NATIONAL UNIVERSITY OF SINGAPORE

GER1000 QUANTITATIVE REASONING

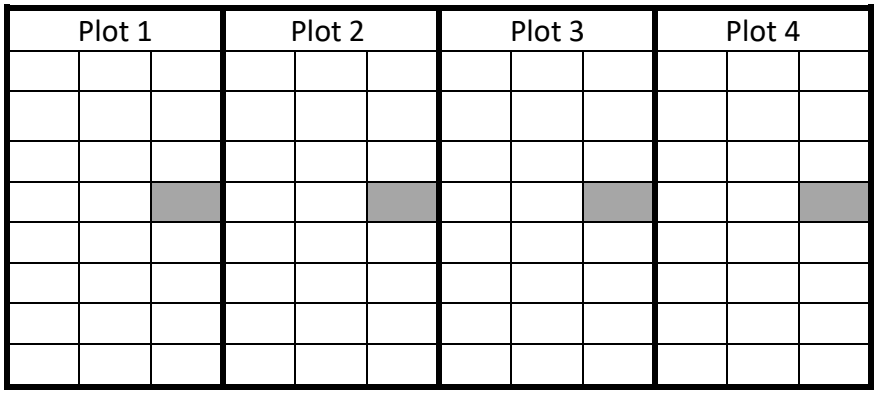
Practice Paper

1. A gourmet food magazine wants to know how its readers feel about serving beer with various types of food. Note that not all readers are subscribers of the magazine. The magazine sends surveys to 1,000 randomly selected names from its list of subscribers. Which of the difficulties in sampling is the magazine most likely to face?

   (i) Using the wrong sampling frame.
   (ii) A low response rate.

   (A) (i) only
   (B) (ii) only
   (C) (i) and (ii)
   (D) None of the above

2. Suppose a large piece of land is divided into several plots. Each plot is further divided into several smaller subplots. A sample of subplots is to be selected in the following way: (a) A subplot is randomly selected from one particular plot; (b) For the remaining plots the subplot that is in the same position as the selected subplot in (a) is selected. For example, in the diagram below, say the shaded subplot in Plot 1 is randomly selected, then the shaded subplots in the remaining plots are also selected.



   This sampling scheme is called

   (A) Simple random sampling
   (B) Stratified sampling
   (C) Systematic sampling
   (D) None of the above

3. Which of the following are advantages of a case-control study over a randomized experiment?

   (i) It saves time and money.
   (ii) It makes establishing causal relationship between the exposure and response variables easier.
   (iii) It presents fewer ethical issues.

   (A) (i) and (ii) are correct.
   (B) (i) and (iii) are correct.
   (C) (ii) and (iii) are correct.
   (D) All are correct.

4. Without random sampling, which of the following can happen?

   (i) The results may not be extend to a larger population.
   (ii) The researcher will have a much easier time getting participants for the study, resulting in a larger sample size, and more accurate data.

   (A) (i) only
   (B) (ii) only
   (C) (i) and (ii)
   (D) None of the above

5. The mid-term test of a certain module has a high (30%) failure rate. The instructor of the module conducted a few special sessions for the entire class after the test. He then randomly selected ten students to sit for another test (with similar scope and level of difficulty). Nine of these students passed the second test. To determine whether the special sessions are effective (in helping students to pass the test), the instructor carried out a hypothesis test. Which of the following is correct?

   (A) At 5% level of statistical significance, the instructor is certain that the special sessions will reduce the failure rate.
   (B) At 5% level of statistical significance, the instructor is certain that the special sessions will not reduce the failure rate.
   (C) The probability that the null hypothesis is true is about 0.15.
   (D) None of the above.

6. A certain infectious disease is spread only by direct contact. Suppose the probability for a healthy person to get infected by a single direct contact is 0.02. What is the chance that a healthy person gets infected with the disease by the fourth time he comes into contact with an infected person? (You may assume the outcome of coming into contact with an infected person is independent from each other.)

   (A) 0.0776
   (B) 0.0188
   (C) 0.08
   (D) None of the above.

