Bài giảng Xác suất Thống kê và ứng dụng

Phan Thanh Hồng

Bộ môn Toán-Đại học THĂNG LONG

Ngày 18 tháng 11 năm 2009

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần III

Tóm tắt và trình bày dữ liệu bằng bảng và đồ thị

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần III

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần III

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần III

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần III

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần III

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần III

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần III

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần III

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần III

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần III

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Tóm tắt dữ liệu bằng bảng tần số

1. Bảng tần số cho dữ liệu định tính
2. Bảng tần số cho dữ liệu định lượng
3. Cách lập bảng tần số bằng R

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Tóm tắt dữ liệu bằng bảng tần số

Bảng tần số dùng để tóm tắt dữ liệu, trong đó các quan sát trong tập dữ liệu được phân chia vào các nhóm, sau đó số lượng quan sát trong mỗi nhóm được xác định và gọi là tần số của nhóm đó.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định tính

Ví dụ: Giả sử ý kiến của 115 khách hàng về bao bì mới của một loại sản phẩm như sau: 45 người rất thích, 37 người thích, 25 người không thích, 8 người rất không thích.
Vậy ý kiến của những khách hàng này bao gồm 4 trường hợp: rất thích, thích, không thích, rất không thích. Ta tạo ra một bảng với cột đầu tiên liệt kê các trường hợp này. Cột thứ hai là cột tần số điền số quan sát của từng trường hợp nói trên. Cột thứ ba là cột tần suất điền tỷ lệ của tần số trên tổng số quan sát.

Ý kiến	Tần số (người)	Tần suất(%)
Rất thích	45	39.13
Thích	37	32.17
Không thích	25	21.74
Rất không thích	8	6.96

Bảng: Bảng tần số ý kiến của 115 khách hàng

Bảng tần số cho dữ liệu định tính

Ý kiến
Rất thích
Thích
Không thích
Rất không thích

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định tính

Ý kiến	Tần số (người)
Rất thích	45
Thích	37
Không thích	25
Rất không thích	8

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định tính

Ý kiến	Tần số (người)	Tần suất(%)
Rất thích	45	39.13
Thích	37	32.17
Không thích	25	21.74
Rất không thích	8	6.96

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định tính

Cách lập bảng tần số cho dữ liệu định tính

1. Cột thứ nhất liệt kê các biểu hiện của đối tượng theo đặc điểm ta muốn lập bảng.
2. Cột hai là cột tần số điền số quan sát của từng biểu hiện trong tập dữ liệu.
3. Cột thứ ba là cột tần suất điền tỷ lệ của tần số trên tổng số quan sát (nếu cần làm tròn thì phải đảm bảo tổng các giá trị trong cột này bằng 100%).

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định lượng

- Nếu đặc điểm ta cần lập bảng có ít biểu hiện thì ta lập bảng tần số tương tự như trường hợp dữ liệu định tính trong đó các biểu hiện là những giá trị có thể có của tập dữ liệu.

Ví dụ: Một cửa hàng bán sữa theo dõi số lượng một loại sữa hộp bán ra trong một ngày trong một tháng như sau

9	10	9	11	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	

Bảng: Số lượng hộp sữa bán trong 30 ngày

Bảng tần số cho dữ liệu định lượng

Vậy có 6 ngày bán được 9 hộp, 8 ngày bán được 10 hộp, 9 ngày bán được 11 hộp, 7 ngày bán được 12 hộp.
Bảng tần số cho số hộp sữa bán một ngày như sau

Số hộp sữa
9
10
11
12

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định lượng

Vậy có 6 ngày bán được 9 hộp, 8 ngày bán được 10 hộp, 9 ngày bán được 11 hộp, 7 ngày bán được 12 hộp.
Bảng tần số cho số hộp sữa bán một ngày như sau

Số hộp sữa	Tần số (hộp)
9	6
10	8
11	9
12	7

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định lượng

Vậy có 6 ngày bán được 9 hộp, 8 ngày bán được 10 hộp, 9 ngày bán được 11 hộp, 7 ngày bán được 12 hộp.
Bảng tần số cho số hộp sữa bán một ngày như sau

Số hộp sữa	Tần số (hộp)	Tần suất(%)
9	6	20
10	8	26.67
11	9	30
12	7	23.33

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định lượng

- Nếu đặc điểm ta cần lập bảng có nhiều biểu hiện thì việc lập bảng tần số như trên sẽ dài. Vì vậy ta tiến hành phân tổ dữ liệu để đưa dữ liệu vào các nhóm có phạm vi giá trị nhất định.
  Việc phân tổ dữ liệu tiến hành dựa trên quy tắc:
  1. Các tổ rời nhau để mỗi quan sát chỉ rơi vào một tổ nào đó
  2. Tất cả các tổ phải bao quát hết mọi giá trị có trong tập dữ liệu
  3. Tránh không để tổ rỗng do không có quan sát nào rơi vào tổ đó

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định lượng

- Nếu đặc điểm ta cần lập bảng có nhiều biểu hiện thì việc lập bảng tần số như trên sẽ dài. Vì vậy ta tiến hành phân tổ dữ liệu để đưa dữ liệu vào các nhóm có phạm vi giá trị nhất định.
  Việc phân tổ dữ liệu tiến hành dựa trên quy tắc:
  1. Các tổ rời nhau để mỗi quan sát chỉ rơi vào một tổ nào đó
  2. Tất cả các tổ phải bao quát hết mọi giá trị có trong tập dữ liệu
  3. Tránh không để tổ rỗng do không có quan sát nào rơi vào tổ đó

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định lượng

- Nếu đặc điểm ta cần lập bảng có nhiều biểu hiện thì việc lập bảng tần số như trên sẽ dài. Vì vậy ta tiến hành phân tổ dữ liệu để đưa dữ liệu vào các nhóm có phạm vi giá trị nhất định.
  Việc phân tổ dữ liệu tiến hành dựa trên quy tắc:
  1. Các tổ rời nhau để mỗi quan sát chỉ rơi vào một tổ nào đó
  2. Tất cả các tổ phải bao quát hết mọi giá trị có trong tập dữ liệu
  3. Tránh không để tổ rỗng do không có quan sát nào rơi vào tổ đó

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Các bước phân tổ

- **Bước 1**: Xác định số tổ cần chia k, thông thường người ta chọn k nằm trong khoảng từ 5 đến 15 tổ. Có thể dùng công thức có tính chất tham khảo: k là số làm tròn của $(2n)^{1/3}$, hoặc k là số nhỏ nhất mà 2^k lớn hơn n trong đó n là số quan sát trong tập dữ liệu.
- **Bước 2**: Xác định khoảng cách tổ h bởi công thức $h = \frac{X_{max} - X_{min}}{k}$ Trong đó X_{min} giá trị nhỏ nhất và X_{max} là giá trị lớn nhất của tập dữ liệu. Khoảng cách tổ h cũng được làm tròn cho tiện lợi.
- **Bước 3**: Xác định giới hạn trên và dưới của các tổ: giới hạn dưới của tổ đầu tiên phải nhỏ hơn hay bằng X_{min}, giới hạn trên của tổ cuối cùng phải lớn hơn hay bằng X_{max}. Giá trị giới hạn trên của tổ này là giá trị giới hạn dưới của tổ liền kề sau đó.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Các bước phân tổ

