







# Bài giảng Xác suất Thống kê và ứng dụng

Phan Thanh Hồng

Bộ môn Toán-Đại học THĂNG LONG

Ngày 16 tháng 11 năm 2009

-  Hoàng Trọng, Chu Nguyễn Mộng Ngọc, *Thống kê ứng dụng trong kinh tế xã hội*, NXB Thống kê, 2007.
-  Nguyễn Cao Văn, Trần Thái Ninh, *Giáo trình lý thuyết xác suất và thống kê toán*, NXB Thống kê, 2005.
-  Nguyễn Văn Tuấn, *Phân tích số liệu và tạo biểu đồ bằng R hướng dẫn thực hành*, Tài liệu trên mạng Internet.
-  Thống kê I, II, Tài liệu lưu hành nội bộ, Hà Nội, 2003.
-  Emmanuel Paradis, *R for beginner*, Tài liệu trên mạng Internet.
-  John Verzani, *Simple R - Using R for Introductory Statistics*, Tài liệu trên mạng Internet.

# Phần I

## Giới thiệu môn học

- 1 Thống kê
- 2 Thu thập dữ liệu và thống kê mô tả
- 3 Thống kê suy diễn và mô hình xác suất
- 4 Sơ lược lịch sử thống kê
- 5 Một số khái niệm
  - Tổng thể và mẫu
  - Đặc điểm thống kê
  - Thang đo

# Phần I

- 1 Thống kê
- 2 Thu thập dữ liệu và thống kê mô tả
- 3 Thống kê suy diễn và mô hình xác suất
- 4 Sơ lược lịch sử thống kê
- 5 Một số khái niệm
  - Tổng thể và mẫu
  - Đặc điểm thống kê
  - Thang đo

# Phần I

- 1 Thống kê
- 2 Thu thập dữ liệu và thống kê mô tả
- 3 Thống kê suy diễn và mô hình xác suất
- 4 Sơ lược lịch sử thống kê
- 5 Một số khái niệm
  - Tổng thể và mẫu
  - Đặc điểm thống kê
  - Thang đo

# Phần I

- 1 Thống kê
- 2 Thu thập dữ liệu và thống kê mô tả
- 3 Thống kê suy diễn và mô hình xác suất
- 4 Sơ lược lịch sử thống kê
- 5 Một số khái niệm
  - Tổng thể và mẫu
  - Đặc điểm thống kê
  - Thang đo

# Phần I

- 1 Thống kê
- 2 Thu thập dữ liệu và thống kê mô tả
- 3 Thống kê suy diễn và mô hình xác suất
- 4 Sơ lược lịch sử thống kê
- 5 Một số khái niệm
  - Tổng thể và mẫu
  - Đặc điểm thống kê
  - Thang đo



# Phần I

- 1 Thống kê
- 2 Thu thập dữ liệu và thống kê mô tả
- 3 Thống kê suy diễn và mô hình xác suất
- 4 Sơ lược lịch sử thống kê
- 5 Một số khái niệm
  - Tổng thể và mẫu
  - Đặc điểm thống kê
  - Thang đo

# Phần I

- 1 Thống kê
- 2 Thu thập dữ liệu và thống kê mô tả
- 3 Thống kê suy diễn và mô hình xác suất
- 4 Sơ lược lịch sử thống kê
- 5 Một số khái niệm
  - Tổng thể và mẫu
  - Đặc điểm thống kê
  - Thang đo

# Phần I

- 1 Thống kê
- 2 Thu thập dữ liệu và thống kê mô tả
- 3 Thống kê suy diễn và mô hình xác suất
- 4 Sơ lược lịch sử thống kê
- 5 Một số khái niệm
  - Tổng thể và mẫu
  - Đặc điểm thống kê
  - Thang đo

- ♣ **Thống kê** là thuật ngữ được dùng để chỉ ngành khoa học nghiên cứu những vấn đề liên quan đến việc thu thập, phân tích và biểu diễn dữ liệu. Dữ liệu ở đây là những con số, sự kiện mà dựa vào đó người ta đưa ra những kết luận nhất định.
- ♣ Thống kê được ứng dụng trong nhiều lĩnh vực như sinh học, y học, tâm lý học, .... Trong lĩnh vực xã hội nói chung và kinh tế nói riêng, thống kê đóng vai trò là công cụ chính, quan trọng trong nhận định tình hình và hỗ trợ việc ra quyết định. Thống kê được dùng để nhận ra và hiểu các biến thiên có hệ thống khi đo lường các hiện tượng kinh tế - xã hội, để tóm tắt dữ liệu và ra quyết định dựa trên dữ liệu đó.

