常用統計技術名詞

1. 全體 Universe, population

全體（或稱母體）為研究對象的全部。全體的大小視研究者的需要而定。由現實存在事物所構成的全體，稱為實在全體；由假想事物所構成的全體，稱為假想全體。例如連續擲一枚銅元，其各次出象所構成的全體即為一個假想全體。全體中所含的個骿數若為有限，則此全體稱為有限全體，否則即為無限全體。

2. 樣本 Sample

由全體中抽出的部分個體，相對於全體而言，稱之為全體的樣本。被抽的全體相對於樣本而言，稱之為樣本的母全體。

3. 抽樣 Sampling

由全體中抽取樣本的過程，稱之為抽樣。抽樣方式以其是否為隨機性的可分為兩種，即隨機抽樣與立意選樣。由截略母體抽取樣本的過程，稱為截略抽樣。所謂截略母體，即是一個全體被切去一部分，通常是切除尾部， 其剩餘部分所構成的母體。

4. 誤差 Error

誤差為實測結果與直實情況之差。由於真實情況恒為未知，故真正的誤差亦恒為未知。真實情況可根據有關資料加以估計，其結果稱為估計真況。實測結果與估計真況之差，稱為估計誤差。誤差的特性有二：就橫的方向而言，不同原因所產生的誤差，累積的機會多，抵銷的機會少；就縱的方向而言，前一階段所產生的誤差將傳播至後一階段，中途很荏消滅。誤差按其是否為隨機性的可分為隨機誤差與非隨機誤差兩種，按其是否因抽樣而產生可分為抽樣誤差與非抽樣誤差兩

種，按其是否偏向分可分為偏誤與非偏誤兩種，按判斷的結果分可分為第一型誤差與第二型誤差兩種。

5. 抽樣誤差 Sampling error

由抽樣而產生的誤差，稱為抽樣誤差。抽樣誤差主要由抽樣的隨機性所產生，此種誤差是一種隨機性的誤差。由於抽取方式的不當亦可能產生抽樣誤差，此種誤差不為隨機性的。抽樣誤差的分配狀況可以一個機率分配表示之，稱為抽樣分配。抽樣誤差可以抽樣設計的方法加以控制。一般說來，抽樣誤差不能完全避免，除非能觀察母體的全部。

6. 非抽樣誤差 Non-sampling error

抽樣以外其他一切過程所產生的誤差，稱為非抽樣誤差。非抽樣誤差沒有任何規律。控制非抽樣誤差的方法，為小心做好每一項工作，不使誤差發生。由於誤差發生後中途很難消滅，故晚近有所謂「無缺點計畫」，即為針對此特質所設計的方法。

7. 偏誤 Bias

偏誤為偏向一方的誤差，其發生的原因有二：其一為研究人員的偏見，另一為儀器設備的偏差。控制偏誤的方法，就人為的偏見而言，研究人員應有接受別人批評的雅量，及時糾正；此外能以工具代行的操作儘量以工具行之，因人不是一種額觀的動物也。就儀器的偏差而言，研究者應當時時檢修儀器，使其經常保持正常狀態。

8. 隨機抽樣 Random sampling

母全體中每個個體皆有同等被抽可能的抽樣方式，稱為隨機抽樣。所抽出的樣本稱為隨機樣本。由於每個個體皆有同等被抽可能，故隨機樣本能用以代表全體，亦即根據隨機樣本可以推論其母全體的性狀。隨機抽樣的內容包括兩方面：其一為隨機抽出法，其目的在抽得「隨機性」的樣本。另一為抽樣設計，其目的在控制抽樣誤差，抽得代表性大的樣本，或同時取得有關影響因素的相關資料，作事後分析。

9. 抽樣設計 Sampling design

對樣本的出處作適當的安排，稱為抽樣設計，其目的在增加樣本的代表性。欲進行抽樣設計，必須對母體加以分類，如此才能引用有關訊息，而增強抽樣的效果。由於受人力物力以及客觀環境皂限制，母體的分類不能完全按照研究者的意志進行，於是乃有多種不同的抽樣設計以配合不同的分類標準，而使抽樣的效果儘可能的加大。常用的抽樣設計有三種，即分層隨機抽樣法、集團隨機抽樣法及兩段隨機抽樣法。不含任何抽樣設計在內的抽樣方法，稱為單純隨機抽樣法。抽樣設計如用之不當，其效果反不如單純隨機抽樣法為高。抽樣設計的原則是何處差異大，何處比例上多抽；何處差異小，何處比例上少抽。

