Exercise 1

1. Siti conducted an investigation by randomly assigning each subject from a randomly selected sample of 50 participants, either to watch Netflix for 1 hour 4 times a week, or to listen to Symphony 924FM for 1 hour 4 times a week. After 6 months, the changes in the subjects' blood pressure readings over the same period were recorded. The changes were compared for the two groups. Which of the following is true?

   (A) This is a randomised experiment because blood pressure was measured at the beginning and end of the study.

   (B) This is an observational study because the two groups were compared at the end of the study.

   (C) This is a randomised experiment because the participants were randomly assigned to either activity.

   (D) This is a randomised experiment because a random sample of participants was used.

   Answer is (C). There is random assignment of subjects to either activity. Thus, this is a randomised experiment.

2. In a study on the relationship between watching television and obesity, 3000 individuals were recruited via an advertisement put up by ABC newspaper in Singapore. Afterwards, the investigator of the study separated the participants into two groups. The participants in one group consists of people who, on average, watch television for at least 4 hours a day, while the other group consists of people who, on average watch television strictly less than 4 hours a day. The investigator then records how many participants are obese in each group. Which of the following is/are true?

   (I) The result of this study is generalisable to the population who reads ABC newspaper.

   (II) This is an observational study.

   (A) Only (I).

   (B) Only (II).

   (C) Neither (I) nor (II).

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

   Answer is (B). The result is not generalisable because the investigator uses non-probability sampling (volunteer sampling). The study is clearly an observational study because the researchers are not involved in the assignment of the subjects to either group, but rather is done by self assignment.

3. Consider a study that intends to examine whether the colour red makes children act impulsively. A group of 500 children were assigned into two groups by the expert opinion of a child psychologist; group Red if the psychologist pointed to the child, and group Green if the psychologist did not. Each child is then led into a room that has a big button in the colour of their group and labelled "DO NOT PRESS ME!". It is then recorded whether the child presses the button within 10 minutes. Each child was given a candy at the end for participating in the study.

   Which of the following best describes the design of the study?

   (A) Observational study without random assignment.

   (B) Observational study with random assignment.

   (C) Controlled experiment without random assignment.

   (D) Controlled experiment with random assignment.

Answer is (C). This is a controlled experiment rather than an observational study, since participants do not self-select themselves into the two groups. The assignment of each child is determined by the expert opinion, which is not random. On a side note, in observational studies, participants cannot be assigned randomly into the groups, simply because the participants self-select into the groups by their inherent characteristics or existing behaviours.

4. In a large scale experiment, a researcher randomly assigned 6000 subjects to receive either a drug or a placebo. 4000 patients were assigned to receive the drug, and the other 2000 patients received the placebo. The researcher did a quick headcount in the drug-receiving group and noted that there were 3002 males who received the drug. The researcher does not have time to do a headcount in the placebo group. Which is the most reasonable number of males to be expected in the placebo group?

   (A) 1000.

   (B) 1500.

   (C) 2000.

   (D) 2998.

   Answer is (B). Randomised assignment of a large number of subjects tend to produce groups which are similar in all aspects (including the proportion of males in each group). 6000 is a reasonably large number, so we would expect the proportion of males and females in each group to be similar. Among the 4000 subjects who received the drug, 3002 (about 75%) were males. Hence among the 2000 patients who received the placebo, about 75% (1500) of them should be male.

5. Is a randomised experiment or an observational study more appropriate to investigate whether women are more likely than men to suffer from anxiety?

   (A) Observational study; it is unethical to subject study participants to anxiety over a long period of time.

   (B) Observational study; it would be easier to control for confounders such as the number of hours worked per week.

   (C) Observational study; gender cannot be assigned as a treatment.

   (D) Randomised experiment; an observational study would take too long.

   Answer is (C). Here we are investigating the relationship between gender and anxiety, with gender as the treatment or exposure variable. It is not possible to assign gender as a treatment in this study.

