tập dữ liệu định tính bằng bảng tần số.
- 2 Tóm tắt tập dữ liệu định lượng bằng bảng tần số.
- 3 Phân một tập dữ liệu nhiều giá trị biểu hiện bằng các tổ.
- 4 Tóm tắt, mô tả một tập dữ liệu định lượng bằng các kiểu biểu đồ: *biểu đồ phân phối tần số, đa giác tần số, biểu đồ thân và lá.*
- 5 Tóm tắt, mô tả một tập dữ liệu định tính bằng các kiểu biểu đồ: *biểu đồ hình tròn, biểu đồ thanh, biểu đồ pareto.*
- 6 **Vẽ các loại biểu đồ bằng R.**

Những kiến thức cần nắm được trong chương 3

- Biết cách lập bảng tần số cho dữ liệu định tính.
- Biết cách lập bảng tần số cho dữ liệu định lượng.
- Biết được khi nào cần phải phân tổ dữ liệu và các điều kiện khi phân tổ.
- Biết cách minh họa phân phối của một tập dữ liệu định lượng bằng biểu đồ phân phối tần số, đa giác tần số, biểu đồ thân lá. Phân biệt được với kiểu dữ liệu nào thì nên dùng loại biểu đồ nào.
- Biết cách minh họa phân phối của một tập dữ liệu định tính bằng biểu đồ hình tròn, biểu đồ thanh, biểu đồ Pareto. Phân biệt được với kiểu dữ liệu nào thì nên dùng loại biểu đồ nào.
- Vẽ được các loại biểu đồ bằng **R**.

Những kiến thức cần nắm được trong chương 3

- Biết cách lập bảng tần số cho dữ liệu định tính.
- Biết cách lập bảng tần số cho dữ liệu định lượng.
- Biết được khi nào cần phải phân tổ dữ liệu và các điều kiện khi phân tổ.
- Biết cách minh họa phân phối của một tập dữ liệu định lượng bằng biểu đồ phân phối tần số, đa giác tần số, biểu đồ thân lá. Phân biệt được với kiểu dữ liệu nào thì nên dùng loại biểu đồ nào.
- Biết cách minh họa phân phối của một tập dữ liệu định tính bằng biểu đồ hình tròn, biểu đồ thanh, biểu đồ Pareto. Phân biệt được với kiểu dữ liệu nào thì nên dùng loại biểu đồ nào.
- Vẽ được các loại biểu đồ bằng **R**.

1 Tóm tắt và trình bày dữ liệu bằng bảng tần số

- Lập bảng tần số cho dữ liệu định tính
- Lập bảng tần số cho dữ liệu định lượng
- Bảng tần số chéo

2 Tóm tắt và trình bày dữ liệu bằng biểu đồ và đồ thị

- Một số biểu đồ và đồ thị minh họa cho tập dữ liệu định lượng
 - Biểu đồ phân phối tần số
 - Đa giác tần số
 - Biểu đồ thân và lá
- Biểu đồ cho dữ liệu định tính
 - Biểu đồ hình tròn
 - Biểu đồ thanh
- Biểu đồ mô tả mối quan hệ giữa hai biến
 - Biểu đồ tán xạ
 - Biểu đồ chéo mô tả phân phối của biến định tính này theo biến định tính khác

Khái niệm

Khái niệm

Bảng tần số là một bảng tổng hợp các biểu hiện có thể có của đặc điểm quan sát, hoặc các khoảng giá trị mà trong đó dữ liệu có thể rơi vào và số quan sát (tần số), tỉ lệ phần trăm chiếm (tần suất) tương ứng với mỗi biểu hiện hoặc khoảng giá trị dữ liệu.

Khái niệm

Bảng tần số là một bảng tổng hợp các biểu hiện có thể có của đặc điểm quan sát, hoặc các khoảng giá trị mà trong đó dữ liệu có thể rơi vào và số quan sát (tần số), tỉ lệ phần trăm chiếm (tần suất) tương ứng với mỗi biểu hiện hoặc khoảng giá trị dữ liệu.

Bảng tần số thường có 2 loại:

- 1 Bảng tần số cho dữ liệu định tính
- 2 Bảng tần số cho dữ liệu định lượng

Ví dụ

Dãy sau đây liệt kê xếp loại tốt nghiệp của 20 sinh viên, trong đó "G" là giỏi, "K" là khá, "TB" là trung bình, "XS" là xuất sắc.

G	XS	G	XS	K	G	K	XS	XS	K
TB	K	XS	K	XS	K	G	G	TB	TB

Ví dụ

Dãy sau đây liệt kê xếp loại tốt nghiệp của 20 sinh viên, trong đó "G" là giỏi, "K" là khá, "TB" là trung bình, "XS" là xuất sắc.

G XS G XS K G K XS XS K
TB K XS K XS K G G TB TB

Khi đó ta có bảng tần số như sau:

Xếp loại	Tần số	Tần suất	Tần số tích lũy	Tần suất tích lũy
TB	3	0.15	3	0.15
K	6	0.30	9	0.45
G	5	0.25	14	0.7
XS	6	0.3	20	1

1 Tóm tắt và trình bày dữ liệu bằng bảng tần số

- Lập bảng tần số cho dữ liệu định tính
- Lập bảng tần số cho dữ liệu định lượng
- Bảng tần số chéo

2 Tóm tắt và trình bày dữ liệu bằng biểu đồ và đồ thị

- Một số biểu đồ và đồ thị minh họa cho tập dữ liệu định lượng
 - Biểu đồ phân phối tần số
 - Đa giác tần số
 - Biểu đồ thân và lá
- Biểu đồ cho dữ liệu định tính
 - Biểu đồ hình tròn
 - Biểu đồ thanh
- Biểu đồ mô tả mối quan hệ giữa hai biến
 - Biểu đồ tán xạ
 - Biểu đồ chéo mô tả phân phối của biến định tính này theo biến định tính khác

Lập bảng tần số cho dữ liệu định tính

Bảng tần số của một dữ liệu định tính có 3 cột dạng:

Biểu hiện	Tần số	Tần suất
Biểu hiện 1	f_1	$\frac{f_1}{n}$
Biểu hiện 2	f_2	$\frac{f_2}{n}$
...
Biểu hiện k	f_k	$\frac{f_k}{n}$

Lập bảng tần số cho dữ liệu định tính

Bảng tần số của một dữ liệu định tính có 3 cột dạng:

Biểu hiện	Tần số	Tần suất
Biểu hiện 1	f_1	$\frac{f_1}{n}$
Biểu hiện 2	f_2	$\frac{f_2}{n}$
...
Biểu hiện k	f_k	$\frac{f_k}{n}$

Đôi khi thêm cột tần số tích lũy và cột tần suất tích lũy:

- Tần số tích lũy của biểu hiện thứ i là tổng các số từ dòng dòng đầu đến dòng thứ i của cột tần số;
- Tần suất tích lũy của biểu hiện thứ i là tổng các số từ dòng dòng đầu đến dòng thứ i của cột tần suất.