# Thu thập dữ liệu và thống kê mô tả

Đôi khi việc phân tích thống kê bắt đầu với tập dữ liệu có sẵn: chẳng hạn tỷ lệ thất nghiệp, tổng thu nhập quốc dân, tỷ lệ lạm phát, động đất, . . . . Sử dụng chúng nhà thống kê có thể miêu tả, tổng kết và phân tích dữ liệu. Trong trường hợp dữ liệu không có sẵn, lý thuyết thống kê được sử dụng để tạo ra những thử nghiệm thích hợp nhằm thu thập dữ liệu. Dữ liệu sau khi thu thập sẽ được mô tả. Nội dung này của thống kê liên quan tới việc mô tả, tổng kết dữ liệu gọi là **thống kê mô tả**.

♣ **Thống kê mô tả** là khoa học mô tả những mặt quan trọng của một tập dữ liệu.

# Thu thập dữ liệu và thống kê mô tả

Đôi khi việc phân tích thống kê bắt đầu với tập dữ liệu có sẵn: chẳng hạn tỷ lệ thất nghiệp, tổng thu nhập quốc dân, tỷ lệ lạm phát, động đất, . . . . Sử dụng chúng nhà thống kê có thể miêu tả, tổng kết và phân tích dữ liệu. Trong trường hợp dữ liệu không có sẵn, lý thuyết thống kê được sử dụng để tạo ra những thử nghiệm thích hợp nhằm thu thập dữ liệu. Dữ liệu sau khi thu thập sẽ được mô tả. Nội dung này của thống kê liên quan tới việc mô tả, tổng kết dữ liệu gọi là **thống kê mô tả**.

♣ **Thống kê mô tả** là khoa học mô tả những mặt quan trọng của một tập dữ liệu.

# Thu thập dữ liệu và thống kê mô tả

Đôi khi việc phân tích thống kê bắt đầu với tập dữ liệu có sẵn: chẳng hạn tỷ lệ thất nghiệp, tổng thu nhập quốc dân, tỷ lệ lạm phát, động đất, . . . . Sử dụng chúng nhà thống kê có thể miêu tả, tổng kết và phân tích dữ liệu. Trong trường hợp dữ liệu không có sẵn, lý thuyết thống kê được sử dụng để tạo ra những thử nghiệm thích hợp nhằm thu thập dữ liệu. Dữ liệu sau khi thu thập sẽ được mô tả. Nội dung này của thống kê liên quan tới việc mô tả, tổng kết dữ liệu gọi là **thống kê mô tả**.

♣ **Thống kê mô tả** là khoa học mô tả những mặt quan trọng của một tập dữ liệu.

# Thống kê suy diễn và mô hình xác suất

Sau khi thử nghiệm hoàn thành và dữ liệu được mô tả và tổng hợp, người nghiên cứu hi vọng có thể rút ra những kết luận nhất định trên tập dữ liệu. Nội dung này của thống kê liên quan tới việc đưa ra kết luận gọi là **thống kê suy diễn**.

Để có thể đưa ra kết luận logic từ tập dữ liệu, ta thường cần một số giả thiết về khả năng xảy ra của các giá trị khác nhau trong tập dữ liệu. Tất cả những giả thiết này gọi là **mô hình xác suất** của dữ liệu.



# Thống kê suy diễn và mô hình xác suất

Sau khi thử nghiệm hoàn thành và dữ liệu được mô tả và tổng hợp, người nghiên cứu hi vọng có thể rút ra những kết luận nhất định trên tập dữ liệu. Nội dung này của thống kê liên quan tới việc đưa ra kết luận gọi là **thống kê suy diễn**.

Để có thể đưa ra kết luận logic từ tập dữ liệu, ta thường cần một số giả thiết về khả năng xảy ra của các giá trị khác nhau trong tập dữ liệu. Tất cả những giả thiết này gọi là **mô hình xác suất** của dữ liệu.

# Sơ lược lịch sử thống kê

Việc thu thập dữ liệu có hệ thống về dân số và xã hội bắt đầu ở Italy trong suốt thời kỳ Phục hưng. Thuật ngữ **thống kê (statistics)** bắt nguồn từ chữ **state**, hàm ý việc thu thập dữ liệu theo yêu cầu của chính quyền. Việc này sau đó mở rộng ra các nước Tây Âu. Thực ra do điều kiện sức khỏe cộng đồng kém, dịch bệnh thường xảy ra (đặc biệt là bệnh dịch hạch), nên người dân phải khai báo việc sinh, tử và kết hôn như một yêu cầu bắt buộc.

