True/False: （每題 2 分）

1. The main hardware components of a personal computer are CPU, Motherboard, Memory, and OS.

2. Bus, Ring, and Star are three basic LAN topologies.

3. CSMA/CD is used in the wireless communication protocol, like IEEE 802.11.

4. One of the major differences between ODBC and JDBC is that JDBC is based on XML technology.

5. HTTP/TCP/IP is the one and only one protocol stack used in the Internet.

6. If a CPU can execute 1 million instructions per second, we call the speed of the CPU is 1 RISC.

7. The difference between RAM and ROM is that ROM is used in the CD-ROM while the data in RAM is unchangeable.

8. Electronic mail and file transfer are typical functions of application layer in OSI model.

9. RIP and IP are routed protocol, and IPX and BGP are routing protocols.

10. A machine cycle consists of two segments of time: Instruction time and Execution time. ALU is used in the Instruction time.

11. Switches can be used for removing the broadcast storm in a LAN.

12. To allocate a record entry from an index tree of a database, it takes $O(\log_2 N)$, assuming that there are N records.

13. An $m \times m$ matrix denotes the social relationship between m persons. John is one of these persons. We can find John's friends from this matrix. Persons are in the same social circle if they all know each other. For example, John knows A, B, and C three persons. John, A, B, and C are in the same social circle because they all know each other. To find John's social circle from this matrix takes $O(m)$.

14. For a program to be executed across different platforms, re-compiling the program to execution codes is necessary.

15. The "quick sort" sorting algorithm estimated by the big O measure to rank n objects, at the best case, takes $O(n\log_2 n)$.

Single Choice: (每題 2 分)

1. Which one of the following tool or language is not suitable for design a CGI program in the Web application? (a) C, (b) PHP, (c) SDL, (d) ASP.

2. In developing a multimedia application, which of the following terms is not essential? (a) stream synchronization, (b) buffer management, (c) user interface, (d) code division multiple access.

3. Which of the following sorting methods is not based on "comparison" technique? (a) Quick sort, (b) Merge sort, (c) Radix sort, (d) Insertion sort.

4. In transmission of data, if the uplink and downlink data rates are identical, then the transmission is symmetric; otherwise, the transmission is asymmetric. Which of the following whose transmission must be symmetric? (a) GSM, (b) ATM AAL5, (c) Cable Modem, (d) none of the above.

5. Which of the following operations is not included in the SQL? (a) Insert, (b) Create, (c) Select, (d) Map.

6. The result of 10011011 XOR 10110110 is (a) 10111111, (b) 10010000, (c) 00101101, (d) none of the above.

7. Which of the following I/O device is not "interrupt" based? (a) Keyboard, (b) Monitor Display, (c) Ethernet LAN card, (d) none of the above.

8. Which of the following is true? (a) Computer virus can replicate itself. (b) A PC will not infect computer virus if it is standalone without connecting to any networks. (c) Embedded systems will not infect virus. (d) None of the above.

9. Which type(s) of cabling is used for 10BASE-T? (a) fiber optical and unshielded twisted pair, (b) fiber optical and coaxial, (c) twisted pair, (d) coaxial.

10. In RIP, the maximum allowable hop count is (a) 15, (b) 31, (c) 63, (d) 127.

11. How many host addresses are available to a Class C network? (a) 254, (b) 256, (c)128, (d) 524.

12. What is the maximum number of bits that can be borrowed to create subnets in a Class C network? (a) 2, (b) 4, (c) 6, (d) 8.

13. If three bits are borrowed from host field of a Class C network, what will the range of usable hosts be in the first subnet? (a) .32 through .63 (b) .0 through .32 (c) .33 through .62 (d) .1 through .31.

14. If you fail to link to the web page after typing the URL in a web browser, what

would be the problems? Notably, you can successfully ping the web site at this moment. (a) the network is down. (b) the web server of the web site is down. (c) the router fails. (d) all of the above are possible.