- **Bước 1**: Xác định số tổ cần chia k, thông thường người ta chọn k nằm trong khoảng từ 5 đến 15 tổ. Có thể dùng công thức có tính chất tham khảo: k là số làm tròn của $(2n)^{1/3}$, hoặc k là số nhỏ nhất mà 2^k lớn hơn n trong đó n là số quan sát trong tập dữ liệu.
- **Bước 2**: Xác định khoảng cách tổ h bởi công thức $h = \frac{X_{max} - X_{min}}{k}$ Trong đó X_{min} giá trị nhỏ nhất và X_{max} là giá trị lớn nhất của tập dữ liệu. Khoảng cách tổ h cũng được làm tròn cho tiện lợi.
- **Bước 3**: Xác định giới hạn trên và dưới của các tổ: giới hạn dưới của tổ đầu tiên phải nhỏ hơn hay bằng X_{min}, giới hạn trên của tổ cuối cùng phải lớn hơn hay bằng X_{max}. Giá trị giới hạn trên của tổ này là giá trị giới hạn dưới của tổ liền kề sau đó.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Các bước phân tổ

- **Bước 1**: Xác định số tổ cần chia k, thông thường người ta chọn k nằm trong khoảng từ 5 đến 15 tổ. Có thể dùng công thức có tính chất tham khảo: k là số làm tròn của $(2n)^{1/3}$, hoặc k là số nhỏ nhất mà 2^k lớn hơn n trong đó n là số quan sát trong tập dữ liệu.
- **Bước 2**: Xác định khoảng cách tổ h bởi công thức $h = \frac{X_{max} - X_{min}}{k}$ Trong đó X_{min} giá trị nhỏ nhất và X_{max} là giá trị lớn nhất của tập dữ liệu. Khoảng cách tổ h cũng được làm tròn cho tiện lợi.
- **Bước 3**: Xác định giới hạn trên và dưới của các tổ: giới hạn dưới của tổ đầu tiên phải nhỏ hơn hay bằng X_{min}, giới hạn trên của tổ cuối cùng phải lớn hơn hay bằng X_{max}. Giá trị giới hạn trên của tổ này là giá trị giới hạn dưới của tổ liền kề sau đó.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Các bước phân tổ

Chẳng hạn ta có thể phân tổ:
Tổ 1: X_{min} — $X_{min} + h$
Tổ 2: $X_{min} + h$ — $X_{min} + 2h$
Tổ 3: $X_{min} + 2h$ — $X_{min} + 3h$
...
Trong các trường hợp cụ thể bước này có thể linh động để đảm bảo tính khoa học và mỹ thuật.

- **Bước 4**: Chia các quan sát vào mỗi tổ theo quy tắc giá trị x_i rơi vào một tổ nếu giới hạn dưới $\leqslant x_i <$ giới hạn trên.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Các bước phân tổ

Chẳng hạn ta có thể phân tổ:
Tổ 1: X_{min} — $X_{min} + h$
Tổ 2: $X_{min} + h$ — $X_{min} + 2h$
Tổ 3: $X_{min} + 2h$ — $X_{min} + 3h$
...
Trong các trường hợp cụ thể bước này có thể linh động để đảm bảo tính khoa học và mỹ thuật.

- **Bước 4**: Chia các quan sát vào mỗi tổ theo quy tắc giá trị x_i rơi vào một tổ nếu giới hạn dưới $\leqslant x_i <$ giới hạn trên.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định lượng

Ví dụ: Xét mẫu 65 thời gian thanh toán. Ta tiến hành phân tổ theo các bước như trên

1. Chọn $K = 7$ vì $2^7 > 65$
2. $X_{min} = 10$, $X_{max} = 29$, $h = \frac{29-10}{7}$ làm tròn thành 3
3. Tổ 1: 10–13
   Tổ 2: 13–16
   Tổ 3: 16–19
   Tổ 4: 19–22
   Tổ 5: 22–25
   Tổ 6: 25–28
   Tổ 7: 28–31

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định lượng

Đếm số quan sát rơi vào từng tổ ta có bảng tần số cho thời gian thanh toán như sau

Thời gian (ngày)
10-13
13-16
16-19
19-22
22-25
25-28
28-31

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định lượng

Đếm số quan sát rơi vào từng tổ ta có bảng tần số cho thời gian thanh toán như sau

Thời gian (ngày)	Tần số
10-13	3
13-16	14
16-19	23
19-22	12
22-25	8
25-28	4
28-31	1

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định lượng

Đếm số quan sát rơi vào từng tổ ta có bảng tần số cho thời gian thanh toán như sau

Thời gian (ngày)	Tần số	Tần suất(%)
10-13	3	4.62
13-16	14	21.54
16-19	23	35.38
19-22	12	18.46
22-25	8	12.31
25-28	4	6.15
28-31	1	1.54

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định lượng

Đếm số quan sát rơi vào từng tổ ta có bảng tần số cho thời gian thanh toán như sau

Thời gian (ngày)	Tần số	Tần suất(%)	Điểm giữa
10-13	3	4.62	11.5
13-16	14	21.54	14.5
16-19	23	35.38	17.5
19-22	12	18.46	20.5
22-25	8	12.31	23.5
25-28	4	6.15	26.5
28-31	1	1.54	29.5

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Bảng tần số cho dữ liệu định lượng

Vì trong ví dụ trên các quan sát đều là những số nguyên nên ta cũng có thể phân tổ bằng cách xếp các quan sát vào những khoảng thời gian như sau: 9.5-12.5 ngày, 12.5-15,5 ngày, . . . 27.5-30.5 ngày.

Lớp
9.5-12.5
12.5-15.5
15.5-18.5
18.5-21.5
21.5-24.5
24.5-27.5
27.5-30.5

Bảng tần số cho dữ liệu định lượng

Vì trong ví dụ trên các quan sát đều là những số nguyên nên ta cũng có thể phân tổ bằng cách xếp các quan sát vào những khoảng thời gian như sau: 9.5-12.5 ngày, 12.5-15,5 ngày, . . . 27.5-30.5 ngày.

Lớp	Tần số
9.5-12.5	3
12.5-15.5	14
15.5-18.5	23
18.5-21.5	12
21.5-24.5	8
24.5-27.5	4
27.5-30.5	1

Bảng tần số cho dữ liệu định lượng

Vì trong ví dụ trên các quan sát đều là những số nguyên nên ta cũng có thể phân tổ bằng cách xếp các quan sát vào những khoảng thời gian như sau: 9.5-12.5 ngày, 12.5-15,5 ngày, . . . 27.5-30.5 ngày.

Lớp	Tần số	Điểm giữa
9.5-12.5	3	11
12.5-15.5	14	14
15.5-18.5	23	17
18.5-21.5	12	20
21.5-24.5	8	23
24.5-27.5	4	26
27.5-30.5	1	29

Bảng tần số cho dữ liệu định lượng

Chú ý: Trong trường hợp tập dữ liệu có một số quan sát bất thường (quá nhỏ hay quá lớn so với những giá trị còn lại, ta có thể xếp chúng vào tổ đầu tiên nếu nhỏ và tổ cuối cùng nếu lớn, và độ rộng của hai tổ này vẫn coi như bằng các tổ khác. Chẳng hạn trong ví dụ trên có một quan sát là 40 thì giá trị này được xếp vào lớp cuối. Việc làm này để tránh việc có quá nhiều số tổ và có tổ rỗng trong khi phân tổ.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Tần số tích lũy

Đối với dữ liệu định lượng hoặc định tính dạng thứ bậc, khi lập bảng tần số, ta có thể xây dựng thêm cột tần số tích lũy và tần suất tích lũy để cung cấp thêm thông tin.

Tần số tích lũy là số liệu tổng cộng thể hiện số quan sát có giá trị bé hơn (hoặc bằng) giới hạn trên của tổ mà nó nằm cùng hàng. Trong bảng tần số người ta đưa thêm cột tần số tích lũy, các giá trị của cột này được tính bằng cách cộng dồn các tần số từ trên cùng cho tới hàng ứng với biểu hiện ta muốn tính tần số tích lũy.