1 Tóm tắt và trình bày dữ liệu bằng bảng tần số

- Lập bảng tần số cho dữ liệu định tính
- Lập bảng tần số cho dữ liệu định lượng
- Bảng tần số chéo

2 Tóm tắt và trình bày dữ liệu bằng biểu đồ và đồ thị

- Một số biểu đồ và đồ thị minh họa cho tập dữ liệu định lượng
 - Biểu đồ phân phối tần số
 - Đa giác tần số
 - Biểu đồ thân và lá
- Biểu đồ cho dữ liệu định tính
 - Biểu đồ hình tròn
 - Biểu đồ thanh
- Biểu đồ mô tả mối quan hệ giữa hai biến
 - Biểu đồ tán xạ
 - Biểu đồ chéo mô tả phân phối của biến định tính này theo biến định tính khác

Lập bảng tần số cho dữ liệu định lượng

Lập bảng cho dữ liệu định lượng tương tự như dữ liệu định tính.

- Khi là dữ liệu định lượng ít biểu hiện thì ta lập bảng như dữ liệu định tính.

Lập bảng tần số cho dữ liệu định lượng

Lập bảng cho dữ liệu định lượng tương tự như dữ liệu định tính.

- Khi là dữ liệu định lượng ít biểu hiện thì ta lập bảng như dữ liệu định tính.
- Khi là dữ liệu định lượng nhiều biểu hiện thì ta phải phân tổ.

Lập bảng tần số cho dữ liệu định lượng

Lập bảng cho dữ liệu định lượng tương tự như dữ liệu định tính.

- Khi là dữ liệu định lượng ít biểu hiện thì ta lập bảng như dữ liệu định tính.
- Khi là dữ liệu định lượng nhiều biểu hiện thì ta phải phân tổ.
Khi phân tổ cần:
 - ① các tổ phải rời nhau.

Lập bảng tần số cho dữ liệu định lượng

Lập bảng cho dữ liệu định lượng tương tự như dữ liệu định tính.

- Khi là dữ liệu định lượng ít biểu hiện thì ta lập bảng như dữ liệu định tính.
- Khi là dữ liệu định lượng nhiều biểu hiện thì ta phải phân tổ.

Khi phân tổ cần:

- ① các tổ phải rời nhau.
- ② hợp các tổ phải phủ kín các giá trị của tập dữ liệu.

Lập bảng tần số cho dữ liệu định lượng

Lập bảng cho dữ liệu định lượng tương tự như dữ liệu định tính.

- Khi là dữ liệu định lượng ít biểu hiện thì ta lập bảng như dữ liệu định tính.
- Khi là dữ liệu định lượng nhiều biểu hiện thì ta phải phân tổ.

Khi phân tổ cần:

- ① các tổ phải rời nhau.
- ② hợp các tổ phải phủ kín các giá trị của tập dữ liệu.
- ③ không có tổ trống

Cách phân tổ đều

Trong phân tổ đều, ta có 2 giá trị cần chú ý nhất: Số tổ và Độ rộng tổ. Giả sử tập dữ liệu có n quan sát với giá trị lớn nhất, nhỏ nhất là $X_{max}; X_{min}$, ta có

- 1 Tính số tổ $k = \sqrt[3]{2(X_{max} - X_{min})}$ hoặc $k = 1 + \log_2(X_{max} - X_{min})$.

Cách phân tổ đều

Trong phân tổ đều, ta có 2 giá trị cần chú ý nhất: Số tổ và Độ rộng tổ. Giả sử tập dữ liệu có n quan sát với giá trị lớn nhất, nhỏ nhất là $X_{max}; X_{min}$, ta có

- 1 Tính số tổ $k = \sqrt[3]{2(X_{max} - X_{min})}$ hoặc $k = 1 + \log_2(X_{max} - X_{min})$.
- 2 Tính độ rộng tổ $h = \frac{X_{max} - X_{min}}{k}$.

Cách phân tổ đều

Trong phân tổ đều, ta có 2 giá trị cần chú ý nhất: Số tổ và Độ rộng tổ. Giả sử tập dữ liệu có n quan sát với giá trị lớn nhất, nhỏ nhất là $X_{max}; X_{min}$, ta có

- 1 Tính số tổ $k = \sqrt[3]{2(X_{max} - X_{min})}$ hoặc $k = 1 + \log_2(X_{max} - X_{min})$.
- 2 Tính độ rộng tổ $h = \frac{X_{max} - X_{min}}{k}$.
- 3 Xác định cận dưới và cận trên của các tổ.

Cách phân tổ đều

Trong phân tổ đều, ta có 2 giá trị cần chú ý nhất: Số tổ và Độ rộng tổ. Giả sử tập dữ liệu có n quan sát với giá trị lớn nhất, nhỏ nhất là $X_{max}; X_{min}$, ta có

- 1 Tính số tổ $k = \sqrt[3]{2(X_{max} - X_{min})}$ hoặc $k = 1 + \log_2(X_{max} - X_{min})$.
- 2 Tính độ rộng tổ $h = \frac{X_{max} - X_{min}}{k}$.
- 3 Xác định cận dưới và cận trên của các tổ.
- 4 Cận trên của tổ trước trùng với cận dưới của tổ liền sau.

Cách phân tổ đều

Trong phân tổ đều, ta có 2 giá trị cần chú ý nhất: Số tổ và Độ rộng tổ. Giả sử tập dữ liệu có n quan sát với giá trị lớn nhất, nhỏ nhất là $X_{max}; X_{min}$, ta có

- 1 Tính số tổ $k = \sqrt[3]{2(X_{max} - X_{min})}$ hoặc $k = 1 + \log_2(X_{max} - X_{min})$.
- 2 Tính độ rộng tổ $h = \frac{X_{max} - X_{min}}{k}$.
- 3 Xác định cận dưới và cận trên của các tổ.
- 4 Cận trên của tổ trước trùng với cận dưới của tổ liền sau.
- 5 Phân chia các quan sát vào các tổ.

Cách phân tổ đều

Trong phân tổ đều, ta có 2 giá trị cần chú ý nhất: Số tổ và Độ rộng tổ. Giả sử tập dữ liệu có n quan sát với giá trị lớn nhất, nhỏ nhất là $X_{max}; X_{min}$, ta có

- 1 Tính số tổ $k = \sqrt[3]{2(X_{max} - X_{min})}$ hoặc $k = 1 + \log_2(X_{max} - X_{min})$.
- 2 Tính độ rộng tổ $h = \frac{X_{max} - X_{min}}{k}$.
- 3 Xác định cận dưới và cận trên của các tổ.
- 4 Cận trên của tổ trước trùng với cận dưới của tổ liền sau.
- 5 Phân chia các quan sát vào các tổ.

Chẳng hạn, có thể phân tổ như sau ($x_0 \leq min, x_0 + k.h \geq max$)

Tổ 1: $[x_0, x_0 + h]$

Tổ 2: $(x_0 + h, x_0 + 2h]$

Tổ 3: $(x_0 + 2h, x_0 + 3h]$

...

