Exercise 1

1. *(A, 2120)* The following is a research question from a scientific journal.

   What percentage of Singaporeans are keen to take vaccine X?

   What type of research question is this?

   (A) Make an estimate about the population.

   (B) Test a claim about the population.

   (C) Compare two sub-populations.

   (D) All of the other options.

   Answer is (A). The population of interest is Singaporeans and the estimate is the percentage of Singaporeans who are keen to take vaccine X.

2. *(A, 2120, 2210*)* Drug A is a new drug created to treat Congenital Amegakaryocytic Thrombocytopenia (CAMT). An experiment was done to evaluate survival outcomes of drug A, against the current standard, Rituximab, for treating CAMT. 150 CAMT patients were assigned to the drug A group, and 150 CAMT patients were assigned to the Rituximab group. A portion of the experimental design is summarised in the table below.

	Drug A	Rituximab
Male	82	85
Female	68	65
Total	150	150

   Determine if the statement below is true or false:

   "The table shows that random assignment was not done, because the two groups (drug A and Rituximab) do not have the same number of females."

   The statement is false. Random assignment does not guarantee that the number of females will be exactly the same in both groups (drug A and Rituximab). If random assignment was done on a large number of subjects, the two groups will tend to be similar (not necessarily the same) in all aspeccts.

3. *(A, 2120*)* Drug X is a new drug created. It is intended to be taken as a tablet by people who have skin allergy reactions. However, before pushing it into the market, researchers need to test the effectiveness of drug X. Thus, they designed a study, with two groups - a treatment and control group. Subjects with skin allergy reactions were invited to the study and placed into either of the two groups. The subjects in the treatment group received a drug X tablet to consume. Subjects were studied to see if their allergic reactions were successfully alleviated and were marked as either 'successful' or 'unsuccessful'.

   What are the possible tablets to give to the control group subjects, for comparing drug X's effectiveness, assuming all tablets look and taste the same? Select all that apply.

   (A) An empty tablet.

   (B) A tablet containing glucose. It is a definite known fact that glucose has a 2% better success rate than an empty tablet.

   (C) A tablet containing salt, with an unknown success rate.

   Only (A) and (B) are correct. A control group acts as a baseline for comparison with the treatment group. The control group can simply be not receiving treatment, receiving a placebo, or receiving an existing treatment. (C) is incorrect because it cannot provide any results that can be compared with the treatment group to show drug X's effectiveness.

4. *(A, 2120, 2210)* Which of the following scenarios is an example of random assignment in a controlled experiment?

   (I) For each subject, Peter throws a fair die of six sides. Subjects are assigned based on the number shown on the top surface of the die. Before the start of the assignment, Peter determines that the numbers "1", "2" and "6" will be assigned to the treatment group, and "3", "4" and "5" to the control group.

   (II) James lists all subjects in the experiment by alphabetical order, and selects subjects whose last name starts with "A", "B", "C" till "M" to place in the treatment group, while subjects whose last name starts with "N", "O", "P" till "Z" are placed in the control group.

   (A) Only (I).

   (B) Only (II).

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

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

   Answer is (A). Note that only (I) involves the use of chance.

5. *(A, 2120, 2210)* Which of the following best describes the function of a placebo in an experiment?

   (A) It is used to assign participants into treatment and control groups.

   (B) It is used to randomly assign the participants into treatment and control groups.

   (C) It is used to blind the participants to which group they belong to.

   (D) It is used to select participants into the experiment.

   Answer is (C). A placebo is a substance which has no effect but is designed to look identical to the treatment and therefore a person receiving it will think that he/she has received treatment. In other words, the people are unable to tell correctly which group they belong to which is precisely the idea of blinding.

6. *(A, 2120, 2210*)* A new drug Y was created to reduce stomach-aches. Researchers wanted to test the effectiveness of drug Y. Thus, they conducted a study containing 2000 subjects. The subjects have the following characteristics:

	Young	Old	Row total
Female	700	700	1400
Male	400	200	600
Column total	1100	900	2000

   Random assignment was conducted to assign the 2000 subjects into the treatment and the control groups. 1200 subjects were assigned to the treatment group, while the remaining 800 subjects were assigned to the control group. How many young females would we expect to see in the treatment group?

   (A) About 210.

   (B) About 420.

   (C) About 500.

   (D) About 700.

   Answer is (B). There are 700 young females in the study plan. If the random assignment ratio is 1200 : 800 for the treatment and control groups respectively, we would expect to see about $(700 \times 1200)/2000 = 420$ young females in the treatment group.

7. *(A, 2120, 2210)* Which of the following statements is/are always true about controlled experiments and observational studies?

   (I) There is no control group in observational studies.

