Tóm tắt chương I: Giới thiệu tổng quan

- **population**

- **Sample**

- **descriptive statistic: (**Tables, graphs, or numerical summary tool, identification of patterns in the data, the population or sample of interest**)**

- **Statistical inference: (**draw conclusion about population parameters**)**

- **Collecting data** (có 3 cách: Retrospective; observation; experiment)

- **Categorise data** (có hai loại : qualitative and quantitative).

Tóm tắt chương II: **Các công thức tính xác suất**

* **Sample space = S** (Không gian mẫu là tập hợp các kết cục (outcome) có thể có khi thực hiện một phép thử).
* **Event:** là một tập con của sample space.
*  (nếu các outcome là đồng khả năng (**equally likely**))
* .
* **Addition rule**: .
* Nếu  thì A và B được gọi là xung khắc (**mutually exclusive**).
* Conditional probability: 
* **Multiple rule**: 
* Nếu  thì A và B được gọi là độc lập (**independence**).
* **Total probability**: Nếu  là hệ biến cố đầy đủ và xung khắc từng đôi thì .
* **Bayes theorem**: 

Tóm tắt chương III: **Biến ngẫu nhiên rời rạc (Discrete random variable)**

1. **Kiến thức chung:**
2. Hàm xác suất (**probabiliy mass function** ): .

Nếu biết hàm xác suất f(x) thì có thể tính 

1. Hàm phân phối tích lũy: (**Cumulative dist function**):



Nếu biết hàm phân phối tích lũy thì có thể tính được  hoặc làm theo cách khác là tìm hàm xác suất f(x) rồi dựa vào đó để tính.

1. Kì vọng (**mean or expected value**) và phương sai (**Variance**)



Chú ý: Nếu cần tính 

1. **Các biến ngẫu nhiên rời rạc thường gặp.**
2. Biến ngẫu nhiên đều (**Uniform random variable**).

Nếu X nhận n giá trị  thì .

1. Biến ngẫu nhiên nhị thức (**Binomial** )

ĐN: Một thí nghiệm gồm n phép thử (**trial**)Bernoulli thỏa mãn *3* điều kiện:

+ Các PT độc lập (**independent**)

+ Mỗi PT chỉ có 2 kết quả ký hiệu là “thành công” (**success**) và “thất bại” (**failure**)

+ Xác suất của kết quả “thành công” trong một PT luôn không đổi và bằng *p*

**Bnn X nhận giá trị bằng số phép thử có kết quả là thành công trong dãy n PT trên được gọi là bnn nhị thức với tham số n, p.**

* **Hàm xác suất: P(X = x) = f(x) = Cxn px (1-p)n-x, x = 0, 1,…,n**
* **Trung bình: µ = np**
* **Phương sai: σ2 = np(1-p)**

1. Phân phối **geometric** và **Negative Binomial**

ĐN: Gọi bnn X : **số phép thử cần thiết cho đến khi có 1 phép thử cho kết quả “thành công”** trong dãy n phép thử Bernoulli , X là bnn **Geometric**

**Hàm xác suất: P(X=x) = f(x) = (1-p)x-1p**

**Trung bình: µ = 1/p**

**Phương sai: σ2 = (1-p)/p2**

ĐN: Gọi bnn X : **số phép thử cần thiết cho đến khi có r phép thử cho kết quả “thành công”** trong dãy n phép thử Bernoulli , X là bnn **Negative Geometric**

**Hàm xác suất: P(X=x) = f(x) = Cx-1r-1(1-p)x-r pr**

**Trung bình: µ = r/p**

**Phương sai: σ2 = r(1-p)/p2**

1. Phân phối **Poisson**

ĐN: Bnn X chỉ **số ‘event’ xảy ra trong một ‘interval’**  được gọi là bnn Poisson.