Tần suất tích lũy là số liệu tổng cộng thể hiện tỷ lệ số quan sát có giá trị bé hơn (hay bằng) giới hạn trên của tổ mà nó nằm cùng hàng. Dữ liệu của cột tần suất tích lũy được tính bằng cách cộng dồn các tần suất từ trên cùng cho tới hàng ứng với biểu hiện ta muốn tính tần suất tích lũy.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Tần số tích lũy

Đối với dữ liệu định lượng hoặc định tính dạng thứ bậc, khi lập bảng tần số, ta có thể xây dựng thêm cột tần số tích lũy và tần suất tích lũy để cung cấp thêm thông tin.

Tần số tích lũy là số liệu tổng cộng thể hiện số quan sát có giá trị bé hơn (hoặc bằng) giới hạn trên của tổ mà nó nằm cùng hàng. Trong bảng tần số người ta đưa thêm cột tần số tích lũy, các giá trị của cột này được tính bằng cách cộng dồn các tần số từ trên cùng cho tới hàng ứng với biểu hiện ta muốn tính tần số tích lũy.

Tần suất tích lũy là số liệu tổng cộng thể hiện tỷ lệ số quan sát có giá trị bé hơn (hay bằng) giới hạn trên của tổ mà nó nằm cùng hàng. Dữ liệu của cột tần suất tích lũy được tính bằng cách cộng dồn các tần suất từ trên cùng cho tới hàng ứng với biểu hiện ta muốn tính tần suất tích lũy.

Tần số tích lũy

Đối với dữ liệu định lượng hoặc định tính dạng thứ bậc, khi lập bảng tần số, ta có thể xây dựng thêm cột tần số tích lũy và tần suất tích lũy để cung cấp thêm thông tin.

Tần số tích lũy là số liệu tổng cộng thể hiện số quan sát có giá trị bé hơn (hoặc bằng) giới hạn trên của tổ mà nó nằm cùng hàng. Trong bảng tần số người ta đưa thêm cột tần số tích lũy, các giá trị của cột này được tính bằng cách cộng dồn các tần số từ trên cùng cho tới hàng ứng với biểu hiện ta muốn tính tần số tích lũy.

Tần suất tích lũy là số liệu tổng cộng thể hiện tỷ lệ số quan sát có giá trị bé hơn (hay bằng) giới hạn trên của tổ mà nó nằm cùng hàng. Dữ liệu của cột tần suất tích lũy được tính bằng cách cộng dồn các tần suất từ trên cùng cho tới hàng ứng với biểu hiện ta muốn tính tần suất tích lũy.

Tần số tích lũy

Thời gian	Tần số
10-13	3
13-16	14
16-19	23
19-22	12
22-25	8
25-28	4
28-31	1

Tần số tích lũy

Thời gian	Tần số	Tần suất(%)
10-13	3	4.62
13-16	14	21.54
16-19	23	35.38
19-22	12	18.46
22-25	8	12.31
25-28	4	6.15
28-31	1	1.54

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Tần số tích lũy

Thời gian	Tần số	Tần suất(%)	Tần số tích lũy
10-13	3	4.62	3
13-16	14	21.54	17
16-19	23	35.38	40
19-22	12	18.46	52
22-25	8	12.31	60
25-28	4	6.15	64
28-31	1	1.54	65

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Tần số tích lũy

Thời gian	Tần số	Tần suất(%)	Tần số tích lũy	Tần suất tích lũy(%)
10-13	3	4.62	3	4.62
13-16	14	21.54	17	26.16
16-19	23	35.38	40	61.54
19-22	12	18.46	52	80
22-25	8	12.31	60	92.31
25-28	4	6.15	64	98.46
28-31	1	1.54	65	100

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Lập bảng tần số bằng R

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

`table(x)`	bảng tần số
`prop.table(table(x))`	bảng tần suất
`cumsum(table(x))`	bảng tần số tích lũy
`cumsum(prop.table(table(x)))`	bảng tần số tích lũy

Trong đó

`x`: vectơ dữ liệu

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Lập bảng tần số bằng R

Trong ví dụ về tập dữ liệu số hộp sữa bán trong một tháng của cửa hàng.

9	10	9	11	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	

Bảng: Số lượng hộp sữa bán trong 30 ngày

Việc lập bảng tần số cho tập dữ liệu được tiến hành như sau

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Lập bảng tần số bằng R

```
> SoHopSua=c(9,10,9,11,...)        # nhập dữ liệu vào vectơ tên SoHopSu
> table(SoHopSua)                  # lập bảng tần số
SoHopSua
 9  10  11  12
 6   8   9   7
> prop.table(table(SoHopSua))      # lập bảng tần suất
SoHopSua
         9          10          11          12
 0.2000000  0.2666667  0.3000000  0.2333333
```

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Lập bảng tần số bằng R

```
> cumsum(table(SoHopSua))          # lập bảng tần số tích lũy
 9  10  11  12
 6  14  23  30
> bang=prop.table(table(SoHopSua))
> cumsum(bang)                     # lập bảng tần suất tích lũy
        9        10        11        12
0.2000000 0.4666667 0.7666667 1.0000000
```

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Lập bảng tần số bằng R

Khi tập dữ liệu định lượng mà các quan sát có nhiều biểu hiện khác nhau, trước khi lập bảng tần số ta tiến hành phân tổ bằn lệnh sau

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

Trong đó

`x`:	véctơ dữ liệu
`breaks`:	là một vectơ 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`)

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Lập bảng tần số bằng R

Sau khi chia tổ ta có thể lập bảng tần số (phân tổ) tương tự như trường hợp trên.
Với mẫu dữ liệu thời gian thanh toán của 65 hóa đơn ta thực hiện lập bảng tần số như sau

22	29	16	15	18	17	12	13	17	16	15
19	17	10	21	15	14	17	18	12	20	14
16	15	16	20	22	14	25	19	23	15	19
18	23	22	16	16	19	13	18	24	24	26
13	18	17	15	24	15	17	14	18	17	21
16	21	25	19	20	27	16	17	16	21	

Bảng: Mẫu 65 thời gian thanh toán

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Lập bảng tần số bằng R

```
> HoaDon=c(22,29,16,15,...)                    # nhập dữ liệu
> ChiaTo=cut(HoaDon,breaks=seq(10,31,3),right=F)
                                               # chia tổ tập dữ liệu
> table(ChiaTo)                                # lập bảng tần số
ChiaTo
 [10,13)  [13,16)  [16,19)  [19,22)  [22,25)  [25,28)  [28,31)
       3       14       23       12        8        4        1
```

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Nội dung trình bày

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ thân và lá (Stem and leaf)

Biểu đồ thân và lá là loại biểu đồ được John Tukey, một nhà thống kê phát triển và sử dụng vào năm 1977. Nó là loại biểu đồ tiện lợi để trình bày dữ liệu, và có thể sử dụng để lập bảng tần số cho tập dữ liệu.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ thân và lá (Stem and leaf)

Ví dụ: Để tìm hiểu lượng xăng tiêu thụ của những chiếc xe mới sản xuất tại một nhà máy, người ta chọn ra 49 chiếc xe và ghi lại lượng xăng tiêu thụ trên 500 km như sau:

30.8	30.9	32.0	32.3	32.6
31.7	30.4	31.4	32.7	31.4
30.1	32.5	30.8	31.2	31.8
31.6	30.3	32.8	30.6	31.9
32.1	31.3	32.0	31.7	32.8
33.3	32.1	31.5	31.4	31.5
31.3	32.5	32.4	32.2	31.6
31.0	31.8	31.0	31.5	30.6
32.0	30.4	29.8	31.7	32.2
32.4	30.5	31.1	30.6	

Bảng: Mẫu 49 lượng xăng tiêu thụ

Biểu đồ thân và lá

Các giá trị trong mẫu nằm giữa 29.8 và 33.3. Ta lấy các số $29, 30, 31, 32, 33$ làm "thân", chúng được biểu diễn ở bên trái một cột:

29
30
31
32
33

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ thân và lá

Các chữ số sau dấu phẩy được ghi bên phải trên dòng của thân tương ứng và tạo thành các "lá"

29	8
30	13445666889
31	00123344455566777889
32	000112234455678 8
33	3

Bảng: Biểu diễn thân và lá của 49 lượng xăng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ thân và lá

Nếu số lá trên thân quá dài ta có thể tách thân đó thành hai thân trong đó các lá được tách giữa số 4 và số 5.