7. An insurance company sells a certain health insurance policy that has a maximum payout of $30,000 when a customer makes a claim. Suppose that, each year, 2% of the customers receive the full claim of $30,000, 3% receive a claim of $20,000, not more than 5% receive a claim of $10,000, while the rest of the customers do not submit a claim.

   In order not to make a loss for sure, the minimum amount that the insurance company should charge per year for this policy is

   (A) below $1,500.
   (B) between $1,500 and $1,800.
   (C) between $1,800 and $2,000.
   (D) more than $2,000.

8. Among 100,000 women with negative mammograms (screening for breast diseases), 20 will be diagnosed with breast cancer in 2 years, whereas 1 woman in 10 with positive mammograms will be diagnosed with breast cancer in 2 years. Suppose that 10% of the general population of women will have a positive mammogram. What is the probability that a woman who develops breast cancer over the next 2 years has a negative mammogram?

   (A) 0.0177
   (B) 0.0196
   (C) 0.00018
   (D) 0.0002

$P(D|-) = 0.0002$

$P(D|+) = 0.1$

$P(+) = 0.1$

D	−D
+1000	9000
−18	90000

10000

$P(-|D) = \frac{18}{1018}$

9. For any two events A and B with both P(A) and P(B) > 0, the conditional probability P(B|A)

   (A) is greater than P(B) if A and B are independent events.

   (B) is equal to P(B) if A and B are mutually exclusive events.

   (C) can be greater than or less than P(B).

   (D) None of the above.

10. The organizer of a running event says that the run will be cancelled in the event that either (i) it rains on the run day; or (ii) the response for the run is poor.

    Consider the following statements:

    (I) The probability that the run is cancelled is the sum of the probability that it rains on the run day and the probability that the response for the run is poor.

    (II) The probability that it rains on the run day and the run is cancelled is equal to the probability that it rains on the run day.

    (A) Both (I) and (II) are correct.

    (B) Only (I) is correct.

    (C) Only (II) is correct.

    (D) Both (I) and (II) are incorrect.

11. Suppose u and v are two adjacent vertices in a network whose order is 50. If the degree centrality measures of u and v are 1/49 and 2/49 respectively, which of the following statements is true?

    (A) *Bcen*(v) is twice *Bcen*(u).

    (B) *Bcen*(u) is twice *Bcen*(v).

    (C) *Ccen*(v) will always be greater than *Ccen*(u).

    (D) *Ccen*(u) will always be greater than *Ccen*(v).

12. A network has the following adjacency matrix:

	u	v_1	v_2	v_3	v_4	v_5
u	0	1	1	1	0	0
v_1	1	0	0	0	0	0
v_2	1	0	0	0	1	0
v_3	1	0	0	0	1	0
v_4	0	0	1	1	0	1
v_5	0	0	0	0	1	0

Using the definition of closeness centrality measure of a vertex u, namely,

$$Ccen(u) = \frac{\sum_{i=1}^{n-1} d(u, v_i)}{n-1},$$

what is the value of *Ccen*(u)?

(A) 9/5

(B) 2

(C) 11/5

(D) 8/5

13. Recall that in our discussion of the movie network in Chapter 5, two actors are adjacent in the graph if they have featured in a movie together. The Bacon number of an actor is the distance of this actor's vertex in the network from the Kevin Bacon vertex.

Suppose that the Bacon number of an actor A was initially computed to be 3. However, it was later discovered that there were two errors in the database. They are:

(i) There is a movie, featuring Kevin Bacon but not actor A, that was omitted from the database.

(ii) There is a movie which was initially thought to have featured actor A but actually did not feature him.

After correcting these errors, which of the following statements is correct?

(A) The Bacon number of actor A is now 1.

(B) The Bacon number of actor A now could be 2, 3 or bigger than 3.

(C) The Bacon number of actor A is now 2.

(D) The Bacon number of actor A is still 3.