**Hàm xác suất: P(X=x) = f(x) = e-λλx/x! (**λ là số trung bình các event trong interval đó**)**

**Trung bình: µ = λ**

**Phương sai: σ2 = λ**

Tóm tắt chương IV: **Biến ngẫu nhiên liên tục (Continuous random variable)**

1. **Kiến thức chung:**
2. Hàm mật độ (**probabiliy densty function** ): .

Nếu biết hàm mật độ f(x) thì có thể tính 

1. Hàm phân phối tích lũy: (**Cumulative dist function**):



Nếu biết hàm phân phối tích lũy thì có thể tính được 

Nếu biết hàm phân phối tích lũy thì có thể tìm hàm mật độ f(x) = F’(x) .

1. Kì vọng (**mean or expected value**) và phương sai (**Variance**)



Chú ý: Nếu cần tính 

1. **Các biến ngẫu nhiên liên tục thường gặp.**
2. Biến ngẫu nhiên liên tục đều trên đoạn [a, b](**Uniform continuous random variable**).

Là biến ngẫu nhiên ***Z*** có **hàm mật độ f(x) =1/(b-a) với a ≤ x ≤ b.**

* **Trung bình µ = E(X) = (a+b)/2,**
* **phương sai: σ2 = (b - a)2 / 12.**

1. Biến ngẫu nhiên tiêu chuẩn (**Standard normal** )

* **Hàm mật độ:** ;
* **Hàm phân phối tích lũy**  (giá trị của hàm này được tra tại table III-A6- textbook)
* **Trung bình: µ = 0**
* **Phương sai: σ2 = 1**
* **Tính xác suất**  (tra bảng table III-A6- textbook).
* **Tìm a để**  thì a sẽ là số thỏa mãn 

1. Biến ngẫu nhiên chuẩn (**Normal** ) với trung bình µ và phương sai σ2

* **Tính xác suất** 
* **Tìm a để**  thì a sẽ là số thỏa mãn 

1. Phân phối mũ **( Exponential)**

Là biến ngẫu nhiên **chỉ khoảng thời gian giữa hai ‘event’** trong quy trình Poisson.

Nếu λ là số trung bình các ‘event’ trên một interval có độ dài là 1 đơn vị thì

**Hàm mật độ xác suất: f(x) = λe-λx**

**Hàm phân phối tích lũy: F(x) = 1 – e-λx**

**Trung bình: µ = 1/λ**

**Phương sai: σ2 = 1/λ2**

1. Xấp xỉ chuẩn của phân phối nhị thức và phân phối Poisson.
2. **Phân phối nhị thức**



1. **Phân phối Poisson**



Tóm tắt chương VI: **Thống kê mô tả (Statistical Descriptive)**

**Kiến thức chung:**



* **Sample range(**Khoảng biến thiên mẫu**):**
* **Median (**trung vị mẫu**):** Là số chia mẫu ra làm hai phần bằng nhau.

**+ Nếu kích thước mẫu là lẻ thì median chính là số đứng ở vị trí chính giữa của mẫu**

**+ Nếu kích thước mẫu là chẵn thì median là trung bình của hai số đứng giữa mẫu.**

* **Mode:** Là số xuất hiện nhiều nhất trong mẫu.
* Q**uartiles (**Tứ phân vị**):** làba số mà chúng chia mẫu thành 4 phần bằng nhau, kí hiệu là q1, q2, q3
* Khoảng tứ phân vị**: IQR = q3 – q1**
* Phân phối tần số **(Frequency distribution)**
* Phân phối tần suất **(Relative frequency distribution)**
* Phân phối tích lũy **(Cumulative distribution)**
* **Stem and leaf.**
* **Box plot**

Tóm tắt chương VII: **Point Estimate and The Central Limit Theorem**

**(Ước lượng điểm và định lí giới hạn trung tâm)**

1. **Point estimate**

* For *μ* , the point estimate is the  - sample mean.
* For *σ*2 , the point estimate is  - the sample variance.
* For *p*, the point estimate is  - the sample proportion.
* For *μ1 – μ2* , the estimate is .
* For *p1 – p2 ,* the estimate is 

1. **The Central Limit Theorem**

*  (khi n đủ lớn )

(với µ là trung bình tổng thể ; σ2 là phương sai tổng thể, n là kích thước mẫu)

Ứng dụng để tính 

* ) (Khi n1, n2 đủ lớn)

(với µ1, µ2 là trung bình tổng thể 1, 2 ; ; σ2 1 ; σ2 2 là phương sai tổng thể 1, 2, n1; n2 là kích thước mẫu 1 và 2).