29	8
30	1344
30	5666889
31	001233444
31	55566777889
32	0001122344
32	556788
33	3

Biểu đồ thân và lá

Cách xây dựng biểu đồ "thân và lá"

1. Xác định đơn vị sử dụng cho "thân" và "lá" sao cho số "thân" nằm giữa 5 và 20.
2. Đặt các "thân" dọc theo một cột, độ lớn tăng dần từ trên xuống.
3. Điền các "lá " của thân tương ứng. Các "lá" nên chỉ có một chữ số (nếu cần thì làm tròn).
4. Sắp xếp các lá theo thứ tự tăng dần độ lớn.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ thân và lá (Stem and leaf)

1. Biểu đồ thân và lá biểu diễn hình dáng phân phối của tập dữ liệu (bằng phẳng hay không, đối xứng hay nghiêng phải hay nghiêng trái...), và sử dụng thích hợp khi tập dữ liệu có không nhiều quan sát.
2. Một số ưu điểm nổi bật: dễ xây dựng; dữ liệu thô có thể khôi phục từ biểu đồ, các quan sát được sắp xếp mạch lạc.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Vẽ biểu đồ thân và lá trong R

Biểu đồ thân và lá được vẽ trong R bằng lệnh

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

Trong đó

`x`:	véctơ dữ liệu
`scale`:	điều khiển số thân của biểu đồ, `scale=2` thì biểu đồ có số thân gấp 2 lần so với mặc định

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Nội dung trình bày

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ tần số (Histogram)

Biểu đồ thân và lá trên đây là một cách biểu diễn dữ liệu mà cho ta thấy được phân phối (sự phân bố) của của các quan sát trong tập dữ liệu. Đối với tập dữ liệu có nhiều quan sát, việc biểu diễn bằng biểu đồ thân và lá trở nên khó khăn. Trong trường hợp đó sau khi tóm tắt dữ liệu trong bảng tần số, người ta dùng biểu đồ tần số để biểu diễn dữ liệu một cách trực quan đồng thời phản ánh được hình dáng của phân phối.

Khi vẽ loại biểu đồ này trục ngang biểu diễn các tổ dữ liệu, còn trục đứng biểu diễn tần số hay tần suất của tổ tương ứng.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ tần số (Histogram)

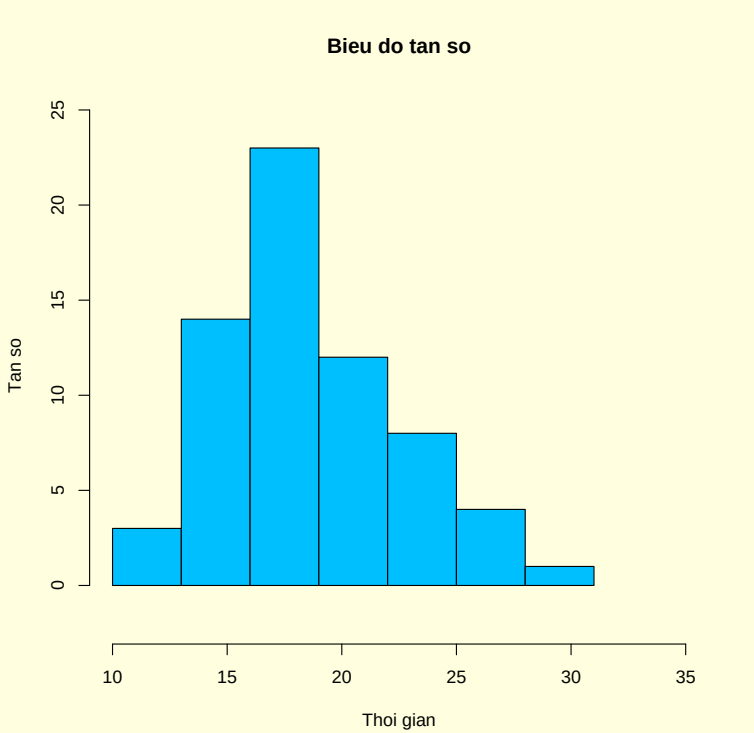
Biểu đồ thân và lá trên đây là một cách biểu diễn dữ liệu mà cho ta thấy được phân phối (sự phân bố) của của các quan sát trong tập dữ liệu. Đối với tập dữ liệu có nhiều quan sát, việc biểu diễn bằng biểu đồ thân và lá trở nên khó khăn. Trong trường hợp đó sau khi tóm tắt dữ liệu trong bảng tần số, người ta dùng biểu đồ tần số để biểu diễn dữ liệu một cách trực quan đồng thời phản ánh được hình dáng của phân phối.

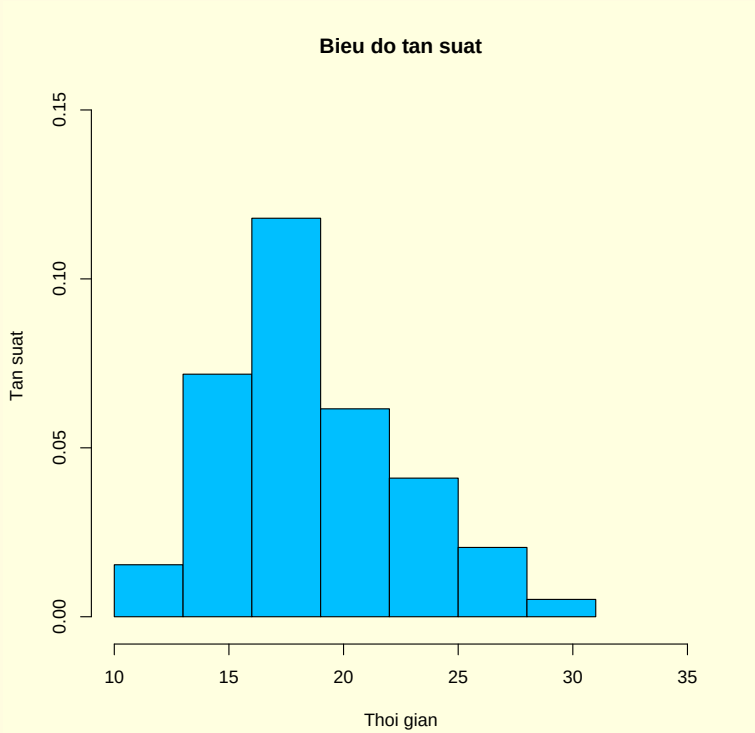
Khi vẽ loại biểu đồ này trục ngang biểu diễn các tổ dữ liệu, còn trục đứng biểu diễn tần số hay tần suất của tổ tương ứng.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ tần số (Histogram)

Sau đây là biểu đồ tần số của tập 65 quan sát về thời gian thanh toán sau khi đã có bảng phân phối tần số.





Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ tần số (Histogram)

Biểu đồ tần số mang những thông tin sau

1. Hình dáng tổng quát của dữ liệu (chuẩn hay khi-bình phương,...)
2. Phân phối của dữ liệu là đối xứng hay nghiêng
3. Dữ liệu có một hay hai hay đa mode

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Vẽ biểu đồ tần số trong R

Biểu đồ tần số được vẽ trong R bằng lệnh như sau

```
hist(x, breaks=TRUE, right = TRUE, include.lowest = TRUE,
xlab=, ylab=, main=,col=,...)
```

Trong đó

`x`:	vectơ dữ liệu
`breaks`:	xem lệnh cut ()
`right`:	xem lệnh cut ()
`include.lowest`:	xem lệnh cut ()
`xlab`:	tên trục hoành
`ylab`:	tên trục tung
`main`:	tên biểu đồ
`col`:	màu sắc của các thanh trong biểu đồ

ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Vẽ biểu đồ tần số trong R

Biểu đồ tần số về mẫu các hóa đơn được vẽ như sau

```
hist(HoaDon, breaks=seq(10,31,3), right=F, xlab="Thoi
gian", ylab="Tan so", main= "Bieu do tan so")
```

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Vẽ biểu đồ tần số trong R

Biểu đồ tần số về mẫu các hóa đơn được vẽ như sau

```
hist(HoaDon, breaks=seq(10,31,3), right=F, xlab="Thoi
gian", ylab="Tan so", main= "Bieu do tan so")
```

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Nội dung trình bày

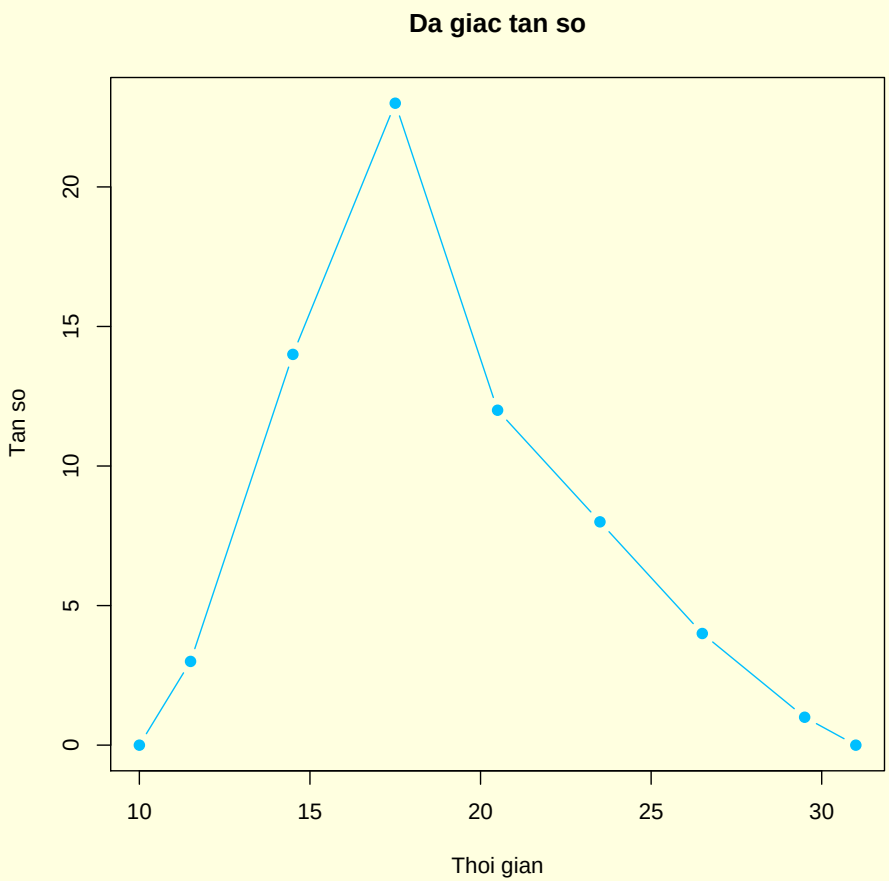
1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Đa giác tần số (Frequency Polygons)

Đa giác tần số là loại biểu đồ tương tự như biểu đồ tần số, và đặc biệt hữu ích khi ta muốn so sánh phân phối của hai tập dữ liệu trên cùng một biểu đồ Ngoài cách dùng biểu đồ tần số ta còn có thể dùng đa giác tần số để biểu diễn phân phối tần số bằng đồ thị. Đa giác tần số được vẽ bằng cách nối các trung điểm của đoạn thẳng trên đỉnh các cột trong biểu đồ tần số bằng những đoạn thẳng.
Ngoài ra ta có thể thêm vào hai điểm có hoành độ lần lượt là giới hạn dưới của tổ đầu, giới hạn trên của tổ cuối và tung độ bằng 0 để đồ thị không có vẻ bị treo lơ lửng.
Đa giác tần số có thể vẽ riêng hay kết hợp với biểu đồ tần số như trong hình sau.

Đa giác tần số



Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Đa giác tần số

Với bảng tần số của tập dữ liệu định lượng không phân tổ ta cũng vẽ đa giác tần số bằng cách nối các điểm mà hoành độ là giá trị của tập dữ liệu (đã xếp tăng dần) và tung độ là tần số tương ứng.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Đa giác tần số

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

Trong đó

`x`:	vectơ các hoành độ (các biểu hiện của tập dữ liệu) đã xếp theo thứ tự tăng dần
`y`:	vectơ các tung độ (tần số tương ứng với từng giá trị của `x`)
`type`:	xác định kiểu vẽ: dạng điểm `"p"`, dạng đoạn thẳng `"l"`, dạng cả điểm và đường thẳng `"b"`, dạng các thanh đứng `"h"`, dạng bậc thang `"s"`, ,...
`main, xlab, ylab`:	xem `hist`.

Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Mô tả dữ liệu định tính

Để biểu diễn trực quan dữ liệu định tính ta sử dụng đồ thị tròn (pie chart), đồ thị thanh (bar chart) trên cơ sở dữ liệu đã được tóm tắt trong bảng tần số.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Nội dung trình bày

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ tròn (Pie Chart)

Trong đồ thị tròn, diện tích hình tròn được chia thành nhiều phần hình quạt, diện tích từng phần này tương ứng với tỷ lệ phân loại mà nó đại diện trong toàn thể.

Ví dụ: Năng lương điện của một quốc gia được sản xuất từ than đá chiếm 55%, từ năng lượng hạt nhân chiếm 22%, từ khí ga tự nhiên chiếm 11%, từ thuỷ điện chiếm 10%, từ dầu hoả chiếm 2%. Thông tin này có thể minh hoạ bằng biểu đồ tròn như sau

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ tròn (Pie Chart)

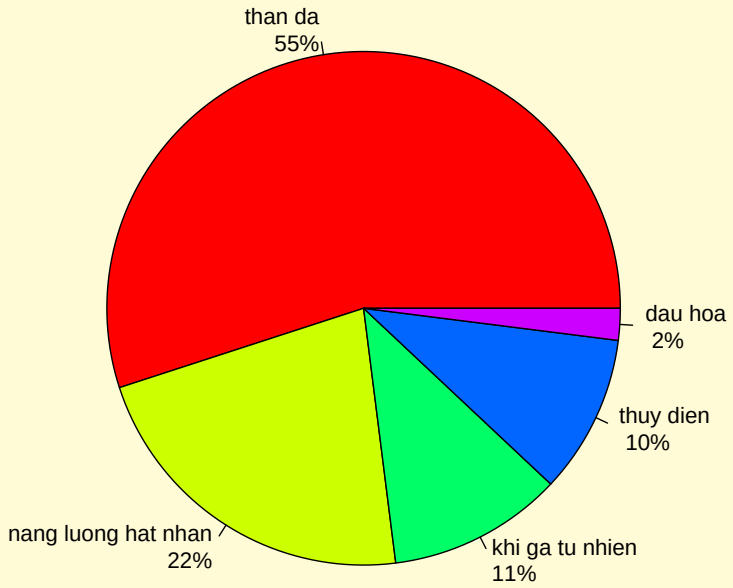
Trong đồ thị tròn, diện tích hình tròn được chia thành nhiều phần hình quạt, diện tích từng phần này tương ứng với tỷ lệ phân loại mà nó đại diện trong toàn thể.