   (II) Randomised assignment of subjects does not occur in observational studies.

   (III) There are no confounders in controlled experiments.

   (A) Only (I) and (II).

   (B) Only (I) and (III).

   (C) Only (II) and (III).

   (D) Only (II).

   (E) None of the other given options is correct.

   Answer is (D). Observational studies generally have control groups. As subjects are self-assigned to the different treatment and control groups in observational studies, it is not possible to assign subjects randomly into groups. There can be confounders in controlled experiments, especially if subjects are not randomly assigned into the different groups. Hence only statement (II) is true.

8. *(A, 2120)* 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.

9. *(A, 2120)* 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.

10. *(A, 2120)* 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.

11. *(A, 2210)* A researcher in University X wanted to conduct a survey to find out the average amount of time spent studying a week, by students in the university. He obtained the list of email addresses of all 2000 students in the university and sent out a survey form to everyone. As a token of appreciation, students who filled up the form received a "10% off" coupon from the university's bookshop. 300 students responded to the survey.

    Which of the following statements is/are correct? Select all that apply.

    (A) The study is likely to contain non-response bias.

    (B) The study is likely to contain selection bias.

    (C) The study uses a census.

    (A) and (C) are correct. Since only 300 out of 2000 students are willing to respond to the survey, the study is likely to contain non-response bias. Selection bias arises from poor sampling method or frame, neither of which is described in the question. The researcher's population of interest is all students in University X, and he sent a survey to all students in University X, so the study uses a census.

12. *(A)* 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.

13. *(A)* 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). It is impossible to conduct a randomised experiment and randomly assign subjects to treatment and control groups differentiated by gender.

14. *(A, 2210)* A publication is estimated to have about 20000 subscribers. A survey was sent to a random sample of 5000 of its subscribers. 300 of them returned the survey. Which of the following statements is true?

    (A) The sample results may not be generalisable to the population of subscribers because they used a self-selected sample.

    (B) The sample results may not be generalisable to the population of subscribers because there is likely to be non-response bias.

    (C) The sample results will be generalisable to the population of subscribers because they used a random sample.

    Answer is (B). The sample was not self-selected, it was a random sample of 5000 subscribers. However, due to a high percentage of non-responses, the results may not be generalisable as there is likely to be non-response bias. In other words, the 4700 subscribers that did not respond are likely to have different opinions on what is being surveyed compared to those who responded.

15. *(A, 2210)* Virus X has been known to cause very severe symptoms in its patients. Previously there has been no anti-viral medicine to treat virus X. Recently, researchers have finally managed to produce a trial drug in the form of a tablet. Researchers want to investigate if the trial drug helps to reduce the duration of symptoms (number of days) in patients. 1000 patients were sampled for the study, and all consented to join the study.

    Which of the following statements is/are true? Select all that apply.

    (A) Random sampling should be done to ensure that the subjects' demographics/characteristics are similar (in the treatment and control groups).

    (B) Blinding the researchers to the subjects' assigned groups (treatment or control group) is important because the researchers may have certain bias for/against the drug.

    (C) If the study randomly assigns 400 subjects into the treatment group, and 600 subjects into the control group, the result of the study will be biased due to the unequal number in the two groups.

    Only (B) is correct. Random assignment (not random sampling) should be done to ensure that the subjects' demographics/characteristics are similar in the treatment and control groups. Furthermore, random assignment does not require the same number of subjects in both treatment and control groups. As long as the number of subjects is large, the treatment and control groups will likely have similar demographics/characteristics. In fact, since we are comparing rates and not numbers, it does not matter if we have unequal group sizes.

16. *(A)* A medical researcher assigned 80000 patients to receive either a new drug or an old drug randomly. Among the 40123 patients who received the new drug, 24007 were male. Among the 80000 patients, what is the most likely proportion of females?

(A) 20%.

(B) 30%.

(C) 40%.

(D) 50%.

Answer is (C). Random assignment of a large number of patients tend to produce groups which are similar in all respects. This applies to the proportion of females here, since 80000 is large.

Among the 40123 patients who received the new drug, 16116 (approximately 40%) were females. Based on the above reasoning, among the 39877 patients who received the old drug, 16017 (approximately 40%) should be female. Therefore, in the 80000 patients, the number of females is likely to be 32133 (approximately 40%).

17. *(A)* 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.

18. *(A)* 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.

19. *(B, 2120, 2210)* (This is a multiple response question.) From the options given, select all possible words that can be used to complete the sentence below.

    Probability sampling refers to a sampling process whereby the probability of selection of individuals within the sampling frame must be _____.

(A) non-zero

(B) known

(C) high

Both (A) and (B) are correct. See definition of probability sampling.

