**GER1000 2020 Sem 1**

**Quiz 8 and solutions**

**Section 1: Questions presented to students**

1. A standard deck of 52 playing cards comprises 4 suits (Clubs, Diamonds, Hearts, Spades), each suit with 13 cards of distinct ranks (A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K).

What is the probability that a card randomly selected from this deck is a ‘2’?

* 1. 1/4
  2. 1/13
  3. 1/52
  4. None of the above

Explanation: Since every card is equally likely, and there are four ‘2’s in the deck, one of each suit, the probability is 4/52 = 1/13.

1. A standard deck of 52 playing cards comprises 4 suits (Clubs, Diamonds, Hearts, Spades), each suit with 13 cards of distinct ranks (A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K).

What is the probability that a card randomly selected from this deck is a ‘2’, given that it is a ‘Spade’?

* 1. 1/4
  2. 1/13
  3. 1/52
  4. None of the above

Explanation: Since there are 13 Spades, which are equally likely, of which there is one ‘2’, the probability is 1/13. A more formal solution is as follows. Let A be the event that the card selected is a ‘2’, and B be the event that the card selected is a ‘Spade’. P(A | B) = P(A and B) / P(B) = (1/52) / (13/52) = 1/13.

1. A large-scale medical experiment enrolled 10000 male subjects and 10000 female subjects. Suppose the researchers were to randomly assign all 20000 subjects into two groups, X and Y, such that group X has 4000 subjects and group Y has 16000 subjects. Which of the following must be true?
2. The number of male subjects in group X is not equal to the number of male subjects in group Y.
3. The rate of male subjects in group X is exactly the same as the rate of male subjects in group Y.
   1. (I) only
   2. (II) only
   3. Both (I) and (II)
   4. Neither (I) nor (II)

Explanation: For the number of male subjects in the two groups to be equal, each group needs to have 5000 male subjects. However, group X has only 4000 subjects, so (I) must be true. On the other hand, even though random assignment of subjects into two large groups should result in similar rates of males in the groups, the rates may not be exactly the same.

1. Which of the following statements about confidence intervals is/are correct?
2. A confidence interval gives us information about the amount of random error present in our estimation of the population parameter.
3. A confidence interval gives us information about the amount of bias present in our estimation of the population parameter.
   1. (I) only
   2. (II) only
   3. Both (I) and (II)
   4. Neither (I) nor (II)

Explanation: See the last page of the lecture notes on ‘Sampling’.

1. Tom selects a child at random from a population of children. Let
   * A be the event a child of age < 3 is selected, and
   * B be the event a child of age < 5 is selected.

It is known that P(A) > 0. Which of the following must be true?

1. P(A or B) < P(A) + P(B)
2. P(A) ≤ P(B)
   1. (I) only
   2. (II) only
   3. Both (I) and (II)
   4. Neither (I) nor (II)

Explanation: Every child of age < 3 is also of age < 5, so B must occur if A occurs. This means P(A) ≤ P(B). It also implies A and B are far from being mutually exclusive. In fact, if we know one of A or B occurs, then we know B occurs, so P(A or B) = P(B). Checking whether (I) is true is then equivalent to checking whether the statement “P(B) < P(A) + P(B)” is true. Since P(A) > 0, “P(B) < P(A) + P(B)” must be true, and so must (I).

6. According to healthxchange.com, depression affects 6% of the aged population (above 65 years old) in Singapore. To find out the exact overall rate of depression in Singapore, what other information is needed?

a) Only the proportion of aged population out of the entire country (e.g., 1/4 of Singapore is aged).

b) Only the rate of depression for the non-aged population (65 years old and below).

c) Both the proportion of aged population in Singapore and the rate of depression of non-aged population.

d) No other information is needed. The rate of depression in Singapore is 6%.

*Explanation: Refer to the unit on Basic rule on rates. Let Rate(D) be the rate of depression. Let Rate(A) be the rate of aged population in Singapore and Rate(B) be the rate of non-aged population in Singapore. Rate(D) = Rate(A)\*Rate(D|A) + Rate(B)\*Rate(D|B).*

7. In Jolie’s new Thai restaurant, she observes the age of her customers and notes down which of the groups they belong to – ’18 years old and below’ or ‘above 18 years old’, and whether they spent more than $30 on a single meal - ‘big spender’ or ‘non-big spender’.