Tổ k: $(x_0 + (k - 1)h, x_0 + k.h]$

Ví dụ 1: Dữ liệu định lượng ít biểu hiện

Ví dụ

Giả sử ta có dữ liệu về số môn còn nợ của 30 bạn sinh viên lớp TH04.1 như sau:

5	1	3	2	0	1	2	4	2	4	4	2	5
5	1	5	4	4	2	5	5	0	1	0	5	0
0	0	3	5									

Ví dụ 1: Dữ liệu định lượng ít biểu hiện

Ví dụ

Giả sử ta có dữ liệu về số môn còn nợ của 30 bạn sinh viên lớp TH04.1 như sau:

5 1 3 2 0 1 2 4 2 4 4 2 5
5 1 5 4 4 2 5 5 0 1 0 5 0
0 0 3 5

Ta có bảng tần số như sau:

Số môn nợ	0	1	2	3	4	5
Tần số	6	4	5	2	5	8
Tần suất	0.2	0.13	0.17	0.07	0.17	0.27
Tần số TL	6	10	15	17	22	30
Tần suất TL	0.2	0.33	0.5	0.57	0.74	1

Ví dụ 2: Dữ liệu định lượng nhiều biểu hiện

Ví dụ

Giả sử có tập dữ liệu về tuổi của người lao động trong một phân xưởng như sau:

48	30	35	31	21	28	34	43	36	45	41
33	47	47	30	47	44	45	32	46	47	23
30	23	49	20	24	20	40	50			

Ví dụ 2: Dữ liệu định lượng nhiều biểu hiện

Ví dụ

Giả sử có tập dữ liệu về tuổi của người lao động trong một phân xưởng như sau:

48	30	35	31	21	28	34	43	36	45	41
33	47	47	30	47	44	45	32	46	47	23
30	23	49	20	24	20	40	50			

Ta có $X_{\min} = 20$, $X_{\max} = 50$, $k = \log_2(30) + 1 \approx 6$, $h = (50 - 20)/6 = 5$.

Ví dụ 2: Dữ liệu định lượng nhiều biểu hiện

Ví dụ

Giả sử có tập dữ liệu về tuổi của người lao động trong một phân xưởng như sau:

48	30	35	31	21	28	34	43	36	45	41
33	47	47	30	47	44	45	32	46	47	23
30	23	49	20	24	20	40	50			

Ta có $X_{\min} = 20$, $X_{\max} = 50$, $k = \log_2(30) + 1 \approx 6$, $h = (50 - 20)/6 = 5$.

Các tổ là:

$[20,25]; (25,30]; (30,35]; (35,40]; (40,45]; (45,50]$

Bảng tần số như sau:

Khoảng tuổi	[20, 25]	(25,30]	(30,35]	(35,40]	(40,45]	(45,50]
Tần số	6	4	5	2	5	8
Tần suất	0.2	0.13	0.17	0.07	0.17	0.27
Tần số TL	6	10	15	17	22	30
Tần suất TL	0.2	0.33	0.5	0.57	0.74	1

Lập bảng tần số bằng R

Trường hợp dữ liệu cần lập bảng tần số là định tính hoặc định lượng nhưng **số các biểu hiện không nhiều**, ta lập bảng tần số, bảng tần suất, bảng tần số tích lũy, bảng tần suất tích lũy bằng các lệnh sau:

<code>table(x)</code>	bảng tần số
<code>prop.table(table(x))</code>	bảng tần suất
<code>cumsum(table(x))</code>	bảng tần số tích lũy
<code>cumsum(prop.table(table(x)))</code>	bảng tần suất tích lũy

trong đó

x: vec tơ dữ liệu

Lập bảng tần số bằng R

Trường hợp dữ liệu cần lập bảng tần số là định tính hoặc định lượng nhưng số các biểu hiện không nhiều, ta lập bảng tần số, bảng tần suất, bảng tần số tích lũy, bảng tần suất tích lũy bằng các lệnh sau:

<code>table(x)</code>	bảng tần số
<code>prop.table(table(x))</code>	bảng tần suất
<code>cumsum(table(x))</code>	bảng tần số tích lũy
<code>cumsum(prop.table(table(x)))</code>	bảng tần suất tích lũy

trong đó

x: vec tơ dữ liệu

Lập bảng tần số bằng R

Trường hợp dữ liệu cần lập bảng tần số là định tính hoặc định lượng nhưng số các biểu hiện không nhiều, ta lập bảng tần số, bảng tần suất, bảng tần số tích lũy, bảng tần suất tích lũy bằng các lệnh sau:

<code>table(x)</code>	bảng tần số
<code>prop.table(table(x))</code>	bảng tần suất
<code>cumsum(table(x))</code>	bảng tần số tích lũy
<code>cumsum(prop.table(table(x)))</code>	bảng tần suất tích lũy

trong đó

x: vec tơ dữ liệu

Lập bảng tần số bằng **R** (không phân tổ)

Ví dụ

Lập bảng tần số bằng **R** (không phân tổ)

Ví dụ

Bảng số liệu về số hộp sữa bán được của một cửa hàng

	9	10	10	11	12	12	
10	11	10	9	11	9	12	11
	9	12	10	9	12	10	11
	11	12	11	10	12	10	11

Lập bảng tần số cho dữ liệu trên.

Lập bảng tần số bằng **R** (có phân tổ)

Khi lập bảng dữ liệu định lượng nhiều biểu hiện, trước khi lập bảng tần số, ta tiến hành phân tổ bằng lệnh sau:

```
cut(x, breaks, right=TRUE, include.lowest=FALSE)
```

trong đó

x: vector dữ liệu

breaks: một vector gồm các điểm chia tổ hoặc một số cho biết số tổ cần chia

right: TRUE (FALSE) nếu các tổ có dạng (a,b] ([a,b))

include.lowest: TRUE (FALSE) đưa giá trị nhỏ nhất (lớn nhất) vào tổ đầu tiên (cuối cùng) khi right=T (right=F).

Lập bảng tần số bằng **R** (có phân tổ)

Khi lập bảng dữ liệu định lượng nhiều biểu hiện, trước khi lập bảng tần số, ta tiến hành phân tổ bằng lệnh sau:

```
cut(x, breaks, right=TRUE, include.lowest=FALSE)
```

trong đó

x: vector dữ liệu

breaks: một vector gồm các điểm chia tổ hoặc một số cho biết số tổ cần chia

right: TRUE (FALSE) nếu các tổ có dạng $(a,b]$ ($[a,b)$)

include.lowest: TRUE (FALSE) đưa giá trị nhỏ nhất (lớn nhất) vào tổ đầu tiên (cuối cùng) khi $right=T$ ($right=F$).