6. May, an owner of a tuition center, wishes to find out if using iPads during tuition class improves her students' academic performance. She decided to conduct an experiment as follows:

   1. She groups all the students in her center according to the day they come for tuition. For simplicity's sake, we can assume each student only goes for tuition once per week, there is at least one class of tuition every day in her center, and no student drops out halfway.
   2. Every student who goes for tuition on weekends will be given an iPad to use during class. The students who go for tuition on weekdays will not be given an iPad.
   3. She then keeps track of all her students' academic performance for the next 6 months.

   Which of the following statements is/are true?

   (I) She used a probability sampling method.

   (II) This is a controlled experiment without random assignment.

   (A) Only (I).

(B) Only (II).

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

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

Answer is (B). Statement (I) is incorrect. Probability is not used in the selection of students into treatment/control. In fact, a census, not sampling, is conducted in this case. Statement (II) is correct. The students who go for tuition on weekends will be in the treatment group, and those who go on weekdays will be in the control group. There is no random assignment involved here.

7. In order to investigate the relationship between height and weight of people in a country, a researcher draws a simple random sample of the country's population, and records how height varies with weight in the sample. What kind of study is this?

   (A) An experiment.

   (B) An observational study.

   Answer is (B). There is no attempt to manipulate the height or the weight of any subject; data along both variables are collected by observation. It is thus an observational study.

8. Alex wants to investigate the effect of listening to music on time taken to solve a maths problem. Which of the following scenarios is an example of random assignment?

   (I) Subjects are assigned based on the time they arrive at the lab. Alex assigns the first 50% of the participants to arrive to the treatment group, and the remaining 50% to the control group.

   (II) For each subject, Alex asks him/her to choose head or tails. He then tosses a fair coin. If the result matches the subject's choice, the subject is assigned to the treatment group. If not, the subject is assigned to the control group.

   (A) Only (I).

   (B) Only (II).

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

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

   Answer is (B). Assignment based on arrival time is not based on a random mechanism. Hence, it could be that more hardworking subjects would arrive earlier and be able to concentrate to solve the problem quicker, whereas latecomers might be more flustered and take a longer time. In scenario (II), we can calculate the exact probability of assigning each subject to the treatment or control group.

9. In a double-blind study, 5000 patients with advanced colorectal cancer were randomly assigned to either receive a high-dose vitamin C (10 g daily) or a placebo.

   The study concluded that Vitamin C was not effective in the treatment of advanced colorectal cancer in the two groups because none of the patients with disease showed any improvement. The "cause and effect" conclusion

   (A) might be justified because randomisation and double-blinding were used.

   (B) cannot be justified because anecdotal evidence from other patients not in the study, shows that some people who take Vitamin C do recover from their disease.

   (C) cannot be justified because there could have been other differences between the two treatment groups, such as age and gender, so that the two groups cannot be compared to each other.

   (D) might be justified because the study is an observational study.

Answer is (A). (B) is not true as anecdotal evidence is unreliable; there could be factors not accounted for that results in the recovery of those patients. (C) is not true as random assignment will make both treatment and control groups similar. (D) is not true because the study is not an observational study. Thus the answer is (A).

10. Professor Lim would like to find out if including a peer review component would affect students' final grades. He decided to get a sample of the students in his tutorial classes and place them into 2 groups. He assigned his Monday, Tuesday and Wednesday morning classes into the "assessment with peer review" group and his Wednesday afternoon, Thursday and Friday classes in the "assessment without peer review" group.

    Which of the following best describes the type of sampling employed?

    (A) Cluster sampling.

    (B) Systematic sampling.

    (C) Volunteer sampling.

    (D) None of the other options.

    Answer is (D). Probability sampling will require deliberate use of chance in the sampling process. In this case, the assignment of individuals have been pre-determined by the Professor. Within the types of non-probability sampling methods, this is not an example of volunteer sampling, as all students from both sub-groups were selected by the Professor to do the study and not self-selected.

11. A researcher is trying to study the happiness level of all current NUS students. Which of the following is/are (an) example(s) of probability sampling methods?