20. *(B, 2120, 2210*)* A study was conducted to understand the average amount of sleep that current hall ABC students get. The hall has five levels, each with 50 rooms on each floor. Currently in hall ABC, every room is occupied by one student.

    For the study, every room is labelled with a specific number from 1 to 250. Identical slips of paper numbered from 1 to 250 are then placed in a box, and the researcher drew random 60 slips without replacement. These 60 numbers indicate the chosen students for the study. A survey form was then sent out to these 60 students.

    Which of the following must be true about the study?

    (A) There will not be any selection bias in this study.

    (B) There will not be any non-response bias in this study.

    (C) The sample will not be representative of the 250 residents in the hall.

    (D) If the 60 individuals selected did respond, the sample average cannot be used as an estimate of the average studying hours for all 250 residents.

    Answer is (A). There is a possibility of non-response bias from the 60 sampled individuals surveyed, if not many of them responded. Since simple random sampling is used, and the sampling frame covers the population of interest (the current students in hall ABC), the sample may be representative of the 250 residents in the hall, if the response rate is sufficiently high. Thus, if all the 60 students responded, the sample results may be generalisable to the population of the current hall ABC students. There will not be any selection bias becuase a probability sampling method was used and the sampling frame covers the population of interest itself.

21. *(B, 2120, 2210*)* Paracetamol company NAS owns a tablet press machine that produces Paracetamol tablets. On one shift, 3000 batches of tablets were manufactured. Each batch contains 10 tablets - a total of 30,000 tablets were manufactured. A researcher wants to ensure the dosage in the tablets is correct but has no time to check every single tablet. Hence, she decides to sample some of the tablets instead.

    Which of the following describes a probability sampling method? Select all that apply.

    (A) Select 3000 tablets at random.

    (B) Label all the tablets in each batch from 1 to 10, select a number from 1 to 10 at random, and select the unit from every batch that corresponds to that number.

    (C) Select 300 batches at random, and then sample all tablets in every selected batch.

    (D) Select the first 3000 tablets that were manufactured.

    (A), (B) and (C) are correct. (A) is an example of simple random sampling. (B) is an example of systematic sampling. (C) is an example of cluster sampling.

22. *(B, 2120)* 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.

23. *(B, 2120, 2210)* A recent study revealed that Singapore is "the most tired country in the world, due to work and internet." A researcher decided to conduct a further study on internet usage behaviour and working hours among all Singaporean adults in Singapore. Data was collected by interviewing commuters alighting from Pasir Ris MRT (East), Woodlands MRT (North), Redhill MRT (South) and Jurong East MRT (West) from 8am to 11pm over a period of 7 days.

Which of the following statements is **necessarily true**?

(A) As data was collected from different parts of Singapore, it is generalisable to the population of Singapore.

(B) Due to the equal representation of Northern, Southern, Eastern and Western parts of Singapore, selection bias is minimised.

(C) In this example, non-response bias exists because of a bad sampling plan.

(D) None of the other options.

Answer is (D). The selection of MRT stations was not done by using a randomised mechanism. Therefore, data collected is not generalisable to the population of Singapore, and it is not a forgone conclusion that selection bias is minimised. There is a possibility of non-response bias existing in this example. However, the reason for its existence is not because of a bad sampling plan.

24. *(B, 2120, 2210*)* Patch Z is a new medicine created to remove muscle soreness. A study was done to investigate the effectiveness of patch Z. The population of interest was Singaporean adult males. For this study, the researchers requested the Singapore Sports Association to sample all male athletes who reported for training over the week. 200 male athletes were sampled. There was no non-response. The 200 subjects had their identity tags randomly shuffled in a box. The first 100 tags picked from the box were assigned to the treatment group - administered patch Z. The remaining 100 were administered a placebo. 72% of the group that received patch Z had their muscle soreness alleviated, while 34% of the other group had their soreness alleviated. Which of the following is/are true?

(I) We are not able to generalise these results to the population of interest.

(II) Random assignment was not conducted.

(A) Only (I).

(B) Only (II).

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

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

Answer is (A). The result cannot be generalised, because of the sampling method and the fact that the sampling frame does not contain the population of interest. The method is non-probability sampling, and the sampling frame contains only male athletes, or those who went for training that week, but the population of interest is all Singaporean adult males. Take note that random assignment was actually conducted.

25. *(B, 2120, 2210)* The United States government conducts a Census of Agriculture every five years. The census comprises of farmland usage in all the 50 states in the country. John generated a sample of 3000 counties across all states from this census. He then collected data on the number of acres of land space these counties in the sample devoted to farms, and summarised his findings in a report as follows:

    (I) Of the 3000 counties selected, 25 counties were selected more than once in the sampling process.