10. 普查 Census

對研究對象的全體作全面的調查，稱之為普查。普查的優點為精確，缺點為耗費不貲。故本法適用於全體範圍較小，或國家重要事項的調查研究問題上，常見的如戶口普查、農漁業普查及工商普查等。由於普查多為國家重要事項的調查，其結果顯示一個國勢的強弱，故普查亦稱之為國勢調查。

11. 抽樣調查 Sampling survey

由全體中抽取部分個體查問的方法，稱為抽樣調查。抽樣調查的優點為節省人力、物力及時間，缺點為精確度較低，故本法適用於爭取時效或蒐集費用非常昂貴的調查研究問題上。抽樣調查與普查為兩種性能互補的方法。

12. 統計資料 Statistical data

在某特定時間及空間內，依據個體的特性去度量或數計社會現象或自然現象的群體所得的資料，稱為統計資料。統計資料的要素有三：即時間、空間及特性，因此統計資料必須註明其發生的時點或時間、地點或地區，以及其特性的名稱與單位，缺一不可。此外統計資料亦有三項特質：即客觀性、數字性與群體性。統計資料其原始必須由實地搜集而來，任何臆測估計的數字不得稱為統計資料。統計資料由度量或數計而來，均為數字的資料，任何非數字的資料不得稱為統計資料。統計資料為度量或數計社會現象或自然現象的群體而得，具有群體性，任何表示單一個體特性的數字資料，亦不得稱為統計資料。

13. 原始資料 Primary data

蒐集者直接由資料來源處搜得而未經整理簡化的統計資料，稱為原始資料。

14. 靜態資料 Static data

表示某一現象在某一特定時點靜止狀態的統計資料，稱為靜態資料。靜態資料多數由調查而來。

15. 動態資料 Dynamic data

表示某一現象在某一特定時間內演變情形的統計資料，稱為動態資料。動態資料實際為按時間先後排列的一連串靜態資料。動態資料多數由登記而來。

16. 分類 Classification

分類為將事物按時間、地域或特性的不同予以分門別類之謂。分類需要有分類標準，否則分類工作無法進行。分類的原則有二，即周延與互斥。所謂周延，是指一種分類標準下所分各類能包含資料的全部，否則即為不周延，周延的目的在不遺漏。所謂互斥，昃指一個資料只能歸入一類，不能同時歸入他類，否則即為不互斥，互斥的目的在不重複。分類為歸納法的一種，是一種重要的科學方法，事物不經分類，不能綱舉目張、條理分明。

17. 時間數列 Time series

同一地域同一特性不同時間的統計資料按發生時間先後排列者，稱為時間數列。時間數列的功用在能顯示一個事項在時間上的變化情形。時間數列的分析方法有二，一種是將一個時間數列按變動發生的原因拆開為數個時間數列，此類方法稱為時間數列分析法。另一種是將數個同時期且性質相近的時間數列融合為一個具有代表性的時間數列，此類方法稱為指數分析法。

18. 統計表 Statistical table

原始資料經分類歸類後按特定規則所作成的表格，稱為統計表。統計表的最大功能在使統計資料系統化。統計表包含三部分。即表頭、表身及表尾。表頭包含表號及標題兩部分，標題說明表內數字發生的時間、地域及其特性，缺一不可。表身包括分類標準及有關統計數字以

及其單位。表尾說明表內資料的來源及有關附註。統計表以其所含分類重數的多寡可分為一重表、二重表及多重表，一、二重表適宜作文章內的說明表，多重表適宜作統計書表大的大量統計表。多重表中分類標準的安排，表內部位的重要性依次為同行(column)、同列(row)、同行隔列及同列隔行。

19. 統計圖及其類型 Statistical chart and types

按照特定的規則以不同圖式顯示統計表內資料的特質所繪成的圖形，稱為統計圖。統計圖的最大功能在使讀者能一目了然統計表內數字的特質，而收宣傳與說服之效，其原因為看圖比識字容易。正因為如此，故統計圖的內容以簡單為宜，通常根據一重表繪製，最多根據較簡單的二重表繪製。統計圖的內容亦包含三部分，即圖頭、圖身及圖尾。圖頭包括圖號及標題兩部分，標題說明圖形內容的時間、地域及特性，缺一不可。圖身包括分類標準及圖示線，必要時尚有圖例。圖尾說明繪圖資料的來源及有關附註。統計圖常用的有歷史曲線圖、統計地圖、次數直方圖或次數曲線圖、長條圖、圓形圖等，依各種數列性質的不同，其最適合的統計圖式亦不同，在通常情況下，時間數列宜繪製歷史曲線圖；空間數列宜繪製統計地圖；變量數列宜繪製次數宜方圖或次數曲線圖；屬性數列如為絕對數宜繪製長條圖，如為相對數宜繪製圖形圖。