    (I) The researcher gets a list of all current NUS students' emails from the administrative office and randomly selects 100 students' emails. He then emails them a link to a short e-survey. 50 students replied to his survey.

    (II) The researcher invites all final year NUS students in his faculty to visit his lab for a psychological test to determine their happiness level, with a promise to compensate them for their time with a $10 voucher. 200 students turned up for the test.

    (A) Only (I).

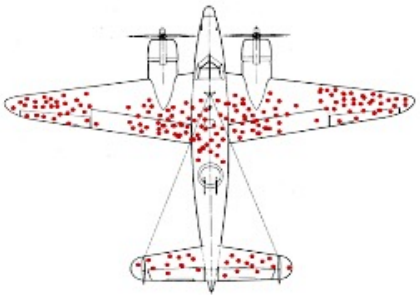
    (B) Only (II).

    (C) Neither (I) nor (II).

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

    Answer is (A). Randomly selecting 100 emails from the list of student emails is a probability sampling method, even if the response rate is not high. On the other hand, the process of sending an email to only final year students in the researchers' faculty is done out of convenience, and convenience sampling is a non-probability sampling method.

12. A military officer was interested in reducing the number of casualties sustained in aerial battle. His population of interest was all planes under his charge. He tasked his men to examine the planes that returned from the war front, and then take note of which parts of the planes sustained ammunition damage. He collated all the data and presented it on a single blueprint of the plane, as shown below (the dots denote where ammunition damage occurred):



The officer then concludes: "Based on my sample data, I propose to fortify the plane armour for regions where ammunition damage was concentrated (using the above blueprint as a guide), so as to help these planes survive better." Would you agree with his assessment and why?

(A) Yes. The sample collected came from a good sampling frame.

(B) No. The sample collected came from an imperfect sampling frame.

(C) Yes. The sample size is big enough.

(D) No. The sample size is too small.

Answer is (B). The sample collected here came from an imperfect sampling frame, only comprising of planes that managed to make it back from the war - but what about those that did not manage to return? When one considers the officer's eventual aim for his statistical analysis (to improve the survivability of planes), it seems that the units that were excluded from this sample (planes that did not survive) are of much more pertinence than the units included in the sample (planes that survived).

13. If a sampling frame is _____ the target population, it will not lead to a loss in the generalisability of the results from the sample to the population.

Which of the following can be used to fill the blank appropriately? Select all that apply.

(A) equal to

(B) smaller than

(C) larger than

(A) and (C) are correct. Refer to the definition of sampling frame. Note that a sampling frame should cover, that is, be equal to or larger than, the target population to achieve good coverage.

14. To find out the employment status of fresh graduates from University ABC, a questionnaire was sent to all of them. 30% of fresh graduates responded to the survey. The employment rate was calculated from the responses. Which of the following is likely to cause the calculated rate to differ from the population rate?

(I) Selection bias.

(II) Non-response bias.

(A) Only (I).

(B) Only (II).

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

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

Answer is (B). 30% of fresh graduates may not be representative of the whole population. It is possible that those who did not respond may have different employment status from those who responded. Although the response rate is low, it is still a census. There is no selection bias in a census.

15. Suppose I wish to find the average intelligence quotient (IQ) of all Primary 5 children studying in local schools in Singapore. I first selected a random sample of 10 schools out of all local primary schools in Singapore. Then I asked all the Primary 5 children in these chosen 10 schools to take an IQ test. Finally, I obtained the average value of all the IQ scores of children who took the test, which was 106. Which of the following statements is/are correct?

    (I) The parameter in this study is the average IQ of all Primary 5 children who took the IQ test.

    (II) Stratified sampling was employed in this study.

    (A) Only (I).

    (B) Only (II).

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

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

    Answer is (D). The parameter in this study is the average IQ of all Primary 5 children **studying in local schools in Singapore**. 106 is a sample estimate of the actual parameter. Hence statement (I) is incorrect. In stratified sampling, the population is divided into groups (strata) and then we randomly obtain a sample from each group. In cluster sampling, the population is first divided into groups (clusters). Then we take a random selection of clusters from all clusters, and include all units in the chosen clusters to comprise our sample. Here, cluster sampling is employed, where each school is a cluster. Hence statement (II) is also incorrect.