    (II) 18% of the counties selected in this sample were from the state of Virginia, while none were from the states of Alaska, Arizona, Connecticut, Delaware, Hawaii, Rhode Island, Utah or Wyoming.

    John claimed that he obtained the sample of 3000 counties by Stratified Sampling **with replacement**, with the stratum being every state in the United States. Assuming that statements (I) and (II) are true, which of the statements do not/does not support John's claim on his sampling method?

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

    Answer is (B). In the case where Stratified Sampling is employed, and sampling with replacement is done within each stratum (in this case, the stratum is the 50 states in the US), there is a chance of obtaining a repeated observation. Also, if Stratified Sampling is done such that every state is represented, it is not possible for the case where some states are not represented. Therefore, statement (II)'s observation is indicative that the sampling method is not Stratified Sampling with the different states being the different strata.

26. *(B, 2120)* 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.

27. *(B, 2120, 2210)* For the following two cases, determine which sampling plan was used.

    Case 1: In an opinion poll, an airline company made a list of all its flights on 1 Jan 2022 and then selected a simple random sample of 30 flights. All the passengers on those flights selected were asked to fill out a questionnaire form.

    Case 2: A departmental store wanted to find out if customers would be willing to pay slightly higher prices for their products in order to have a smartphone app which customers can use to help them locate items in the store. The store hired an interviewer John and placed him at the only entrance on a particular day. John was asked to collect a sample of 100 opinions by interviewing the next person who came through the entrance each time he finishes an interview.

(A) Case 1: Cluster sampling plan; Case 2: Systematic sampling plan.

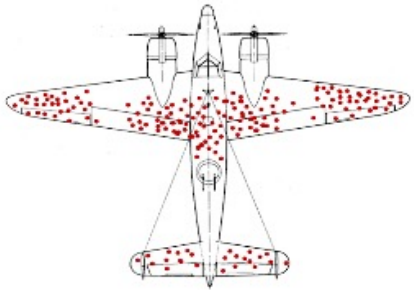
(B) Case 1: Stratified sampling plan; Case 2: Systematic sampling plan.

(C) Case 1: Stratified sampling plan; Case 2: Non-probability sampling.

(D) Case 1: Cluster sampling plan; Case 2: Non-probability sampling.

Answer is (D). The sampling plan in Case 1 is cluster sampling because for each randomly selected flight, all passengers were asked to complete the questionnaire form. The sampling plan in Case 2 is non-probability sampling because the next customer selected was not randomly selected.

28. *(B, 2120, 2210)* 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).

29. *(B, 2210)* 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.

30. *(B)* 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.

31. *(B)* 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.

32. *(B)* A researcher wishes to study procrastination and social anxiety levels amongst students majoring in architecture in University X. He wanted to collect a sample of 100 students out of the 1000 students majoring in architecture. He took a name list of all architecture students in University X. The researcher rolled a **fair** six-sided die, which landed on 3.

    He then decided to pick the 3^{rd} student in the name list, and every 10^{th} student afterwards until he collected his desired sample size of 100 students. That is, he selects the 3^{rd} student, 13^{th} student, 23^{rd} student ... until he gets his desired sample size of 100 students.

    What kind of sampling method did the researcher emply?

    (A) Systematic sampling.

    (B) Simple random sampling.

    (C) Non-probability sampling.

    (D) Stratified sampling.

Answer is (C). Even though the number 3 was chosen randomly using a fair die, the 7^{th} to the 10^{th} person on the name list has a zero chance of being selected, as a fair is only six-sided. This means the 17^{th} to 20^{th}, 27^{th} to 30^{th}, ..., 997^{th} to 1000^{th} student, has no chance of being chosen. As some of the architecture students have a zero chance of being selected, this sampling method is a non-probability sampling method.

33. *(B)* An airline would like to find out the quality of service provided by their staff on one of its flights. The airline posted an interviewer right after the plane landed and told the interviewer to start her interview when the third passenger came out from the plane. Thereafter, she would interview the next customer who came out from the plane each time she had finished interviewing the previous customer. What is the sampling method employed by the interviewer?

    (A) Systematic sampling.

    (B) Simple random sampling.

    (C) Cluster sampling.

    (D) None of the other given options.

    Answer is (D). This sampling method is not a probability sampling method since there is no probability involved in selecting passengers to be interviewed. All the methods given in the other options are probability sampling methods.

34. *(B)* A group of students in NUS is interested to find out about the relation between eating junk food and having migraine among adults in Singapore. Which of the following is the most appropriate sampling frame that can be used for the study?

    (A) All adults on Earth.

