

MÔN HỌC

**THỐNG KÊ ỨNG DỤNG - XD (KC107)**



GIÁO VIÊN PHỤ TRÁCH

**ĐẶNG THẾ GIA**

Bộ môn Kỹ Thuật Xây Dựng  
Khoa Công Nghệ, Trường Đại Học Cần Thơ

## Nội dung chương

1. Định nghĩa
2. Thống kê vs Xác suất
3. Các khái niệm trong thống kê
4. Phương pháp & Quá trình thống kê
5. Dữ liệu thống kê & Thang đo

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Chương 1:

### GIỚI THIỆU THỐNG KÊ HỌC (Introduction to Statistics)

Bộ môn Kỹ thuật xây dựng

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Các định nghĩa thống kê

- Theo nhà thống kê học Arthur Lyon Bowley, thống kê là “**báo cáo bằng số của các dữ kiện** trong bất kỳ bộ phận của cuộc điều tra được đặt trong mối quan hệ với nhau” (numerical statements of facts in any department of inquiry placed in relation to each other).
- Theo từ điển Merriam-Webster, thống kê là “**những dữ kiện được phân loại** đại diện cho các điều kiện của một người trong một nhà nước - đặc biệt là các dữ kiện mà nó có thể được định nghĩa bằng các con số hoặc bất kỳ dạng bảng hoặc phân loại sắp xếp khác” (classified facts representing the conditions of a people in a state – especially the facts that can be stated in numbers or any other tabular or classified arrangement).

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thống kê học

- **Thống kê học** là ngành nghiên cứu của tập hợp nhiều lĩnh vực khác nhau bao gồm **thu thập** (collection), **phân tích** (analysis), **giải thích** (interpretation), **trình bày** (presentation), và **tổ chức** (organization) dữ liệu.
- **Thống kê học** là môn khoa học nghiên cứu hệ thống các **phương pháp thu thập, xử lý và phân tích** các con số (mật lượng) của những hiện tượng số lớn để tìm hiểu bản chất và tính quy luật (mặt chất) vốn có của nó trong những điều kiện thời gian và địa điểm cụ thể.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thống kê làm những việc gì?

Thống kê học là một ngành học giúp ta thực hiện những công việc sau:

1. **Thu thập dữ liệu** (data) và **thiết kế** các nghiên cứu định lượng
2. **Tóm tắt thông tin** nhằm hỗ trợ quá trình tìm hiểu về một vấn đề hoặc đối tượng nào đó
3. **Kiểm định** để đưa ra những kết luận (information) dựa trên các số liệu giả thuyết, và
4. **Ước lượng** hiện tại hoặc dự báo tương lai (information)

Thống kê thường đi kèm với môn học song hành là **Xác suất**, là ngành học nhằm đưa ra các mô hình toán học về **sự ngẫu nhiên** và cho phép tính toán về **sự ngẫu nhiên** trong những trường hợp phức tạp.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thống kê học

- Việc áp dụng thống kê trong các lĩnh vực như khảo học, công nghiệp, hoặc các vấn đề xã hội, thật cần thiết để bắt đầu nghiên cứu với một quần thể thống kê (statistical population) hoặc một tiến trình (statistical model process).
- **Quần thể** là một đa dạng các chủ đề như "*tổng số dân của một châu lục*" hay "*số nguyên tử hydro trong tự nhiên*". Thống kê đề cập tới tất cả các khía cạnh của dữ liệu bao gồm việc lập kế hoạch thu thập dữ liệu (mẫu) cho các cuộc khảo sát và thí nghiệm.
- Khi không thể khảo sát trên một quần thể/tổng thể, chúng ta sẽ tiến hành khảo sát/thí nghiệm trên **mẫu**.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## THỐNG KÊ & XÁC SUẤT

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thống kê vs Xác Suất [1]

**Xác suất** là dự đoán khả năng xảy ra của những dữ kiện trong tương lai, trong khi **thống kê** là việc phân tích tần suất của những dữ kiện đã xảy ra (*Probability deals with predicting the likelihood of future events, while statistics involves the analysis of the frequency of past events*).

