# Engineering Statistics

## Probability

**Dr. Vvn Weian Chao (趙韋安)**
https://ce.nctu.edu.tw/member/teachers/23
Department of Civil Engineering, National Yang Ming Chiao Tung University, Taiwan

# Purpose

# 描述統計
# 進入
# 推論統計

Random Sampling

Probability

# Purpose

-Experiments
-Chance experiments
-Probability
-Disjoint Event
-Conditional probability
-Independent Events

# Data from Experiments

實驗設計取得樣本資料，用以探討實驗變數與一般結論成果之間的關係。

- The choice of a data collection method is dictated by how we intend to use the data.
- If our work involves applying standards and codes, we use operational definitions.
- We ensure that the results will be directly *comparable* to similar tests and measurements made by ourselves and others.

# Data from Experiments

不同廠商提供的塑料樹脂，誰可以提供最佳的硬度需求?

不同注射機器? 不同的塑料樹脂?

| Brand A | | Brand B | |
|---|---|---|---|
| $x_1$ | $x_2$ | $x_3$ | $x_4$ |

(a)

| | Brand A | Brand B |
|---|---|---|
| Machine 1 | $x_1$ | $x_3$ |
| Machine 2 | $x_2$ | $x_4$ |

(b)

# Purpose Design of Experiments

$$Y=a_1x_1+a_2x_2+a_3x_3+\ldots \quad \text{+-error}$$

| | |
|---|---|
| 研究因果關係 | 參數之間的相互影響關係 |
| 增加額外的有效數據 | 產品最佳化 |
| 研究特定因子對於某現象的影響 | 測量實驗誤差 |

# Experimental Design Terminology

| | |
|---|---|
| Response variable (應變數) | Experimental Error (估計實驗誤差) |
| Independent variable (自變數) | Randomization (隨機量測) |
| Replication (重複性量測) | Blocking (去除外部影響因子) |

# Blocking

**Block what you know, randomize what you don’t know**

# Measurement Systems

- **Metrology (計量學):**
  - the study of measurement
  - deals with our ability to produce measurements of sufficient **accuracy** (**準確度**) and **precision** (**精準度**) to support any analyses based on these measurements
- **Calibration:**
  - addresses the various systematic errors that can cause an instrument’s readings to be in error
  - instruments are said to be “in calibration” if they give true readings, that is, if their offset is zero

# Measurement Systems

- Accuracy and Precision
- Repeatability and Reproducibility

# Accuracy and Precision

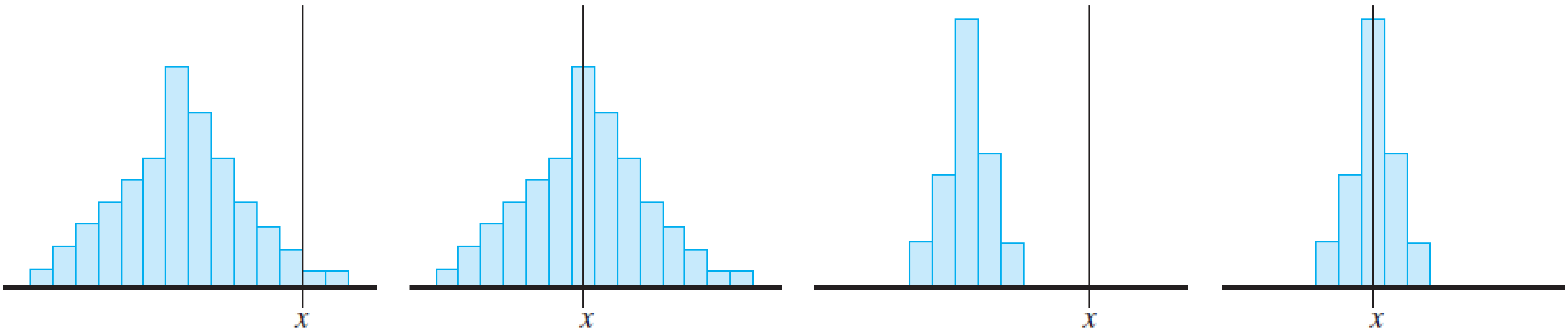
- **Accuracy 準確性** refers to the degree to which repeated measurements of a known quantity *x* tend to agree with *x*.
- Given several repeated measurements $x_1, x_2, x_3, \ldots, x_n$ of some known value *x*, the accuracy of the readings:

$$\boxed{\text{accuracy} = \bar{x} - x}$$

- **Precision 精確性** is a measure of variation between repeated measurements and is estimated by the sample standard deviation of *n* repeated measurements:

$$\boxed{\text{precision} = s = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}}$$

# Accuracy and Precision




(a) Inaccurate and imprecise (b) Accurate, but imprecise (c) Precise, but inaccurate (d) Accurate and precise

# Repeatability & Reproducibility

- Repeatability

(可重複性，可以減少外部因子造成的測量誤差):
  - the amount of variation expected when almost all external sources of measurement error have been controlled and held fixed
  - is often conducted by the same person using a single instrument to repeatedly measure a single item
  - is a measure of the best that one can hope to achieve from a measuring instrument
- Reproducibility (可重現性):
  - allows several factors to vary at the same time
  - uses several operators and several instruments to measure several production items
  - is usually based on simple experimental designs that allow us to break measurement variation into distinct components that estimate the contribution of the various noise factors

# Chance Experiments

-一樣的試驗如果每次做都無法得到相同結果，則稱為「Chance Experiments」: 氣象預報、藥物藥效反應

- A **chance experiment**, also called a **random experiment**, is simply an activity or situation whose outcomes, to some degree, depend on chance.
- While doing experiments, we often have questions: Will I get exactly the same result if I repeat the experiment more than once?
- If the answer is "no", the experiment follows a chance experiment.
- Examples: Weather forecast, measuring the yield of a chemical reaction, assessing the potency of a pharmaceutical product.

