



UNIVERSITY OF GHANA

(All rights reserved)

B.SC INFORMATION TECHNOLOGY, FIRST SEMESTER EXAMINATIONS: 2015/

2016

CSIT 107: STATISTICS FOR IT PROFESSIONALS (3 CREDITS)

INSTRUCTION:

Section A: Answer ALL Questions in the Answer Booklet

Section B:

TIME ALLOWED:

TWO AND A HALF (2½) HOURS

SECTION A (40 marks)

Each question carries 1 mark.

1. The South African government is concerned about the high incidence of HIV/AIDS in South Africa. They wish to estimate the true number of people in South Africa who are HIV positive. A random sample of 5000 people was tested and 1980 of them were HIV positive. What is the parameter of interest?

A. The number of people in South Africa

- B. The 5000 people sampled
- C. The number of people in South Africa who are HIV positive
- D. The 1980 people sampled who were HIV positive
- E. None of the above

1. The South African government is concerned about the high incidence of HIV/AIDS in South Africa. They wish to estimate the true number of people in South Africa who are HIV positive. A random sample of 5000 people were tested and 1980 of them were HIV positive.

What is the statistic?

- A. The 5000 people sampled
- B. All the people in South Africa
- C. The 1980 people sampled who were HIV positive
- D. Impossible to calculate from the given information
- E. None of the above

1. The South African government is concerned about the high illiteracy rates amongst adults in South Africa. They wish to estimate the true number of adults (over 18 years of age) in South Africa who are illiterate (that is, they cannot read or write in at least one language). A random sample of 10000 adults were interviewed and 1107 of them were found to be illiterate.

What is the parameter of interest in this situation?

- A. The number of adults in South Africa
- B. The number of adults in South Africa who are illiterate
- C. The number of adults of the 10000 sampled who are illiterate
- D. The total number of people in South Africa

1. The South African government is concerned about the high illiteracy rates amongst adults in South Africa. They wish to estimate the true number of adults (over 18 years of age) in South Africa who are illiterate (that is, they cannot read or write in at least one language). A random sample of 10000 adults were interviewed and 1107 of them were found to be illiterate.

What is the statistic in this situation?

- A. The number of adults in South Africa
- B. The number of adults in South Africa who are illiterate
- C. The number of adults of the 10000 sampled who are illiterate
- D. The total number of people in South Africa
- E. None of the above

1. The science faculty is trying to estimate the proportion of their students who have access to a computer at home. A random sample of 300 science students were interviewed.

The population of interest in this experiment is:

- A. All registered UCT students who are in the science faculty
- B. All registered UCT students who have computers at home
- C. All registered UCT students
- D. The 300 science faculty students who were interviewed
- E. None of the above

1. The science faculty is trying to estimate the proportion of their students who have access to a computer at home. A random sample of 300 science students were interviewed. The statistic obtained in this experiment is:

- A. The proportion of the 300 science students interviewed who have computers at home
- B. The proportion of all UCT students who have a computer at home
- C. The proportion of all UCT students who were interviewed
- D. The proportion of the interviewed students who are in the science faculty
- E. None of the above

1. The commerce faculty is trying to estimate the proportion of their students who have access to the internet at home. A random sample of 500 commerce students were interviewed.

The population of interest in this experiment is:

- A. All registered UCT students who have computers at home
- B. All registered UCT students
- C. The 500 commerce faculty students who were interviewed
- D. All registered UCT students who are in the commerce faculty
- E. None of the above

1. The commerce faculty is trying to estimate the proportion of their students who have access to the internet at home. A random sample of 500 commerce students were interviewed.

The statistic obtained in this experiment is:

- A. The proportion all UCT students who have a computer at home
- B. The proportion of the 500 commerce students interviewed who have access to the internet at home
- C. The proportion of all UCT students who were interviewed
- D. The proportion of the interviewed students who are in the commerce faculty

1. Which of the following is an example of a ranked (ordinal) variable?

- A. Your favourite breakfast cereal
- B. Your monthly cell phone expenditure
- C. The rating (excellent, good, fair or poor) that you would give your stats lecturer
- D. The first three digits of your car's registration number
- E. Your favourite make of car

1. Which of the following is an example of a ranked (ordinal) variable?

- A. The rating (excellent, good, fair or poor) that you would give the service at a particular fast food chain
- B. Your favourite TV programme
- C. Your annual expenditure on clothing
- D. The first three digits of your cellphone number

E. Your monthly cellphone expenditure

2. A bimodal histogram is often an indication that:

- A. the data are incorrect
- B. the data are inconclusive
- C. the sample is not representative of the population
- D. the data come from two or more distinct populations