20. 歷史曲線圖（時間數列曲線圖） Curves of time series

歷史曲線圖為以曲線的升降表示事項在時間上變化的圖形，又稱時間數列曲線圖。由於時間有連續性，曲線亦有連續性，故以曲線圖表示時間數列在時間上的變化情形，最為適宜。通常橫坐標表示時間，縱坐標表示特性的量數。以算術比度繪製者稱為簡單歷史曲線圖；以對

數化度繪製者稱為對數歷史曲線圖；以分量累加而得總量的資料繪製同時以不同線紋表示不同分量者，稱為帶紋歷史曲線圖；帶紋歷史曲線圖中的總量及分量均以百分數表示者，稱為距限歷史曲線圖。

21. 統計地圖 Statistical map

統計地圖為以地圖為基礎並以有關圖式表示事項在地域上分布情形的圖形 。空間數列著重地域間的比較，故宜用統計地圖。統計地圖主要有三種，即多點地圖、密點地圖及線紋地圖。統計事項僅知其發生於某一地區而不知其分布狀況時，宜用多點地圖。統計事項如約略知其分布狀況時，可用密點地圖以顯示其分布狀況。多點地圖所用的點較大，密點地圖所用的點較小，不論點的大小，同一圖形中所用的點大小應相同。表示密度的統計事項如人口密度，宜用線紋地圖，以不同的線紋表示不同的密度，並蓋滿整個地區。

22. 長條圖 Bar charts

長條圖為以若干長條的長短以表示事項數量大小的圖形。屬性數列各屬性間無連貫性，長條圖中各長條亦互相分開，故屬性數列絕對數的圖示宜用長條圖。繪製長條圖時有兩點應特別注意，其一基線不可遺漏，因各長條長度皀比較均以基線為準也；另一為各長條的寬度應相等。含有一種分類標準的長條圖，稱為單式長條圖；含有兩種分類標準的長條圖，稱複式長條圖；將長條分段以表示大類中各分類量數的長條圖，稱為單式分段長條圖；單式分段長條圖中的大類及各分類量數以百分數表示者，稱為距限分段長條圖。

23. 圓形圖 Area chart

圓形圖為以全圓的面積代表事項的全體，並將全圓劃分為若干個扇形以表示各類別量較大小的圖形。由於圓形圖能顯示各類別量數在總量中所佔比例的大小，故屬性數列相對數的圖示宜用圓形圖。含有一種分類標準的圓形圖，稱為單式圓形圖；含有兩種分類標準的圓形圖，稱為複式圓形圖。複式圓形圖的繪製方式有兩種，其一為以同心圓的方式繪製，另一為分繪為數個大小相同圓形的方式繪製。

24. 次數分配 Frequency distribution

屬於同一變數的一群資料，具有同一數值變量的個數，或將變量分組，屬於同一組變量的個數，稱為次數。各組次數按變量大小順序排列者稱之為次數分配。次數分配最大的功能在能顯示一個現象的分配狀況。次數分配按其型態可分為單峰對稱分配、單峰右偏分配、單峰左偏分配、J 型分配、倒 J 型分配、U 型分配、均等分配及不規則分配等。次數分配的分析方法計有四種，對任何分配皆適用的為計算其平均數及差量，僅適用於單峰分配的為計算其偏態係數及峰度係數。

25. 次數分配表 Frequency distribution table

作成或顯示次數分配的表格，稱為次數分配表。變量的分組以 6 至 15 組為 宜。每一組有兩個組限，較小的一個稱為下限，較大的一個稱為上限。前一組的上限若與後一組的下限相等，一個變量恰當組限時，依習慣歸入後一組。同一組的上限與下限之差，稱為組距。各組的組距最好能相等，如此顯示次數分配的分配狀況以及進行分析時均較方便。同一組上限與下限的平均稱為組中點，組中點可用以代表組內各變量的數值。屬於同一組資料的個數稱為次數，各組次數的總和應等於資料的總個數。

26. 累加次數分配 Cumulative frequency distribution

將次數分配表中各組的次數由上向下順次予以累加，其結果稱為以下累加次數分配。以下累加次數分配的功能在顯示較各組上限為小變量的總個數。將次數分配表中各組的次數由下向上順次予以累加，其結果稱為以上累加次數分配。以上累加次數的功能在能顯示較各組下限為大變量的總個數。以下累加次數分配及以上累加次數分配合稱為累加次數分配。累加次數分配不但可顯示較某組下限為大或上限為小變量的個數，同時尚能據以求算中位數及四分位數。