Cho đến thế kỷ 18, thống kê được hiểu là khoa học mô tả trạng thái. Mặc dù sang thế kỷ 19, lý thuyết xác suất đã phát triển nhưng ứng dụng trong thống kê còn hạn chế. Và do đó, suy diễn xác suất từ mẫu lên tổng thể chưa phát triển.

Từ cuối thế kỷ 19, các nhà thống kê quan tâm tới việc đưa ra kết luận từ dữ liệu và thống kê suy diễn bắt đầu có những bước tiến mạnh mẽ với nhiều kết quả nghiên cứu quan trọng bởi nhiều nhà khoa học.

# Sơ lược lịch sử thống kê

Việc thu thập dữ liệu có hệ thống về dân số và xã hội bắt đầu ở Italy trong suốt thời kỳ Phục hưng. Thuật ngữ **thống kê (statistics)** bắt nguồn từ chữ **state**, hàm ý việc thu thập dữ liệu theo yêu cầu của chính quyền. Việc này sau đó mở rộng ra các nước Tây Âu. Thực ra do điều kiện sức khỏe cộng đồng kém, dịch bệnh thường xảy ra (đặc biệt là bệnh dịch hạch), nên người dân phải khai báo việc sinh, tử và kết hôn như một yêu cầu bắt buộc.

Cho đến thế kỷ 18, thống kê được hiểu là khoa học mô tả trạng thái. Mặc dù sang thế kỷ 19, lý thuyết xác suất đã phát triển nhưng ứng dụng trong thống kê còn hạn chế. Và do đó, suy diễn xác suất từ mẫu lên tổng thể chưa phát triển.

Từ cuối thế kỷ 19, các nhà thống kê quan tâm tới việc đưa ra kết luận từ dữ liệu và thống kê suy diễn bắt đầu có những bước tiến mạnh mẽ với nhiều kết quả nghiên cứu quan trọng bởi nhiều nhà khoa học.

Đầu thế kỷ 20, hai lĩnh vực ứng dụng thống kê quan trọng nhất là nông nghiệp và sinh học. Ngày nay, thống kê có ứng dụng trong rất nhiều lĩnh vực. Trong đó thống kê suy diễn không thể thiếu trong nghiên cứu về sức khỏe, y học, giáo dục, kinh tế, dự báo khí tượng, thể thao, bảo hiểm, ...

# Nội dung trình bày

- 1 Thống kê
- 2 Thu thập dữ liệu và thống kê mô tả
- 3 Thống kê suy diễn và mô hình xác suất
- 4 Sơ lược lịch sử thống kê
- 5 Một số khái niệm
  - Tổng thể và mẫu
  - Đặc điểm thống kê
  - Thang đo

♣ **Tổng thể:** là một tập hợp các phần tử thường là người, vật hay sự kiện.

**Ví dụ:** Tổng thể bao gồm tất cả các sinh viên tốt nghiệp năm ngoài của một trường đại học, tất cả những chiếc xe ô tô được sản xuất trong một năm của một công ty, tất cả các đám cháy tại một thành phố trong một tháng.

♣ **Mẫu** là một tập con của tổng thể.

**Ví dụ:** Xét tổng thể các sinh viên mới tốt nghiệp nói trên và giả sử có 2675 sinh viên trong tổng thể đó, vì số lượng sinh viên quá nhiều nên khó có thể điều tra được hết nên người ta chọn ra một mẫu và thu thập số liệu trên mẫu đó.







♣ **Tổng thể:** là một tập hợp các phần tử thường là người, vật hay sự kiện.

**Ví dụ:** Tổng thể bao gồm tất cả các sinh viên tốt nghiệp năm ngoái của một trường đại học, tất cả những chiếc xe ô tô được sản xuất trong một năm của một công ty, tất cả các đám cháy tại một thành phố trong một tháng.

♣ **Mẫu** là một tập con của tổng thể.

**Ví dụ:** Xét tổng thể các sinh viên mới tốt nghiệp nói trên và giả sử có 2675 sinh viên trong tổng thể đó, vì số lượng sinh viên quá nhiều nên khó có thể điều tra được hết nên người ta chọn ra một mẫu và thu thập số liệu trên mẫu đó.

# Nội dung trình bày

- 1 Thống kê
- 2 Thu thập dữ liệu và thống kê mô tả
- 3 Thống kê suy diễn và mô hình xác suất
- 4 Sơ lược lịch sử thống kê
- 5 Một số khái niệm
  - Tổng thể và mẫu
  - Đặc điểm thống kê
  - Thang đo

- ♣ Mỗi tính chất của các phần tử trong tổng thể được gọi là một **đặc điểm thống kê** (hay **biến**).