15. If you have the ADSL (256/64kbps) connection at home and you want to download a file (1MByte) from a remote Web site, how long will it take? (a) 16 seconds, (b) 32 seconds, (c) 64 seconds, (d) 128 seconds.

Programming and/or Calculation Problems:

1. We can sort a given set of n numbers by first building a binary search tree containing these numbers and then printing the numbers by an in-order tree walk.

   (a) Please give an example by drawing a binary search tree for the following set: {3, 12, 4, 6, 20, 45, 11}. (5%)

   (b) What are the worst-case and best-case running times for this sorting algorithm, when a set of n numbers is given. (5%)

2. To break in a computer system, a person needs to "guess" the password for a given user name. Suppose the "hacker" knows that (1) there exists a user name: **root** in the system that he tries to break in, (2) the maximum length of the password is 10 characters with the possible character set ranging from a to z, and (3) the login shell will disconnect and reset upon 3 unsuccessful trials. Assume that (1) it takes 1 second to key in a guessing password (2) it takes 10 seconds for the login shell to disconnect and reset, and (3) all other waiting times are negligible. Please calculate, in the worst case, how many *years* does he need to take to break in the system if the brutal force approach is taken. (Please show the formula without solving it.) (10%)

3. Compare and contrast Interior and Exterior routing protocols (5%).

4. Compare and contrast the OSI and TCP/IP models for data communications (5%).

5. Given a string of English letters, please write pseudocodes to find the longest substrings which occur equal to or more than three times in the given string. For example, in the string, S={ACDEFGACDEKJGSACDEWKK}, the output we will obtain from your pseduocodes is {ACDE} because this substring appears in S three times. (10%)
   Input: ACDEFGACDEKJGSACDEWKK
   Output: ACDE
   Process: (write your codes starting from here) (※答案請寫在答案紙上)

簡答題:(每題 5 分,請以 50 個字(含)以內回答下列各題,超過 50 個字以上,該題不給分)

1. What is the difference between circuit switching and packet switching?

2. What is 2-phase locking protocol in a database management system?

3. What is flow control in a computer network?

4. 請簡述決策支援系統和知識管理在本質上的差異。

5. 請簡述系統外包的優缺點。

6. 何謂 Intranet?

問答題:

7. 在過去的時代,當我們想依自己需要訂製一個產品時,常常要付較高的價格、等較長的時間才可獲取,在現在資訊科技發達的時代,透過資訊整合,廠商常可以較低的價格,迅速地提供給顧客客製化的產品,請問必須做到哪些資訊整合才能達到?為什麼?(20 分)

8. 請依據 Michael Porter 所提的五力分析模式說明電子商務對企業經營所帶來的衝擊。(15 分)

9. Data quality is essential to an information system. However, quality data in operational databases can hardly be ensured. For example, according to a Federal Trade Commission study of 17,000 items in retail databases, 5% of product prices were incorrect. As a result, wrong price data in retail databases may cost American consumers as much as several millions in overcharges annually.

Besides data entry errors, please describe two possible sources for such low-quality data in an information system. For each source, please propose a solution to improve the data quality in the information system. (15 分)

10. When you transform a specialization relationship in the ER diagram into relations in the relational data model, you can choose one from the following three options:

Option A) a corresponding relation for each node in the specialization hierarchy
Option B) a corresponding relation for each leaf node in the specialization hierarchy
Option C) a corresponding relation for the superclass with one or more "type" attribute

For example, students can be specialized into undergraduate students or graduate students (as shown in Figure 1).
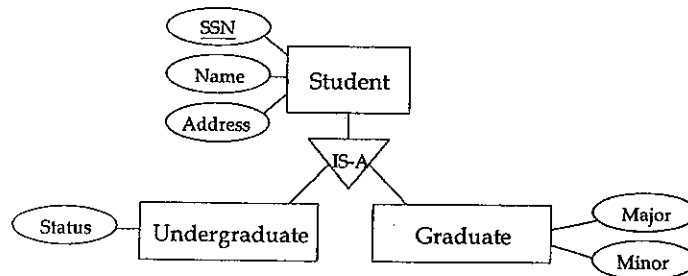

