GER1000 QUANTITATIVE REASONING

(Semester 2 : AY2015/2016)

TEST

5 March 2016

1:30—2:30 pm

Instructions:

(i) This test paper contains 7 pages (including this page) and 14 questions. Answer all questions. Give your answers in the bubble answer sheet provided. Follow the instructions on the sheet. Fill in the bubbles neatly and completely, using a 2B pencil.

(ii) Only the bubble answer sheet will be marked. Make sure you have written and bubbled in your correct matriculation number (the number starting with "A") in the answer sheet.

(iii) Only one answer per question is allowed. Each correct answer is given one (1) mark. An incorrect/incomplete/missing answer receives 0 mark. The maximum total mark for this test is 14.

(iv) This is a closed book exam. Calculators are allowed. Computers, tablets and mobile devices are not allowed.

1. 62,000 women were randomly divided into two equal groups. The treatment group was encouraged to do annual screening for breast cancer; 20,200 accepted, but 10,800 refused. The control group was offered the usual health care. After 5 years, death rates from breast cancer were 1.1 for the women who were screened, 1.5 for the women who refused, and 2.0 for the control group. It is known that poorer women were less likely to accept screening than richer ones, and that breast cancer affects the rich more than the poor.

   1.5 – 1.1 = 0.4 is _____ of the reduction in death rate from breast cancer due to screening.

   (A) an underestimate

   (B) a reasonable estimate

   (C) an overestimate

2. The Zika virus is spread by Aedes aegypti, the same mosquito that spreads the Dengue virus. 2015 saw a serious outbreak of the virus in Brazil. The Brazilian Ministry of Health thought the virus has something to do with microcephaly (abnormal smallness of the head) in new-borns. The ministry's position was mainly based on the following observation. Before 2015, Brazil had about 160 microcephaly cases each year on average. In 2015, more than 400 cases have been confirmed. These data come from a

   (A) controlled experiment, with no control group.

   (B) controlled experiment, with historical control.

   (C) controlled experiment, with randomised control.

   (D) observational study.

3. To find out how methods of eating rice (chopsticks vs spoon) affect the amount of glucose in the blood, an investigator enrolled 12 healthy volunteers in Singapore, who were randomly assigned to treatment and control groups of equal size. Two limitations of the study are

   (I) The subjects may not be representative of all healthy people in Singapore.

   (II) The treatment and control groups are likely to be different in important ways at the beginning of the study.

   Suppose the investigator went out to enroll more healthy volunteers. He now has 1,000 subjects who are randomly assigned to two groups of equal size. Which limitation still holds for the new study?

   (A) (I) only.

   (B) (II) only.

   (C) Both (I) and (II).

   (D) Neither (I) nor (II).

4. All families in a country live in 3 types of homes: large (5 or more bedrooms), medium (1 to 4 bedrooms), and small (no bedroom). Overall, 40% of families own their homes. In the large group, 70% of families own their homes; in the medium group, 50% of families own their homes. The percentage of families who own their homes in the small group is

   (A) more than 70%.

   (B) between 50% and 70%.

   (C) between 40% and 50%.

   (D) less than 40%.

5. In a population, every person is either exposed or unexposed to a certain risk factor, and every person is either diseased or healthy. A cohort study is conducted, by randomly sampling from the exposed and unexposed groups separately. A case-control study is conducted, by randomly sampling from the diseased and healthy groups separately. In the following two statements, "ratio" refers to contrasting the exposed group to the unexposed group, and the event of interest is "diseased".

   (I) The population risk ratio can be estimated reliably from the cohort study, but not the population odds ratio.

   (II) The population odds ratio can be estimated reliably from the case-control study, but not the population risk ratio.

   (A) Only (I) is true.

   (B) Only (II) is true.

   (C) Both (I) and (II) are true.

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

6. In a large department, it was discovered that, out of 64 people who have disease X, 46 had eaten tuna casserole from the canteen. Further investigation found that the same department also has 65 people who do not have disease X, although 25 of those had eaten tuna casserole from the canteen. There are another 10 people in the department whose disease status is unclear. What is the odds ratio of X (for those who have eaten tuna casserole, relative to those who have not eaten)?