Lập bảng tần số bằng **R** (có phân tổ)

Khi lập bảng dữ liệu định lượng nhiều biểu hiện, trước khi lập bảng tần số, ta tiến hành phân tổ bằng lệnh sau:

```
cut(x, breaks, right=TRUE, include.lowest=FALSE)
```

trong đó

x: vector dữ liệu

breaks: một vector gồm các điểm chia tổ hoặc một số cho biết số tổ cần chia

right: TRUE (FALSE) nếu các tổ có dạng $(a,b]$ ($[a,b)$)

include.lowest: TRUE (FALSE) đưa giá trị nhỏ nhất (lớn nhất) vào tổ đầu tiên (cuối cùng) khi $right=T$ ($right=F$).

Lập bảng tần số bằng R

Ví dụ

Giả sử có tập dữ liệu về tuổi của người lao động trong một phân xưởng như sau:

*48 30 35 31 21 28 34 43 36 45 41 33 47 47 30 47 44 45 32 46 47 23 30 23
49 20 24 20 40 50.*

Lập bảng tần số, tần suất, tần số tích lũy, tần suất tích lũy cho bảng số liệu trên (có phân tổ).

1 Tóm tắt và trình bày dữ liệu bằng bảng tần số

- Lập bảng tần số cho dữ liệu định tính
- Lập bảng tần số cho dữ liệu định lượng
- Bảng tần số chéo

2 Tóm tắt và trình bày dữ liệu bằng biểu đồ và đồ thị

- Một số biểu đồ và đồ thị minh họa cho tập dữ liệu định lượng
 - Biểu đồ phân phối tần số
 - Đa giác tần số
 - Biểu đồ thân và lá
- Biểu đồ cho dữ liệu định tính
 - Biểu đồ hình tròn
 - Biểu đồ thanh
- Biểu đồ mô tả mối quan hệ giữa hai biến
 - Biểu đồ tán xạ
 - Biểu đồ chéo mô tả phân phối của biến định tính này theo biến định tính khác

Bảng tần số chéo

Dùng mô tả mối tương quan giữa hai biến định tính, dạng đơn giản nhất là:

Biến 1	Biến 2		
	Biểu hiện 1	...	Biểu hiện n
Biểu hiện 1	Tần số	Tần số	Tần số
...
Biểu hiện k	Tần số	Tần số	Tần số

Bảng tần số chéo

Dùng mô tả mối tương quan giữa hai biến định tính, dạng đơn giản nhất là:

Biến 1	Biến 2		
	Biểu hiện 1	...	Biểu hiện n
Biểu hiện 1	Tần số	Tần số	Tần số
...
Biểu hiện k	Tần số	Tần số	Tần số

Khi thay Tần số bởi Tần suất thì có bảng tần suất. Ngoài ra còn bảng tần suất theo dòng, theo cột.

Bảng tần số, tần suất nhiều tầng không xét ở đây.

Ví dụ

Giả sử ta có bảng dữ liệu sau về giới tính và nghề nghiệp của 10 người được điều tra.

<i>TT</i>	<i>Giới tính</i>	<i>Nghề nghiệp</i>
1	Nam	CongNhan
2	Nam	QuanLy
3	Nam	CongNhan
4	Nam	NVVP
5	Nam	NVVP
6	Nam	NVVP
7	Nu	QuanLy
8	Nam	CongNhan
9	Nu	NVVP
10	Nam	CongNhan

Ví dụ

Bảng tần số chéo giữa giới tính và nghề nghiệp của 10 người được điều tra nói trên có thể được mô tả như sau:

<i>Nghề nghề nghiệp</i>	<i>Giới tính</i>	
	<i>Nam</i>	<i>Nữ</i>
<i>CongNhan</i>	4	

Ví dụ

Bảng tần số chéo giữa giới tính và nghề nghiệp của 10 người được điều tra nói trên có thể được mô tả như sau:

<i>Nghề nghề nghiệp</i>	<i>Giới tính</i>	
	<i>Nam</i>	<i>Nữ</i>
<i>CongNhan</i>	<i>4</i>	<i>0</i>

Ví dụ

Bảng tần số chéo giữa giới tính và nghề nghiệp của 10 người được điều tra nói trên có thể được mô tả như sau:

Nghề nghiệp	Giới tính	
	Nam	Nữ
CongNhan	4	0
NVVP	3	

Ví dụ

Bảng tần số chéo giữa giới tính và nghề nghiệp của 10 người được điều tra nói trên có thể được mô tả như sau:

<i>Nghề nghề nghiệp</i>	<i>Giới tính</i>	
	<i>Nam</i>	<i>Nữ</i>
<i>CongNhan</i>	<i>4</i>	<i>0</i>
<i>NVVP</i>	<i>3</i>	<i>1</i>
<i>QuanLy</i>	<i>1</i>	<i>1</i>

1 Tóm tắt và trình bày dữ liệu bằng bảng tần số

- Lập bảng tần số cho dữ liệu định tính
- Lập bảng tần số cho dữ liệu định lượng
- Bảng tần số chéo

2 Tóm tắt và trình bày dữ liệu bằng biểu đồ và đồ thị

- Một số biểu đồ và đồ thị minh họa cho tập dữ liệu định lượng
 - Biểu đồ phân phối tần số
 - Đa giác tần số
 - Biểu đồ thân và lá
- Biểu đồ cho dữ liệu định tính
 - Biểu đồ hình tròn
 - Biểu đồ thanh
- Biểu đồ mô tả mối quan hệ giữa hai biến
 - Biểu đồ tán xạ
 - Biểu đồ chéo mô tả phân phối của biến định tính này theo biến định tính khác

1 Tóm tắt và trình bày dữ liệu bằng bảng tần số

- Lập bảng tần số cho dữ liệu định tính
- Lập bảng tần số cho dữ liệu định lượng
- Bảng tần số chéo

2 Tóm tắt và trình bày dữ liệu bằng biểu đồ và đồ thị

- Một số biểu đồ và đồ thị minh họa cho tập dữ liệu định lượng
 - Biểu đồ phân phối tần số
 - Đa giác tần số
 - Biểu đồ thân và lá
- Biểu đồ cho dữ liệu định tính
 - Biểu đồ hình tròn
 - Biểu đồ thanh
- Biểu đồ mô tả mối quan hệ giữa hai biến
 - Biểu đồ tán xạ
 - Biểu đồ chéo mô tả phân phối của biến định tính này theo biến định tính khác

Biểu đồ phân phối tần số

Cách lập biểu đồ phân phối tần số:

- Phân chia tập dữ liệu thành các tổ;
- Xác định tần số trong mỗi tổ;
- Vẽ các cột đứng đặt cạnh nhau với độ rộng là khoảng cách trong mỗi tổ và độ cao tương ứng là tần số trong mỗi tổ.