Figure 1: ER Diagram

According to Option A), three relations are resulted:
Student(SSN, Name, Address)
Undergraduate(SSN, Status)
Graduate(SSN, Major, Minor)
   where SSN in Undergraduate or Graduate is a foreign key that references the Student relation.
According to Option B), two relations are resulted:
Undergraduate(SSN, Name, Address, Status)
Graduate(SSN, Name, Address, Major, Minor)
Finally, according to Option C), only one relation is derived:
Student(SSN, Name, Address, Type, Status, Major, Minor)
   where Type is either "U" for undergraduate or "G" for graduate students.

What factors need to be considered when you decide which option to use for transforming a specialization relationship in an ER model into the relational model? Please elaborate the advantages and disadvantages of each transformation option. (20 分)
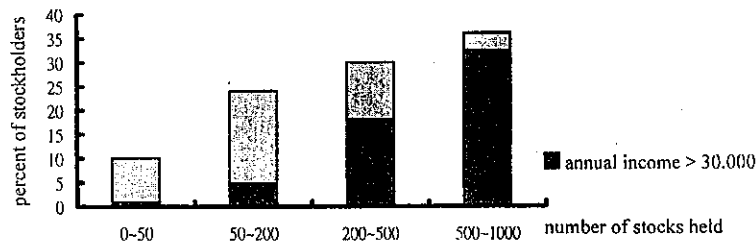
1. 複選題 (choose the correct statement(s)) [10]
   (a) A company's stockholders are classified into four classes according to their number of stocks held: 0 to under 50, 50 to under 200, 200 to under 500, and 500 to 1000. The percentage of each class is shown in the following graph. Furthermore, the shaded area in each bar refers to those stockholders whose annual income is above $30,000.

① The modal class in terms of a histogram is the number of stocks held from 500 to 1000.
② The total percentage of stockholders with annual income > 30,000 is above 50%.
③ The median number of stocks held by all stockholders is less than that held by those stockholders whose annual income > $30,000.
④ Probability that the company's stockholder, given his annual income > $30,000, holds 200 to 500 stocks is around 0.3.
⑤ Probability that the company's stockholder, given his annual income ≤ $30,000, holds 50 to 200 stocks is around 0.3.

2. The policy of a particular bank branch is to have its ATM stocked with enough cash to for customers' withdrawing cash over the weekend. It is expected that the average amount of money withdrawn per customer transaction over the weekend is $160 with an expected standard deviation of $30. However, this branch gets more and more complaints about running out of cash in the ATM over the weekend. An employee who takes charge of justifying these complaints selects a random sample of 36 customer transactions during weekends. Assume he performs this procedure several times. The level of significance is set to be 0.05.
   (1) State the hypotheses. [2]
   (2) Assume that among the tests, the rejection rate of the null hypothesis is just 50%. Which hypothesis do you think the employee should believe? Why? [5]
   (3) Assume that among the tests, the rejection rate of the null hypothesis is 90%. Based on this information, estimate the average amount of money withdrawn per customer transaction over the weekend. [5]
   (4) Continued from (3). Another approach to estimate the average amount is to take the average of the sample means but excluding those falling within the acceptance region. Is this approach appropriate? Why? [5]
   (5) What other information does the employee need to collect in order to justify the complaints? [3]

3. Trend analysis is important for a company to recognize costumers' purchasing trend and products' selling trend over time. For example, we may find certain customers have similar consuming patterns or certain products have similar selling patterns. Or we may be interested in finding whether a certain product keeps similar trend compared to its historical selling patterns.