    (B) All residents of Singapore.

    (C) All adults working in NUS.

    (D) All students in Singapore.

    Answer is (B). Characteristics of a good sampling frame involves "good coverage" and up-to-date and complete units. If the sampling frame is too big and covers more than what we want, then the cost of getting the right units would be high.

35. *(B)* 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 results 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) Straified sampling.

    (C) Cluster sampling.

    (D) Simple random sampling.

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

36. *(B, 2210)* Clothes retailer G&N wishes to find out from all their visitors how receptive they are in terms of recycling used clothing. The management decides to survey a sample from customers paying for purchases at the cashier. They interview every fifth paying customer during retail hours from 11am to 9.30pm. Which of the statements is/are correct?

    (I) The above is an example of systematic sampling.

    (II) The sampling frame is the same as the target population.

(A) Only (I).

(B) Only (II).

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

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

Answer is (D). The sampling method seems like systematic sampling except for the crucial point that the number 5 (and thus sampling every fifth paying customer) is not the result of a random choice, that is, the decision of sampling every fifth (rather than sixth or seventh and so on) paying customer did not come about due to a random process. So (I) is incorrect. The target population is all visitors to the retailer but the sampling frame only includes those paying customers. Since the target population includes non-paying visitors, it is larger than the sampling frame.

37. *(B)* Assume you have a sampling frame of your entire population of interest, which comprises of 100 people's names. Which of the following methods can be used to select a simple random sample of 10 people from this population?

(I) Assign, without replacement, to each person a random number from 1 to 100 such that none of them share the same number. Choose the people assigned numbers 1 to 10.

(II) Write the names on equal sized pieces of paper, put the papers in a hat. Shake the hat, mix the papers well and draw out 10 names.

(A) Only (I).

(B) Only (II).

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

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

Answer is (C). In simple random sampling, every unit in the population of interest must have the same chance of being selected in the sample.

38. *(B)* The Registrar of a University has the list of all students in the University, sorted by the students' matriculation number in increasing order. Among the 25000 students in the University, a researcher proposes to choose a number at random from 1 to 100. Starting from that number, every 100^{th} person is included in the sample. Which of the following statements is/are correct?

(I) This is an example of systematic sampling.

(II) Since the names are sorted in order of matriculation number, this is a non-probability sampling method.

(A) Only (I).

(B) Only (II).

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

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

Answer is (A). This is an example of systematic sampling. It is still a probability sampling method because the starting number is drawn at random (between 1 to 100) and the order of names will not affect the probability method of sampling.

39. *(B, 2210)* A researcher is interested in drawing a sample from town X that has a population of 2000. He has a sampling frame of all 2000 townfolks' names. Which of the following methods can be used to select a simple random sample of 100 people from this population? Select all that apply.

(A) Sort the peoples' names by alphabetical order (A to Z) and place the names in a list. Choose the people whose names appear at the top 100 of the list.

(B) Assign each townsfolk a unique random integer from 1 to 2000. Choose the people assigned numbers 1 to 100.

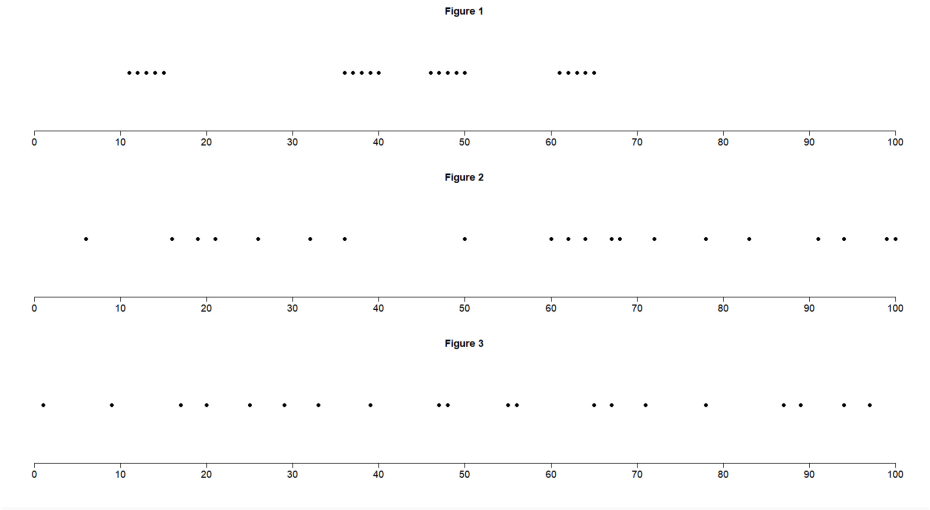
(C) Write the 2000 peoples' names on equal sized pieces of paper, mix the papers in a box, and draw out 100 pieces of paper in one go. Choose the people whose names appear on the drawn papers.

