**Tổng hợp lý thuyết MAS291**

**Chương 1:**

1. Lý thuyết về thống kê
2. Descriptive Statistics (Thống kê mô tả) vs Inferential Statistics (thống kê suy luận)

* **Descriptive Statistics:** Involves organizing, summarizing, and displaying data. Those methods involving the collection, presentation, and characterization of a set of data to properly describe the various features of that set of data. **Key work:** Descriptiion, presentation, display – mô tả: ở đây để hoạt động mô tả lại các đặc điểm tính chất của dữ liệu, không hề rút ra kết luận, hướng phát triển, hay suy luận ra tính chất nào khác.
* **Inferential Statistics:** Involves using sample data to draw conclusions about a population. **Key word:** conclusions, inference – những cụm từ mang nghĩa suy luận, kết luận

1. Một số định nghĩa thường gặp trong thống kê

* **Population**: the complete collection of all individuals to be studied.
* Là tập **lớn nhất**, phạm vi lớn nhất bao chứa các tập con bên trong. Ví dụ: Toàn bộ người trên hành tinh/ toàn bộ nữ giới/ tất cả nam giới
* **Sample**: *Sub-collection* of members selected from a population.
* **Tập con** được chọn từ **population**. Ví dụ: Vietnamese là sample của human, nữ giới trong fpt là sample của toàn bộ nữ giới
* **Data**: consist of information coming from observations, counts, measurements, or responses.
* Lưu **thông tin** của population/sample. Ví dụ chiều cao, giới tính, … của bạn X
* **Parameter**: a numerical measurement describing some characteristic of a population.
* **Parameter** – tham **số** là những số liệu dạng **số** của **population**
* Statistic: a numerical measurement describing some characteristic of a sample**.**
* **Statistic** – thống kê: dữ liệu dạng số của những sample.

**Cách phân biệt** **Statistic** và **Parameter:** Statistic là thống kê, chỉ áp dụng được với **sample**. population là một phạm vi rất rất rộng ta không thể quan sát/thống kê hết được -> vì vậy thống kê áp dụng lên sample – là một phạm vi nhỏ hơn, dễ đo đạc, tính toán hơn.

1. Lý thuyết về dạng dữ liệu
2. Phân biệt Qualitative và Quantitative

* Qualitative: Liên quan tới **Quality** (Chất lượng) là những thứ không thể đo đạc được. Ví dụ màu da, giới tính, …
* Quantitative: Liên quan tới **Quantity** (Số lượng) là những thứ có thể cân đo đong điếm được
  + Chia ra 2 loại là rời rạc (đếm được) và liên tục (không đếm được)

1. Một số phương pháp thu thập dữ liệu

* **Retrospective study:** Phương pháp/ nghiên cứu về dữ liệu trong quá khứ. Key word: Historical, History, Day/Month/Year in the past, …
* **Observational study:** Liên quan tới **Observation** (sự quan sát) để chỉ phương pháp thu thập dữ liệu bằng cách quan sát, đo đạc hay tính toán (không tác động, làm biến dạng dữ liệu)
* **Experimental study:** Liên quan tới **Experiment** (thí nghiệm) để chỉ những phương pháp thu thập dữ liệu thông qua các thí nghiệm, các phương pháp tác động trực tiếp lên sample

**Một ví dụ phân biệt giữa Observational study và Experimental study:** Đo đạc chiều cao, chiều rộng, khối lượng một chiếc xe là **Observational study.** Cho xe va chạm để kiểm tra độ bền, chà xát để kiểm tra chất lượng sơn, gập mở nhiều lần cửa xe là **Experimental study**