(1) Assume a company sells five products of A, B, C, D, E. Their sales amount in four quarters of year 2000 is shown in the following table (in units of thousand dollars). Which product has the most similar trend to product A? Test whether they are really similar in the statistical sense. (α=0.05) [6]

	Spring	Summer	Fall	Winter
A	10	52	28	9
B	3	48	23	17
C	22	35	43	50
D	4	27	13	6
E	23	25	27	24

(2) Assume product A's sales amount in four quarters of the past three years is shown in the following table (in units of thousand dollars). Do you believe that the product has similar selling trend in the past three years? (α=0.05) [7]

	Spring	Summer	Fall	Winter
1998	6	23	16	4
1999	9	35	21	12
2000	10	52	28	9

(3) If it is desired to predict the sales amount of all quarters in this year based on the table given in (2), one possible approach is to predict the total sales amount this year based on regression. Proportions of the sales amount in each quarter are then estimated. The predicted sales amount of all quarters is thus determined by multiplying the predicted total sales amount with their estimated proportions. Find the predicted sales amount of all quarters this year based on this approach. [7]

4. As part of a marketing program, a city block containing 4 households was selected and a sample of 2 households was sampled as follows: John, the research assistant identified the households and numbered each household from 1 to 4. He then was to list all combinations of the 4 households 2 at a time. These are:

1 and 2	1 and 3	1 and 4
2 and 3	2 and 4	3 and 4

Unfortunately, John forgot the combination 3 and 4. He chose a random number between 1 and 5 (it turned out to be 2) which corresponded to the combination 1 and 3. Thus households 1 and 3 were sampled. The variable of interest was out of pocket medical expenses incurred by the household. These were as follows for the 4 households

Household	Expenses (NT)
1	10,300
2	4,500
3	18,200
4	32,500

(1) Based on John's sampling procedure, what are the mean, standard deviation, and standard error of the estimated mean out of pocket medical expenses? [6]

(2) Is the estimate derived from this sampling procedure unbiased? [6]

(3) Does every household have the same chance of appearing in the sample? Why or why not? [6]

5. The following data is a part of records from the physical examination for undergraduate students in the beginning of a spring semester. Now the researchers would like to explore the possible association relationships among those variables. Please write down two problem statements associated with hypotheses, such that the researchers can follow your statements to solve the questions according to their understanding of statistics. (Hint: Two possible statistic analyses can be ANOVA and Comparison of two samples) [16]

Students	Grade	Sex	Weight (kg)	Students	Grade	Sex	Weight (kg)
00101	So	F	52	00108	F	F	48
00102	So	F	61	00109	So	F	57
00103	F	M	70	001010	J	M	72
00104	J	F	53	001011	J	M	62
00105	Se	M	59	001012	F	F	65
00106	Se	M	68	001013	J	M	80
00107	F	M	80	001014	F	F	51

Grade: F, Freshman; So, Sophomore; J, Junior; Se, Senior

Sex: F, Female; M, Male

6. Describe the relationship between the two variables in the following four plots. [每小題 4 分，共 16 分]
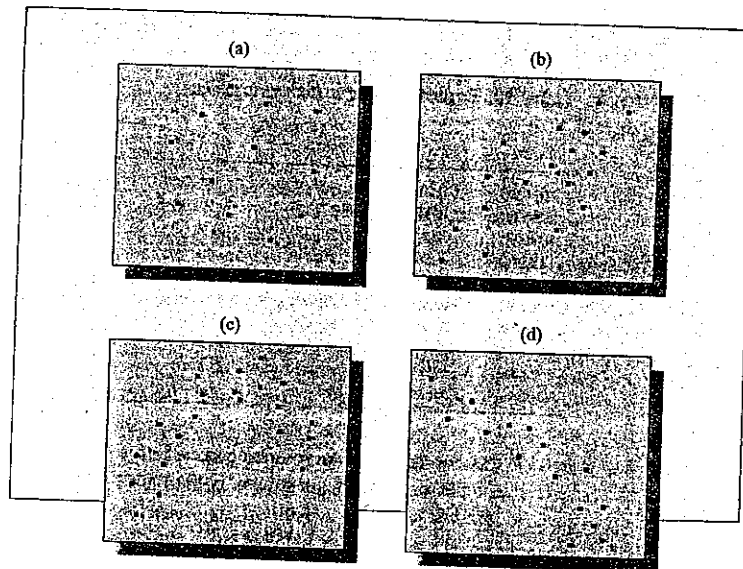
α=0.05	Degree of freedom	2	3	4	5	6	7	8
	t distribution	2.92	2.353	2.132	2.015	1.943	1.895	1.860
	χ^2 distribution	5.991	7.815	9.488	11.07	12.59	14.07	15.51