**Ví dụ:** Khi xem xét về tổng thể các sinh viên mới tốt nghiệp, đặc điểm thống kê có thể là: giới tính, tuổi, lương khởi điểm, ... Đối với tổng thể những chiếc xe ô tô, đặc điểm thống kê có thể là: màu sắc, chủng loại, lượng xăng tiêu hao trên 100 km, ...

♣ Mỗi tính chất của các phần tử trong tổng thể được gọi là một **đặc điểm thống kê** (hay **biến**).

**Ví dụ:** Khi xem xét về tổng thể các sinh viên mới tốt nghiệp, đặc điểm thống kê có thể là: giới tính, tuổi, lương khởi điểm, ... Đối với tổng thể những chiếc xe ô tô, đặc điểm thống kê có thể là: màu sắc, chủng loại, lượng xăng tiêu hao trên 100 km, ...

♣ Đặc điểm thống kê gồm hai loại:

- ♣ **Đặc điểm định lượng** là tính chất của các phần tử mà có thể biểu hiện bằng con số, chẳng hạn như: chiều cao, cân nặng, lương, ... Đặc điểm định lượng gọi là rời rạc nếu các giá trị của nó là hữu hạn hay đếm được, và gọi là liên tục nếu các giá trị của nó lấp đầy một khoảng trên trục số.
- ♣ **Đặc điểm định tính** là tính chất của các phần tử không biểu hiện trực tiếp bằng các con số. Ví dụ: giới tính, dân tộc, màu sắc, ...

♣ Đặc điểm thống kê gồm hai loại:

- ♣ **Đặc điểm định lượng** là tính chất của các phần tử mà có thể biểu hiện bằng con số, chẳng hạn như: chiều cao, cân nặng, lương, ... Đặc điểm định lượng gọi là rời rạc nếu các giá trị của nó là hữu hạn hay đếm được, và gọi là liên tục nếu các giá trị của nó lấp đầy một khoảng trên trục số.
- ♣ **Đặc điểm định tính** là tính chất của các phần tử không biểu hiện trực tiếp bằng các con số. Ví dụ: giới tính, dân tộc, màu sắc, ...

- ♣ Chúng ta tiến hành việc đo lường để gán giá trị của đặc điểm thống kê đối với từng phần tử của tổng thể. Giá trị này là biểu hiện của đặc điểm thống kê trên phần tử đó. Nếu ta đo lường trên mọi phần tử của tổng thể ta sẽ có một tập hợp các số đo gọi là **tổng thể các quan sát** hay **tổng thể các số đo**.
- ♣ Khi đo lường trên mẫu đã chọn ra ta được một mẫu những quan sát hay mẫu những số đo.

# Nội dung trình bày

- 1 Thống kê
- 2 Thu thập dữ liệu và thống kê mô tả
- 3 Thống kê suy diễn và mô hình xác suất
- 4 Sơ lược lịch sử thống kê
- 5 Một số khái niệm
  - Tổng thể và mẫu
  - Đặc điểm thống kê
  - Thang đo



Các đặc điểm thống kê mô tả các đặc trưng của tổng thể có thể là định tính hay định lượng. Để biểu diễn các giá trị của các số đo trong tổng thể hay mẫu ta dùng các thang đo khác nhau.

♣ **Thang đo định danh:** dùng để đánh số các tính chất hay phạm trù cùng loại.

Chẳng hạn thang giới tính gồm hai phạm trù là Nam (0) và Nữ (1).  
Màu sắc sản phẩm có thể là Xanh (1), Đỏ (2), Tím (3), Vàng (4),...  
Các con số ở đây không có quan hệ hơn kém. Thang định danh thường dùng để đếm tần số các hiện tượng xảy ra.

- ♣ **Thang đo định danh:** dùng để đánh số các tính chất hay phạm trù cùng loại.  
Chẳng hạn thang giới tính gồm hai phạm trù là Nam (0) và Nữ (1).  
Màu sắc sản phẩm có thể là Xanh (1), Đỏ (2), Tím (3), Vàng (4),...  
Các con số ở đây không có quan hệ hơn kém. Thang định danh thường dùng để đếm tần số các hiện tượng xảy ra.

- ♣ **Thang đo định danh:** dùng để đánh số các tính chất hay phạm trù cùng loại.  
Chẳng hạn thang giới tính gồm hai phạm trù là Nam (0) và Nữ (1).  
Màu sắc sản phẩm có thể là Xanh (1), Đỏ (2), Tím (3), Vàng (4),...  
Các con số ở đây không có quan hệ hơn kém. Thang định danh thường dùng để đếm tần số các hiện tượng xảy ra.