# Chance Experiments

Events

Depicting Events

Forming New Events

# Events

## -Simple events, Events, Sample space

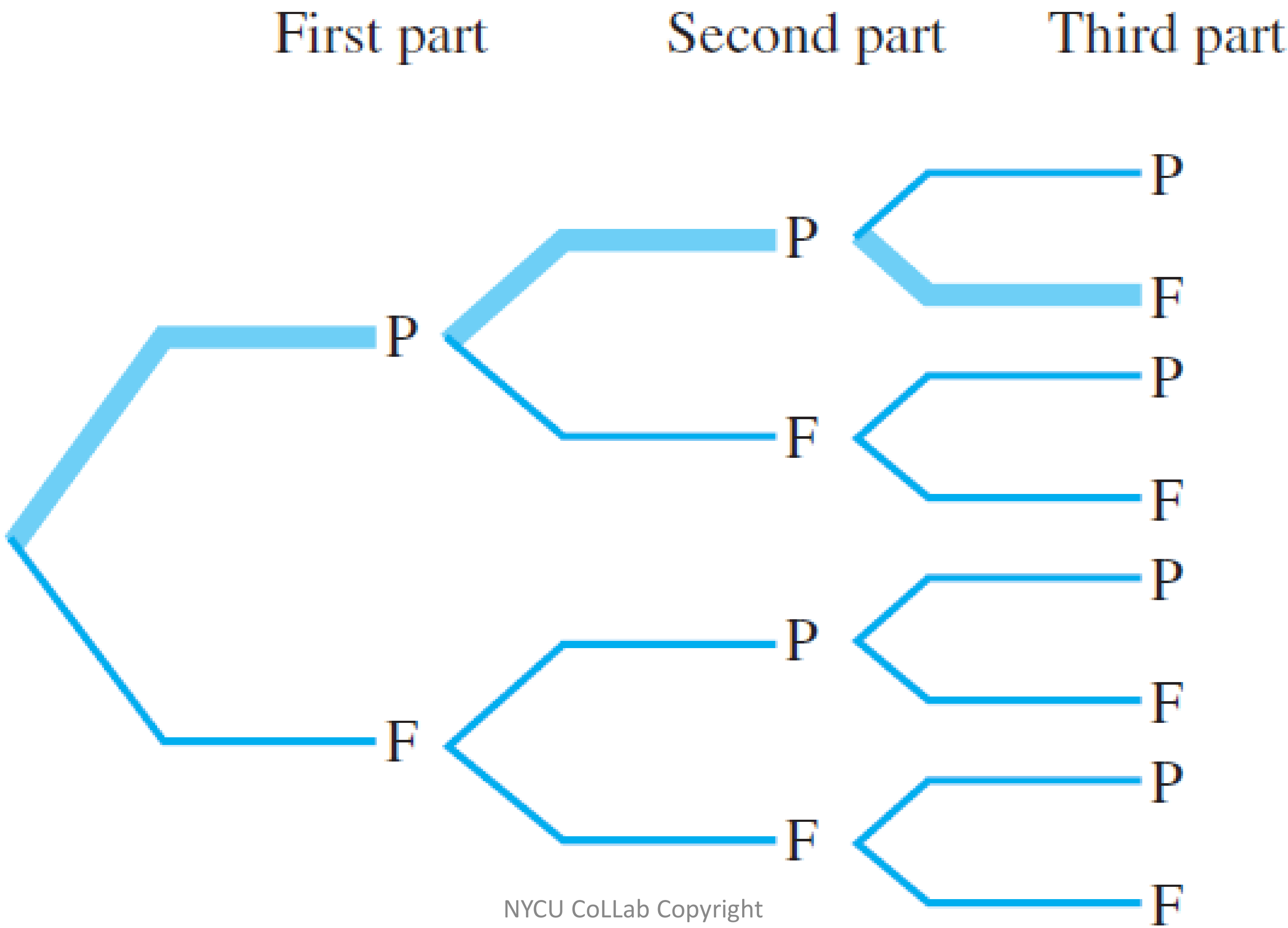
- The outcomes of chance experiments can be:
  - **Simple events** – the individual outcomes of an experiment
  - **Events** – consist of collections of simple events
- Events are often denoted by single uppercase letters
- For example: The event that at least 2 out of 3 metal parts pass a stress test corresponds to the set of outcomes {*PPP, PPF, PFP, FPP*}, and is written by $A$ = {*PPP, PPF, PFP, FPP*}, $P$ for "pass" and $F$ for "fail"
- All the possible outcomes of the chance experiment, which, taken together, form the **sample space** of the experiment.

# Depicting Events

- **Tree diagrams**
  - useful for depicting experiments that are conducted in a sequence of steps
  - each step in the sequence is given its own set of branches, which themselves form the starting points for all branches to their right
- **Venn diagram 凡氏圖**
  - useful for depicting ***relationships*** between events
  - simple two-dimensional figures, whose enclosed regions are intended to depict a collection of simple events, called *points,* in a sample space

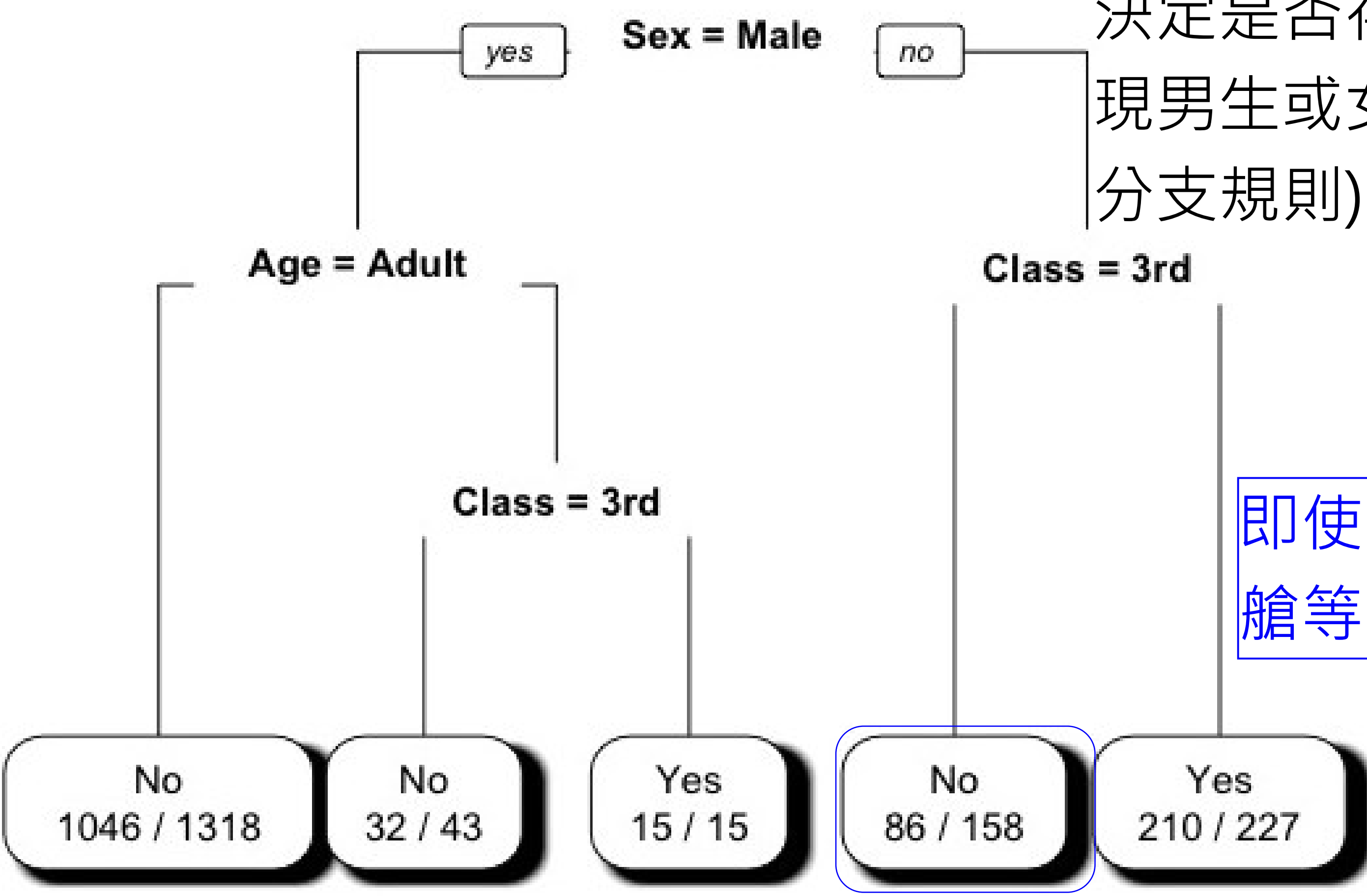
# Depicting Events

## -Tree Diagrams



# Titanic data example: Decision Tree

## rpart()決策樹套件




決定是否存活，我們可以發現男生或女生最為重要(第一分支規則)；其次為艙等。

即使為女性，且為最低艙等的死亡率可達54%

number of correct classifications/number of observations in that node

# R: Decision Tree-CART

rpart(
Survived ~. ,
data = train,
).