Ví dụ

Ta trở lại Ví dụ về Tuổi của đội công nhân ở trên, ta đã xây dựng được bảng tần số như sau:

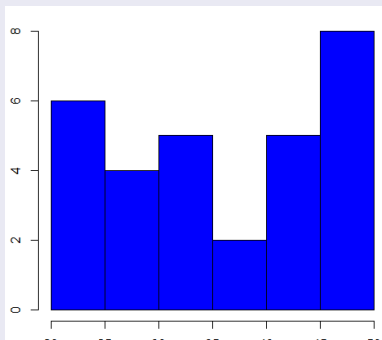
Độ tuổi	$[20, 25]$	$(25,30]$	$(30,35]$	$(35,40]$	$(40,45]$	$(45,50]$
TS	6	4	5	2	5	8

Ví dụ

Ta trở lại Ví dụ về Tuổi của đội công nhân ở trên, ta đã xây dựng được bảng tần số như sau:

Độ tuổi	$[20, 25]$	$(25,30]$	$(30,35]$	$(35,40]$	$(40,45]$	$(45,50]$
TS	6	4	5	2	5	8

Biểu đồ phân phối tần số (histogram) tương ứng là:



Nhận xét về biểu đồ cần nói đến những gì?

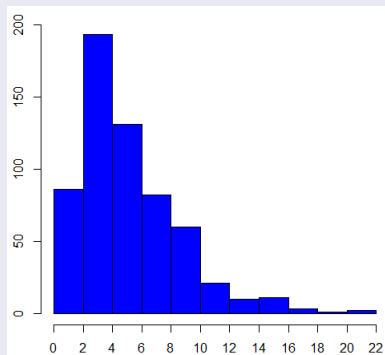
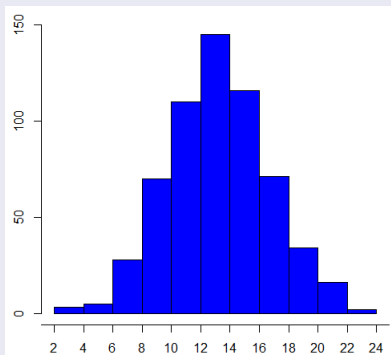
- Mức độ tập trung tương đối của phân phối của tập dữ liệu;
- Hình dạng tương đối của phân phối của tập dữ liệu là bằng phẳng, lệch hay cân đối.

Nhận xét về biểu đồ cần nói đến những gì?

- Mức độ tập trung tương đối của phân phối của tập dữ liệu;
- Hình dạng tương đối của phân phối của tập dữ liệu là bằng phẳng, lệch hay cân đối.

Ví dụ

2 tập dữ liệu thu nhập:

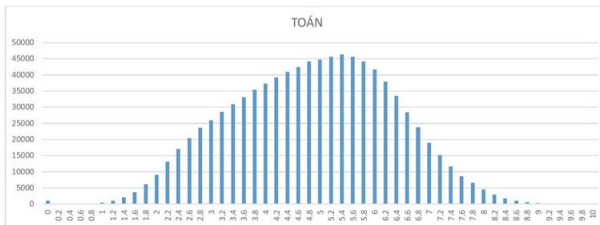


Ví dụ

Ví dụ về phổ điểm thi THPT Quốc Gia môn Toán

Điểm	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.2	4.4	4.6	4.8	5
Số lượng	951	7	7	40	139	414	960	1981	3620	6016	9017	13015	17017	20306	23574	25948	28533	30836	33075	35327	37206	39112	40887	42286	44071	44729

Điểm	5.2	5.4	5.6	5.8	6	6.2	6.4	6.6	6.8	7	7.2	7.4	7.6	7.8	8	8.2	8.4	8.6	8.8	9	9.2	9.4	9.6	9.8	10
Số lượng	45578	46228	45613	44060	41633	37831	33390	28375	23700	18842	15127	11640	8622	6485	4444	2904	1765	1045	567	301	151	70	30	7	2



Vẽ biểu đồ tần số trong R

Lệnh Hist

```
hist(x, breaks = "Sturges", include.lowest = TRUE, right = TRUE, col =  
, border = , labels = FALSE, main = "", xlab = "", ylab = "")
```

Vẽ biểu đồ tần số trong R

Lệnh Hist

```
hist(x, breaks = "Sturges", include.lowest = TRUE, right = TRUE, col =  
, border = , labels = FALSE, main = "", xlab = "", ylab = "")
```

trong đó

- x: vector dữ liệu dạng số dùng để vẽ biểu đồ,
- breaks: như lệnh cut
- right: như lệnh cut
- include.lowest: như lệnh cut
- xlab: tên trục ngang
- ylab: tên trục dọc
- main: tên biểu đồ
- col: màu sắc các thanh

Vẽ biểu đồ tần số trong R

Lệnh Hist

```
hist(x, breaks = "Sturges", include.lowest = TRUE, right = TRUE, col =  
, border = , labels = FALSE, main = "", xlab = "", ylab = "")
```

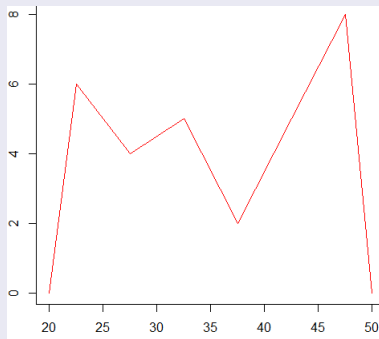
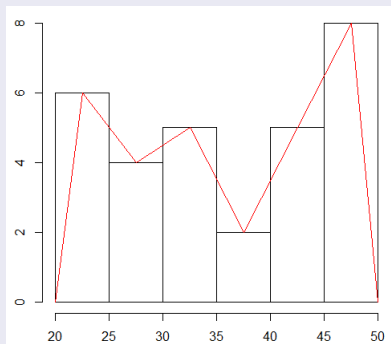
trong đó

- x: vector dữ liệu dạng số dùng để vẽ biểu đồ,
- breaks: như lệnh cut
- right: như lệnh cut
- include.lowest: như lệnh cut
- xlab: tên trục ngang
- ylab: tên trục dọc
- main: tên biểu đồ
- col: màu sắc các thanh

Đa giác tần số

Đa giác tần số là đồ thị gồm các đoạn thẳng nối các điểm với nhau dùng để miêu tả phân phối tần số của tập dữ liệu định lượng.

Ví dụ



Hình thứ nhất vẽ cùng với histogram, hình thứ 2 vẽ riêng đa giác tần số.

Vẽ đa giác tần số trong R

Lệnh plot

```
plot(x, y, type = "b", main = "", sub = "", xlab = "", ylab = "")
```

Vẽ đa giác tần số trong R

Lệnh plot

```
plot(x, y, type = "b", main = "", sub = "", xlab = "", ylab = "")
```

trong đó

- x, y: hoành độ, tung độ của các điểm trong đồ thị,
- type = "b": chỉ kiểu vẽ trong đồ thị là điểm nối các đoạn thẳng
- xlab: tên trục ngang
- ylab: tên trục dọc
- main, sub: tên chính, tên phụ của đồ thị
- clo: màu sắc các thanh

Vẽ đa giác tần số trong R

Lệnh plot

```
plot(x, y, type = "b", main = "", sub = "", xlab = "", ylab = "")
```

trong đó

- x, y: hoành độ, tung độ của các điểm trong đồ thị,
- type = "b": chỉ kiểu vẽ trong đồ thị là điểm nối các đoạn thẳng
- xlab: tên trục ngang
- ylab: tên trục dọc
- main, sub: tên chính, tên phụ của đồ thị
- clo: màu sắc các thanh

Ví dụ

Ví dụ về cách vẽ kết hợp biểu đồ phân phối tần số và đa giác tần số trên cùng một hình trong **R**

Khái niệm

Biểu đồ thân và lá được xây dựng bằng cách chia các chữ số trong mỗi số của tập dữ liệu định lượng thành hai phần: phần thân và phần lá để miêu tả phân phối của tập dữ liệu.