(B) and (C) are correct. A simple random sample of size n consists of n units from the population chosen in such a way that every set of units have equal chance to be the sample actually selected. The method involving sorting the peoples' names by alphabetical order and picking the first 100 does not ensure every set of units has an equal chance to form the sample, hence it is not considered as simple random sampling.

40. *(B, 2210)* 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 $\{1, 2, \ldots, 10\}$, $\{11, 12, \ldots, 20\}$, ..., $\{91, 92, \ldots, 100\}$; and a SRS of 2 numbers was drawn from each of the 10 strata.
    3. cluster sampling: the population was divided into 20 clusters $\{1, 2, 3, 4, 5\}$, $\{6, 7, 8, 9, 10\}$, ..., $\{96, 97, 98, 99, 100\}$; 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.

41. *(C, 2120, 2210)* Select the correct word from the list for the respective blank in the sentence.

    "The __(1)__ is used to measure how widely spread the data points are about its __(2)__."

    List: *interquartile range, standard deviation, mean, mode.*

    The answer for (1) is standard deviation and the answer for (2) is mean. Refer to the definition of mean, standard deviation, median, interquartile range and mode to see that 'the standard deviation is used to measure how widely spread the data points are about its mean.'

42. *(C, 2120, 2210)* The Registry of Marriages is interested to see the relationship between the ages of husbands and wives in City X. They randomly sampled 1000 pairs of husbands and wives from the population of City X and obtained data of their ages (in years). Looking through the data, they found that men always marry women who are younger than them.

    Based only on the information given above, which of the following statements must be true?

    (I) The average age of the husbands is more than the average age of the wives.

    (II) The standard deviation of husband's age is more than the standard deviation of wife's age.

    (A) Only (I).

    (B) Only (II).

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

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

    Answer is (A). We cannot tell anything about the spread of either the husbands' ages or the wives' ages just by knowing that men are marrying women younger than them. If all husbands are older than their wives, then it follows that the average age of the husbands is going to be more than the average age of the wives.

43. *(C, 2120, 2210*)* 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.

44. *(C, 2120, 2210)* The CEO of a company wishes to find out the level of job satisfaction that his employees have. His company has 876 employees. He administers an anonymised survey in which there are questions that ask about the employees' satisfaction with regards to various aspects of their jobs, including welfare, remuneration, career progression, learning etc.

    For each question, employees are to rate their satisfaction on a scale of 1-9. In this scale, 1 represents the lowest level of satisfaction whilst 9 represents the highest level of satisfaction. Broadly, any score below 5 for a question is regarded as "not satisfied" and any score above 5 for a question is regarded as "satisfied" while a score of 5 represents being "neutral".

    Every employee fills in the survey based on how he/she feels about the job and the data is collected. Assume that every employee is honest in the response and is fully aware of what the values on the scale represent.

    Which of the statements below describes an appropriate type of analysis for the data on the satisfaction scores?

(I) For each question, one can perform a summary statistics calculation on the satisfaction scores to obtain the mean, standard deviation, median, as well as the interquartile range. From these numerical summary statistics, we can conclude meaningful information on the employees' satisfaction as a whole.

(II) For each question, one can calculate the proportion of employees who gave each satisfaction score and determine what percentage of employees are "not satisfied" and what percentage are "satisfied".

(A) Only (I).

(B) Only (II).

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

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

Answer is (B). As satisfaction scores can be subjective and differences in satisfaction levels need not be the same, this variable cannot be treated as a numerical variable and is an ordinal variable. Hence, while we can compute summary statistics and perform arithmetic, after doing so, we are unable to establish any numerical meaning from these statistics. Thus, the type of analysis described in (I) is not appropriate.

For each question, it is appropriate to calculate the proportion of responses for each of the satisfaction scores and then determine the overall proportion of employees who are "satisfied" and "not satisfied" which is generally how we should deal with categorical variables. Thus, the type of analysis described in (II) is appropriate.

45. *(C, 2120, 2210)* We have learnt that the standard deviation and interquartile range (IQR) are examples of summary statistics that help us to quantify the spread of data points. However, they are not the only ways of quantifying spread and there are other summary statistics that can also help us to do this. For a numerical variable x, we can define the Mean Absolute Deviation (commonly abbreviated as MAD) using the formula

$$\text{Mean Absolute Deviation of } x = \frac{|x_1 - \overline{x}| + |x_2 - \overline{x}| + \cdots + |x_n - \overline{x}|}{n},$$

where $x_1, x_2, \ldots, x_n$ are values for the variable in a data set and n is the number of data points in the data set. The MAD is sometimes used in place of the standard deviation as a measure of quantifying the spread of the data. Based on the above formula, which properties **must** the MAD possess? Select all that apply.