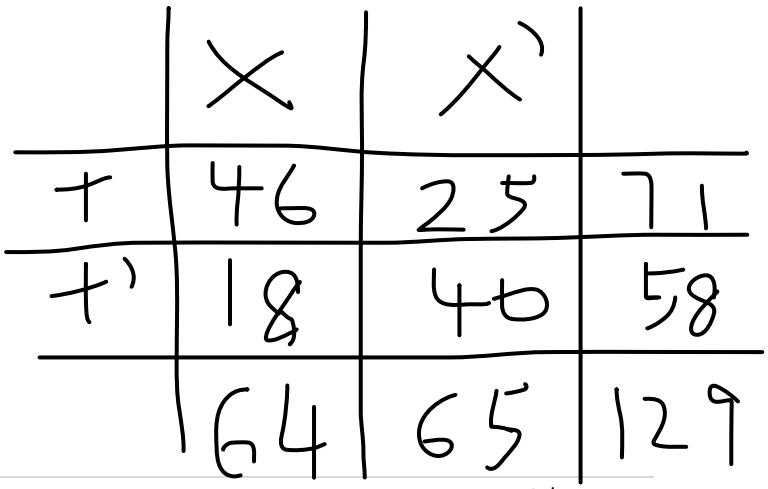
   (A) 1.9

   (B) 2.1

   (C) 4.1

   (D) 0.98

   (E) 1.2



Let A and B be two characteristics in a population. We say that they are positively associated if the rate of A among people with B, rate(A|B), is larger than the rate of A among people without B, rate(A|not B). Conversely, A and B are negatively associated if rate(A|B) < rate(A|not B). The next two questions use these definitions.

7. How does "forgiveness" (being forgiving) and empathy go together? The study of Toussaint and Webb on 45 men and 82 women are summarised in the following hypothetical tables:

	Empathy	No empathy	Row total
Forgiving	10	10	20
Not forgiving	9	16	25
Column total	19	26	45

Distribution of 45 men

	Empathy	No empathy	Row total
Forgiving	30	31	61
Not forgiving	12	9	21
Column total	42	40	82

Distribution of 82 women

(I) Forgiveness and empathy are positively associated among men.

(II) Forgiveness and empathy are positively associated among women.

(A) Only (I) is true.

(B) Only (II) is true.

(C) Both (I) and (II) are true.

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

8. Suppose A and B are positively associated among men, but negatively associated among women. The statement "When the men and women are combined, there is no association between A and B." is

(A) always true.

(B) always false.

(C) None of the above.

9. The *Sports Illustrated* jinx states that "Successful individuals or teams who appear on the cover of the *Sports Illustrated* magazine will subsequently experience bad luck in their careers." This is an example of

   (A) regression fallacy.

   (B) Simpson's paradox.

   (C) odds ratio indicating an association.

   (D) relative risk suggesting an association.

   (E) baseline risk for an exposed group.

10. The correlation coefficient for 126 pairs of (X,Y) values is 0.86. X ranges from just above 0 to just below 1. It was decided that half the points with smaller X values should be deleted. What will happen to the correlation coefficient of the remaining 63 points?

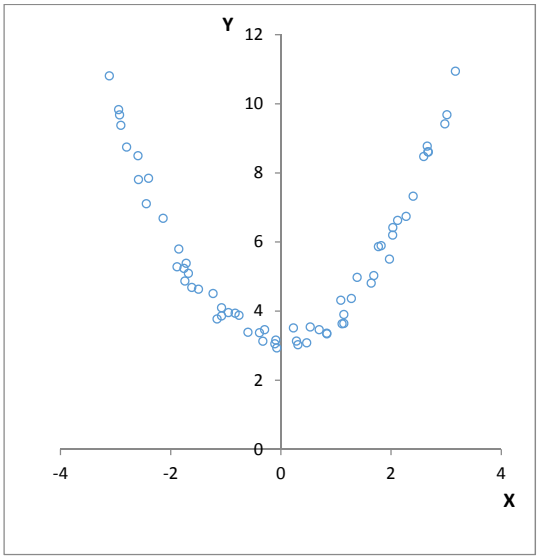
    (A) The correlation coefficient will remain about the same, unless X and Y are not linearly related.

    (B) The correlation coefficient will generally increase.

    (C) The correlation coefficient will be 0.43.

    (D) None of the above.

11. The following scatterplot shows that 2 industrial processes, X and Y, are closely related:



    Which of the following statements is the best description of the plot?

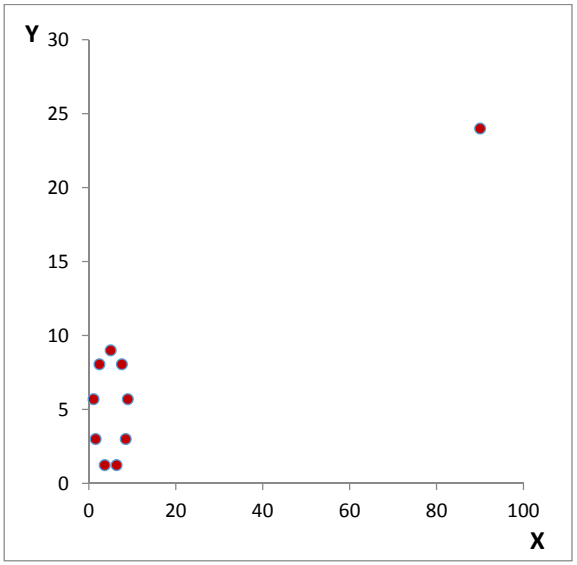
    (A) The correlation coefficient between X and Y is nearly 0.