1. A histogram that has exactly two peaks is called a distribution.

- A. unimodal
- B. bimodal
- C. skewed
- D. scatter

2. An opinion variable expressed numerically on a 1-5 scale is a(n):

- A. histogram
- B. opinion poll
- C. categorical scale
- D. Likert scale

1. A variable is classified as ordinal if:

- A. there is a natural ordering of categories
- B. there is no natural ordering of categories
- C. the data arises from continuous measurements
- D. we track the variable through a period of time

1. A population includes:

- A. only people
- B. only households
- C. all objects of interest in a particular study
- D. only machines

2. Researchers try to gain insight into the characteristics of a population by examining a

_____ of the population.

- A. model
- B. sample
- C. description
- D. replica

3. Numerical variables can be subdivided into which two types?

- A. Diverse and categorical
- B. Discrete and continuous
- C. Nominal and progressive
- D. Cross-sectional and discrete

4. Individual observations within each category may be found in a frequency table.

- A. True
- B. False

5. In the term “frequency table,” frequency refers to the number of data values falling within each category.

A. True

B. False

6. In order for the characteristics of a sample to be generalized to the entire population, it should be _____ of the population.

A. symbolic

B. typical

C. representative

D. illustrative

7.

8. 20. A bimodal histogram is often an indication that:

9. a. the data are incorrect

10. b. the data are inconclusive

11. c. the sample is not representative of the population

12. d. the data come from two or more distinct populations

13.

14. 21. The median can also be described as:

15. a. the middle observation when the data is arranged in ascending order

16. b. the second quartile

17. c. the 50th percentile

18. d. all of the above

19. e. none of the above

20.

21. 22. For a boxplot, the box itself represents how many percent of the observations.

22. a. 25

23. b. 50

24. c. 75

25. d. 90

26. e. 100

27.

28. 23. Which of the following best describes a data warehouse?

29. a. A physical facility full of databases

30. b. A storage facility for government information

31. c. A database specifically structured to enable data mining

32. d. A database used to track warehousing information

33. e. None of the above

34.

35. 24. If all values in a data set are negative, the value of the standard deviation may be either positive or negative.

36. a. True

37. b. False

38.

39. 25. Assume that the histogram of a data set is symmetric and bell shaped, with mean of 72 and standard deviation of 10. Then, using the “rules of thumb”, we can say that 95% of the data values were between 52 and 92.

40. a. True

41. b. False

42.

43. 26. If a histogram has a single peak and looks approximately the same to the left and right of the peak, we should expect no difference in the values of the mean, median, and mode.

44. a. True

45. b. False

46.

47. 27. The mean is a measure of central location.

48. a. True

49. b. False

50.

51. 28. In a negatively skewed distribution, the mean is larger than the median and the median is larger than the mode.

52. a. True

53. b. False

54.

55. 29. Since the population is always larger than the sample, the population mean is always larger than the sample mean.

56. a. True

57. b. False

58.

59. 30. The length of the box in the boxplot portrays the interquartile range.

60. a. True

61. b. False

62.

63. 31. Express the indicated degree of likelihood as a probability value:

64. "There is a 40% chance of rain tomorrow."

65. a. 40

66. b. 0.60

67. c. 0.40

68. d. 4

69.

70. 32 A bag contains 6 red marbles, 3 blue marbles, and 7 green marbles. If a marble is randomly selected from the bag, what is the probability that it is blue?

71. a. $\frac{1}{3}$

72. b. $\frac{1}{7}$

73. c. $\frac{3}{16}$

74. d. $\frac{1}{13}$

75.

76. 33. Assume that one student in a class of 27 students is randomly selected to win a prize. Would it be "unusual" for you to win? (Assume "unusual" is a probability less than or equal to 0.05)

77. a. True

78. b. False

79.

80. 34. A bag contains 8 red marbles, 4 blue marbles, and 1 green marble. Find $P(\text{not blue})$.

81. a. $9/13$

82. b. 9

83. c. $13/9$

84. d. $4/13$

85.

86. 35. If A and B are dependent events, then $P(A \text{ and } B)$ is

87. a. $P(A) \cdot P(B|A)$

88. b. $P(A) \cdot P(B)$

89. c. $P(A) \cdot P(A|B)$

90.

91. 36. A company manufactures glass bottles. The probability that a randomly selected bottle is defective is 0.002. What is the probability that a randomly selected bottle is not defective?

92. a. 0.998

93. b. 0.980

94. c. 0.002

95. d. None of the above is correct.

96.

97. 37. Find ${}_{10}C_2$

98. a. 80,640

99. b. 40,320

100. c. 45

101. d. 5

102.

103. 38. There is a 65% chance of John passing mathematics. There is a 35% chance that John will pass both mathematics and statistics. There is a 70% chance that he will pass either mathematics or statistics or both. John has been informed that he has passed statistics, what is the probability that he will pass mathematics?