**Xác suất** cơ bản là một ngành lý thuyết của toán học trong đó nghiên cứu hệ quả của các định nghĩa toán học. **Thống kê** cơ bản là một ngành ứng dụng của toán học trong đó làm cho các quan sát trong thế giới thực trở nên có ý nghĩa (*Probability is primarily a theoretical branch of mathematics, which studies the consequences of mathematical definitions. Statistics is primarily an applied branch of mathematics, which tries to make sense of observations in the real world*).

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thống kê và Xác Suất - Ví dụ

### Quan hệ mưa rào và dòng chảy lũ

- Khi nghiên cứu quan hệ mưa rào và dòng chảy lũ của lưu vực A, ta thấy sự xuất hiện lũ trên các sông rất **ngẫu nhiên**. Cùng một lượng mưa rơi trên lưu vực nhưng ở những lần khác nhau sẽ cho lưu lượng đỉnh lũ khác nhau.
- Khi quan sát **hiện tượng ngẫu nhiên** một ít lần thì không thấy theo quy luật nào! Nhưng quan sát nhiều lần ta thấy có một quy luật nhất định, gọi là **quy luật đám đông**.
- Trong tự nhiên các hiện tượng ngẫu nhiên rất phức tạp, do đó ta phải thống kê nhiều số liệu ngẫu nhiên này để tính toán **xác suất xuất hiện**, Vì vậy muốn ứng dụng **lý thuyết xác suất** vào thực tế ta phải sử dụng **lý thuyết thống kê** (thống kê xác suất).

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thống kê vs Xác Suất [2]

**Xác suất** là ngôn ngữ toán học của sự ngẫu nhiên trong đó cho phép bạn lý giải hoặc lập báo cáo dự đoán về kết quả của các hệ thống hoặc quá trình vật lý có tính ngẫu nhiên hoặc không chắc chắn (*Probability is the mathematical language of randomness which enables you to reason about or make predictive statements about outcomes of physical systems or processes that have randomness or uncertainty*).

**Thống kê** làm việc theo cách khác: nó mô tả và tóm tắt đặc điểm của các kết quả quan sát, hoặc dữ liệu; chiết xuất mẫu/mô hình từ nó, và báo cáo về cơ chế nền tảng hoặc cấu trúc bản chất của hệ thống hay quy trình vật lý (*Statistics works the other way: it describes and summarizes characteristics of observed outcomes, or data, extracts patterns from it, and makes statements about underlying mechanisms or intrinsic structures of the physical systems or processes*).

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## CÁC KHÁI NIỆM TRONG THỐNG KÊ

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Các khái niệm trong thống kê [1]

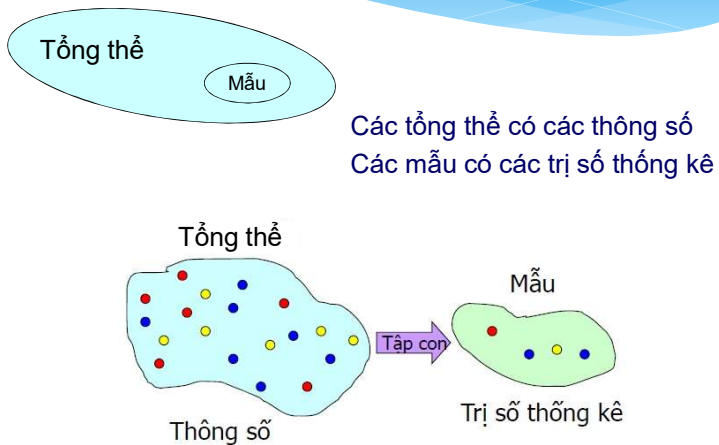
- **Tổng thể** (quần thể, population): một nhóm gồm các thành phần mà nhà thực hành thống kê quan tâm. VD: Tất cả các nhà thầu ở Hòa An
- **Mẫu** (sample): là một tập hợp các dữ liệu thu thập được lựa chọn từ một tổng thể thống kê bằng một quy tắc rõ ràng. VD: Một mẫu 75 nhà thầu ở Hòa An
- **Thông số** (parameter): một đại lượng mô tả tổng thể. VD: doanh thu bình quân của tất cả các nhà thầu là 100 tỷ đồng
- **Trị số (tham số) thống kê** (statistic): một đại lượng mô tả mẫu. VD: doanh thu bình quân của mẫu 75 nhà thầu là 120 tỷ đồng