Cách lập biểu đồ thân lá:

- Xác định phần thân, phần lá trong mỗi dữ liệu;
- Lập phần thân của biểu đồ;
- Xếp phần lá tương ứng với mỗi phần thân đã lập.

Example

Giả sử ta có điểm quá trình môn thống kê của 30 sinh viên như sau:

6.9 3.2 6.4 7.2 7.2 4.8 6.1 7.4 7.8 8.0 6.0 6.6 6.5 7.2 2.2
8.7 7.3 7.8 4.7 4.8 5.7 6.1 6.5 8.7 5.0 7.8 6.5 5.6 7.9 8.8

Ta được biểu đồ thân và lá như sau:

2		2
3		2
4		788
5		067
6		011455569
7		222348889
8		0778

Vẽ biểu đồ thân lá trong R

Hàm stem()

```
stem(x, scale = 1)
```


Vẽ biểu đồ thân lá trong R

Hàm stem()

```
stem(x, scale = 1)
```

trong đó

- x: vector dữ liệu dạng số;
- scale: tham số điều chỉnh chiều dài của biểu đồ.

Vẽ biểu đồ thân lá trong R

Hàm stem()

```
stem(x, scale = 1)
```

trong đó

- x: vector dữ liệu dạng số;
- scale: tham số điều chỉnh chiều dài của biểu đồ.

Xét tập dữ liệu về tuổi của người lao động

48 30 35 31 21 28 34 43 36 45 41 33 47 47 30 47 44 45 32 46 47 23 30 23
49 20 24 20 40 50

1 Tóm tắt và trình bày dữ liệu bằng bảng tần số

- Lập bảng tần số cho dữ liệu định tính
- Lập bảng tần số cho dữ liệu định lượng
- Bảng tần số chéo

2 Tóm tắt và trình bày dữ liệu bằng biểu đồ và đồ thị

- Một số biểu đồ và đồ thị minh họa cho tập dữ liệu định lượng
 - Biểu đồ phân phối tần số
 - Đa giác tần số
 - Biểu đồ thân và lá
- Biểu đồ cho dữ liệu định tính
 - Biểu đồ hình tròn
 - Biểu đồ thanh
- Biểu đồ mô tả mối quan hệ giữa hai biến
 - Biểu đồ tán xạ
 - Biểu đồ chéo mô tả phân phối của biến định tính này theo biến định tính khác

Biểu đồ hình tròn

Khái niệm

Biểu đồ hình tròn miêu tả dữ liệu định tính dưới dạng hình tròn, trong đó diện tích của toàn bộ hình tròn đại diện cho 100% các phần tử của tập dữ liệu và diện tích các hình quạt tròn đại diện cho phần trăm các tập con cần biểu diễn.

Biểu đồ hình tròn

Khái niệm

Biểu đồ hình tròn miêu tả dữ liệu định tính dưới dạng hình tròn, trong đó diện tích của toàn bộ hình tròn đại diện cho 100% các phần tử của tập dữ liệu và diện tích các hình quạt tròn đại diện cho phần trăm các tập con cần biểu diễn.

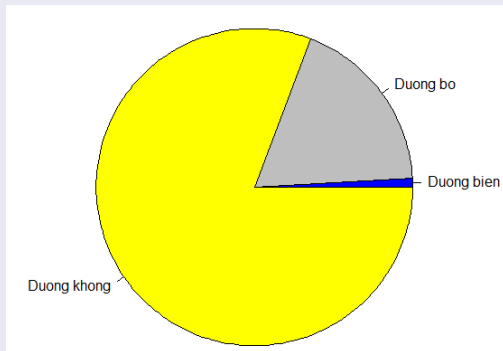
Cách vẽ biểu đồ hình tròn:

- Xác định các biểu hiện trong tập dữ liệu (định tính);
- Xác định tỉ lệ cho từng biểu hiện;
- Tính góc của hình quạt tương ứng với từng biểu hiện
- Dựa trên số đo góc cho hình quạt để vẽ phần diện tích cho mỗi biểu hiện tương ứng.

Ví dụ

Theo ước tính, lượng khách quốc tế đến Việt Nam vào tháng 2 năm 2015:

Phương tiện	Lượng khách	Tỉ lệ	Góc (độ)
Đường biển	7 021	0.009287	3.34
Đường bộ	138 146	0.182733	65.78
Đường không	610 834	0.80798	290.87



Biểu đồ tròn cho ta biết gì?

Biểu đồ tròn cho ta biết

- Mức phân phối tương đối của mỗi biểu hiện so với toàn thể;

Biểu đồ tròn cho ta biết gì?

Biểu đồ tròn cho ta biết

- Mức phân phối tương đối của mỗi biểu hiện so với toàn thể;
- Biểu hiện nào chiếm nhiều nhất, ít nhất.

Biểu đồ tròn cho ta biết gì?

Biểu đồ tròn cho ta biết

- Mức phân phối tương đối của mỗi biểu hiện so với toàn thể;
- Biểu hiện nào chiếm nhiều nhất, ít nhất.

Chẳng hạn, biểu đồ hình tròn ở ví dụ trên cho ta thấy:

+ lượng khách quốc tế đến nước ta chủ yếu qua đường hàng không, chiếm hơn 80%,

+ đường biển rất ít, có thể nói là không đáng kể so với lượng đến từ đường hàng không.

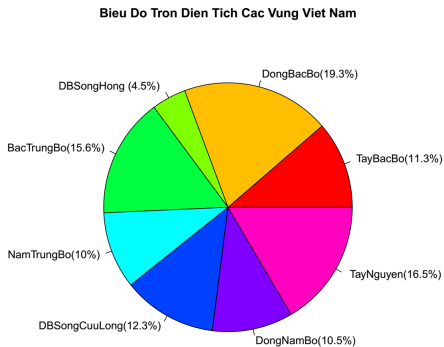
Ví dụ về biểu đồ hình tròn

Cho bảng số liệu về diện tích các vùng trong lãnh thổ Việt Nam.