27. 次數直方圖 Histogram

將各組的組限順次定在橫軸上，次數定在縱軸上。每一組以橫軸為基線在適當位置上豎立一個長方形，該長方形的寬度相當於該組的組距，高度相當於該組的次數。如此即繪成一個次數直方圖。次數直方圖所包圍的面積與資料的總個數成正比。一個次數分配如採用不等組距的方式，則長方形的高度即應加以調整，如某一組的組距較其他各組大一倍，則該組長方形的高度即應降低一倍，如此該長方形的面積始與該組的次數成比例，而確保全部直方圖所包圍的面積與資料總個數成比例。

28. 次數曲線圖 Frequency curve

將次數直方圖中每個長方形頂線的中點順序予以連接，再將第一組長方形頂線的中點與前一假想組的中點連接，最後一組長方形頂線的中點與後一假想組的中點連接，如此即繪成一個次數曲線圖。第一與最後兩組長方形頂線中點須與前後兩假想組中點連接的原因，為如此所繪成的次數曲線其所包圍的面積與次數直方圖所包圍的面積完全相等，而與資料的總次數成正比。

29. 累加次數曲線圖 Cumulative frequency curve

將各組的組限順次定在橫軸上，次數定在縱軸上。根據各組的上限及該組的以下累加次數決定若干點，連綴之即成以下累加次數曲線。根據各組的下限及該組的以上累加次數決定若干點，連綴之即成以卜累加次數曲線。以下累加次數曲線為不減曲線，以上累加次數曲線為不增曲線。

30. 統計量 Statistic

樣本各變量的一價函數(single-valued function)，稱為樣本統計量，或簡稱統計量。統計量可以是樣本的表徵數，也可以不是。

31. 參數 Parameter

理論機率分配中所含的未知常數，稱為參數。例如常態分配中的 μ 及 σ2 即為其兩個參數。參數的數值不同，機率分配不同，但仍屬於同一族。例如常態分配中的 μ，其不同的數值，決定一個不同的常態分配，儘管其分配不同，但仍是一個常態分配，亦即屬於同一族。理論分配的參數可以是該分配的表徵數，但多數情況不是。參數雖能顯示一個理論分配的性狀，但不同分配其參數所顯示的性狀不同，故理論分配除參數外尚有計算共同表徵數的必要。由於參數與表徵數均能顯示一個理論分配的性狀，故表徵數必為參數的函數。

32. 平均數 Mean

一群資料的代表值稱為平均數。該代表值多數以平均方式求得，故稱為平均數。平均數表示一個次數分配的中心所在位置，故又稱為地位常數(measure of location) 。又平均數為一個適中的數值，是一種現象各個體所趨向的中心，例如平均所得，故平均數又稱之為集中趨勢

(central tendency) 。平均數的功能為能以一個簡單的數值以代表全體資料。平均數主要有五種，即算術平均數、幾何平均數、調和平均數、中位數及眾數。此外尚有一種特殊的平均數，即加權算術平均數。

33. 算術平均數 Arithmetic mean

算術平均數為變量的總和除以變量個數之商。其計算公式如下：

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n} X_i$$

式中 $\bar{X}$ 代表算術平均數

X_i 代表第 i 個變量

n 代表變量總個數

算術平均數適用於算術級數資料之平均數。

34. 中位數 Median

一群資料中大小居中的一個數值，稱為中立數。其法先將資料由小至大順序排列，令其結果為Y1 ，Y2 ，....Yn 。若變量的個數 n 為奇數，則第（n＋1）／2個變量即為中位數，即

$$Me = Y_{(n+1)/2}$$

式中Me 代表中位數。

若 n 為偶數，則取第 n/2 個變量與第（n＋2）／2個變量的平均為中位數，即

$$Me = \frac{1}{2}\left[Y_{n/2} + Y_{(n+2)/2} \right]$$

35. 四分位數 Ouartiles

一群數值中，一個數值較其為小的數值佔全部資料的四分之一，為大的數值佔四分之三，則此數值即為第一四分位數。如為小及為大的數值各佔一半，則此數值即為第二四分位數，亦即中位數。如為小的數值佔四分之三，為大的數值佔四分之一，則此數值即為第三四分位數。四分位數的功能，第一四分位數能指出較其為小數值佔四分之一及為大數值佔四分之三的分界點所在，第三四分位數能指出較其為小數值佔四分之三及為大數值佔四分之一的分界點所在。其計算公式列示如下：

$$Q_i = L_{Qi} + \frac{\frac{i.n}{4} - n_{Qi}}{f_{Qi}} \times h_{Qi} \qquad i = 1,2,3$$