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## PHƯƠNG PHÁP THỐNG KÊ & QUÁ TRÌNH THỐNG KÊ

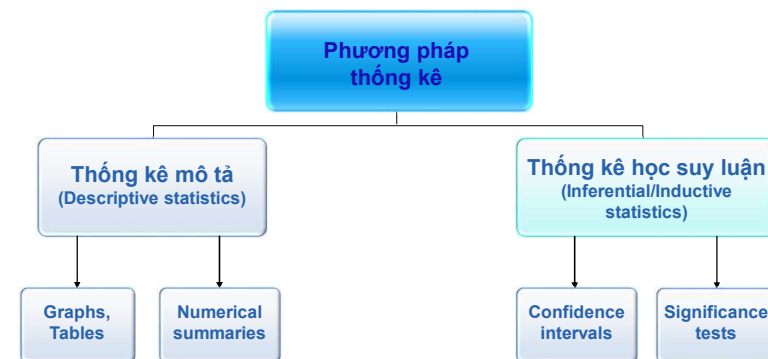
Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Các khái niệm trong thống kê [2]



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Phương pháp thống kê



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

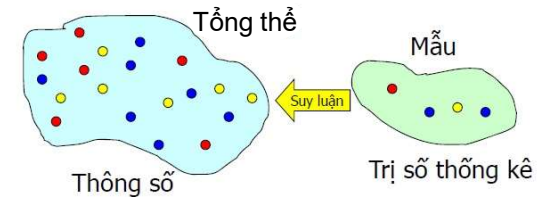
## Thống kê mô tả

- **Thống kê mô tả** được sử dụng để mô tả những đặc tính cơ bản của dữ liệu thu thập được từ nghiên cứu thực nghiệm qua các cách thức khác nhau.
- Thu thập số liệu
  - VD: khảo sát, quan sát, thí nghiệm
- Trình bày dữ liệu
  - VD: biểu đồ và đồ thị
- Xác định đặc điểm của dữ liệu
  - VD: trung bình của mẫu

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thống kê suy luận

Khi sử dụng mẫu, yếu tố ngẫu nhiên đã tác động vào mẫu, do vậy thống kê mô tả cũng không được chắc chắn. Để rút ra kết luận có ý nghĩa về toàn bộ tổng thể, thống kê suy luận là cần thiết.



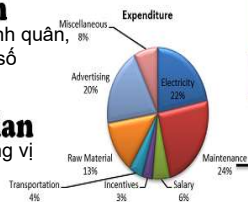
Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thống kê mô tả

### Descriptive Statistics

#### Mean

Giá trị bình quân,  
Kỳ vọng số



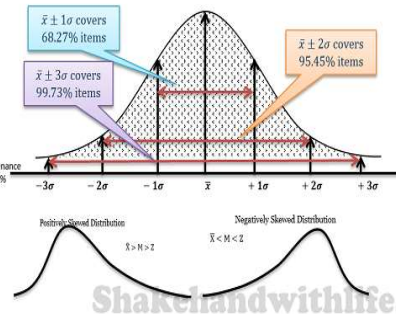
#### Median

Số trung vị

#### Mode

$$\text{Std. Dev. } \sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$$

Giá trị thường gặp của phần tử có số lần xuất hiện lớn nhất



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

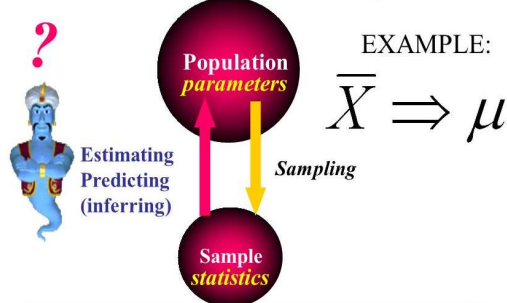
## Thống kê suy luận

- Thống kê suy luận là quá trình tạo ra một ước lượng, dự đoán, hay quyết định về tổng thể dựa trên mẫu.
- Những suy luận có thể mang hình thức trả lời có hoặc không các câu hỏi về dữ liệu (kiểm định giả thuyết), ước tính số lượng dữ liệu (ước lượng), mô tả các liên kết của dữ liệu (tương quan) và các mối quan hệ của các mẫu trong dữ liệu (ví dụ sử dụng phân tích hồi quy).
- Suy luận có thể mở rộng để dự báo, tiên đoán và ước tính giá trị không được chú ý đến hoặc sự liên kết với tổng thể được nghiên cứu. Nó có thể bao gồm các biến ngoại suy hoặc biến nội suy của chuỗi thời gian hoặc dữ liệu không gian, và khai thác dữ liệu.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thống kê suy luận