    (B) The correlation coefficient between X and Y is nearly 0.5.

    (C) The correlation coefficient between X and Y is nearly 0.9.

    (D) The plot strongly suggests that X is caused by Y.

12. The following plot shows an outlier in both the X and Y directions. What will happen if we remove the outlier?



(A) The correlation coefficient between X and Y will remain roughly the same.

(B) The correlation coefficient between X and Y will decrease.

(C) The correlation coefficient between X and Y will increase.

(D) The correlation coefficient may increase or decrease, depending on the scales of measurement for X and Y.

(E) It is not possible to know what will happen to the correlation coefficient between X and Y.

13. When analyzing categorical variables, which of the following is false?

(A) When the odds ratio is nearly 1 between two variables, we may conclude that they are not associated.

(B) In a cohort study, if the risk ratio is nearly 1 between two variables, we may conclude that they are not associated.

(C) Case control studies are not controlled experiments.

(D) Correlation coefficient can be used to gauge the relationship between the variables.

(E) The risk ratio is not appropriate for case control studies.

14. A regression line is fitted to relate the Winning Time (WT) of a 100-km cycling road race and the Number (N) of participants. It can be expressed as:

    WT = 100 + 0.11 x N

    However, some people observed that "WT is longer for very small and very large N values, compared to intermediate N values". How might this be reconciled with the regression result?

    (A) The observation is wrong, since the regression equation shows that the timing increases with larger number of participants.

    (B) This might be the result of Ecological correlation.

    (C) Possibly, the relationship between WT and N is non-linear.

    (D) This demonstrates the atomistic fallacy.

1A

2D

3A

4D

5B

6C

7A

8C

9A

10D

11A

12B

13D

14C

Brief solutions to GER1000 Quantitative Reasoning test, AY2015/16 Semester 2

1 A. Reason: women who agreed to screening were more at risk than those who refused. This is like comparing the treatment and control groups in the NFIP study; see Chapter 1 Unit 3, and Unit 4 slide 8.

2 D. People were not deliberately infected with the virus in 2015, so this is not a controlled experiment, even though the control group is historical.

3 A. People who volunteer can be very different from those who do not, and getting many volunteers will not help much; see Chapter 1 Unit 10 slide 10. But a large randomised experiment is very likely to have similar treatment and control groups at the beginning of the study; see Chapter 1 Unit 4.

4 D. This is similar to Quiz 2 question 4.

5 B. The OR can be estimated from both kinds of studies.

6 C. OR is (46x40) / (25x18) ~ 4.1.

7 A. Among men, rate(empathy|forgiving) = 10/20 = 0.50, rate(empathy|not forgiving) = 9/25 ~ 0.36: positive association. Among women, the corresponding rates are 0.49 and 0.57: negative association.

8 C. This is Simpson's paradox; see Chapter 1 Unit 9, particularly slide 5.

9 A. This can be explained by the concept of regression towards mediocrity. When we think about it from test and re-test point of view, those who has done very well in their careers are going to appear on the cover. Next season being the re-test for them, they may have a slightly worse performance. See chapter 2 unit 9 slide 12 for more elaboration.

10 D. This example of range restriction may bring about the attenuation effect (of reduced correlation) if both the X and Y distributions are normal. Generally, however, anything may happen when data is halved via range restriction. Consider two scenarios:

i) For $0 < X < 0.5$, X and Y are perfectly linear and for $0.5 < x < 1$, X and Y has no linear relation. Let's assume overall correlation for such a scenario is 0.86. If we delete the data for $x < 0.5$ then r will go to 0.

ii) For $0 < X < 0.5$, X and Y has no linear relation and for $0.5 < X < 1$, X and Y are perfectly linear. Let's assume overall correlation for such a scenario is 0.86. If we delete the data for $x < 0.5$, then r will go to 1.

You may want to look into chapter 2, Unit 9, Slides 2-5.

11 A. It is obvious that B and C cannot be the answer. Why is D not the answer? Simply because association is not causation. You may look into Chapter 2, Unit 5, slide 13.

12 B. With the outlier, there is a linear relation between X and Y. When the outlier is removed, r value will be reduced as there is no apparent linear relation between X and Y. Chapter 2, Unit 7, Slide 6.

13 D. In this module, two types of variables are introduced. When we are dealing with measurement variables, we have scatter diagram and we use correlation coefficient to study the association. For categorical data, we have contingency table and we use Odds Ratio or Risk Ratio to establish the association depending on the design of the study. Chapter 2, Unit 1, Slide 2.

14 C. Please pay attention that a regression line is being "fitted" to relate the two variables. The two variables may not have a linear relationship in the first place. B & D should not be considered because there was no aggregation of individuals into groups. Chapter 2, Unit 7, Slides 11&12.