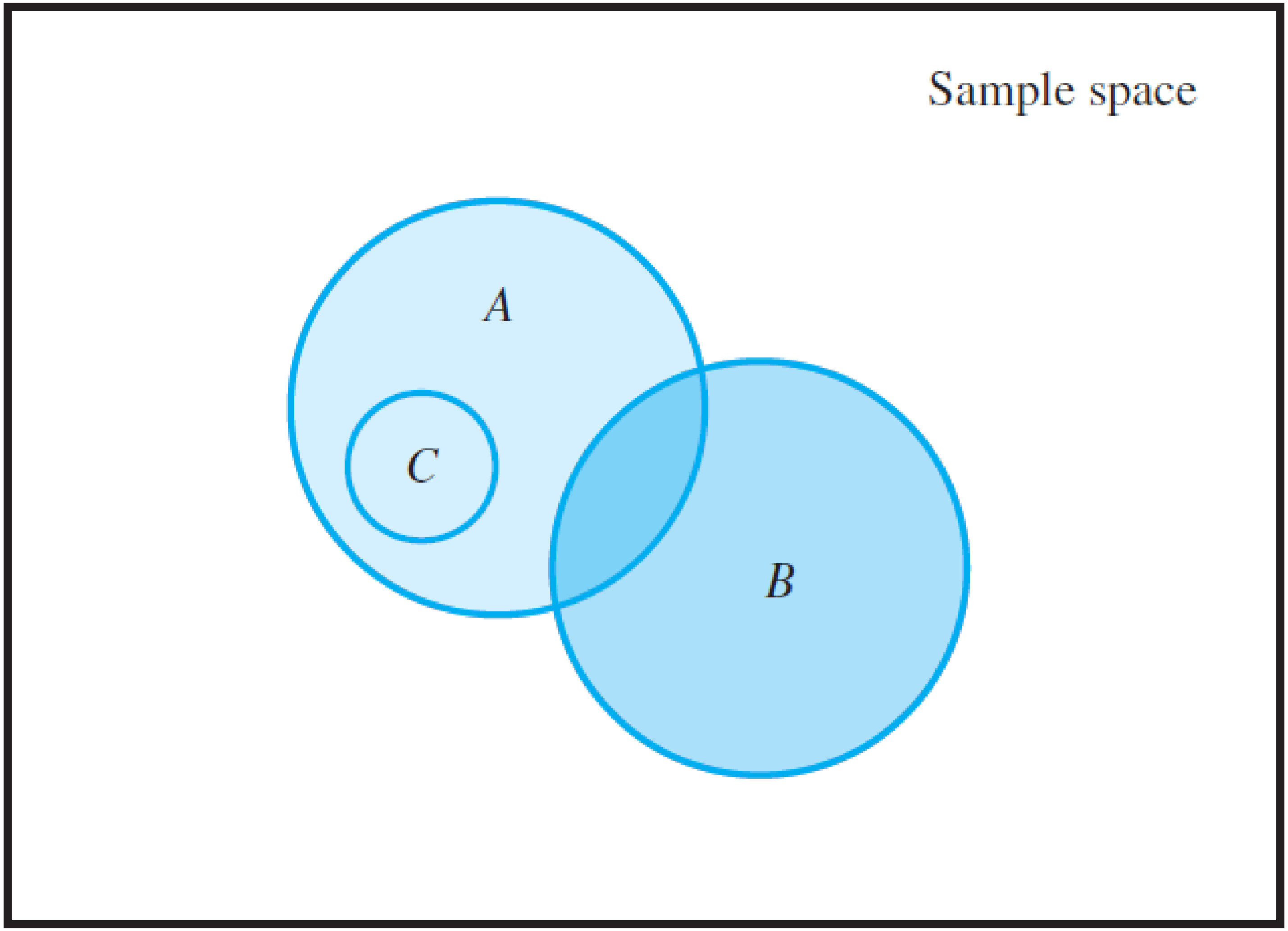
# TRY
# it
# in
# R

# R: Decision Tree-CART

titanic.R

# Depicting Events

## -Venn Diagrams: relationships



# Forming New Events

- New complex events are built from old events through the use of the words *and, or,* and *not.*
- For a chance experiment and any two events *A* and *B*:
  1. The event ***A or B*** consists of all simple events that are contained in either *A* or *B*. ***A or B*** can also be described as the event that *at least one* of *A* or *B* occurs.
  2. The event ***A and B*** consists of all simple events common to both *A* and *B*. ***A and B*** can be described as the event that *both A* and *B* occur.
  3. The event ***A′***, called the **complement of *A***, consists of all simple events that are *not* contained in *A*. ***A′*** is the event that *A* does not occur.

# Forming New Events: Example

**Example 5.3: The experiment of selecting and testing four metal parts**

- Consider the events $A$ and $B$ as described in Example 5.1
- Then $A$ *or* $B$ = {*PPPP, PPPF, PPFP, PFPP, FPPP, PPFF, PFPF, PFFP, FPPF, FPFP, FFPP, PFFF, FPFF, FFPF, FFFP, FFFF*}
- $A$ *and* $B$ = {*PPFF, PFPF, PFFP, FPPF, FPFP, FFPP*}
- $A'$ = {*PFFF, FPFF, FFPF, FFFP, FFFF*}

**Observation:**

- The event *A or B* contains *all* 16 sample space points.
- *A and B* = exactly two parts pass (and, hence, two fail) the stress test.
- $A'$ = the event that at most one part passes the test.

# Forming New Events: Example

-Mutually exclusive or disjoint: 兩個事件中沒有共同的部分

- When two events $A$ and $B$ have no simple events in common, they are ***mutually exclusive*** or ***disjoint***.
- Given a chance experiment and any events $A_1, A_2, A_3, \ldots, A_k$:
  1. The event ***$A_1$ or $A_2$ or $A_3$ or . . . or $A_k$*** consists of all the simple events that are contained in at least one of the events $A_1, A_2, A_3, \ldots,$ or $A_k$. It can be described as the event that *at least one* of the events $A_1, A_2, A_3, \ldots,$ or $A_k$ occurs.
  2. The event ***$A_1$ and $A_2$ and $A_3$ and . . . and $A_k$*** consists of all simple events common to *all* the events $A_1, A_2, A_3, \ldots,$ and $A_k$. It can be described as the event that *all* of the events $A_1, A_2, A_3, \ldots,$ and $A_k$ occur.
  3. Several events $A_1, A_2, A_3, \ldots,$ and $A_k$ are said to be **mutually exclusive** or **disjoint** if no two of them have any simple events in common.