Ví dụ: Năng lương điện của một quốc gia được sản xuất từ than đá chiếm 55%, từ năng lượng hạt nhân chiếm 22%, từ khí ga tự nhiên chiếm 11%, từ thuỷ điện chiếm 10%, từ dầu hoả chiếm 2%. Thông tin này có thể minh hoạ bằng biểu đồ tròn như sau

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ tròn (Pie Chart)

Nguồn nguyên liệu	Tỷ lệ
Than đá	55%
Năng lượng hạt nhân	22%
Khí ga tự nhiên	11%
Thuỷ điện	10%
Dầu hoả	2%



Vẽ biểu đồ tròn bằng R

```
pie(x, labels = names(x), col = , main = , ...)
```

Trong đó

`x`:	một vector những số thực không âm xác định diện tích các hình quạt
`labels`:	một vector những xâu chữ cho biết tên các hình quạt
`col`:	màu sắc các hình quạt
`main`:	xem `hist`

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Nội dung trình bày

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ thanh (Bar Chart)

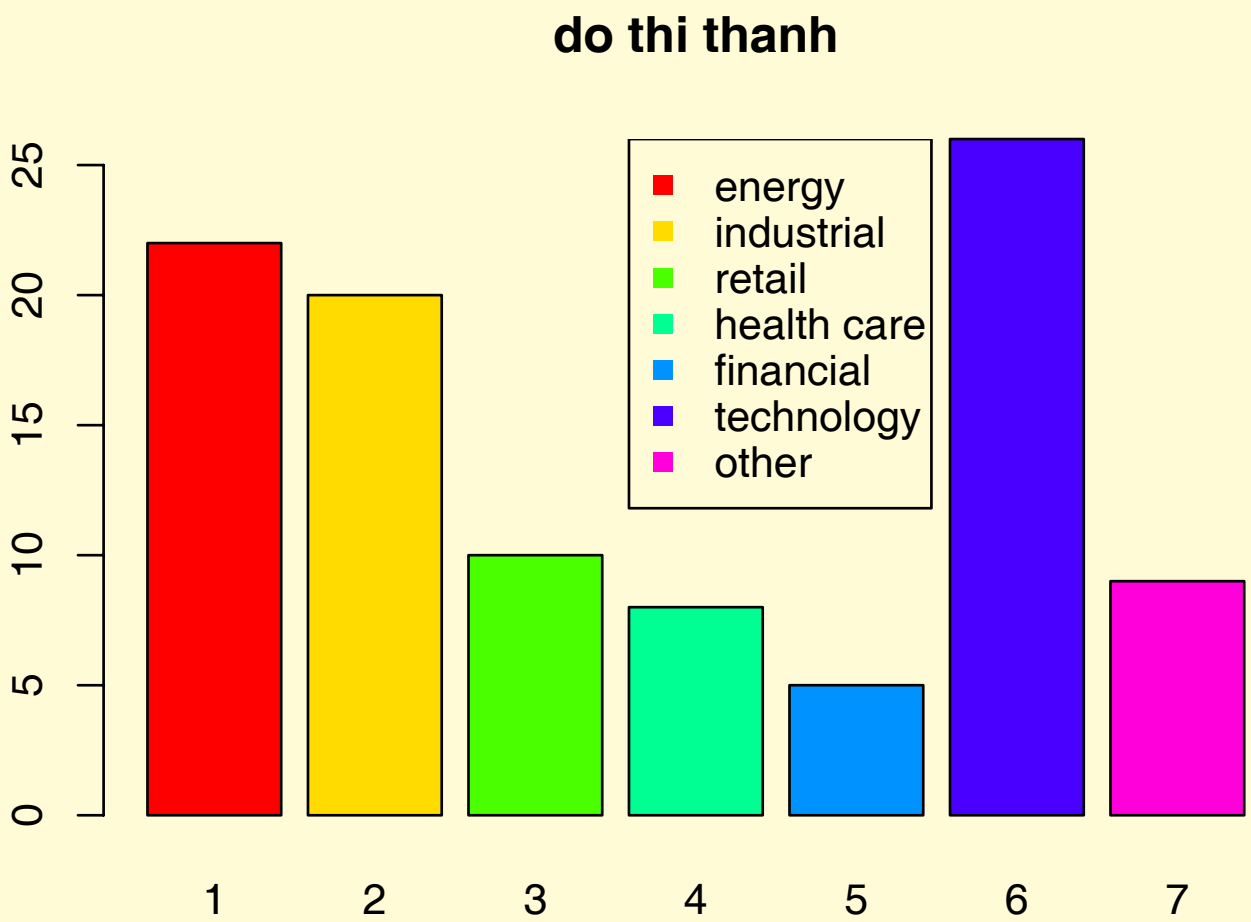
Đồ thị thanh mô tả tần số hay tỷ lệ mỗi phân loại theo chiều dài của thanh. Ta có thể dùng đồ thị thanh đứng hoặc ngang.

Ví dụ: Trong số 100 công ty tăng trưởng nhanh nhất của Mỹ năm 1998 có 22 công ty năng lượng, 20 công ty công nghiệp, 10 công ty bán lẻ, 8 công ty chăm sóc sức khỏe, 5 công ty tài chính, 26 công ty công nghệ, và 9 công ty thuộc các lĩnh vực khác.