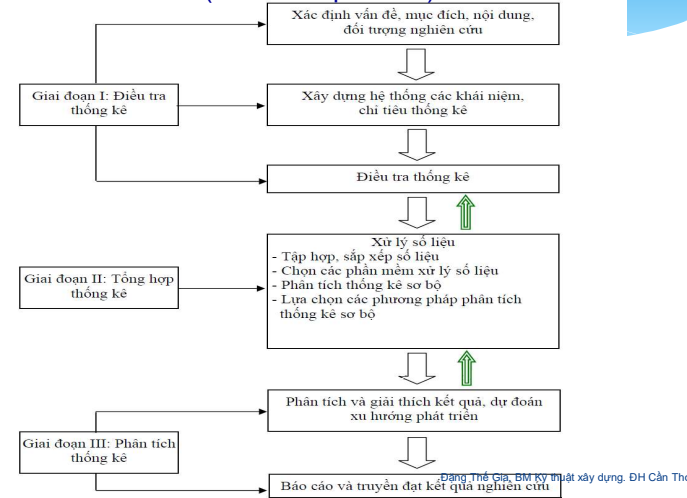
Inferential statistics describes a population of data using the information contained in a sample



A sample is a portion, or part, of the population of interest

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Quá trình thống kê (Statistical process)



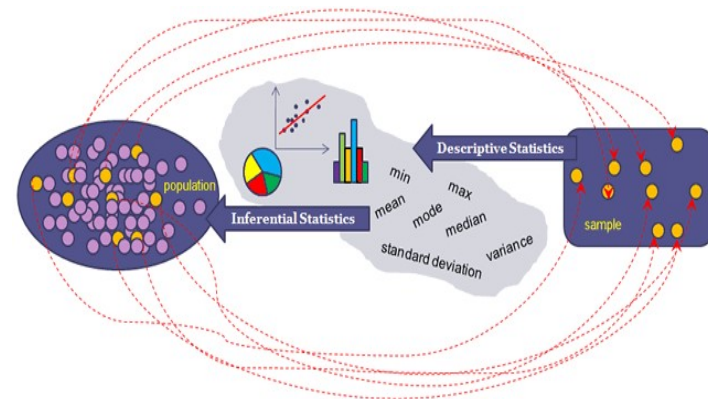
Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Quá trình thống kê (Statistical process)

- 1) Lập kế hoạch nghiên cứu, bao gồm việc tìm kiếm số liệu để trả lời cho các nghiên cứu. Xem xét việc lựa chọn đối tượng nghiên cứu, lưu ý tầm quan trọng của đạo đức nghiên cứu.
- 2) Thiết kế nghiên cứu: ngăn sự ảnh hưởng của các biến gây nhiễu, tránh lỗi xử lý và lỗi thực nghiệm.
- 3) Tiến hành thực nghiệm và phân tích dữ liệu theo các giao thức (experimental protocol) đã định.
- 4) Kiểm tra bộ dữ liệu thông qua các phân tích thứ cấp (secondary analysis) để đưa ra các giả thuyết mới cho những nghiên cứu trong tương lai.
- 5) Lập hồ sơ và trình bày kết quả nghiên cứu.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Quá trình thống kê (Statistical process)