# R: Venn Diagram

**venn.diagram(**

x, filename,

imagetype,

fill,

cat.col, cat.cex,

margin **).**

# R: Venn Diagram

x: list variable

filename= “venn.png”

imagetype= “png”

fill: 事件類別顏色

cat.col: 事件類別字體顏色

cat.cex: 字體大小

# R: Venn Diagram

install.packages

( “VennDiagram” )

library(grid)

library(futile.logger)

library(VennDiagram)

# TRY it in R

# R: Venn Diagram

R_probability_a.R

# Probability Concepts

-(1)機率➔量化不確定事件的可能性

-(2)機率值➔長時間及多次測試之下的某狀況出現比例

- Probability allows us to quantify the likelihood associated with uncertain events, that is, events that result from chance experiments.
- The probability of an event can be thought of as the proportion of times that the event is expected to occur in the long run.
- Probabilities are reported as:
  - *Proportions* (between 0 and 1) or;
  - *Percentages* (between 0% and 100%)
- $P(A)$ denotes the probability of an event $A$ occurring
- So, $P(A) = .30$ means the probability of event $A$ occurring is .30, or the event $A$ has a 30% chance of occurring

# Probability Concepts

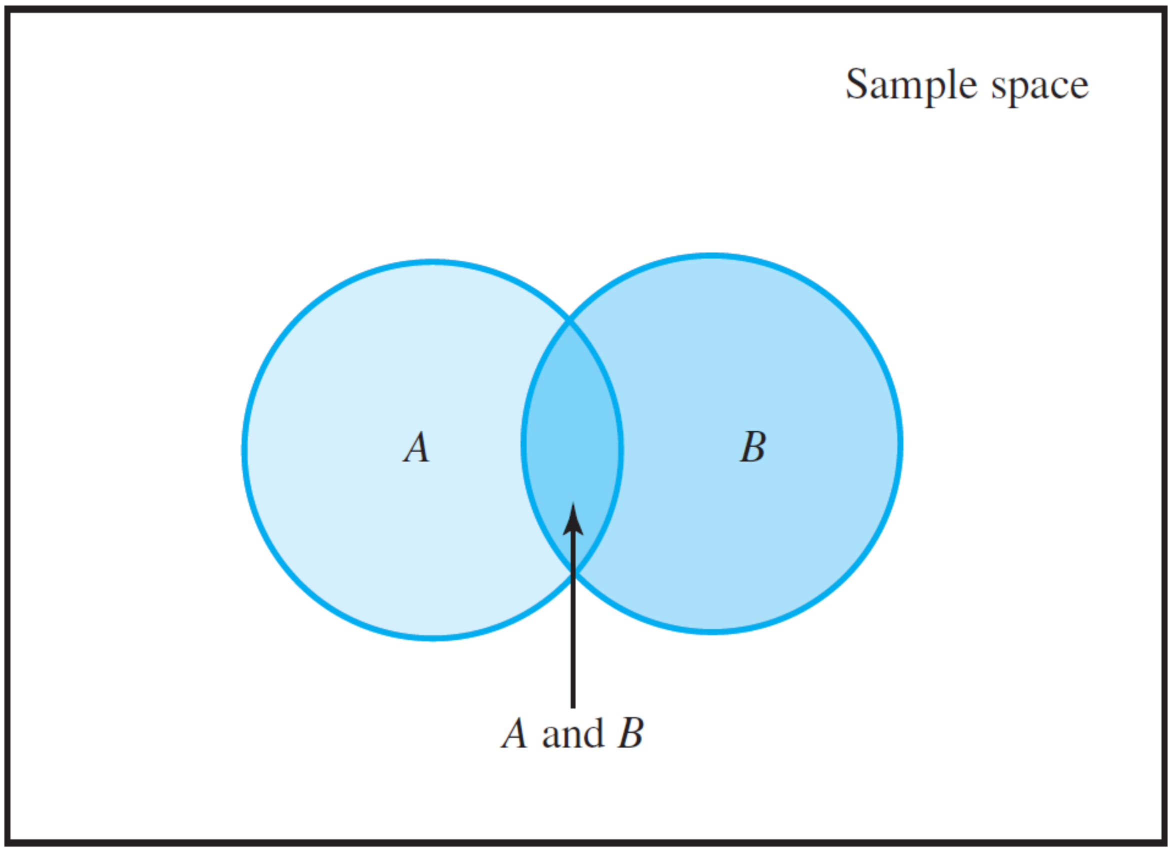
- The General Addition Rule
- The Addition Rule for Disjoint Events
- Complementary Events
- Assigning Probabilities

透過機率的加法準則判斷機率事件是否為不相干

# The General Addition Rule

- For any two events *A* and *B*, which need not be mutually exclusive:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$



# The Addition Rule for Disjoint Events

- **Disjoint**, or **mutually exclusive, events** are events that cannot occur simultaneously.
- For any two disjoint events $A$ and $B$,

$$P(A \text{ or } B) = P(A) + P(B)$$

- More generally, for any collection of disjoint events $A_1, A_2, A_3, \ldots, A_k$,

$$P(A_1 \text{ or } A_2 \text{ or } A_3 \text{ or} \ldots \text{or } A_k) = P(A_1) + P(A_2) + P(A3) + \ldots + P(A_k)$$

# Complementary Events

- The **complement** $A'$ of an event $A$ is the collection of simple events that are *not* in $A$.
- Complementary events are mutually exclusive events: $A$ and $A'$ are disjoint.
- When an event $A$ does *not* occur, we say that its complement $A'$ has occurred, and vice versa.
- The probabilities of $A$ and $A'$ are related by the formula $P(A) = 1 - P(A')$, which is called the *law of complementary events.*

# Complementary Events

## -機率: 計算至少有一個失敗的機率

**Example 5.6**

Refer to Example 5.5. Suppose you want to find the probability that, of the 20 items randomly selected for inspection, at least one item fails to meet quality standards. Denote this event by $D$ = at least one item fails inspection. One approach to finding this probability is to partition $D$ into the events $E_1, E_2, E_3, \ldots, E_{20}$, where, for each $i = 1, 2, 3, \ldots, 20$, the $E_i$ denotes the event that exactly $i$ items fail inspection. Since $E_1$ through $E_{20}$ are disjoint, the addition rule says that $P(D) = P(E_1) + P(E_2) + \cdots + P(E_{20})$. As mentioned in Example 5.5, the binomial mass function could then be used to find each $P(E_i)$ in this summation.

Although the addition rule will give the correct value for $P(D)$, an easier method for [redacted] The complement of the event $D$ = at least one item fails inspection is the event $D'$ = no items fail inspection. As we will see in Section 5.4, finding $P(D')$ requires only one computation with the binomial mass function, whereas the partition method requires 20 separate computations.

# Assigning Probabilities

-做試驗的測數越多，某事件的機率會越來越穩定

**Probability Axioms**

1. The probability of any event must lie between 0 and 1. That is, $0 \leq P(A) \leq 1$ for any event .
2. The total probability assigned to the sample space of an experiment must be 1.

We determine probabilities:

- as frequencies of occurrence

P(A)=number of times A occurs/numbers of times experiment is repeated: as the number of trials increases, we expect this ratio to