式中 Qi ：第 i 四分位數　　f Qi：Qi 所在組之次數

n ：總次數　h Qi：Qi 所在組之組距

L Qi：Qi 所在組下限

n Qi：較 Qi 所在組下限為小各組次數和

f i ：代表 Qi 所在組之次數

36. 眾數 Mode

一群資料中出現次數最多的數值，即為其眾數。為減少偶發性的誤差，眾數通常根據分組資料求算。眾數的計算不但適用於量的資料，同時亦適用於質的資料。眾數的功能在能顯示最普遍的數值或品質。

37. 幾何平均數 Geometric mean

n 個變量的連乘積的 n 次方根，即為其幾何平均數。其計算公式如下：

$$G = \sqrt[n]{X_1 \cdot X_2 \ldots X_n} = \sqrt[n]{\prod_{i=1}^{n} X_i}$$

式中 G代表幾何平均數

Xi 代表第 i 個變量

n 代表變量的個數

幾何平均數適用於幾何級數資料之求平均數。

38. 調和平均數 Harmonic mean

n個變量各變量倒數的平均再取倒數，即為其調和平均數。其計算公式如下：

$$H = \frac{n}{\sum_{j=1}^{n} \frac{1}{X_i}}$$

式中 H代表調和平均數

Xi 代表第 i 個變量

n代表變量的個數

調和平均數適用於調和級數資料之求平均數。

39. 加權算術平均數 Weighted arithmetic mean

變量與其權數乘積之和，除以權數總和所得之商，稱為加權算術平均數。

其計算公式如下：

$$W \cdot M \cdot = \frac{\sum_{j=1}^{n} W_i X_i}{\sum_{j=1}^{n} W_i}$$

式中W．M．代表加權算術平均數

Xi 代表第 i 個變量

Wi 代表第 i 個變量的權數

當各變量的重要性不同時，求平均宜用加權算術平均數，否則不能區別各變量在平均數中所佔的重要性。權數為能區別變量相對重要性的量數。

40. 離差 Dispersion

一群計量資料彼此間的差異，稱為離差。產生離差的原因有一，其一為受某些因素的影響而產生，另一為誤差。離差的測量方式有兩種，其一為以某種中心量數為準加以測量，其結果稱為離中差；另一為測量各變量彼此間的差異，其結果稱為互差。分析離差的方法有兩類，一類為求算一個統計表徵以表示離差的大小，此類統計表徵數稱為差量；另一類為分析其發生的原因，即為所謂的變異數分析。

41. 差量 Measure of dispersion

顯示一群資料離差大小的統計表徵數，稱為差量。差量分為兩大類，一類根據離中差計算，稱為離中差量，包括平均差、四分位差及標準差三種。另一類根據非離中差計算，稱為非離中差量，包括全距及均

互差兩種。計算差量的目的在區別平均數代表性的大小，一群資料的差量愈大，則平均數對全體的代表性即愈低。

42. 全距 Range

一群計量資料中，最大變量與最小變量之差，稱為全距。全距之功能在能用以測知一群資料的全部距離。全距雖然不是一種精確的差量，但因其計算簡單，故在品質管制上有其重要性。

43. 四分位差 Quartile deviation

一群計量資料，其第三及第一四分位數之差的一半，稱為四分位差。其計算公式如下：

$$Q.D.=\frac{(Q3-Me)+(Me-Q1)}{2}=\frac{Q3-Q1}{2}$$

式中　Q.D.代表四分位差

Q1 代表第一四分位數

Q3 代表第三四分位數

Me 代表中位數

由上列公式知，四分位差以中位數為中心，為一種離中差量。

四分位差顯示一群資料大小居中一半變量的差異情形。

44. 平均差 Mean deviation

一群計量資料，各變量與其中位數之差絕對值的平均，稱為平均差。其計算公式如下：

$$M.D.=\frac{1}{n}\sum_{i=1}^{n}|X1-Me|$$

式中 M.D.代表平均差

Xi 代表第 i 個變量

n　代表變量總個數

Me　代表中位數

平均差以中位數為中心，為一種離中差量。平均差的計算須取離中差的絕對值，故不適代數處理。以中位數為中心的平均差較以其他任何中心量數為中心的平均差為小。

45. 標準差 Standard deviation

一群計量資料各變量與其算術平均數之差稱之為離均差，各變量離均差平方的算術平均數的方根稱為標準差。其計算公式如下：

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

式中　S　代表標準差

Xi 代表第 i 個變量

n　代表變異總個數

X　代表算術平均數

標準差以算術平均數為中心，是一種離中差量。以算術平均數為中心的標

準差較以其他任何中心量數為中心的標準差為小。

46. 相對離差 Relative dispersion

相對離差為絕對離差與某種平均數或其他適當量數之比，並取成百分數。皮爾生氏(K. Pearson)稱其為變異係數(Coefficient of variation)。例如