♣ **Thang đo thứ bậc:** là loại thang định danh mà giữa các phạm trù có quan hệ hơn kém.

Chẳng hạn để đặc trưng cho đánh giá của khách hàng về sản phẩm có thể dùng thang đo: Không hài lòng(1), tương đối hài lòng (2), rất hài lòng (3).

♣ **Thang đo thứ bậc:** là loại thang định danh mà giữa các phạm trù có quan hệ hơn kém.

Chẳng hạn để đặc trưng cho đánh giá của khách hàng về sản phẩm có thể dùng thang đo: Không hài lòng(1), tương đối hài lòng (2), rất hài lòng (3).

♣ **Thang đo khoảng:** là thang đo thứ bậc có các khoảng cách đều nhau giữa các bậc.

Dùng thang đo khoảng ta có thể tính toán trên các số đo, từ đó tính được các tham số như trung bình, phương sai, ... Khi sử dụng thang đo này các số đo có thể được xếp vào các lớp có khoảng cách không nhất thiết phải đều nhau.

Chẳng hạn để đặc trưng lứa tuổi có thể dùng thang đo khoảng: Trẻ (30)(dưới 35 tuổi), trung niên (50)(từ 36 đến 60 tuổi), già (70)(trên 61 tuổi). Hay nhiệt độ có thể chia ra thành: dưới  $20^{\circ}\text{C}$ , từ  $20^{\circ}\text{C}$  đến  $30^{\circ}\text{C}$ , trên  $30^{\circ}\text{C}$ .

♣ **Thang đo khoảng:** là thang đo thứ bậc có các khoảng cách đều nhau giữa các bậc.

Dùng thang đo khoảng ta có thể tính toán trên các số đo, từ đó tính được các tham số như trung bình, phương sai, ... Khi sử dụng thang đo này các số đo có thể được xếp vào các lớp có khoảng cách không nhất thiết phải đều nhau.

Chẳng hạn để đặc trưng lứa tuổi có thể dùng thang đo khoảng: Trẻ (30)(dưới 35 tuổi), trung niên (50)(từ 36 đến 60 tuổi), già (70)(trên 61 tuổi). Hay nhiệt độ có thể chia ra thành: dưới  $20^{\circ}\text{C}$ , từ  $20^{\circ}\text{C}$  đến  $30^{\circ}\text{C}$ , trên  $30^{\circ}\text{C}$ .



♣ **Thang đo khoảng:** là thang đo thứ bậc có các khoảng cách đều nhau giữa các bậc.

Dùng thang đo khoảng ta có thể tính toán trên các số đo, từ đó tính được các tham số như trung bình, phương sai, ... Khi sử dụng thang đo này các số đo có thể được xếp vào các lớp có khoảng cách không nhất thiết phải đều nhau.

Chẳng hạn để đặc trưng lứa tuổi có thể dùng thang đo khoảng: Trẻ (30)(dưới 35 tuổi), trung niên (50)(từ 36 đến 60 tuổi), già (70)(trên 61 tuổi). Hay nhiệt độ có thể chia ra thành: dưới  $20^{\circ}\text{C}$ , từ  $20^{\circ}\text{C}$  đến  $30^{\circ}\text{C}$ , trên  $30^{\circ}\text{C}$ .

- ♣ **Thang đo tỷ lệ:** là thang đo khoảng với một điểm gốc tuyệt đối. Chỉ với thang đo này ta mới có thể đo lường các hiện tượng như các đơn vị đo lường vật lý thông thường.  
Ví dụ: tiền tệ, mét, kg, tấn, tạ, ... là những thang đo tỷ lệ.

- ♣ **Thang đo tỷ lệ:** là thang đo khoảng với một điểm gốc tuyệt đối. Chỉ với thang đo này ta mới có thể đo lường các hiện tượng như các đơn vị đo lường vật lý thông thường.  
Ví dụ: tiền tệ, mét, kg, tấn, tạ, ... là những thang đo tỷ lệ.

Hai thang đo đầu tiên cung cấp cho ta các dữ liệu định tính, còn gọi là thang đo định tính. Hai thang đo còn lại cung cấp cho ta các dữ liệu định lượng còn gọi là thang đo định lượng. Dữ liệu định lượng sau khi thu thập có thể chuyển về dạng dữ liệu thứ bậc định tính, tuy nhiên không thực hiện việc chuyển đổi ngược lại được.