**Chương 2:**

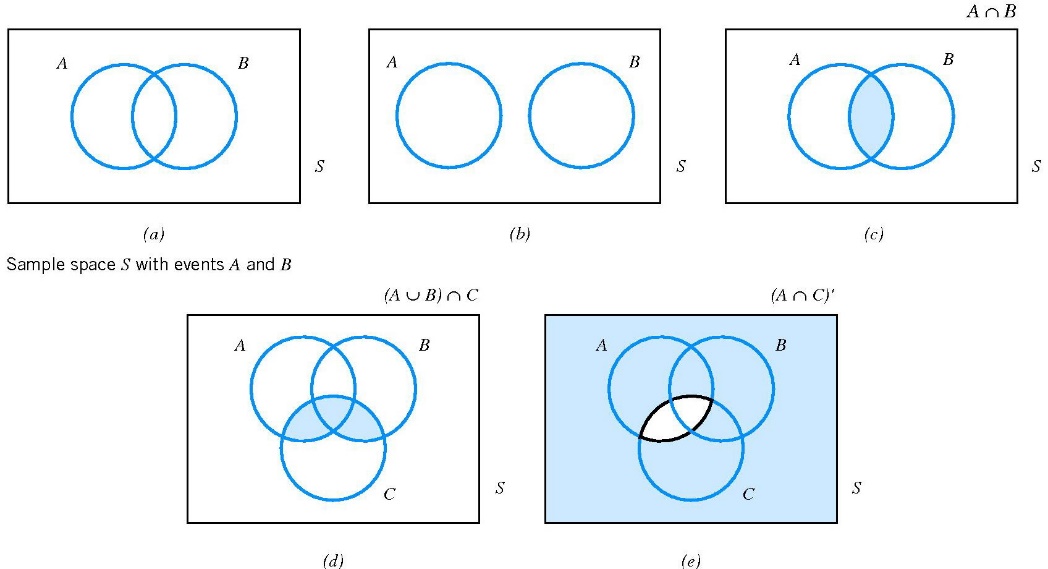
1. Sample space vs Event

* Sample space – không gian mẫu là tất cả các trường hợp có thể xảy ra
* Event – biến cố: là một tập con của sample space

1. Một số thuật ngữ thông dụng về tập hợp

* Intersection:A ∩ B (giao)
* Union:A ∪ B (hợp)
* Complement of event A : A’ (phần bù)
* Mutually exclusive if A ∩ B = ⊘ (xung khắc)

**Mẹo ghi nhớ giao và hợp:** Thần chú lên hợp xuống giao, keywork: **liên hợp**. Lên để chỉ chữ U quay lên hay cục nam châm được dựng lên.



1. Một số tính chất quan trọng của tập hợp (Thay ∪ bằng +, ∩ bằng nhân và ‘ bằng -1)

* A ∪ (B ∪ C) = (A ∪ B) ∪ C
* A ∩ (B ∩ C) = (A ∩ B) ∩ C



* A ∩ (B ∪ C) = (A ∩ B) ∪ (A ∩ C)



* (A ∪ B)’ = A’∩ B’
* (A ∩ B)’ = A’∪ B’
* A = (A ∩ B) ∪ (A ∩ B’)

1. Xác xuất có điền kiện

* Xác xuất của B theo A được biểu diễn bằng:

P(B | A) = P(A ∩ B)/P(A)

* Cách nhớ: A ở dưới B trong ngoặc thì P(A) sẽ ở dưới mẫu. Liên tưởng:

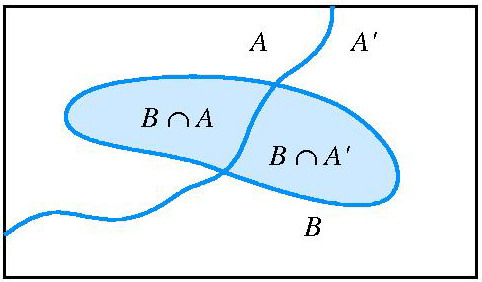
B / A = B \* A / A^(2)