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## DỮ LIỆU & THANG ĐO

### STATISTICAL DATA TYPES & LEVELS OF MEASUREMENT

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

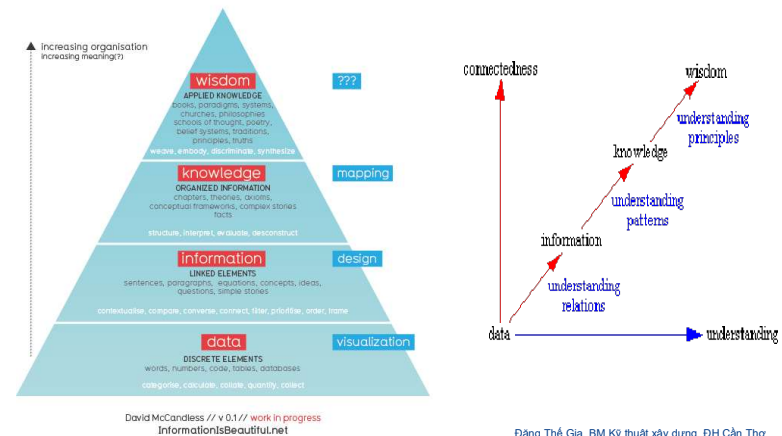
## Dữ liệu

- Dữ liệu (data), thông tin (information), kiến thức (knowledge), và trí tuệ (wisdom) là những khái niệm có liên quan chặt chẽ; mỗi cái đều có vai trò riêng trong mối quan hệ với những cái khác.
- **Dữ liệu** được thu thập và phân tích để tạo ra **thông tin** thích hợp cho việc ra quyết định, trong khi **kiến thức** có nguồn gốc từ một lượng phong phú của kinh nghiệm đối phó với thông tin về một chủ đề.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Dữ liệu

### Hierarchy Of Visual Understanding? Just playing. Something in this?



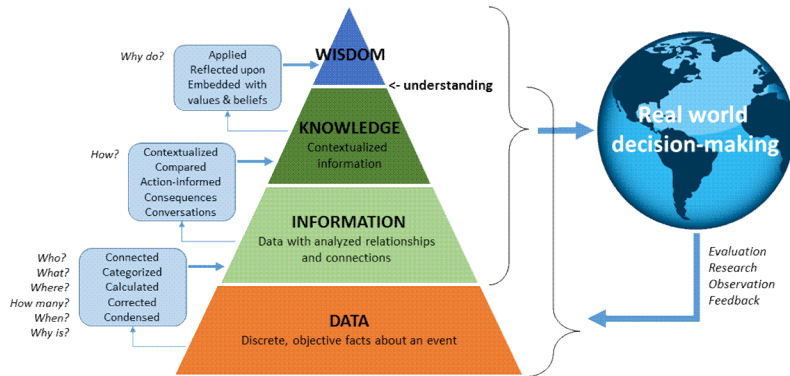
## Dữ liệu

- Dữ liệu (Data): là giá trị đại lượng cùng ngữ cảnh (biến cố hay sự kiện) đi kèm,  
Ngữ cảnh của dữ liệu: “Five W’s”: Who, What, When, Where & Why, Có thể có cả “How”
  - Thử nghiệm cấp độ bền chịu nén của bê tông: các giá trị cường độ chịu nén của tổ mẫu BT,...
  - Thí nghiệm cường độ chống cắt của đất, đá: các giá trị cường độ chống cắt cục bộ của tổ mẫu đất, đá,...
  - Lượng mưa, lưu lượng con sông trên lưu vực,,: các con số đo đạc quan trắc về lượng mưa và lưu lượng,...
- Dữ liệu không có ngữ cảnh đi kèm thì hầu như vô nghĩa

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ



## Dữ liệu



<https://www.climate-eval.org/blog/answer-42-data-information-and-knowledge>

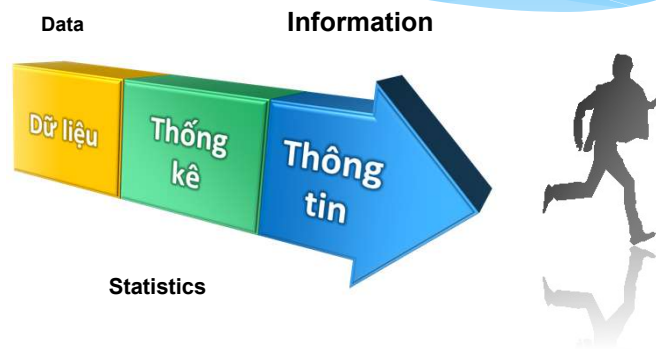
Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Các loại dữ liệu (Type of Data)

- Phân loại theo thuộc tính:
  - Dữ liệu định tính (categorical/qualitative data)
  - Dữ liệu định lượng (quantitative data)
- Phân loại theo cách thu thập:
  - Dữ liệu sơ cấp
  - Dữ liệu thứ cấp