Vùng	Diện tích (m^2)	Tỉ lệ (%)	Góc (độ)
ĐB Sông Hồng	14862.4	4.50	16.15
Đông Bắc Bộ	64025.2	19.3	69.59
Tây Bắc Bộ	37533.9	11.3	40.80
Bắc Trung Bộ	51551.9	15.6	56.03
Nam Trung Bộ	33166.0	10.0	36.05
Tây Nguyên	54659.6	16.5	59.41
Đông Nam Bộ	34807.8	10.5	37.83
ĐB Sông Cửu Long	40604.8	12.3	44.13

Ví dụ về biểu đồ hình tròn

Ta có biểu đồ như sau:



Vẽ biểu đồ hình tròn trong R

Hàm pie()

```
pie(x, labels = names(x), col = NULL, border = NULL, lty = NULL, main = NULL)
```

Vẽ biểu đồ hình tròn trong R

Hàm pie()

```
pie(x, labels = names(x), col = NULL, border = NULL, lty = NULL, main = NULL)
```

trong đó

- **x**: vector dữ liệu dạng số thể hiện giá trị của mỗi hình quạt trong biểu đồ;
- **labels**: tham số chỉ tên những hình quạt trong biểu đồ;
- **col**: tham số chỉ màu của các hình quạt;
- **boder**: tham số chỉ màu của các đường danh giới giữa các hình quạt;
- **main, sub**: tham số chỉ tiêu đề và tiêu đề phụ của biểu đồ.

Khái niệm

Biểu đồ thanh

Khái niệm

Biểu đồ thanh bao gồm thanh đứng (ngang) dùng để miêu tả phân phối tần số của một tập dữ liệu định tính.

Cách vẽ biểu đồ thanh:

Biểu đồ thanh

Khái niệm

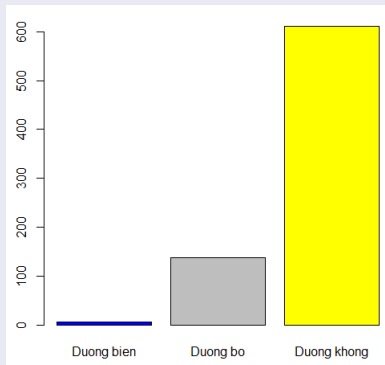
Biểu đồ thanh bao gồm thanh đứng (ngang) dùng để miêu tả phân phối tần số của một tập dữ liệu định tính.

Cách vẽ biểu đồ thanh:

- Xác định các biểu hiện trong tập dữ liệu;
- Xác định giá trị cho từng biểu hiện trong tập dữ liệu;
- Vẽ các thanh đứng cho từng biểu hiện với chiều cao tương ứng với giá trị mỗi biểu hiện.

Ví dụ 1

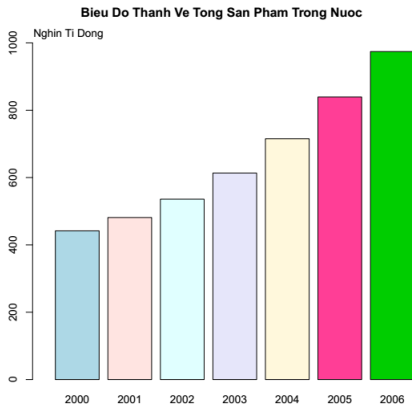
Ví dụ



Ví dụ 2

Ví dụ

Năm	Tổng sản phẩm
2000	441.7
2001	481.3
2002	535.8
2003	613.4
2004	715.3
2005	839.2
2005	974.3



Biểu đồ thanh cho ta biết gì?

Biểu đồ thanh cho ta

- Sự so sánh tương đối giữa các biểu hiện với nhau;

Biểu đồ thanh cho ta biết gì?

Biểu đồ thanh cho ta

- Sự so sánh tương đối giữa các biểu hiện với nhau;
- Biểu hiện nào có chiếm nhiều nhất, ít nhất.

Biểu đồ thanh cho ta biết gì?

Biểu đồ thanh cho ta

- Sự so sánh tương đối giữa các biểu hiện với nhau;
- Biểu hiện nào có chiếm nhiều nhất, ít nhất.

Chẳng hạn, từ biểu đồ thanh thứ nhất trong ví dụ trên ta thấy rằng số lượt khách đến bằng đường hàng không là chiếm lớn nhất, nó gấp 4 lần số lượt khách đến từ đường bộ. Và số lượt khách đến từ đường biển thì cực ít so với hai phương tiện trên.

??

Vậy chọn biểu đồ thanh hay biểu đồ tròn?

Vẽ biểu đồ thanh trong R

Hàm barplot()

```
barplot(height, names.arg = NULL, beside = FALSE, horiz = FALSE, col  
= "", border = "", main = "", sub = "", xlab = "", ylab = "")
```


Vẽ biểu đồ thanh trong R

Hàm barplot()

```
barplot(height, names.arg = NULL, beside = FALSE, horiz = FALSE, col  
= "", border = "", main = "", sub = "", xlab = "", ylab = "")
```

trong đó

- height: vector hoặc ma trận dữ liệu dạng số thể hiện giá trị của mỗi hình quạt trong biểu đồ;
- name.arg: tham số chỉ tên viết dưới mỗi thanh hoặc nhóm thanh trong biểu đồ;
- beside: tham số dạng logic, nếu =TRUE thì các cột của biểu đồ được vẽ chồng lên nhau, nếu = FALSE thì các cột được vẽ cạnh nhau;
- horiz: tham số dạng logic, nếu =TRUE thì các cột được vẽ song song với trục nằm ngang với cột đầu tiên ở dưới cùng, nếu = FALSE thì các cột được vẽ vuông góc với trục nằm ngang với cột đầu tiên nằm bên trái;

Vẽ biểu đồ thanh trong R

Hàm barplot()

```
barplot(height, names.arg = NULL, beside = FALSE, horiz = FALSE, col = "", border = "", main = "", sub = "", xlab = "", ylab = "")
```

trong đó

- height: vector hoặc ma trận dữ liệu dạng số thể hiện giá trị của mỗi hình quạt trong biểu đồ;
- name.arg: tham số chỉ tên viết dưới mỗi thanh hoặc nhóm thanh trong biểu đồ;
- beside: tham số dạng logic, nếu =TRUE thì các cột của biểu đồ được vẽ chồng lên nhau, nếu = FALSE thì các cột được vẽ cạnh nhau;
- horiz: tham số dạng logic, nếu =TRUE thì các cột được vẽ song song với trục nằm ngang với cột đầu tiên ở dưới cùng, nếu = FALSE thì các cột được vẽ vuông góc với trục nằm ngang với cột đầu tiên nằm bên trái;

1 Tóm tắt và trình bày dữ liệu bằng bảng tần số

- Lập bảng tần số cho dữ liệu định tính
- Lập bảng tần số cho dữ liệu định lượng
- Bảng tần số chéo

2 Tóm tắt và trình bày dữ liệu bằng biểu đồ và đồ thị

- Một số biểu đồ và đồ thị minh họa cho tập dữ liệu định lượng
 - Biểu đồ phân phối tần số
 - Đa giác tần số
 - Biểu đồ thân và lá
- Biểu đồ cho dữ liệu định tính
 - Biểu đồ hình tròn
 - Biểu đồ thanh
- Biểu đồ mô tả mối quan hệ giữa hai biến
 - Biểu đồ tán xạ
 - Biểu đồ chéo mô tả phân phối của biến định tính này theo biến định tính khác

Khái niệm

Biểu đồ tán xạ là đồ thị hai chiều bao gồm các điểm mô tả mối quan hệ giữa hai biến định lượng.