$$C.V. = \frac{S}{X} \times 100$$

式中 C.V.代表變異係數

S 代表標準差

X 代表算術平均數

計算變異係數的目的一方面在消除計量單位的影響，另一方面在消除平均數大小的影響，使不同單位或不同均數的兩組資料的分散度可以互相比較。

47. 指數 Index number

指數為表示多種同類現象一般水準並以某一時期為基準的百分數。指數有四種基本性質，即綜合性、代表性、平均性與相對性。所謂綜合性，即指數由多種同類現象綜合而成。所謂代表性，即包含在指數內的項目能代表未包含在內的項目。所謂平均性，即各項目同一時期的數值求算其平均以表示一般水準。所謂相對性，即不同時期的數值求算以某一時期為基準的百分數。指數以基期是否固定分為定基指數與環比指數兩種；以加權與否分為簡單指數與加權指數兩種；以求平均與求比過程的先後分為綜值式指數與平均式指數兩種；以計算的對象分為物價指數、物量指數及物值指數三種。

48. 簡單指數 Simple index number

簡單指數即未加權的指數。簡單指數分為兩類，即簡單綜值式指數與簡單平均式指數。

50. 加權指數 Weighted index number

加權指數為以能表示物品相對重要性的數值為權數加權而得的指數。加權指數分為兩類，即加權綜值式指數及加權平均式指數。

51. 基期 Base period

指數據以為準進行比較的時期，稱為基期。基期分為兩種，即固定基期與移動基期。固定基期顧名思義為以某一固定時間為基準的時期，移動基期為以計算期前一期為基準的時期。以固定基期為準所計算的指數稱為定基指數，以移動基期為準所計算的指數稱為環比指數。定基指數適於作長期比較，環比指數適於作短期比較。以物價指數為例，固定基期宜選擇物價穩定時期為之，長度以一年或一年以上為佳，以一年以下時間為基期不能避免季節變動的影響。

52. 機率 Probability

設一事件有ｎ種同等出現可能的出象，其中含性質Ａ者ｎＡ種，則事象Ａ

出現的機率為：

$$P(A)=\frac{n_A}{n}$$

式中　Ｐ（Ａ）代表事象Ａ出現的機率

此為古典機率理論對機率所下的定義。除此而外，尚有次數比的機率理論

、機率的公理體系及主觀的機率理論，各種理論均對機率有一個定義。

53. 條件機率 Conditional probability

設有Ａ、Ｂ兩事象，在Ａ出現後再出現Ｂ的條件機率為：

$$P(B \mid A)=\frac{P(A \cap B)}{P(A)}, \quad \text{但} P(A)>0$$

式中 Ｐ（Ｂ｜Ａ）代表Ａ出現後再出現Ｂ的條件機率。

P（A∩B）代表A、B積事象發生的機率。

P（A）代表A事象發生的機率。

54. 機率分配 Probability distribution

一個隨機變數各變量發生的機率按變量大小順序排列者，稱為機率分配。機率分配為隨機現象多次試行規律的一個模型，可據以進行分析與判斷。

55. 常態分配 Normal distribution

設有一個連續隨機變數X，若其機率密度函數 f(x) 成為下列形式，即為一個常態分配。

$$f(x)=\frac{1}{\sqrt{2}\pi\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \qquad -\infty< x <\infty$$

$$-\infty<\mu<\infty，-\infty<\sigma^2<\infty$$

式中 μ 及 σ^2 為未知參加。此分配可以 $n(\mu，\sigma^2)$ 符號表示之。常態分配為一個單峰對稱分配，對稱於 $x=\mu$ 處，曲線的轉向點在 $x=\mu\pm\sigma$ 處。