(A) The MAD cannot take a negative value.

(B) The MAD does not change when a constant is added to all the data points.

(C) The MAD does not change when a constant is multiplied to all the data points.

(D) If the MAD is zero, then all the values of $x_1, x_2, \ldots, x_n$ in the data set are the same.

(A), (B) and (D) are correct. Based on the formula above, the MAD behaves very similarly to the standard deviation. Since we are taking absolute values, the MAD can never be negative.

If a constant is added to all the data points, then the constant is also added to the mean of the new data therefore the absolute difference between each point and the mean continues to remain the same.

When we multiply a constant to all the data points, the mean is also multiplied by the same constant, therefore the difference between each point and the mean is also multiplied by the same constant. Hence, the MAD is multiplied by the constant. The constant can be numbers other than 1 or -1, so the MAD can change.

Finally, if the MAD of x is zero, it means

$$\frac{|x_1 - \overline{x}| + |x_2 - \overline{x}| + \cdots + |x_n - \overline{x}|}{n} = 0,$$

which means

$$|x_1 - \overline{x}| + |x_2 - \overline{x}| + \cdots + |x_n - \overline{x}| = 0.$$

Since the absolute value of any number cannot be negative, we are adding numbers which cannot be negative to give us 0. Therefore, each number can only be zero which means every data point is equal to the mean.

46. *(C, 2120, 2210*)* A teacher has finished marking her students' test scripts for a Mathematics test. The maximum mark attainable for the test is 50. She records the following summary statistics for her class.

- Mean: 37.4
- Median: 35
- Standard deviation: 17.22
- Quartile 1: 23
- Quartile 3: 43
- Highest mark: 48
- Lowest mark: 16
- Range: $48 - 16 = 32$.

She returns the test papers to her class and goes through the answers. Whilst going through the answers, she realises that she has marked a question incorrectly for the whole class. She collects her students' scripts back and corrects her mistake. As a result, everyone in the class gets 2 additional marks. Which of the following summary statistics will change for the class? Select all that apply.

(A) Median.

(B) Standard deviation.

(C) Highest mark.

(D) Quartile 1.

(A), (C) and (D) are correct. We have learnt that standard deviation does not change when a constant is added to all the data points and the median will change by the amount that is added to/subtracted from it. Quartile 1 is the 25th percentile and it will also change when constants are added to/subtracted from it. The highest mark will increase from 48 to 50.

47. *(C, 2120, 2210*)* Suppose X is a numerical variable and the following are 10 data points for this variable.

$$4,\ 7,\ 4,\ 14,\ 10,\ 11,\ 17,\ 3,\ 8,\ r,$$

where r is a positive whole number that is unknown. Which of the following statements is/are always correct? Select all that apply.

(A) If the mean is greater than 8 then r must be greater than 2.

(B) If r is greater than 2, then the median must be greater than 8.

(C) The mean is always greater than the median regardless of the value of r.

(D) The mode is always greater than the median regardless of the value of r.

Only (A) is correct. The sum of the 10 data points is

$$4 + 7 + 4 + 14 + 10 + 11 + 17 + 3 + 8 + r = 78 + r.$$

If the mean is greater than 8, then $78 + r$ must be greater than 80, which implies r must be greater than 2. Arranging the 9 numbers excluding r in increasing order, we have

$$3,\ 4,\ 4,\ 7,\ 8,\ 10,\ 11,\ 14,\ 17.$$

Note that the median is a number m where 50% of the numbers are smaller than m. For example, if $r = 3$, the median is 7.5 thus (B) is incorrect. From above, since r is at least 1, the mean is at least 7.9. But if $r = 10$ for example, then the median is 9 which is higher than the mean. So (C) is incorrect. (D) is also incorrect since if $r = 4$, the mode is 4 while the median is 7.5.

48. *(C, 2120, 2210*)* A school consists of 53 classes. During a budget meeting, school board members decided to review class size information to determine budgeting for the classes. Let x be the numerical variable whose values are the number of students among the 53 classes. Summary statistics for x are shown in the table below.

$\overline{x}$	33.39 students
s_x	5.66 students
min	17
Q_1	29
median	33
Q_3	40
max	40

During the meeting, the following budget is set for classroom stationery supplies. Every class receives \$12 plus an additional \$0.75 for each student in the class. For example, a class with one student receives \$12.75, while a class of 40 students receives $\$12 + 40(\$0.75) = \$42$. Define a numerical variable y where

$$y = \$(12 + 0.75x).$$

Basically, y takes values that correspond to the amount of money that classes receive for their stationery supplies. Based on the summary statistics for x, which of the following statements must be true regarding the summary statistics of y? Select all that apply.