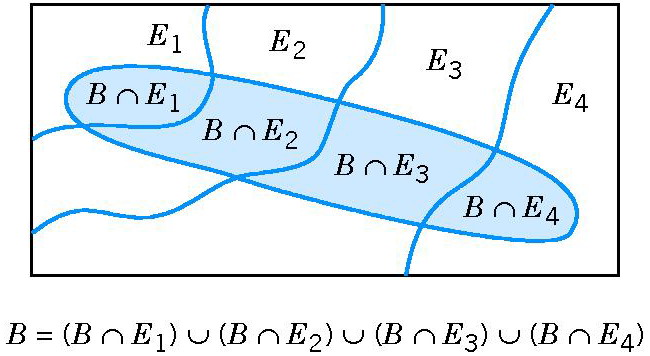
* Một số công thức thêm

P(A ∩ B) = P(B | A)P(A) = P(A | B)P(B)



P(B) = P(B ∩ A) + P(B ∩ A′) = P(B | A)P(A) + P(B | A′)P(A′)





**Link tham khảo chương 2:** <https://vnoi.info/wiki/translate/he/Number-Theory-6.md#c%C3%A1c-ki%E1%BA%BFn-th%E1%BB%A9c-c%C6%A1-b%E1%BA%A3n-v%E1%BB%81-x%C3%A1c-su%E1%BA%A5t-probabilities>

**Chương 3:**

1. Discrete random variables (Biến ngẫu nhiên rời rạc)

* Ôn lại về khái niệm rời rạc: Có thể biểu diễn được dưới dạng số nguyên

1. Probability mass function (Hàm khối xác suất) (**Khối** đá -> đếm được -> discrete)
   1. .
2. Cumulative distribution function (Hàm phân phối tích lũy)
3. Mean and Variance (Ký vọng và Phương sai)
   1. Mean (Trung bình hay Kỳ vọng)

**Tip:** Liên hệ tới trường hợp đặc biệt nhất là tính trung bình các số liên tiếp từ 1 đến 10. Công thức là (1 + 2 + … + 10)/10. Viết lại theo công thức bên trên:

1 x 1/10 + 2 x 1/10 + … + 10 x 1/10

Với xi = 1, 2, … ,10 và p(xi) = f(xi) = 1/10 do là **phân phối đồng nhất**, tất cả xác xuất của các điểm trong phân bố đều bằng nhau. Ở các bài toán thông thường f(xi) thường **khác nhau.**

* 1. Variance (Phương sai)

: **standard deviation** (**độ lệch chuẩn**) bằng căn bậc 2 phương sai, để đo độ phân tán của tập phân phối

**Lưu ý:**

* *E(aX + b) = aE(X) + b*
* *V(aX + b) = a^2.V(X)*

1. Binomial distribution

* Xác suất p, thử n lần, x lần thành công
* *µ = E(X) = np*
* Liên hệ đến công thức ở đây ta mặc định = 1 và chính là p. Khi đó tổng chính bằng công thức np hay + + … + ) nhân p với = 1
* *σ2 = V(X) = np(1-p)*

1. Geometric distribution

* Xác suất p, thử x lần, lần cuối thành công
* Chọn x – 1 phần tử mà biến cố không xảy ra nên sẽ có xác suất là (1-p)^(x-1), phần tử cuối cùng có biến cố xảy ra nên mang xác suất là p

1. Negative Binomial distribution

* Xác suất p, thử x lần cho tới khi r lần thành công
* Giải thích công thức: Trong x-1 lần cho tới khi được r lần thành công này ta chọn ra r-1 lần (tổ hợp chập r – 1 của x – 1), tại sao lại là x – 1 và r – 1? Vì lần thử cuối cùng, lần thứ x cũng là lần thành công cuối cùng, lần thứ r vì vậy ta luôn chừa lại 1 lần chắc chắn thành công mà không cần chọn

1. Hyper-geometric distribution

* Chọn n phần tử từ tập gồm N phần tử n < N. Trong N phần tử này có K phần tử thành công (có biến cố xảy ra) và N – K phần tử thất bại. Xác suất để chọn được x phần tử thành công trong n phần tử được nhặt ra là

1. Poisson distribution

* Cho X là số biến cố xảy ra trong một khoảng thời gian cụ thể (ví dụ 1 ngày, 1 giờ, 1 phút, …)
* µ = *E*(*X*) = λ
* σ2 = *V*(*X*) = λ