56. 標準常態分配 Standardized normal distribution

設有一個連續隨機變數Z，若其機率密度函數 f(z) 成為下列形式，即為標準常態分配。

$$f(x)=\frac{1}{\sqrt{2}\pi}e^{-\frac{1}{2}z^2} \qquad -\infty< z <\infty$$

此分配為一個單峰對稱分配，對稱於 $z=0$ 處，曲線的轉向點在 $z=\pm 1$ 處。一般常態分配 $N(\mu, \sigma^2)$，取

$$Z=\frac{X-\mu}{\sigma}$$

此稱為常態標準值，經變數變換即可化為標準常態分配。

57. 大數法則 Law of large numbers

設隨機變數 X 的機率密度函數為 f(x)，其均數為 μ，變異數為 σ^2，且 $\sigma^2<\infty$；令 $\overline{X}n$ 為次數為 n 的樣本均數；又設 ε 及 δ 為兩個微小數值，且 $\varepsilon>0$，$\delta>0$；若 n 為正整數，且大於 $\sigma^2/\varepsilon^2\delta$，則

$$P\{-\varepsilon<\overline{X}n-\mu<\varepsilon\}\geqq 1-\delta$$

此即為所謂的大數法則。只要代 $\varepsilon=a\sigma/\sqrt{n}$ 及 $\delta=1/a^2$ 入拓拔契夫不等式即獲得本法則。此法則說明樣本次數 n 加大時，樣本均數 $\overline{X}n$ 接近母體均數 μ 的機率趨近於一。簡言之，即觀察的次數愈多，則推測的結果愈為可靠。

58. 中央極限定理 central limit theorem

設隨機變數 X 的機率密度函數為 f(x)，其均數為 μ，變異數為 σ^2。由其中隨機抽取 n 個變量為一組樣本，樣本均數為 $\overline{X}n$，令

$$Yn=\frac{\overline{X}n-\mu}{\sigma/\sqrt{n}}$$

Yn 為 $\overline{X}n$ 的標準值。當 $n\rightarrow\infty$ 時，Yn 的抽樣分配 f(y) 以標準常態分配為其極限，即

$$\lim_{n\to\infty} f(y) = n(0, 1)$$

此定理說明不拘母體分配為何，樣本均數的抽樣分配當樣本次數趨近於無窮大時，以常態分配為其極限。中央極限定理為假定母體為常態分配的重要理論依據。

59. 抽樣分配 Sampling distribution

由母體中抽取所有可能同次數樣本的同一種統計量的機率分配，稱為樣本統計量的抽樣分配，簡稱為抽樣分配。例如假定母體為一個點二項分配，即 $f(x) = p^x q^{1-x}$ ，$x = 0, 1$。由其中隨機抽取 n 個變量為一組樣本，各變量為 X_1，X_2，....，X_n。取 $Y = \Sigma X$，此為樣本變量和。每抽一組樣本即可求得一個樣本變量和，所有可能同次數的樣本即可求得許多樣本變量和。

這許多樣本變量和按其大小即可作成一個機率分配，此機率分配即為點二項母體樣本變量和的抽樣分配，其結果為一個二項分配，即

$$g(y) = \binom{n}{y} p^y q^{n-y} \quad , y = 0, 1, 2, ..., n$$

60. 期望數 Expectation

設隨機變數 X 的機率密度函數為 f(x) ，則 X 的期望數可求得如下：

(1)當 X 為間斷隨機變數時，其期望數為：

$$E(X) = \sum_x x f(x)$$

(2)當 X 為連續隨機變數時，其期望數為：

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

期望數與平均數的計算方式甚為相似，但意義略有不同。即平均數為實現的數值，而期望數則為平均可望之數，含有機率的成分在內。又期望數亦為一套計算工具，可用以求算機率分配的各種表徵數，包括均數、變異數、偏態係數及峰度係數等。

61. 變異數 Variance

變異數為標準差或標準誤的平方，用以表示一群資料或抽樣分配的分散度。表示資料分散度最恰當的統計量為標準差或標準誤，取其平方的目的在便於運算與分析，因時時平方及開方不但不勝其煩，且有時亦無此必要。例如變異數分析即是根據變異數直接進行的分析，根本不需要開方。變異數的計算方式符合誤差法則，故在推論統計上特別重要。

62. 標準誤 Standard error

抽樣分配變異數的方根稱為標準誤。標準誤的計算方式與標準差者甚為相似，但意義略有不同。標準誤根據抽樣分配計算而得，表示平均抽樣誤差的大小。標準差根據次數分配計算而得，表示一群資料的平均差異程度，不限於誤差，特別不是抽樣誤差。

63. 估計 Estimation

根據樣本資料推估母體參數的方法，稱之為估計。估計方法可分為兩大類，其一為推估母體參數的一個可能數值，此類方法稱之為點估計。另一是為母體參數建立一個可能所在範圍，此類方法稱之為區間估計。