16. In a drug factory, pills were manufactured in 1000 batches, with 20 units per batch, forming a total of 20000 units. You decide to sample some of these pills to ensure that the dosage is right. Suppose you randomly sample two batches and then select every unit in these batches to be in your sample. What sampling method did you use?

    (A) Systematic sampling.

    (B) Stratified sampling.

    (C) Cluster sampling.

    (D) Simple random sampling.

    Answer is (C). Since every unit from the randomly selected batches (the clusters) are included in the sample, this is an example of the cluster sampling method.

17. Suppose you are a researcher who is interested in drawing a simple random sample of 200 people from a population of 5000 individuals. Which of the following would be a correct approach? Select all that apply.

    (A) Sort the names of the entire population by alphabetical order (A to Z) and place the names in a list. Select the people whose names appear at the top 200 of the list.

    (B) Assign each individual in the population a unique integer from 1 to 5000 by random assignment. Then choose the people assigned numbers 4801 to 5000.

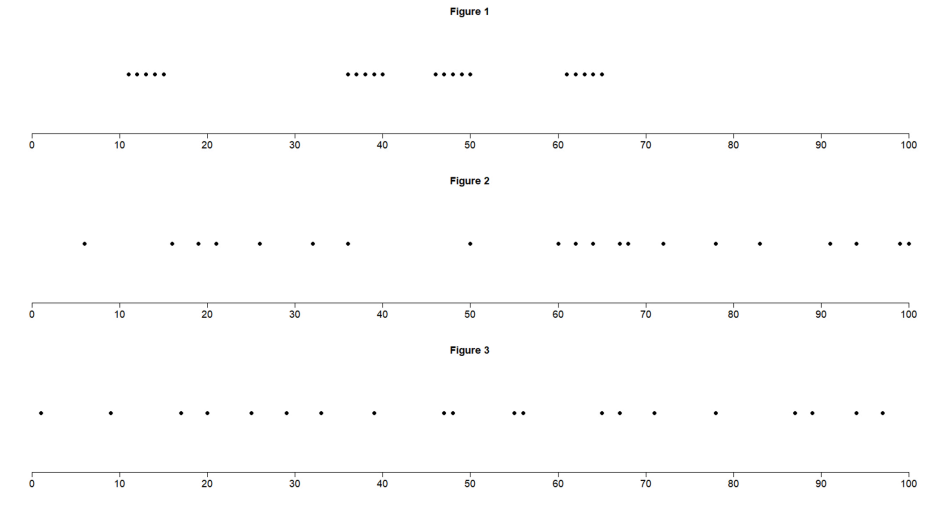
    (C) Write all the names of the entire population on equal-sized pieces of paper, mix the papers in a box and draw out 200 pieces of paper at one go. Choose the people whose names appear on the drawn papers.

    (B) and (C) are correct. A simple random sample of size n consists of n units from the population chosen in a way such that every set of units has equal chance to be the sample being selected. (A) is incorrect as there is an unequal chance of selection between the first 200 individuals in the list, and the remaining 4800 individuals, with the former having absolute certainty of being selected while the latter having no chance of being selected. In other words, (A) does not ensure that every set of units has an equal chance to form the sample.

18. Tom selected 4 samples of 20 integers from the population $\{1, 2, \ldots, 100\}$ using 4 different methods. They are

    1. simple random sampling (SRS).
    2. stratified sampling: the population was divided into the 10 strata $\{1, 2, \ldots, 10\}$, $\{11, 12, \ldots, 20\}$, $\ldots$, $\{91, 92, \ldots, 100\}$; and a SRS of 2 numbers was drawn from each of the 10 strata.
    3. cluster sampling: the population was divided into 20 clusters $\{1, 2, 3, 4, 5\}$, $\{6, 7, 8, 9, 10\}$, $\ldots$, $\{96, 97, 98, 99, 100\}$; and a SRS of 4 of these clusters was selected.
    4. systematic sampling: a random starting point between 1 to 5 was selected; and every 5th unit thereafter was selected too.