1. (5%) Let A={a, b, c, d} and let R be a relation in A defined by R={(a, a), (b, a), (a, c), (c, b), (d, a), (d, d), (c, c), (c, d)}, Find R^+ and R^*.

2. (5%) Construct a transition-system recognizer for the following regular expression

   10*101*1

3. (5%) Find the disjunction normal form of the following function using algebraic techniques,

   $$f(x, y, z) = \overline{(\overline{xy} \vee \overline{xz})(x \vee \overline{yz})}$$

4. (5%) 由文法(T, N, S, P)所標示的語言 L, 其中

   T={a, b, c} terminal

   N={S, A, B} non-terminal

   S start

   P={S->AB, A->ab, A->aAb, B->c, B->Bc} Production Rule.

   試判斷下列各句子是否屬於此語言.

   aabb

   aaabbc

   aaabbbccc

   ababcc

5. (10%) Find the number of different ways to three color the vertices of a full binary tree with seven vertices (do not distinguish left from right, i.e., you should eliminate the duplicate combinations).

6. (10%) 試證明 $1^3 + 2^3 + \ldots + n^3 = (1 + 2 + \ldots + n)^2$.

7. (10%) Let G be a bipartite graph, E denotes the set of edges in G, and W denote the set of edges in a matching from X to Y. A vertex (in X or Y) is said to be unsaturated if no edge incident with it is included in W. A path is said to be an alternate path if it contains edges in W and $E - W$ alternately. The *Hungarian* method claims that a matching is maximal if and only if there is no alternate path connecting two distinct unsaturated vertices. Show that a matching is not maximal if there exists such a path by showing that for each such path, the number of vertices in X matched into Y can be increased by 1.

8. Classes Point 和 Square 定義如下。 其中 Square 繼承 Point。
(8.1)(5%) 請列出 Square 的所有可能資料。
(8.2) (5%) 請列出 Point 的所有可能資料。
(8.3) (5%) 依據 (8.1) and (8.2) 說明為何 Square 之資料數目比 Point 之資料數
目多。但 Square 之資料結構卻是 Point 之資料結構的 subtype。

Class Point {
　　Public：
　　　Enum Weekend {Saturday， Sunday}；
　　Private：
　　　Weekend　a；
}；
Class Square ： Public Point {
　　Public：
　　　Enum Card {Spade， Heart， Diamond， Club}；
　　Private：
　　　Card　b；
}；

9.求下列實變數的實數函數的定義域
(9.1)(5%) f(x) $= 1 / (x - 3)$
(9.2)(5%) f(x) $= \log(x + 3)$

10.系統架構圖 (SAD) 說明一個 module 是 經由那些 modules 組成的。下圖 B
有兩個箭頭分別指向 D 和 E 表示 module B 是 由 D 和 E 兩個 modules 組
成的。
(10.1)(5%)為何下圖不是一個正確的 SAD？
(10.2)(5%)描述一群 modules 之間的關係 if 它們可以形成一個正確的 SAD。
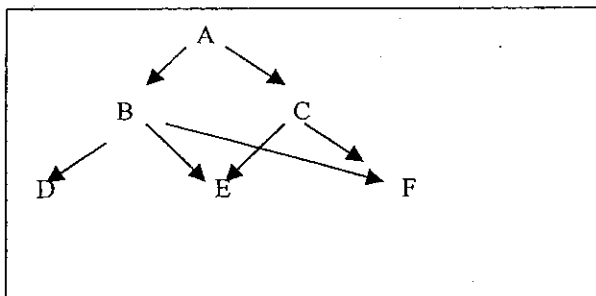
11.(15%) The mathematical notation, C^n_k, read "n choose k", is used to denote the number of distinct subsets of k items that can be chosen from a set of n distinct objects. It can be shown that:

$$C^n_k = n! / ((n-k)! * k!)$$

Although the factorial function could be used in implementing n choose k, this would be inefficient for large values of n and small k. We can algebraically rewrite the definition into the following form for k > 0:

$$C^n_k = (n* (n-1)*...*(n-k+1)) / (1*2*...* k)$$

Prove that the result of the above function is always an integer.