For Questions 14 and 15, consider the following:

In addition to the three centrality measures introduced in the module, there are many other centrality measures proposed by researchers. For a vertex u in the network, let *Acen*(u) be one such other centrality measure. Suppose we have a network of order 500 and the centrality measures *Acen*(u), *Bcen*(u) (betweenness centrality measure) and *Ccen*(u) (closeness centrality measure) were computed for each vertex u.

14. While investigating the linear association between measures *Acen* and *Bcen* using a scatter plot, it was found that miraculously, the correlation coefficient *r* is equal to 1. If v is a vertex in the network with *Acen*(v) = 0.56, what can be said of the value *Bcen*(v)?

(A) It is smaller than 0.56.

(B) It is 0.56.

(C) It is larger than 0.56.

(D) It cannot be determined with the information given.

15. Suppose the 500 vertices in the network represent 300 teenagers (aged 16-20) and 200 adults (aged 21 and above). It was discovered that *Acen* and *Ccen* are negatively correlated among the teenagers but positively correlated among the adults.

When the data for teenagers and adults are combined, the correlation coefficient between *Acen* and *Ccen*

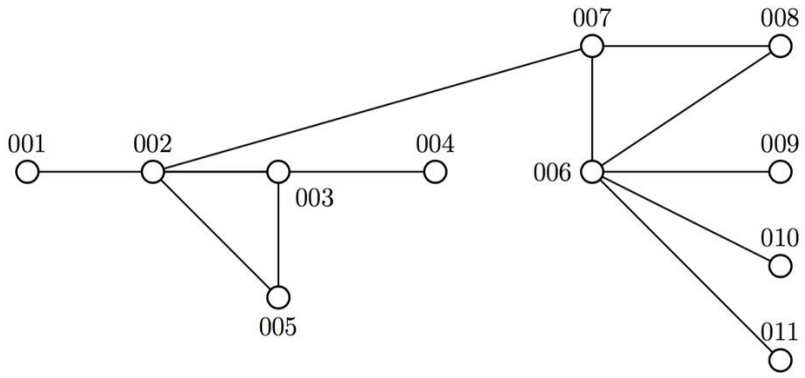
(A) cannot be determined with the information given.

(B) will be positive.

(C) will be negative.

(D) will be zero.

16. The following diagram represents a network of undercover agents that you are overseeing. Each vertex represents an individual and an edge between two vertices represents the existence of a direct communication channel between the two individuals. You may assume that in disseminating information to all the agents, you would always keep the number of communications between individuals to be as small as possible, in order to minimize the risk of exposing the agents' identities.



Your enemy intends to convert one of your agents in the network to working for him and the job of this converted spy is to intercept as many messages and as much information that is being passed around as possible. The degree, closeness and betweenness centrality measures for all the vertices, as defined in the module GER1000, are given in the table below:

Vertex	Dcen	Ccen	Bcen	**Vertex**	Dcen	Ccen	Bcen
001	0.1	2.8	0	**007**	0.3	1.8	0.56
002	0.4	1.9	0.6	**008**	0.2	2.3	0
003	0.3	2.5	0.2	**009**	0.1	2.9	0
004	0.1	3.4	0	**010**	0.1	2.9	0
005	0.2	2.6	0	**011**	0.1	2.9	0
006	0.5	2.0	0.53				

Which vertex would your enemy most likely try to convert?

(A) 006 since it has the largest degree centrality measure.

(B) 007 since it has the smallest closeness centrality measure.

(C) 002 since it has the largest betweenness centrality measure.

(D) 004 since it is in the centre of the network.

Some Definitions

- Let x and y be two vertices in a network (or graph). There can be several paths connecting x and y, including a path with the smallest number of edges. The **distance** between x and y, d(x,y), is defined as the number of edges in this path.

- Let u be a vertex in a network (or graph) of n vertices. The **degree centrality measure** of u is the number of edges attached to it, divided by n-1. The **closeness centrality measure** is the sum of the distances between u and all other vertices, divided by n-1.

[END OF PAPER]

Answers

1. C
2. C
3. B
4. A
5. D
6. A
7. B
8. A
9. C
10. C
11. D
12. D
13. B
14. D
15. A
16. C