    He created dot plots for exactly 3 of the samples generated. Identify the sampling method depicted by each of the following plots.



    Sampling method depicted by Figure 1: _____

    Sampling method depicted by Figure 2: _____

    Sampling method depicted by Figure 3: _____

    Figure 1 depicts cluster sampling, Figure 2 depicts simple random sampling, Figure 3 depicts stratified sampling.

19. Researchers are interested in the average number of hours of study an NUS student does in a day. They take a simple random sample of 100 students from all the students in the Faculty of Law. They asked the 100 students for their self-reported average number of hours of study a day, and all 100 students responded to the researchers.

    Which of the following is true?

    (A) There is selection bias because an imperfect sampling frame is used.

    (B) There is non-response bias because students from the College of Design and Engineering were not selected into the sample.

    (C) A census was conducted.

    (D) The sample is representative of the population of NUS students.

    Answer is (A). There is selection bias because the sampling frame used (Faculty of Law), is smaller than the population of interest (all NUS students). There are students in the population of interest (such as those in faculties other than the Faculty of Law), that have zero chance of being selected

into the sample. Non-response bias is due to lack of response from the subjects in the sample, but all the chosen 100 students responded to the researchers. A census was not conducted since not all NUS students were asked to participate in the study. The sample is not representative since the sample only consists of students from the Faculty of Law.

20. We use random numbers to take a simple random sample of 50 students from a list of 6000 undergraduate students (45% males, 55% females) at the National University of Singapore. 50 random numbers from 1 to 6000 are selected, and the correspondingly numbered students are selected. Which of the following statements is TRUE?

    (A) We would never get the number 1111, because it is not random.

    (B) The draw 1234 is no more or less likely than the draw 1111.

    (C) Since the sample is random, it is impossible that it has only females in the sample.

    (D) Since the sample is random, it is impossible to get the random numbers 0001, 0002, 0003, ..., 0049, 0050.

    Answer is (B). Each number 0001 to 6000 is equally likely to be drawn. Thus all statements are false except (B).

21. There is a population of 200 secondary school students from Singapore that I wish to take a sample from. Among the students,

    (I) There are 130 male students and 70 female students.

    (II) In the name list, students numbered 1 to 60 have last names starting with A to I; students numbered 61 to 130 have last names starting with J to S and students numbered 131 to 200 have the last names starting with T to Z.

    Identify the sampling method employed in each of the following scenarios:

    Scenario 1: 50 numbers between 1 to 200 are randomly generated. Students whose number on the list matches the 50 numbers generated are included in the sample while others are not.

    Scenario 2: A fair six-sided die is tossed once. If it shows 1 or 2, then students whose last name starts with A to I are included in the sample while others are not. If it shows 3 or 4, then students whose last name starts with J to S are included in the sample while others are not. If it shows 5 or 6, then students whose last name starts with T to Z are included in the sample while others are not.

    Scenario 3: If the sampling was done on Mondays, Wednesdays or Fridays, all Male students are included in the sample while the Female students are not included. If the sampling was done on Tuesdays or Thursdays, all Female students are included in the sample while the Male students are not included. (Note that students do not go to school on Saturdays and Sundays.)

    (A) Scenario 1: simple random sampling; Scenario 2: stratified sampling; Scenario 3: cluster sampling.

    (B) Scenario 1: non-probability sampling; Scenario 2: cluster sampling; Scenario 3: stratified sampling.

    (C) Scenario 1: non-probability sampling; Scenario 2: stratified sampling; Scenario 3: non-probability sampling.

    (D) Scenario 1: simple random sampling; Scenario 2: cluster sampling; Scenario 3: non-probability sampling.