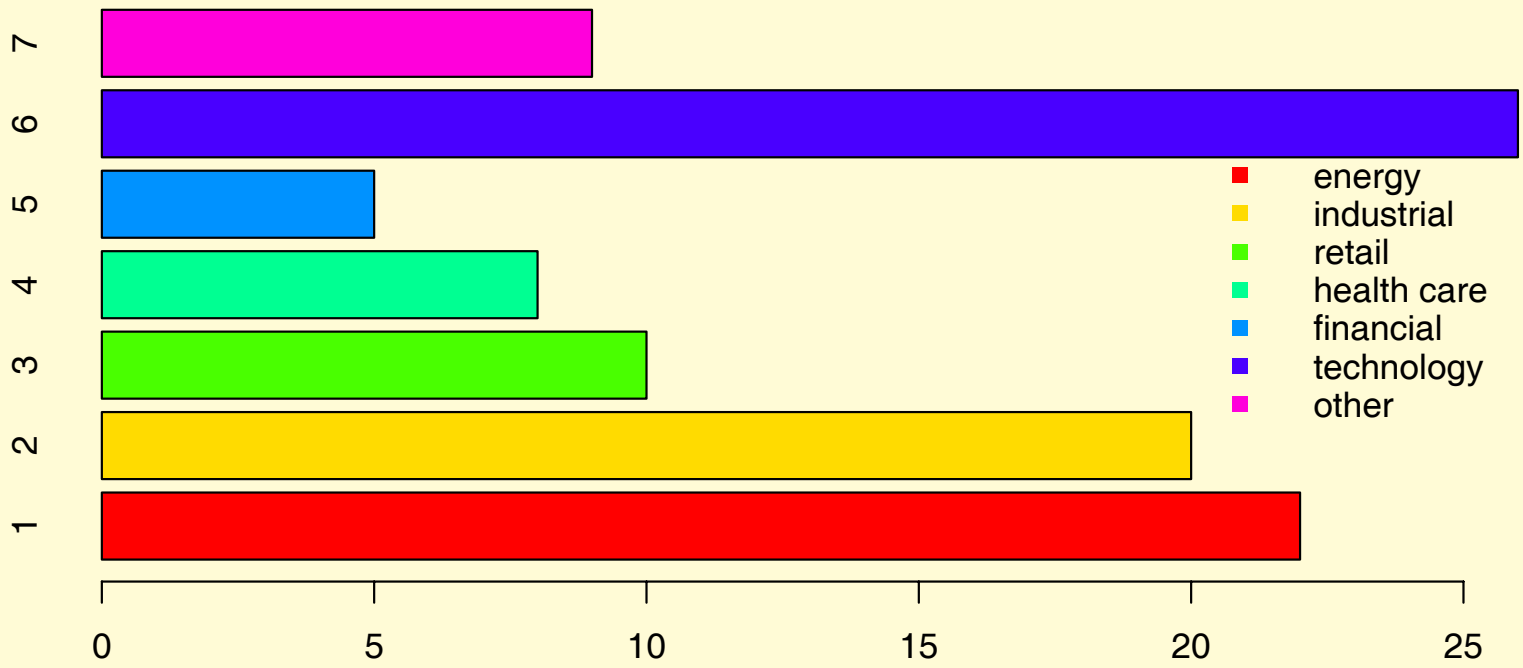
Vậy ta có bảng tần số

Công ty	Tần số
Năng lượng	22
Công nghiệp	20
Bán lẻ	10
Chăm sóc SK	8
Tài chính	5
Công nghệ	26
Lĩnh vực khác	9

Biểu đồ thanh (Bar Chart)



Vẽ biểu đồ thanh bằng R



ThangLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Vẽ biểu đồ thanh bằng R

```
barplot(height, names.arg = , horiz = FALSE, col = , main =
, xlab = , ylab = , ...)
```

Trong đó

`height`	một vector (hoặc một ma trận) những giá trị xác định độ dài các thanh
`names.arg`:	một vector những xâu chữ cho biết tên viết dưới các thanh
`horiz`:	FALSE/TRUE tương ứng với đồ thị thanh đứng/ngang
`col`:	màu sắc các thanh
`main, xlab, ylab`:	xem `hist`

Nội dung trình bày

1. Tóm tắt dữ liệu bằng bảng tần số
2. Biểu đồ mô tả hình dáng phân phối của tập dữ liệu
   - Biểu đồ thân và lá
   - Biểu đồ tần số
   - Đa giác tần số
3. Mô tả dữ liệu định tính
   - Biểu đồ tròn
   - Biểu đồ thanh
   - Biểu đồ Pareto
4. Biểu đồ mô tả mối quan hệ giữa hai biến định lượng

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ Pareto

Biểu đồ Pareto là một loại đồ thị thanh đứng đặc biệt trong đó thông tin về các quan sát được phân loại và đưa lên đồ thị theo thứ tự giảm dần của các tần số kết hợp với đa giác tần số tích lũy.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ Pareto

Tại một nhà máy sản xuất chai nhựa dùng đựng bơ người ta phải loại đi một số vì kém chất lượng. Người ta cũng thấy rằng những chai này bị loại là do: nhựa không tốt, máy dán nhãn làm sai, bạc màu, độ dày không đúng, tay cầm bị vỡ, và một số nguyên nhân khác.
Chọn 500 chai từ số chai bị loại, kết quả về nguyên nhân bị loại của những chai này như sau:

Vấn đề	Số lượng
Bạc màu	32

Biểu đồ Pareto

Tại một nhà máy sản xuất chai nhựa dùng đựng bơ người ta phải loại đi một số vì kém chất lượng. Người ta cũng thấy rằng những chai này bị loại là do: nhựa không tốt, máy dán nhãn làm sai, bạc màu, độ dày không đúng, tay cầm bị vỡ, và một số nguyên nhân khác.
Chọn 500 chai từ số chai bị loại, kết quả về nguyên nhân bị loại của những chai này như sau:

Vấn đề	Số lượng
Bạc màu	32
Độ dày không đúng	117

Biểu đồ Pareto

Tại một nhà máy sản xuất chai nhựa dùng đựng bơ người ta phải loại đi một số vì kém chất lượng. Người ta cũng thấy rằng những chai này bị loại là do: nhựa không tốt, máy dán nhãn làm sai, bạc màu, độ dày không đúng, tay cầm bị vỡ, và một số nguyên nhân khác.
Chọn 500 chai từ số chai bị loại, kết quả về nguyên nhân bị loại của những chai này như sau:

Vấn đề	Số lượng
Bạc màu	32
Độ dày không đúng	117
Vỡ quai	86

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ Pareto

Tại một nhà máy sản xuất chai nhựa dùng đựng bơ người ta phải loại đi một số vì kém chất lượng. Người ta cũng thấy rằng những chai này bị loại là do: nhựa không tốt, máy dán nhãn làm sai, bạc màu, độ dày không đúng, tay cầm bị vỡ, và một số nguyên nhân khác.
Chọn 500 chai từ số chai bị loại, kết quả về nguyên nhân bị loại của những chai này như sau:

Vấn đề	Số lượng
Bạc màu	32
Độ dày không đúng	117
Vỡ quai	86
Nhựa không tốt	221

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ Pareto

Tại một nhà máy sản xuất chai nhựa dùng đựng bơ người ta phải loại đi một số vì kém chất lượng. Người ta cũng thấy rằng những chai này bị loại là do: nhựa không tốt, máy dán nhãn làm sai, bạc màu, độ dày không đúng, tay cầm bị vỡ, và một số nguyên nhân khác.
Chọn 500 chai từ số chai bị loại, kết quả về nguyên nhân bị loại của những chai này như sau:

Vấn đề	Số lượng
Bạc màu	32
Độ dày không đúng	117
Vỡ quai	86
Nhựa không tốt	221
Dán nhãn sai	40