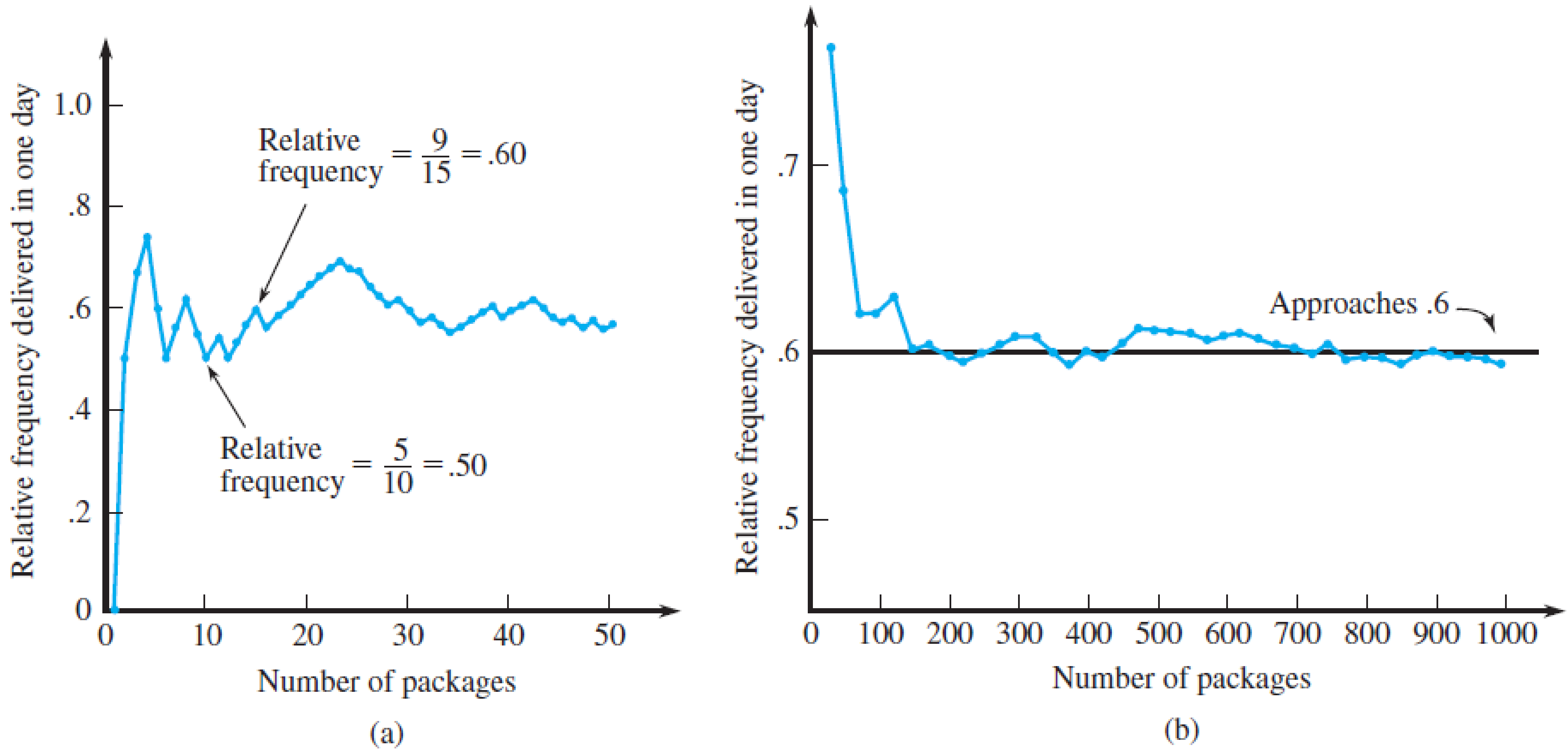
「stabilize and eventually approach a limiting value」

- from subjective estimates (主觀估計)
- by assuming that events are equally likely
- by using **density and mass functions**

# Assigning Probabilities: Example

-機率: 兩天出貨期，但商品可以在一天內抵達的事件

| Package No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Did A occur | N | Y | Y | Y | N | N | Y | Y | N | N |
| Relative frequency of A | 0 | .5 | .667 | .75 | .6 | .5 | .571 | .625 | .556 | .5 |



(a) (b)

TRY

it

in

R

# R: Dice game

R_probability_b.R

# Conditional Probability & Independence

- If the probability that $A$ occurs depends on whether $B$ has occurred, we use the methods of **conditional probability.**
- If the occurrence or nonoccurrence of $B$ has no effect at all on the probability that $A$ occurs, we say that $A$ and $B$ are **independent events.**

# Conditional Probability & Independence

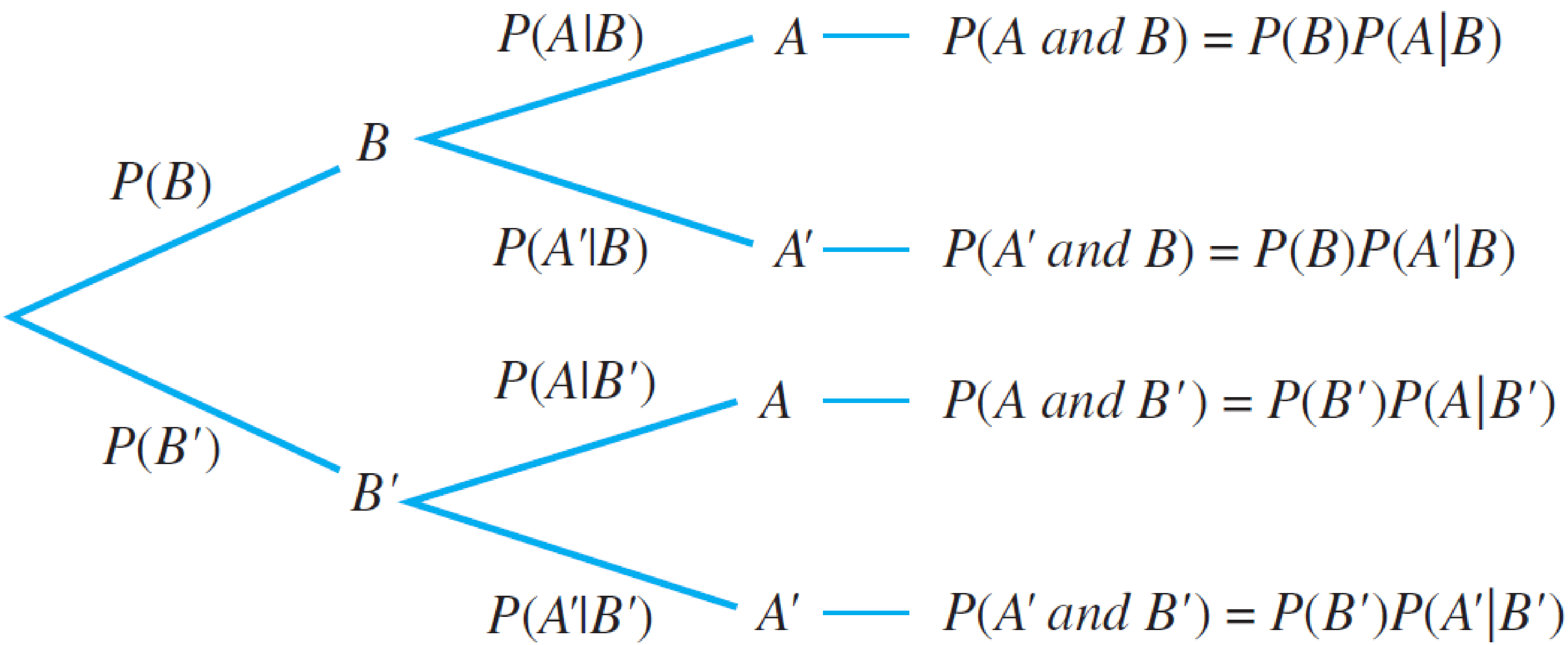
- Conditional Probability
- Independent Events
- Combining Several Concepts

# Conditional Probability

- Let $A$ and $B$ be two events with $P(B) > 0$.
- The **conditional probability of** $A$ **occurring given that event** $B$ **has already occurred** is denoted by $P(A|B)$.
- The computing formula is:

$$P(A \mid B) = P(A \text{ and } B) / P(B)$$

# Conditional Probability



# Independent Events

- Two events, $A$ and $B$, are **independent events** if the probability that either one occurs is not affected by the occurrence of the other:

$$P(A \text{ and } B) = P(A)\,P(B)$$

- Several events, $A_1, A_2, A_3, \ldots, A_k$, are **independent** if the probability of each event is unaltered by the occurrence of any subset of the remaining events.
- The probability that all the events in any *subset* occur equals the product of their individual probabilities of occurring.
- For all $k$ events:

$$P(A_1 \text{ and } A_2 \text{ and } A_3 \text{ and} \ldots \text{and } A_k) = P(A_1)P(A_2)P(A_3) \ldots P(A_k)$$

# Independent Events: Example

-一個串聯系統，只有兩個元件都正常運作下才可以

Example 5.8

One branch of reliability theory, called *topological reliability*, is concerned with calculating the reliability of systems comprising several components connected in specific patterns. One common layout for components is the series system (Figure 5.9), in which the system operates correctly only if *each* of its subcomponents works correctly. A familiar example of such a system is a circuit with two switches, both of which must be closed for the circuit to conduct electricity. It is commonly assumed that the components are independent when performing reliability calculations.