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Data vs Information



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Các loại dữ liệu (Type of Data)

- Chúng ta cần xác định các phương pháp phân tích thích hợp dựa vào mục đích nghiên cứu và bản chất của dữ liệu thống kê.
- Tìm hiểu bản chất của dữ liệu thống kê qua khảo sát các cấp độ đo lường khác nhau vì mỗi cấp độ sẽ chỉ cho phép một số phương pháp nhất định.
- Khái niệm về số đo: là việc gán những giá trị hay ký hiệu cho những hiện tượng quan sát.
- Khái niệm về thang đo: là tạo ra một thang điểm để đánh giá đặc điểm của đối tượng nghiên cứu thể hiện qua sự đánh giá, nhận xét.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ



## Dữ liệu sơ cấp & thứ cấp

(Primary & Secondary Data)

- Dữ liệu sơ cấp là do trực tiếp thu thập nên độ chính xác cao hơn. Dữ liệu sơ cấp giúp giải quyết cấp bách và kịp thời những vấn đề đặt ra. Tuy nhiên, dữ liệu sơ cấp phải qua quá trình thực tế mới có được, vì vậy việc thu thập dữ liệu sơ cấp thường tốn nhiều thời gian và chi phí.
- Vì vậy, các nhà nghiên cứu sẽ phải cân nhắc khi nào sẽ phải thu thập dữ liệu sơ cấp và lựa chọn phương pháp thu thập hiệu quả để hạn chế nhược điểm này.
- Dữ liệu thứ cấp là dữ liệu đã có sẵn, không phải do mình thu thập, đã công bố nên dễ thu thập, ít tốn thời gian & chi phí trong quá trình thu thập, nhưng là loại tài liệu quan trọng trong việc nghiên cứu trong khoa học xã hội.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thu thập dữ liệu sơ cấp

(Data collection)

- Phương pháp thực nghiệm: thu thập dữ liệu trực tiếp qua các thực nghiệm có chủ định trên đối tượng nghiên cứu. Dữ liệu đúng & chính xác, mất thời gian & chi phí,...
- Phương pháp quan sát: được áp dụng khi không thể lấy thông tin trực tiếp từ đối tượng nghiên cứu. Người nghiên cứu sẽ phải dùng các giác quan hoặc máy móc để quan sát đối tượng nghiên cứu trong một khoảng thời gian nhất định. Dữ liệu đúng & chính xác, đòi hỏi nhận lực & thời gian,...
- Phương pháp khảo sát trực tuyến: khảo sát qua thư điện tử hay các website. Ưu điểm của phương pháp này là thu thập dữ liệu với số lượng lớn, tiết kiệm chi phí, nhưng số lượng phản hồi thấp & kết quả chính xác tương đối.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Đặc tính của Dữ liệu thứ cấp

- Dữ liệu thứ cấp chỉ cung cấp các thông tin mô tả tình hình, chỉ rõ qui mô của hiện tượng chứ chưa thể hiện được bản chất hoặc các mối liên hệ bên trong của hiện tượng nghiên cứu.
- Dữ liệu thứ cấp là những thông tin đã được công bố nên thiếu cập nhật, đôi khi thiếu chính xác và không đầy đủ.
- Tuy nhiên, dữ liệu thứ cấp cũng đóng một vai trò quan trọng do các lý do: Các dữ liệu thứ cấp có thể giúp người quyết định đưa ra giải pháp để giải quyết vấn đề trong những trường hợp thực hiện những nghiên cứu mà các dữ liệu thứ cấp là phù hợp mà không cần thiết phải có các dữ liệu sơ cấp.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Các thang đo dữ liệu thống kê

(Levels of Measurement)

- Theo Stanley Smith Stevens (1946):
  - Định tính (qualitative): **định danh** (nominal),
  - Định lượng (quantitative): **thứ tự** (ordinal), **khoảng** (interval), và **tỷ lệ** (ratio scale).
- Mosteller và Tukey (1977):
  1. Names
  2. Grades (e.g. freshmen, sophomores etc.)
  3. Counted fractions bound by 0 and 1
  4. Counts (non-negative integers)
  5. Amounts (non-negative real numbers)
  6. Balances (any real number)