Tóm tắt chương VIII: **Confidence Interval (viết tắt là CI ) for population parameter (Khoảng tin cậy cho tham số của tổng thể)**

* Bài toán tổng quát của chương này:

Dựa vào mẫu, tìm ***ước lượng khoảng (confidence interval = CI)*** của một *tham số* (µ hoặc σ hoặc p) của biến ngẫu nhiên.

* Định nghĩa chung: Ước lượng khoảng của tham số θ với độ tin cậy (**confidence level**) 1- α là khoảng [L, U] sao cho

P(L ≤ θ ≤ U) = 1 – α,

trong đó L, U là các hàm của mẫu ngẫu nhiên (X1,…,Xn).

* **CI on µ: Với độ tin cậy (confidence level) 1- α thì khoảng tin cậy cho µ là:**

**TH1** (tổng thể có phân phối chuẩn, σ đã biết).

* Hai phía: 

với  là số mà .

* Một phía 
* Nếu dùng  để xấp xỉ cho µ thì muốn sai số không vượt quá E thì với độ tin cậy 1-α kích thước mẫu cần tìm là:



**TH2: (**kích thước mẫu lớn) (n > = 30)

* Hai phía: 

với  là số mà .

* Một phía 

**TH3: (**tổng thể có phân phối chuẩn**,** σ chưa biết)

* Hai phía: 

với  là số được tra bởi table V-A9-textbook bằng cách lấy giao của cột α/2 và dòng n-1.

* Một phía 
* **CI on σ2: Với độ tin cậy (confidence level) 1- α thì khoảng tin cậy cho σ2 là:**
* Hai phía 

Với  là giá trị được tra từ Table IV-A8-textbook.

* Một phía 
* **CI on p: Với độ tin cậy (confidence level) 1- α thì khoảng tin cậy cho plà:**
* Hai phía 
* Một phía : 
* **Với độ tin cậy 1 – α, muốn sai số khi xấp xỉ *p* bởi  không vượt quá E thì kích thước mẫu cần thiết là:**



* **Với độ tin cậy ít nhất bằng 1 – α, muốn sai số khi xấpxỉ *p* bởi  không vượt quá E thì kích thước mẫu cần thiết là:**



Tóm tắt chương XI: **Hồi quy tuyến tính ( Linear Regression) và hệ số tương quan (Correlation).**

**Với n cặp quan sát**  lấy từ tổng thể của biến ngẫu nhiên (X, Y) thì

* **The estimated (fitted) linear regression line :**

** với**



* **Sample Correlation** - Hệ số tương quan mẫu



**Ý nghĩa của R**: Đo mức độ tương quan tuyến tính mẫu của X và Y. Giá trị của R càng xấp xỉ lớn thì mức độ tương quan tuyến tính càng mạnh.

* **Tổng bình phương các sai số (SSE).**



* **Test on β1:**

1. H0: β1 = β1,0  H1: β1 ≠ β1,0
2. Test statistic: 
3. Critical values 
4. Reject H0 nếu 

* **Test on β0**

**1)**  H0: β0 = β0,0  H1: β0 ≠ β0,0

1. Test statistic 
2. Critical values 
3. Reject H0 nếu 

* **Test on ρ**: = population correlation

1. H0: ρ = 0 H1: ρ ≠ 0
2. Test statistic 
3. Critical values 
4. Reject H0 nếu 