→ Component A → Component B →

圖 5.9　一種串聯兩元件的系統，只有在兩元件正常運作下才會正常運作。

Suppose that the switches A and B in a two-component series system are closed about 60% and 80% of the time, respectively. If we assume that the closing of switch A occurs independently of switch B, the probability that the entire circuit is closed is

$$\begin{aligned} P(\text{circuit closed}) &= P(\text{A closed } \textit{and} \text{ B closed}) \\ &= P(\text{A closed})\, P(\text{B closed}) \\ &= (.60)(.80) = .48 \end{aligned}$$

That is, the circuit will be closed about 48% of the time.

# Independent Events: Example

-消費品質研究，單一產品有瑕疵的機率為P=0.005

-至少一個有瑕疵的機率0.2217

-代表還是有很高的機率有瑕疵

Example 5.9

In an example demonstrating how vendor quality affects customer quality, H. S. Gitlow and D. A. Wiesner ("Vendor Relations: An Important Piece of the Quality Puzzle," *Quality Progress*, 1988: 19–23) considered a hypothetical product consisting of 50 critical parts, any one of which, if defective, could cause the finished product to be defective. Suppose that each of these parts is purchased from a different vendor. It is therefore reasonable to assume that the condition of each part, created by a different vendor, should be *independent* of the conditions of the others. Furthermore, suppose that about 99.5% of all the parts supplied by a given vendor are good. What is the overall proportion of assembled products that can be expected to be defective?

To answer this question, let $D_i$ denote the event that the part purchased from the $i$th vendor is defective, so that $P(D_i) = .005$ and $P(D_i') = .995$. Then, the probability we seek is

$$\begin{aligned} P(\text{at least one of the 50 parts is defective}) &= 1 - P(D_1')P(D_2')P(D_3') \cdots P(D_{50}') \\ &= 1 - (.995)^{50} = 1 - .7783 = .2217 \end{aligned}$$

This example demonstrates the important point that it is possible for complex systems to have high failure rates even if the quality of their individual components is relatively good.

# Combining Several Concepts

## -並聯系統形成封閉迴路的機率有多高?

Example 5.10
並聯系統

串聯系統

Consider the portion of an electronic circuit diagrammed in Figure 5.10. The circuit is primarily a **parallel system** (i.e., either switch A or *both* switches B and C must function if the current is to flow from left to right). The branch containing switches B and C, however, forms a **series system**. To compute the probability that a closed circuit is made between the left and right sides of the diagram, we must find the probability of the event $\{A \text{ or } \{B \text{ and } C\}\}$. Assuming that the switches function independently of one another and that they are closed with probabilities $P(A) = .80$, $P(B) = .70$, and $P(C) = .90$, we proceed as follows:

$$P(A \text{ or } (B \text{ and } C)) = P(A) + P(B \text{ and } C) - P(A \text{ and } (B \text{ and } C)) \leftarrow$$

The general addition rule applied to the events A and $\{B \text{ and } C\}$

$$= P(A) + P(B)P(C) - P(A)P(B)P(C) \leftarrow$$

Since A, B, and C are independent

$$= .80 + (.70)(.90) - (.80)(.70)(.90) = .926$$

Thus the circuit is closed about 92.6% of the time. Since switch A is closed 80% of the time, the probability that the circuit is closed must certainly exceed 80%, so our answer makes sense.

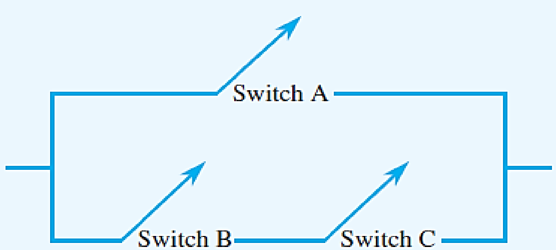
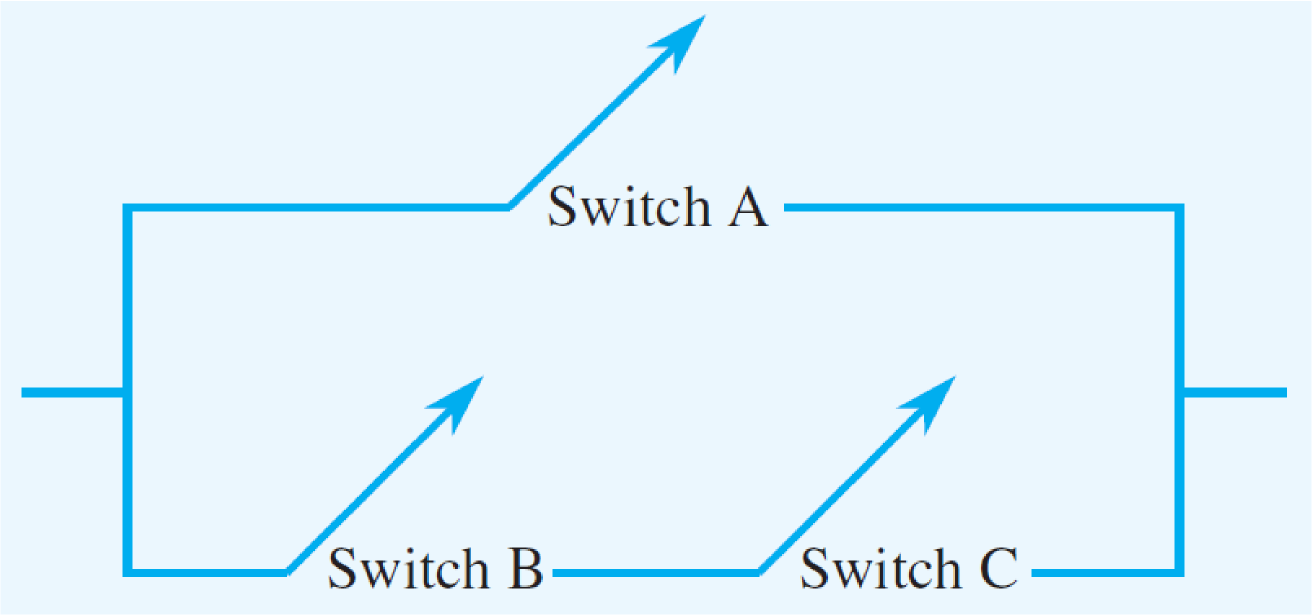