64. 點估計 Point estimation

根據樣本資料推估母體參數的一個可能數值的方法，稱之為點估計。點估計的內容包括兩方面，其一為點估計量的評選，另一為點估計量的求取方法。評定點估計量的準則常用者有五點，即不偏性、一致性、漸近有效性、充分性及最小變異不偏性。求取點估計量的方法常用者有三種，即最概法、貝氏法及最小平方法。

65. 區間估計 Interval estimation

根據樣本資料為母體未知參數提供一個可能所在範圍的方法，稱為區間估計。區間估計通常是根據母體參數的最概估計量及其抽樣分配加以建立的，故區間估計為點估計的引申，進一步說明估計誤差的大小。

66. 機率區間 Probability interval

一個隨機變數出現於某一固定區間可能性的大小，稱為機率，該固定區間即為所謂的機率區間。例如母體為常態分配 f（x）＝n（x；μ，σ^2），式中 μ 為均數，σ^2為變異數。隨機變數X出現於下列區間的機率為 0.95。

（$\mu-1.96\sigma$；$\mu+1.96\sigma$）

μ 與 σ^2 雖為未知，但均為固定數值，故上列區間是一個固定區間。整個機率陳述為：

$$P\{\mu-1.96\sigma < X < \mu+1.96\sigma\}=0.95$$

67. 信賴區間 Confindence interval

可能蓋括母體未知參數在內的隨機區間，稱為信賴區間。信賴區間通常由機率區間解不等式而得。例設母體為常態分配，即 f（x）＝n（x；μ，σ^2），樣本次數為n，樣本均數為X。X的抽樣分配為g （x）＝n（x；μ，σ^2／n）。設機率為 0.95，則 X 的機率區間為：

$$P\left\{\mu-1.96\frac{\sigma}{\sqrt{n}}<X<\mu+1.96\frac{\sigma}{\sqrt{n}}\right\}=0.95$$

解不等式，其結果如下：

$$P\left\{X-1.96\frac{\sigma}{\sqrt{n}}<\mu<X+1.96\frac{\sigma}{\sqrt{n}}\right\}=0.95$$

此即為 μ 的信解區間，即〔X－1.96σ／√n；X＋1.96σ／√n〕。該區間上下限中均含有隨機變數 X 在內，故為一個隨機區間，隨所出現樣本之不同而異。此區間蓋括 μ 在內的可能性為 0.95，稱為信賴係數。

68. 相關 Correlation

兩個或兩個以上變數相互間的關係，稱為相關。兩個或兩個以上變數的資料必須成對或成束出現，才能分析其間的相關。相關分析主要求算表示變數間相關的係數，其次為進一步進行推論，瞭解母體上的相關情形，相關分析因資料性質的不同而異，計量資料的相關即簡稱為相關，計數資料的相關稱為品質相關。計量資料的相關按變數的多寡分為兩類，兩個變數間的相關稱為簡單相關，兩個以上變數間的相關稱為複相關。計數資料的相關若兩種品質均各分為兩類，其所分析的相關稱為相聯；若兩種品質至少有一種分為兩類以上，其所分析的相關稱為列聯。兩個以上變數的品質相關亦稱為列聯，因其分析方法與上一情況完全相同也。此外還有一種相關，即所謂的等級相關。等級相關不限計量資料或計數資料，其唯一條件為每一變數各變量均能排定等級。

69. 簡單相關 Simple correlation

計量資料兩個變數間的相關，稱為簡單相關。簡單相關又可分為兩類，即簡單直線相關與簡單非直線相關。簡單直線相關以直線為準分析兩變數間的相關，簡單非直線相關以曲線為準或以其他適當量數為準分析兩變數間的相關。

70. 相關係數 Correlation coefficient

表示計量資料兩變數間以直線為準的相關方向及程度的係數，稱為相關係數。設有（Ｘ，Ｙ）兩相關變數，其ｎ對成對出現的資料為（Ｘ1，Ｙ1）（Ｘ2，Ｙ2），......，（Ｘn，Ｙn）。求算相關係數的公式為：

$$\gamma=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{S_x}\right)\left(\frac{Y_i-\bar{Y}}{S_y}\right)$$

式中γ代表相關係數，Ｘ代表Ｘ的均數，Ｙ代表Ｙ的均數，Ｓx 代表Ｘ的標準差，Ｓy 代表Ｙ的標準差。

其性質如下：

(1)相關的方向──若γ=＋，兩變數間有正相關；若γ=－，兩變數間有負相關。

(2)相關的程度──若γ=０，兩變數間無相關；｜γ｜愈接近１表示相關 程度愈高。｜γ｜=１，兩變數間為完全相關。

(3)所在範圍── －１≦γ≦＋１。