    Answer is (D). For scenario 1, the ordering of the list is irrelevant. It is simple random sampling since the 50 numbers, corresponding to the 50 students selected, are randomly generated. For scenario 2, the 200 students are grouped into clusters according to their last names. All students from the selected cluster are included in the sample, hence it is cluster sampling. For scenario 3, there is no deliberate use of chance in the sampling process, hence it is non-probability sampling.

22. Select the correct word from the list for the respective blank in the sentence.

    "The __(1)__ is used to quantify the degree of spread relative to the __(2)__ and is a useful statistic for comparing the degree of variation across different variables within a data set."

    List: *Coefficient of variation, interquartile range, standard deviation, mean, median.*

    The answer for (1) is coefficient of variation and the answer for (2) is mean. Refer to the definition of mean and coefficient of variation to see that 'coefficient of variation is used to quantify the degree of spread relative to the mean'.

23. A teacher has just finished marking the final examination scripts for her class of 50 Secondary 1 students. She informs the students that the class average is 67.3. The maximum mark for the examination is 100 and the passing mark is 50. A student receives his examination script and realises his score is 65 which is lower than the average score. Based only on the information given above, which of the following statements must be true?

    (I) The student has performed worse than half the class.

    (II) Everyone in the class has passed the test.

    (A) Only (I).

    (B) Only (II).

    (C) Neither (I) nor (II).

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

    Answer is (C). A score lower than the mean does not imply that the student has performed worse than half the class. Consider the following set of scores for 10 students

    $$45, 55, 60, 62, 64, 64, 65, 85, 86, 87.$$

    The average is 67.3. The student who has scored 65 clearly has not performed worse than half the class. Neither has everyone in the class passed the test since one student has scored 45. A similar data set can also be constructed for 50 students. Therefore, neither statement is true.

24. Consider a data set consisting of values for a numerical variable x. Let the values be $x_1, x_2, \ldots, x_n$ arranged in ascending order. A value y is said to be the **balancing point of x in the data set** if the following condition is satisfied.

    $$(y - x_1) + (y - x_2) + \cdots + (y - x_k) = (x_{k+1} - y) + (x_{k+2} - y) + \cdots + (x_n - y)$$

    where $x_1, x_2, \ldots, x_k$ are the values of x in the data set that are **smaller than or equal to** y and $x_{k+1}, x_{k+2}, \ldots, x_n$ are the values of x in the data set that are **larger than** y. For example consider a small data set $\{1, 3, 5, 5, 5, 7, 9\}$. In this case the value 5 is the balancing point of the data set since

    $$(5 - 5) + (5 - 5) + (5 - 5) + (5 - 3) + (5 - 1) = (7 - 5) + (9 - 5).$$

    Which of the two statements below is/are true?

    (I) The median of x is always the balancing point of x in any data set.

    (II) The mode of x is always the balancing point of x in any data set.

    (A) Only (I).

    (B) Only (II).

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

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

Answer is (D). Neither median nor mode is always the balancing point of a data set. Consider a small data set $\{1, 1, 3, 4, 5\}$. The median of the data set is 3. However, we observe that $(3-1)+(3-1)$ is not the same as $(4-3)+(5-3)$. Similarly, the mode is 1 but once again we see that $(1-1)+(1-1)$ is not the same as $(3-1)+(4-1)+(5-1)$.

25. An examination was given to Class A and Class B, which consisted of 20 students each. The score of each student is between 0 and 100.

    The range of scores in Class A is from 70 to 90. All the students in Class B scored less than 40 marks. Due to manpower shortages, Class A and Class B were combined to form Class C. Hence Class C now contains 40 students, who were previously from Class A and Class B.

    Which of the following statements about the relationship between the mean score in Class C and the mean score in Class A is always true?

    (A) The mean score in Class C must be lower than the mean score in Class A.

    (B) The mean score in Class C must be the same as the mean score in Class A.

    (C) The mean score in Class C must be higher than the mean score in Class A.

    (D) There is insufficient information to deduce the relationship between the mean score of Class C and the mean score of Class A.