圖 5.10　有著三開關都是開的串並聯電路。

# Combining Several Concepts

-單一個A的系統閉合機率為0.8

-因此，全部閉合的機率一定會超過0.8




$$P(A\,or\,(B\,and\,C)) = P(A) + P(B\,and\,C) - P(A\,and\,(B\,and\,C))$$
$$= P(A) + P(B)P(C) - P(A)P(B)P(C)$$
$$= .80 + (.70)(.90) - (.80)(.70)(.90) = .926$$

# 組合數(combination)-choose

-Use the choose(r, n):

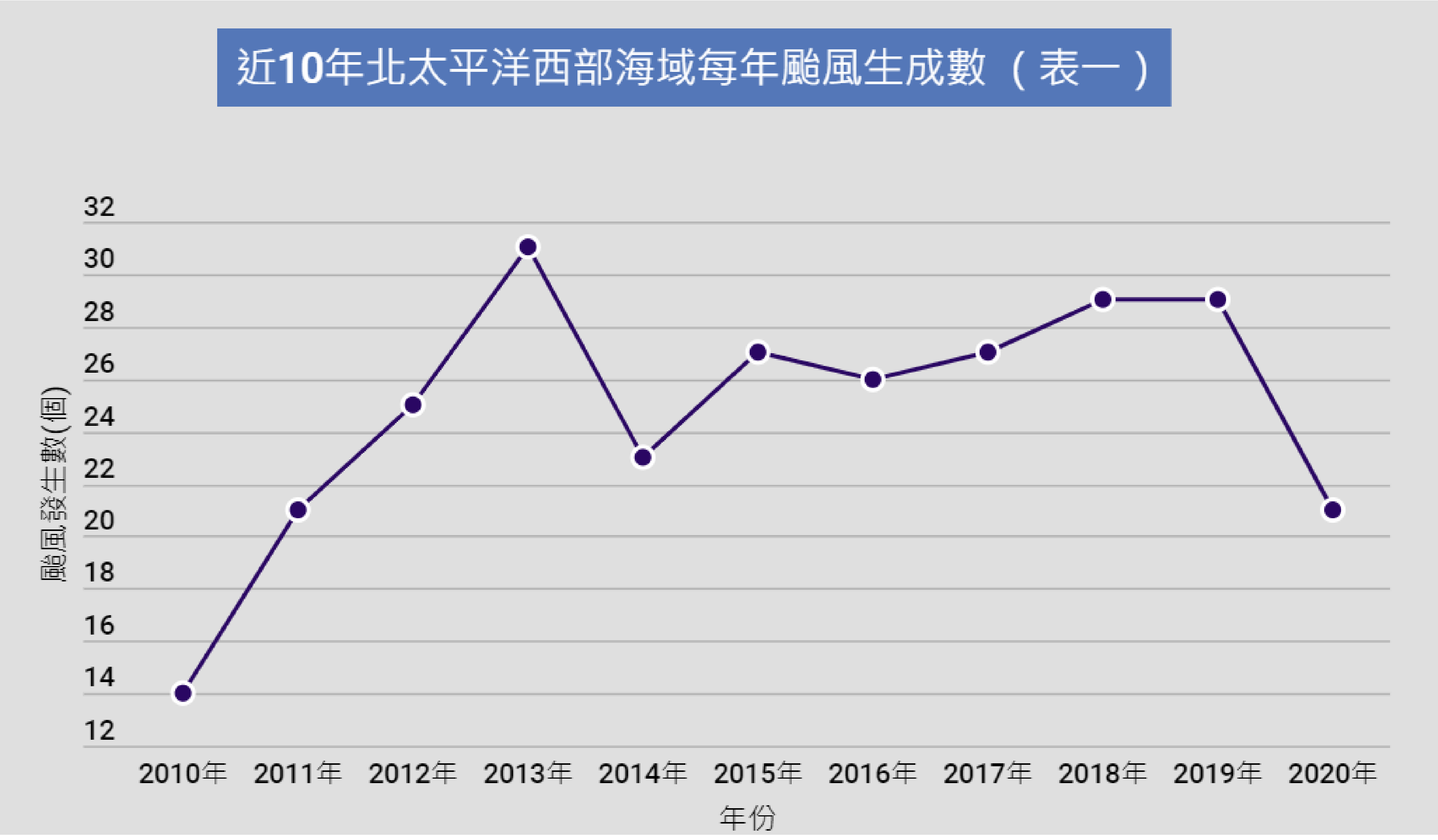
$$C_n^r = \frac{A_n^r}{r!} = \frac{n!}{r!\,(n-r)!}$$