Biểu đồ tán xạ

Khái niệm

Biểu đồ tán xạ là đồ thị hai chiều bao gồm các điểm mô tả mối quan hệ giữa hai biến định lượng.

Cách lập biểu đồ tán xạ:

- 1 Xác định các biểu hiện của biến định lượng thứ nhất: x_1, x_2, \dots, x_n

Biểu đồ tán xạ

Khái niệm

Biểu đồ tán xạ là đồ thị hai chiều bao gồm các điểm mô tả mối quan hệ giữa hai biến định lượng.

Cách lập biểu đồ tán xạ:

- 1 Xác định các biểu hiện của biến định lượng thứ nhất: x_1, x_2, \dots, x_n ;
- 2 Xác định các biểu hiện của biến định lượng thứ hai: y_1, y_2, \dots, y_n ;

Biểu đồ tán xạ

Khái niệm

Biểu đồ tán xạ là đồ thị hai chiều bao gồm các điểm mô tả mối quan hệ giữa hai biến định lượng.

Cách lập biểu đồ tán xạ:

- 1 Xác định các biểu hiện của biến định lượng thứ nhất: x_1, x_2, \dots, x_n ;
- 2 Xác định các biểu hiện của biến định lượng thứ hai: y_1, y_2, \dots, y_n ;
- 3 Xác định các cặp điểm $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$;

Biểu đồ tán xạ

Khái niệm

Biểu đồ tán xạ là đồ thị hai chiều bao gồm các điểm mô tả mối quan hệ giữa hai biến định lượng.

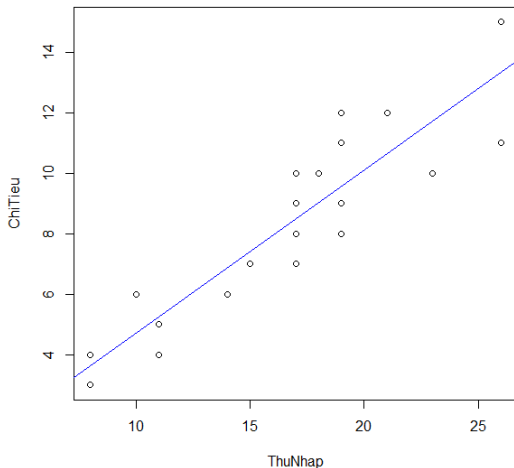
Cách lập biểu đồ tán xạ:

- 1 Xác định các biểu hiện của biến định lượng thứ nhất: x_1, x_2, \dots, x_n ;
- 2 Xác định các biểu hiện của biến định lượng thứ hai: y_1, y_2, \dots, y_n ;
- 3 Xác định các cặp điểm $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$;
- 4 Vẽ các cặp điểm này trên mặt phẳng đề các.

Example

Giả sử ta có thu nhập và chi tiêu của 20 người như sau.

TN	19	17	18	23	14	11	17	26	19	19	8	26
CT	12	7	10	10	6	5	10	15	11	9	3	11



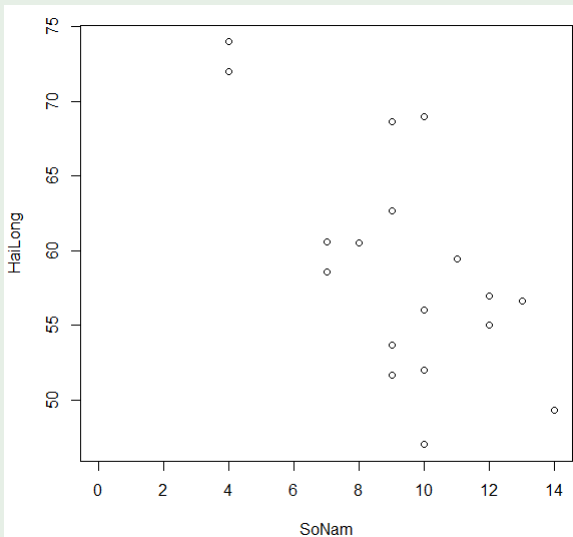
Biểu đồ tán xạ cho biết điều gì?

- Biểu đồ tán xạ cho ta cái nhìn tương đối về mối quan hệ giữa hai biến định lượng.
- Biểu đồ tán xạ cho ta thấy hai biến định lượng có quan hệ dạng tuyến tính hay phi tuyến.

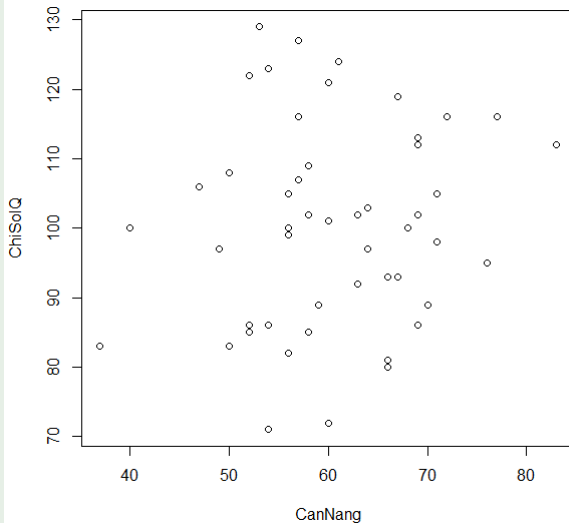
Biểu đồ tán xạ cho biết điều gì?

- Biểu đồ tán xạ cho ta cái nhìn tương đối về mối quan hệ giữa hai biến định lượng.
- Biểu đồ tán xạ cho ta thấy hai biến định lượng có quan hệ dạng tuyến tính hay phi tuyến.

Example



Example



Vẽ biểu đồ tán xạ trong R

Hàm `plot()`

```
plot(x, y, type = "p", main = "", xlab = "", ylab = "", bty = "l")
```

Vẽ biểu đồ tán xạ trong R

Hàm plot()

```
plot(x, y, type = "p", main = "", xlab = "", ylab = "", bty = "l")
```

trong đó

- x,y: vector tọa độ của biến định lượng thứ nhất, thứ hai;
- bty="l": tham số cho kiểu bao quanh biểu đồ giống hai trục tọa độ Oxy.

Vẽ biểu đồ tán xạ trong R

Hàm plot()

```
plot(x, y, type = "p", main = "", xlab = "", ylab = "", bty = "l")
```

trong đó

- `x,y`: vector tọa độ của biến định lượng thứ nhất, thứ hai;
- `bty="l"`: tham số cho kiểu bao quanh biểu đồ giống hai trục tọa độ Oxy.

Biểu đồ chéo mô tả phân phối của biến định tính này theo biến định tính khác

Nghề nghiệp	Giới tính	
	Nam	Nữ
CongNhan	4	0
NVVP	3	1
QuanLy	1	1