1. (25%)

A subsequence of a given sequence is just the given sequence with some elements left out. For example, Z=(B, C, D, B) is a subsequence of X=(A, B, C, B, D, A, B). Now consider the problem of finding the longest common subsequence (LCS) of two sequences. For example, a longest common sequence of X=(A, B, C, B, D, A, B) and Y=(B, D, C, A, A, B) is (B, C, A, B). We have the following observations:

Let $X=(x_1, x_2, …, x_m)$ and $Y=(y_1, y_2, …, y_n)$ be sequences, and let $Z=(z_1, z_2, …, z_k)$ be any LCS of X and Y.

1. If $x_m=y_n$, then $z_k=x_m=y_n$ and $(z_1, z_2, …, z_{k-1})$ is an LCS of $(x_1, x_2, …, x_{m-1})$ and $(y_1, y_2, …, y_{n-1})$.

2. If $x_m \neq y_n$, then $z_k \neq x_m$ implies that Z is an LCS of $(x_1, x_2, …, x_{m-1})$ and Y.

3. If $x_m \neq y_n$, then $z_k \neq y_n$ implies that Z is an LCS of X and $(y_1, y_2, …, y_{n-1})$.

Based on these observations, we would like to compute the length of an LCS of two sequences X and Y. Specifically, let's define $c[i, j]$ as the length of an LCS of the sequences $(x_1, x_2, …, x_i)$ and $(y_1, y_2, …, y_j)$. Please do the following:

1. Justify the above observations by showing examples. (10%)

2. Write the recursive function of $c[i, j]$ based on these observations. (10%)

3. Write a recursive algorithm for computing $c[i, j]$. (5%)

2. （25%）

假設你想去 model 某一族群的人際網路，於是你選擇一些關係較密切的大學生（比如同班或同社團）為族群，並對每一個人詢問其在該族群裡所認為的好朋友，但請注意好朋友是一個非對稱的關係（asymmetric relation），也就是說甲認為乙是她的好朋友並不代表乙也會認為甲是她的好朋友。有了這些資料後，可以用 directed graph 的方式來表達好朋友的關係。假設我們對死黨感興趣，死黨的定義為一群人，其中任意兩人都認為對方是她的好朋友。

1. 請正式描述（formally define）此人際關係的 directed graph。（8%）

2. 假設你想找出某一特定人（a given person）所屬的最大死黨，請正式描述出其在 directed graph 上的問題。(7%)

3. 請寫出一演算法來找出某一特定人(a given person)所屬的最大死黨。（只要寫出即可，不要求高效率）(10%)

3. 解釋下列名詞的意義：盡可能清楚而明確的敘述(30%)
    (a) SAN: Storage Area Network
    (b) RPC: Remote Procedure Call
    (c) DDOS: Distributed Denial Of Service
    (d) NFS: Network File System
    (e) RAID: Redundant Arrays of Inexpensive Disks
    (f) Atomic Transaction
    (g) Virtual Machine
    (h) Process Migration
    (i) Open Source
    (j) Race Condition

4. 某一電腦硬體廠商要開發一款最新的 PDA（掌上型電腦），需要外包軟體公司來開發一個適用於該款 PDA 的作業系統。請你幫這家硬體廠商開一個系統委外需求規格(specification)書。請特別注重在 PDA OS 的相關特性，例如 Kernel 要小, 通訊能力要考慮有線及無線等，還有 memory management, user interface，‧‧‧等(20%)