(A) The maximum value of y is higher than the third quartile of y.

(B) The median of y is $\$12 + 0.75(33) = \$36.75$.

(C) $\overline{y} = 12 + 0.75(33.39) = \$37.04$ (correct to 2 decimal places).

(D) The standard deviation of y is lower than the standard deviation of x.

(E) The IQR for y is the same as the IQR for x.

(B), (C) and (D) are correct. By basic properties of mean and median (Remark 1.4.3 and Remark 1.6.3 in the notes), the mean and median work out to be as what is indicated in (B) and (C). Since we are multiplying x by a factor of 0.75 in the process of obtaining y, the standard deviation and IQR will also change by a factor of 0.75 (see Remark 1.5.4 and Remark 1.6.7 in the notes) hence the standard deviation of y will be lower than that of x and so will the IQR.

49. *(C, 2120)* Let $x_1, x_2, \ldots, x_n$ be values of a numerical variable x within a data set containing n points. Which of the following statements are definitely true with regards to the standard deviation? Select all that apply.

(A) If the standard deviation of x is 0, then $x_i = \overline{x}$ for all i ranging from 1 through n.

(B) If the standard deviation of x is 0, then $x_i = 0$ for all i ranging from 1 through n.

(C) If $x_i = c$, for all i ranging from 1 through n, where c is a constant, then the standard deviation of x is 0.

(D) If the mean of x is 0 in the data set, then the standard deviation of x is also 0 in the data set.

(A) and (C) are correct. Examining the formula for standard deviation, we see that if $s_x = 0$ then the variance of x is 0. But this means

$$\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n - 1} = 0,$$

which translates to

$$(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2 = 0.$$

But since we are adding non-negative terms to give us zero, it means that each term must be zero and therefore every data point is equal to the mean. This is precisely statement (A). However, this does not necessarily mean that every data point must be 0 since the mean can be non-zero. So statement (B) is not always correct. For statement (C), if every data point is a constant c, then the mean is also c which gives us that the variance is 0 and therefore the standard deviation is zero. Finally for statement (D), a mean of zero does not mean a standard deviation of zero. Consider a small data set $\{-2, -1, 0, 1, 2\}$. The mean is zero but the standard deviation is non-zero in this case.

50. *(C, 2210)* A telecommunication company is interested in understanding how many mobile phones people own. Their population of interest is all 2000 people in town X. They took a random sample of 100 people from town X. Assuming there is 100% response rate, which of the following statements is/are correct?

    (I) If among the 100 people sampled, every person has 2 or more mobile phones, then the mean number of mobile phones in the sample will be greater than or equal to 2.

    (II) If the mean number of mobile phones in this sample is greater than or equal to 2, then everyone among the 100 people sampled has 2 or more mobile phones.

    (A) Only (I).

    (B) Only (II).

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

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

    Answer is (A). If all 100 numbers are greater than or equal to 2, the mean will also be greater than or equal to 2. But the opposite is not necessarily true. A large mean can result from some people having very large values, while others have low values. For example, if one person has 101 mobile phones and the rest have only one each, then the mean will still be 2.

51. *(C)* 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 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 $(3-1)+(4-1)+(5-1)$.

52. *(C)* 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.

53. *(C, 2210)* 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 $\times$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 $(50 + 3 \times 4) - (50 + 1 \times 4) = 8$.

54. *(C)* Consider the sample data set WEIGHT comprising the following numerical values:

$$48, 53, 39, 54, 55, 51.$$

Obtain WEIGHT2 and WEIGHT3 by multiplying all values in WEIGHT by 2 and 3 respectively. Which of the following statements is/are true? Select all that apply.

(A) The coefficient of variation of WEIGHT2 is the same as the coefficient of variation of WEIGHT.

(B) The coefficient of variation of WEIGHT3 is the same as the coefficient of variation of WEIGHT.

(C) The coefficient of variation of WEIGHT, correct to 3 decimal places, is 0.108.

(A) and (B) are correct. We compute the coefficient of variation of our sample data WEIGHT to be 0.119, correct to 3 decimal places. As it turns out, WEIGHT2 and WEIGHT3 have the same coefficient of variation. In fact, multiplying all values in WEIGHT by any positive number preserves the coefficient of variation. To see this, note that both the mean and the standard deviation of WEIGHT scale by the same factor r whenever we multiply all values in WEIGHT by a positive number r. Since the coefficient of variation equals standard deviation divided by mean, it remains unchanged.

55. *(C)* 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 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.