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Các thang đo dữ liệu thống kê

(Type of Data)

- Chrisman (1998):
  1. Nominal
  2. Graded membership
  3. Ordinal
  4. Interval
  5. Log-Interval
  6. Extensive Ratio
  7. Cyclical Ratio
  8. Derived Ratio
  9. Counts
  10. Absolute
- Nelder (1990): continuous counts, continuous ratios, count ratios & phân loại (categorical modes)

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thang đo thứ bậc

(Ordinal Scala)

- Thang đo **thứ bậc** (ordinal scala) là loại thang đo dùng cho các dữ liệu thuộc tính. Tuy nhiên, trường hợp này biểu hiện của dữ liệu **có sự so sánh**.
- Có sự **khác biệt không chính xác** giữa các giá trị liên tiếp, nhưng có một **trật tự có ý nghĩa** đối với những giá trị đó, và cho phép bất kỳ sự chuyển đổi duy trì đơn hàng.
- Ví dụ: trình độ thành thạo của công nhân được phân từ thợ bậc 1 đến 7. Phân loại giảng viên trong các trường đại học: Giáo sư, Phó giáo sư, Giảng viên chính, Giảng viên.
- Thang đo này cũng không thực hiện được các phép tính đại số.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thang đo định danh

(Nominal Scala)

- Thang đo **định danh** (nominal scala) là loại thang đo sử dụng cho **dữ liệu định tính**.
- Các biểu hiện của dữ liệu thuộc tính không có sự hơn kém, khác biệt về thứ bậc, không có thứ tự giữa các giá trị.
- Các giá trị (con số, mã hóa) không có mối quan hệ hơn kém, không thực hiện các phép tính đại số.
- Các giá trị chỉ mang tính chất mã hóa để phân loại thuộc tính của dữ liệu và không mang bất kỳ ý nghĩa nào khác. Ví dụ: giới tính ta có thể đánh số 1 là nam, số 2 nữ.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thang đo khoảng

(Interval Scala)

- Thang đo **khoảng** (interval scala) là loại thang đo cho các dữ liệu số lượng. Là loại thang đo cũng có thể dùng để xếp hạng các đối tượng nghiên cứu nhưng khoảng cách bằng nhau; là thang đo thứ bậc **có các khoảng cách đều** nhau.
- Có khoảng cách có ý nghĩa giữa các phép đo được xác định, nhưng giá trị bằng 0 là tùy ý (như trường hợp với kinh độ và đo nhiệt độ theo Celsius hoặc Fahrenheit), và cho phép chuyển đổi tuyến tính.
- Có thể thực hiện các phép tính đại số (trừ phép chia không có ý nghĩa). Ví dụ như điểm môn học của sinh viên. Sinh viên A có điểm thi là 8, sinh viên B có điểm là 4 thì không thể nói rằng sinh viên A giỏi gấp hai lần sinh viên B.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Thang đo tỉ lệ

(Ratio Scala)

- Thang đo **tỷ lệ** (ratio scala) có cả giá trị bằng không có ý nghĩa, khoảng cách giữa các phép đo khác nhau được xác định, và cho phép bất kỳ chuyển đổi định dạng lại.
- Là loại thang đo cũng có thể dùng dữ liệu có đặc tính số lượng.
- Là thang đo cao nhất trong các loại thang đo định lượng.
- Ngoài đặc tính của thang đo khoảng, phép chia có thể thực hiện được. Ví dụ: thu nhập trung bình một tháng của ông A là 2 triệu đồng và thu nhập của bà B là 4 triệu đồng thì ta có thể nói rằng thu nhập trung bình trong một tháng của bà B gấp 2 lần thu nhập của ông A.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## SỬ DỤNG SAI & DIỄN GIẢI SAI

*MISUSE & MISINTERPRETATION*

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## So sánh các thang đo

(Comparison)

Incremental Progress	Measure Property	Mathematical Operators	Advanced Operations	Central Tendency
Nominal	Classification, Membership	=, !=	Grouping	Mode
Ordinal	Comparison, Level	>, <	Sorting	Median
Interval	Difference, Affinity (mối liên quan)	+, -	Yardstick (thước đo)	Mean, Deviation
Ratio	Magnitude, Amount	*, /	Ratio	Geometric Mean, Coeff. of Variation