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

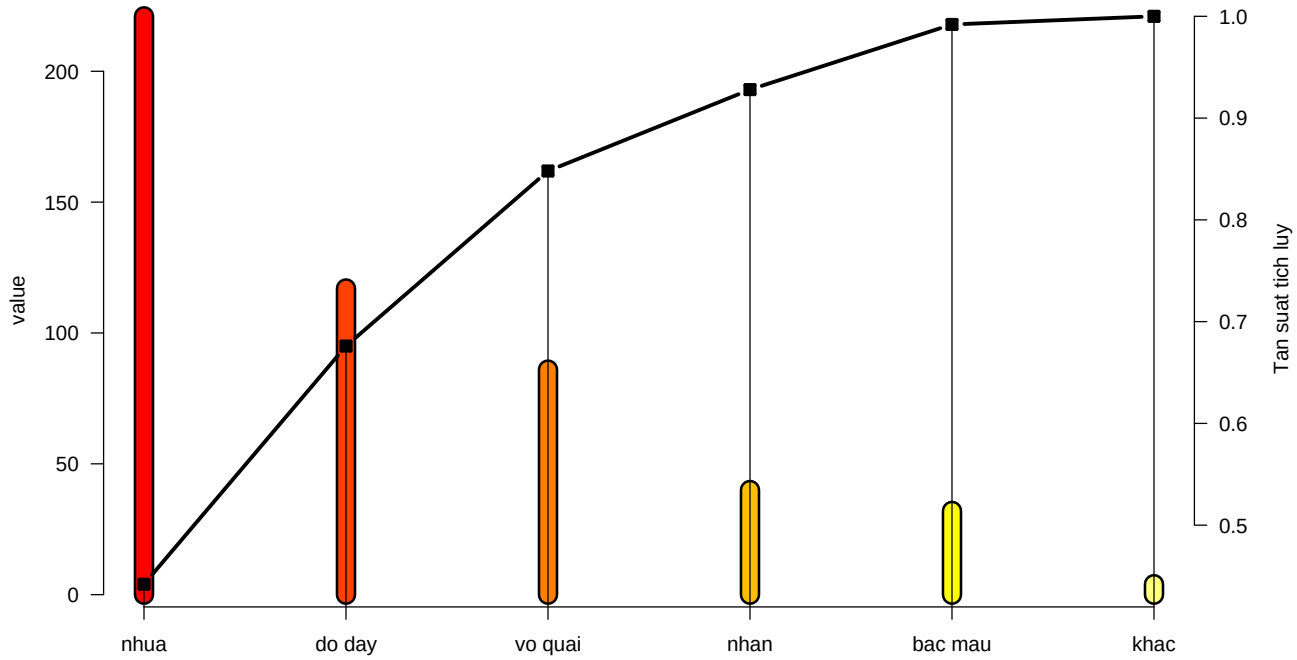
Biểu đồ Pareto

Tại một nhà máy sản xuất chai nhựa dùng đựng bơ người ta phải loại đi một số vì kém chất lượng. Người ta cũng thấy rằng những chai này bị loại là do: nhựa không tốt, máy dán nhãn làm sai, bạc màu, độ dày không đúng, tay cầm bị vỡ, và một số nguyên nhân khác.
Chọn 500 chai từ số chai bị loại, kết quả về nguyên nhân bị loại của những chai này như sau:

Vấn đề	Số lượng
Bạc màu	32
Độ dày không đúng	117
Vỡ quai	86
Nhựa không tốt	221
Dán nhãn sai	40
Khác	4

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ Pareto



Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Biểu đồ tán xạ (Scatter diagram)

Phương pháp tống kê thường được dùng để nghiên cứu và định lượng mối quan hệ giữa các biến. Từ đó mô tả, dự đoán và điều khiển một biến gọi là biến phụ thuộc (ký hiệu là y). Việc này được thực hiện bằng cách liên hệ y với một hay nhiều biến khác được gọi là các biến độc lập. Một phương pháp được sử dụng là phân tích hồi quy. Phương pháp này đưa ra một phương trình liên hệ y với các biến độc lập. Phương pháp này sẽ được giới thiệu ở những chương sau. Trong mục này ta sẽ dùng một cách đơn giản là đồ thị để xem xét mối quan hệ giữa các biến.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Dùng biểu đồ tán xạ nghiên cứu mối quan hệ giữa hai biến

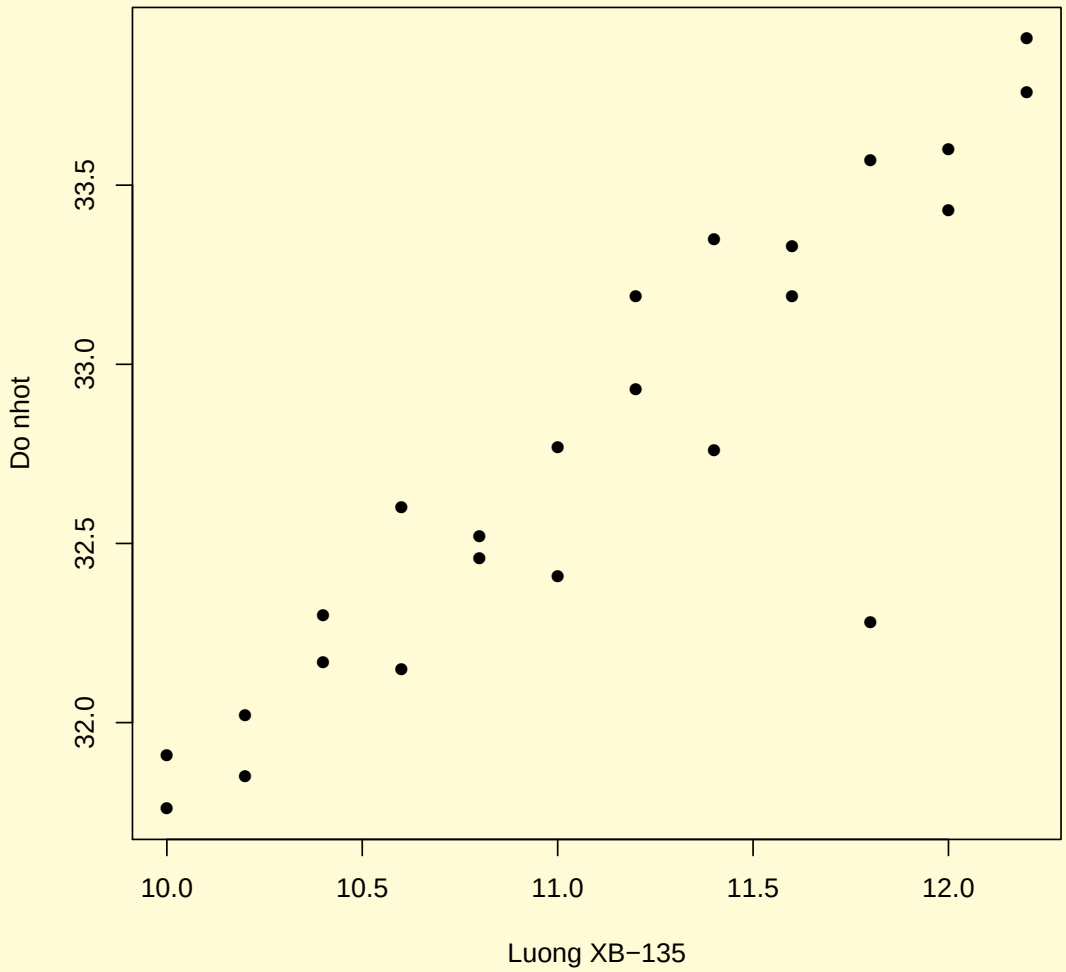
Ví dụ: Tại một nhà máy sản xuất một chất hóa học người ta nhận thấy có mối quan hệ giữa độ nhớt của sản phẩm sản xuất ra với lượng chất XB-135 được dùng trong quá trình sản xuất. Để tìm hiểu về mối quan hệ này, người ta chọn ra 24 mẻ sản phẩm, lượng XB-135 dùng trong từng mẻ và độ nhớt của sản phẩm như sau

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Dùng biểu đồ tán xạ nghiên cứu mối quan hệ giữa hai biến

Lượng XB-135	Độ nhớt	Lượng XB-135	Độ nhớt
10	31.76	11.2	32.93
10	31.91	11.2	33.19
10.2	32.02	11.4	33.35
10.2	31.85	11.4	32.76
10.4	32.17	11.6	33.33
10.4	32.3	11.6	33.19
10.6	32.6	11.8	32.28
10.6	32.15	11.8	35.57
10.8	32.52	12	33.60
10.8	32.46	12	33.43
11	32.41	12.2	33.91
11	32.77	12.2	33.76

Dùng biểu đồ tán xạ nghiên cứu mối quan hệ giữa hai biến



Dùng biểu đồ tán xạ nghiên cứu mối quan hệ giữa hai biến

Nhìn vào đồ thị ta thấy các điểm trên đồ thị phân bố xung quanh một đường thẳng và ta nói có mối quan hệ tuyến tính giữa độ nhớt (y) của sản phẩm và lượng XB-135 (x). Ta cũng thấy rằng đường thẳng nói trên có hướng đi lên, như vậy giữa hai đại lượng có mối quan hệ đồng biến.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Vẽ biểu đồ tán xạ trong R

```
plot(x, y, xlab = , ylab =, main = , ...)
```

Trong đó

`x`	một vector những giá trị biểu diễn trên trục hoành
`y:`	một vector những giá trị biểu diễn trên trục tung
`main, xlab, ylab:`	xem `hist()`

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775