    Answer is (A). Since Class A and Class B have the same number of students, and all students in Class A scored strictly higher than the maximum score of Class B, it implies that the mean for Class A is strictly higher than that of Class B. Therefore when pooling both classes together, the overall mean will be the mean of Class A + the mean of Class B divided by 2 which will be strictly less than the mean of Class A.

26. City planners wanted to know how many people lived in a typical housing unit so they compiled data from hundreds of forms that had been submitted in various city offices. Summary statistics are shown in the table below.

Mean	Standard Deviation	Min	Q_1	Median	Q_3	Max
2.53	1.4	1	1	2	3	8

    The city bases their garbage disposal fee on the occupancy level of the home or apartment. The annual fee is \$50 plus \$4 per person, so a single-occupant home pays \$54 and homes with 10 people pay \$50 + \$4 ×10 =\$90 a year.

    The median fee paid is _____(1)_____ and the IQR of the fee paid is _____(2)_____ .

    Fill in the blanks for the statement above, give your answers correct to 2 decimal places.

    The answer to the first blank is $50 + 2 \times 4 = 58$ since the median number of occupants is 2. The IQR of the fee paid is $(50 + 3 \times 4) - (50 + 1 \times 4) = 8$.

27. Consider the following numerical values:

$$14,\ 15,\ 18,\ 20,\ 24,\ 29,\ 33,\ 34,\ x,$$

    where x is unknown and x may not necessarily be greater than or equal to 34. Which of the following statements is/are necessarily true? Select all that apply.

    (A) Regardless of the value of x, the median can never be higher than 24.

    (B) If the median of the values is less than 24, the mode cannot be 24.

    (C) The range cannot be 24.

(A) and (B) are correct. If x is higher than or equal to 24, the median of the data set is exactly 24 and if x is lower than 24, the median of the data set is less than 24. Suppose that the mode is 24, this means that x has to be 24. However, since the median is less than 24, we cannot have x to be greater than or equal to 24. Thus the mode cannot be 24. Finally, the range can possibly be 24. For example, suppose $x = 10$. This would make the minimum to be 10 and maximum to be 34, so the range is 24.

28. Which of the following can be obtained from nominal categorical data?

    (A) Mean.
    (B) Median.
    (C) Mode.
    (D) Variance.

    Answer is (C). The mode of data along a nominal categorical variable is the most commonly ocurring category (if it exists). Computing median requires an ordering of the data points, whereas mean and variance require arithmetic operations to be carried out on the data, neither of which can be done with nominal categorical data.

29. An examination was given to Class A and Class B, which consisted of 20 students each. The minimum possible score for the examination is 0 and the maximum possible score is 100. The range of scores in Class A is from 75 to 95 marks. All the students in Class B scored less than or equal to 50 marks. Due to shortage of teachers, Class A and Class B were combined to form Class C.

    Consider the following statements:

    (I) The median score in Class C must be greater than the median score in Class B.
    (II) The interquartile range of scores in Class C must be greater than the interquartile range of scores in Class A.

    Which of the statements is/are true?

    (A) Only (I).
    (B) Only (II).
    (C) Both (I) and (II).
    (D) Neither (I) nor (II).

    Answer is (C). Since Class A and Class B have the same number of students, and all students in Class A scored strictly greater than the maximum score of Class B, the median for Class C will be greater than the maximum score of Class B. Hence the median of Class C will be greater than the median of Class B. The largest possible interquartile range for Class A is $95 - 75 = 20$. The smallest possible interquartile range for Class C is $75 - 50 = 25$. Hence the interquartile range for Class C must be greater than the interquartile range for Class A.

30. A pop quiz in a class resulted in the following scores:

    $$0, 60, 66, 78, 82, 96, 98, 100.$$

    The five-number summary for these test scores is

    (A) $66, 78, 82, 96, 98$.
    (B) $0, 66, 82, 98, 100$.
    (C) $0, 25, 50, 75, 100$.
    (D) $0, 63, 80, 97, 100$.

    Answer is (D). Following the steps shown in the lecture notes, the five-number summary for these test scores is found to be (D).