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Sử dụng sai

(Misuse)

- Sử dụng sai mục đích các số liệu thống kê có thể có những kết quả không lường được;
- Những sai số nghiêm trọng trong mô tả và giải thích sai ý nghĩa, ngay cả các chuyên gia có kinh nghiệm;
- Dẫn đến đưa ra quyết định sai. Ví dụ chính sách xã hội, nghề thuốc, và độ tin cậy của cấu trúc dựa trên các số liệu thống kê.
- Ngay cả khi các kỹ thuật thống kê được áp dụng một cách chính xác, kết quả có thể khó để giải thích cho những người thiếu chuyên môn.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Sử dụng sai & Lạm dụng

(Misuse)

- Ý nghĩa thống kê của một phương pháp có thể được gây ra bởi sự thay đổi ngẫu nhiên trong mẫu; có thể hoặc không thể đồng ý với đánh giá trực quan của mức ý nghĩa.
- Có ý kiến cho rằng kiến thức thống kê được cho là bị lạm dụng một cách quá bình thường bằng cách tìm ra hướng để giải thích các dữ liệu có ích cho người trình bày.
- Sự nghi ngờ và hiểu sai về thống kê có liên qua đến câu nói, “có ba loại lừa dối: dối trá, rất dối trá và thống kê”.
- Lạm dụng các số liệu thống kê có thể có được kể cả vô ý và có chủ ý.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Cách tránh sử dụng sai & Lạm dụng

(Prevention)

- Để làm cho dữ liệu thu thập được từ các số liệu thống kê đáng tin cậy và chính xác, mẫu được chọn phải có tính tổng thể;
- Theo Huff, “độ tin cậy của một mẫu có thể bị phá hủy giá trị sai lệch, cho phép một số mức độ hoài nghi”;
- Để hỗ trợ cho sự hiểu biết của các số liệu thống kê, Huff đã đề xuất một loạt các câu hỏi được hỏi trong mỗi trường hợp:
  1. Ai nói vậy?
  2. Làm thế nào để anh/chị biết
  3. Những gì còn thiếu
  4. Có ai thay đổi nội dung
  5. Nó có ý nghĩa không

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Cách tránh sử dụng sai & Lạm dụng

(Prevention)

- Cách để tránh số liệu thống kê bao gồm sử dụng sơ đồ thích hợp và ngăn ngừa sai số.
- Sử dụng sai số có thể xảy ra khi kết luận là sai số quá lớn và yêu cầu có tính đại diện hơn so với giá trị thật, thường là cố ý hay vô ý không nhận thấy ra sai số mẫu.
- Đồ thị dạng cột được cho là biểu đồ đơn giản nhất để sử dụng và hiểu, các biểu đồ này có thể vẽ bằng tay hoặc bằng các chương trình máy tính đơn giản.
- Nhưng hầu hết mọi người đều không nhìn ra giá trị sai lệch hay sai số, vì vậy những lỗi sai này không được sửa chữa. Nên mọi người thường tin vào kết quả ngay cả khi nó không phải là kết quả tốt.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Diễn giải sai tương quan

(Misinterpretation correlation)

- Khái niệm ‘tương quan’ cần đặc biệt lưu ý vì tiềm ẩn những nhầm lẫn. Phân tích thống kê của một bộ số liệu thường cho thấy chỉ hai biến (tính chất) của quần thể biến thiên với nhau như thể chúng có quan hệ với nhau.
- Ví dụ 1: Thực nghiệm con cào cào. Ví dụ 2: một nghiên cứu về thu nhập hàng năm (biến 1) và tuổi thọ (biến 2) có khả năng dẫn đến mối tương quan rằng người nghèo thường có tuổi thọ thấp hơn người giàu.
- Hai biến được xem là tương quan (theo thống kê) đôi khi không phải làm nguyên nhân hay hệ quả của nhau. Mối tương quan này có thể được gây ra bởi một biến (hiện tượng) thứ ba mà trước đây không được xem xét (‘biến ẩn’ hoặc ‘biến gây nhiễu’). → Không thể ngay lập tức suy ra sự tồn tại về một mối quan hệ nhân quả giữa hai biến

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