# 課堂練習: 學號-姓名-ch8-Probability.R

https://topic.udn.com/event/typhoondelay2020

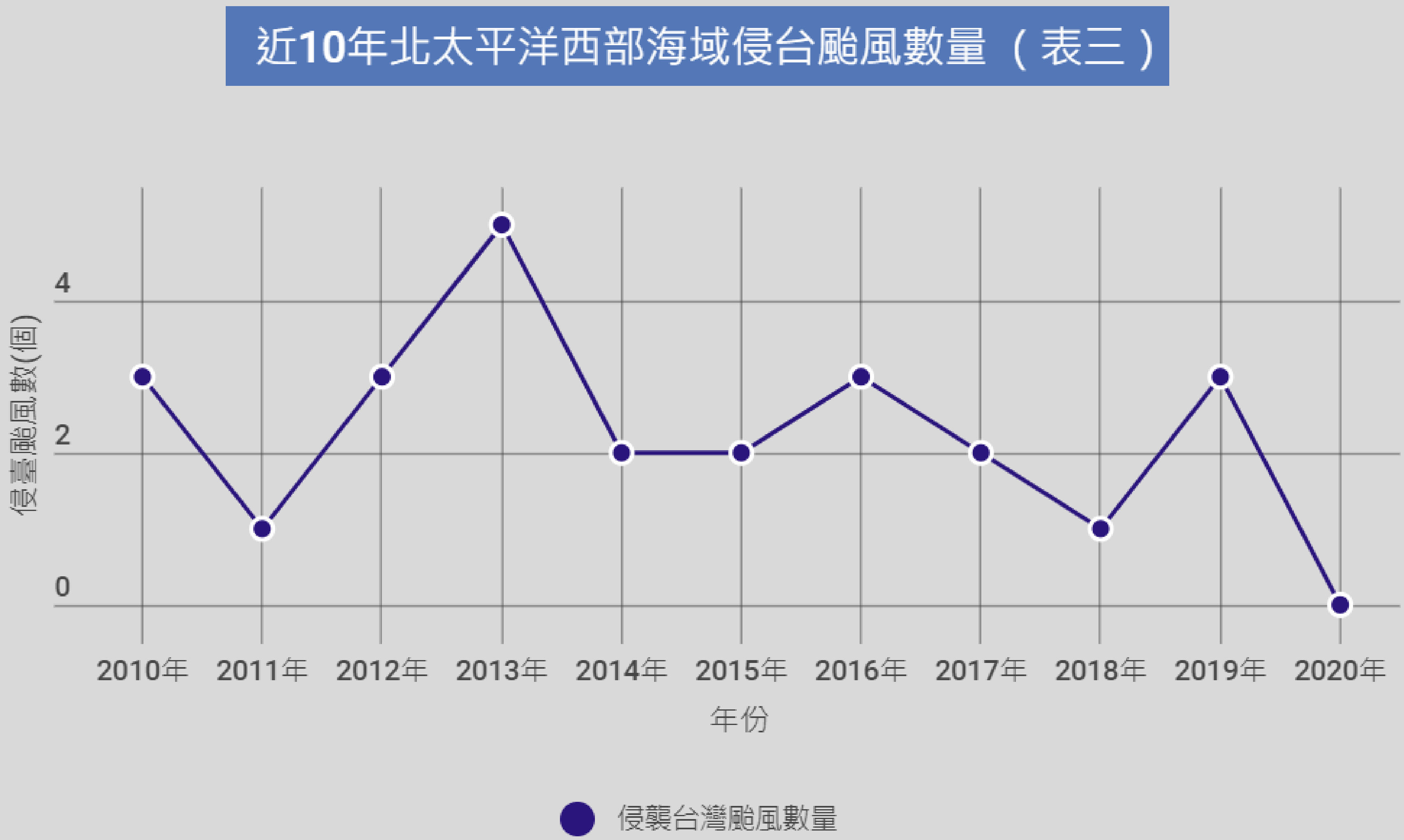
# 課堂練習: 學號-姓名-ch8-Probability.R

https://topic.udn.com/event/typhoondelay2020



# 課堂練習: 學號-姓名-ch8-Probability.R

https://topic.udn.com/event/typhoondelay2020




資料來源：氣象局有發警報颱風列表、2019年北太平洋西部海域颱風之氣候分析。（表三）時間統計截至2020年11月9日。

# 課堂練習: 學號-姓名-ch8-Probability.R

程式模擬算機率: 颱風侵台問題

**若颱風形成個數為20，每一個颱風侵台的機率是0.1且每個颱風侵台是互相獨立的**

**觀念(1)**: 分配隨機數字來代表侵台與非侵台
用雙位數來代表颱風
00,01,02,03,...,09 = 侵台颱風
10,11,12,13,...,99 = 非侵台颱風
**觀念(2)**: 模擬颱風侵台
在20次抽數字中(模擬20個颱風形成, sample(0:99,1)，有n.strike次抽到00~09的數字，即代表有n.strike個颱風侵台。上述的動作重複n次，即可模擬颱風侵台機率問題

# 課堂練習: 學號-姓名-ch8-Probability.R

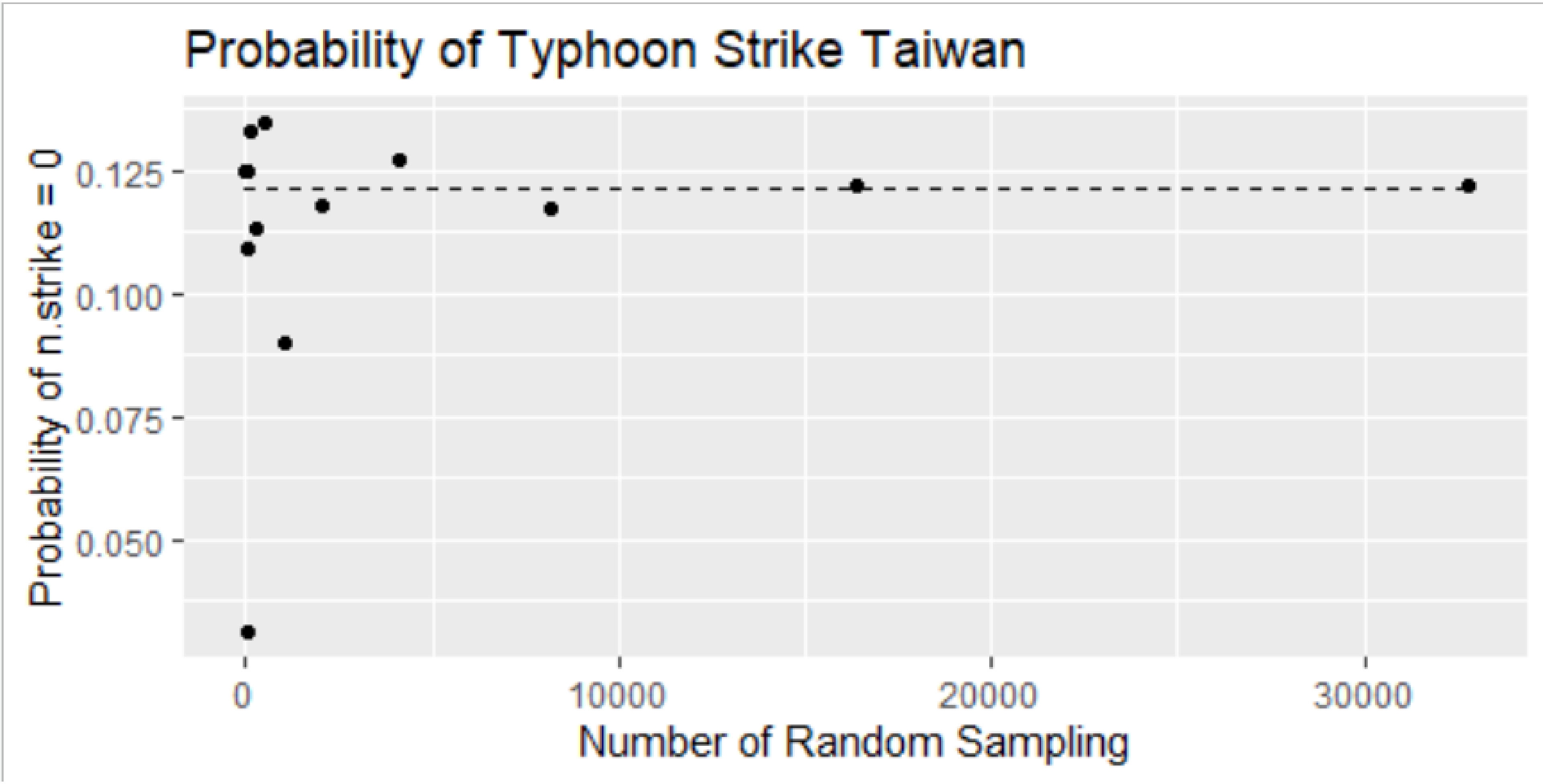
試著回答以下問題:

(1) 若該年度颱風形成數量為20個(n.typhoon)，則無任一颱風侵台的機率為多少? (數學計算，理論值)

(2) 請試著透過R程式設計模擬颱風侵台機率問題，重複8次(n=8)的抽數字(模擬20個颱風形成，單次抽取20個數字)，並計算完全沒有抽到00~09之間數字(代表為非侵台颱風)的次數m，則此次模擬無颱風侵台的機率為多少? (m/8) (注意若要每次計算相同結果，需使用set.seed)

(3) 上述模擬颱風侵台重複次數增加為(n=)$2^4,2^5,2^6,2^7,\ldots,2^{15}$ 請繪製模擬無颱風侵台機率隨著不同重複次數的變化，並觀察及描述機率變化的現象

# 課堂練習: 學號-姓名-ch8-Probability.R



# 課堂練習: 學號-姓名-ch8-Probability.R

Hint:

```
# (2) repeated random sampling n = 8
set.seed(1)
n <- 8
n.typhoon <- 20
m <- 0
pp <- 0
for (i in 1:n){
  n.strike <- 0
  for (j in 1:n.typhoon){
    num <- sample(0:99,1)
    if (num <= 09) {
      n.strike <- n.strike + 1
    }
  }
  if (n.strike == 0) m <- m + 1
}

cat(m,'of',n,'are results of number of non-strike typhoon','\n')
pp <- m / n
cat(pp,' is computed probability',' Sample number is ', n,'\n')
```