Given that **rate (big spender | above 18 years old) > rate (big spender | 18 years old and below),** what conclusion can be made?

a) Being a big spender and being above 18 years old are positively associated.

b) Being a big spender and being above 18 years old are negatively associated.

c) Being above 18 years old causes one to become a big spender.

d) Being above 18 years old is positively associated with visiting her restaurant.

*Explanation: Refer to Chapter 1. By definition, r(A|B) > r(A| not B) shows that A and B are positively associated.*

8. Jolie wishes to open a new Thai restaurant around Clementi and her marketing manager advises her to gather information on food prices from her potential competitors around her area. From Jolie’s data, she finds that the nearer the location of a restaurant is to an MRT station, the more expensive its signature dish would be.

Based on her dataset, which of the following best describes the relationship between the distance of a restaurant to an MRT station and price of its signature dish?

a) Positive correlation

b) Negative correlation

c) No correlation

d) Atomistic fallacy

*Explanation: Distance to MRT and the price of the food move in opposite direction, i.e., the shorter the distance, the more expensive the food, hence it is negative correlation.*

9. Charlie wishes to investigate the effects of gummy bears which have been claimed to help increase the speed of hair growth. After a probability-based sampling plan is used, she randomly assigns half of the participants to ingest the gummies (treatment) and keeps the other half as a control group.

To gauge the positive effect of the gummy, she calculates the ratio of the average speed of hair growth in treatment group (speed1) to the average speed of hair growth in the control group (speed2). Denoting this ratio by R, so R = speed1 / speed2. At the end of the study, she finds that speed1 is equal to 2 times speed2. The null hypothesis is that the speeds are the same (i.e., no effect from ingesting gummies).

She then conducts a hypothesis test to see whether there is significant evidence to conclude that ingesting gummies is associated to hair growth, and obtains a p-value of 0.04.

Which of the following interpretation of the p-value is correct?

a) There is a 4% chance of obtaining an R value of 2 or more, given that the gummies do not work.

b) There is a 96% chance of obtaining an R value of 2 or more, given that the gummies do not work.

c) There is a 4% chance of obtaining an R value of 1 or more, given that the gummies do not work.

d) There is a 96% chance of obtaining an R value of 1 or more, given that the gummies do not work.

*Explanation: Refer to the chapter on Hypothesis Testing. P-value refers to the* probability of obtaining an outcome that is equivalent to or more extreme than the observed *given that the null hypothesis is true.*

10. Charlie wishes to investigate the effects of gummy bears which have been claimed to help increase the speed of hair growth. After a probability-based sampling plan is used, she randomly assigns half of the participants to ingest the gummies (treatment) and keeps the other half as a control group.

To gauge the positive effect of the gummy, she calculates the ratio of the average speed of hair growth in treatment group (speed1) to the average speed of hair growth in the control group (speed2). Denoting this ratio by R, so R = speed1 / speed2. At the end of the study, she finds that speed1 is equal to 2 times speed2. The null hypothesis is that the speeds are the same (i.e., no effect from ingesting gummies).

She then conducts a hypothesis test to see whether there is significant evidence to conclude that ingesting gummies is associated to hair growth, and obtains a p-value of 0.04

What conclusion can be drawn at 5% level of significance?

a) There is sufficient evidence to reject the null hypothesis.

b) There is insufficient evidence to reject the null hypothesis.

c) There is sufficient evidence to reject the alternative hypothesis.

d) There is insufficient evidence to reject the alternative hypothesis.

*Explanation: Refer to the chapter on Hypothesis Testing. When the p-value is below a pre-specified threshold (e.g. 5% or 10%), this gives us sufficient evidence to reject the null hypothesis. C and D are wrong because in hypothesis testing, we do not reject the alternative hypothesis, we only seek to reject the null hypothesis.*