104. a. 0.58

105. b. 0.88

106. c. 0.55

107. d. 0.70

108. e. 0.15

109.

110. 39. There is a 45% chance of John passing mathematics. There is a 30% chance that John will pass both mathematics and statistics. There is a 70% chance that he will pass either mathematics or statistics or both. John has been informed that he has passed statistics, what is the probability that he will pass mathematics?

111. a. 0.58

112. b. 0.88

113. c. 0.55

114. d. 0.70

115. e. 0.15

116.

117. 40. There is a 45% chance of John passing mathematics. There is a 35% chance that John will pass both mathematics and statistics. There is a 60% chance that he will pass either mathematics or statistics or both. John has been informed that he has passed statistics, what is the probability that he will pass mathematics?

118. a. 0.58

119. b. 0.88

120. c. 0.55

121. d. 0.70

SECTION B

Answer Question B1 and any other Two (2) Questions from this section

B.1 (Compulsory)

A. The following are two sets of data, each with a sample of size $n = 9$

Set 1

| | | | | | | | | | |
|-------|----|----|----|----|----|----|----|----|----|
| | 5 | 2 | 4 | 2 | 3 | 2 | 10 | 2 | 12 |
| Set 2 | 15 | 12 | 14 | 12 | 13 | 12 | 20 | 12 | 22 |

- I. For each data, compute the mean, median, and mode (3pts)
- II. Compare your results in the previous question , and summarize your findings(1pts)
- III. Compare the first sampled item in each set, compare the second sampled item in each set, and so on. Briefly describe your findings in light of your summary in Question b.(2pts)
- IV. For each set, compute the range, interquartile range, variance and standard deviation.(3 pts)
- V. Describe the shape of each set.(1pts)

B. An aircraft emergency locator transmitter (ELT) is a device designed to transmit a

signal in the case of a crash. The Altigauge Manufacturing Company makes 80% of the ELTs, the Bryant Company makes 15% of them, and the Chartair Company makes the other 5%. The ELTs made by Altigauge have a 4% rate of defects, the Bryant ELTs have a 6% rate of defects, and the Chartair ELTs have a 9% rate of defects (which helps to explain why Chartair has the lowest market share).

- a. If an ELT is randomly selected from the general population of all ELTs, find the probability that it was made by the Altigauge Manufacturing Company. (3 marks)
- b. If a randomly selected ELT is then tested and is found to be defective, find the probability that it was made by the Altigauge Manufacturing Company.(3 marks)
- c. Write an expression that extends Bayes' theorem so that it can be used to find $P(A|Z)$, assuming that the initial event can occur in one of four ways: A, B, C, D.(3 marks)
- d. Define discrete random variable and continuous random variable(1 marks)

B2.

I. A fair die is tossed twice. Find the probability of getting:

a. a 4,5 or 6 on the first toss and,(2 marks)

b. a 1,2,3 or 4 on the second toss (2 marks)

I. Three balls are drawn from a box containing 6 red balls, 4 white balls, and 5 blue balls.

Find the probability that they are drawn in the order red, white, and blue if each ball is

a. Replaced and , (2 marks)

b. Not replaced (2 marks)

I. Find the probability of a 4 turning up at least once in two tosses of a fair die (2 marks)

I. A and B play 12 games of chess, of which 6 are won by A, 4 are won by B, and 2 end in a draw. They agree to play a match consisting of 3 games. Find the probability that

a. A wins all 3 games (2 marks)

- b. 2 games end in draw (2 marks)
- c. A and B win alternately (2 marks)
- d. B wins at least 2 game (2 marks)

B3.

The final grades in statistics of 80 students at UG Legon are recorded in the accompanying table

With reference to this table, find:

| | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|
| 68 | 84 | 75 | 82 | 68 | 90 | 62 | 88 | 76 | 93 |
| 73 | 79 | 88 | 73 | 60 | 93 | 71 | 59 | 85 | 75 |
| 61 | 65 | 75 | 87 | 74 | 62 | 95 | 78 | 63 | 72 |
| 66 | 78 | 82 | 75 | 94 | 77 | 69 | 74 | 68 | 60 |
| 96 | 78 | 89 | 61 | 75 | 95 | 60 | 79 | 83 | 71 |
| 79 | 62 | 67 | 97 | 78 | 85 | 76 | 65 | 71 | 75 |
| 65 | 80 | 73 | 57 | 88 | 78 | 62 | 76 | 53 | 74 |
| 86 | 67 | 73 | 81 | 72 | 63 | 76 | 75 | 85 | 77 |

- I. The highest grade (2 marks)
- II. The lowest grade (2 marks)
- III. The range (2 marks)
- IV. The grade of the five highest-ranking students (2 marks)
- V. The grade of the five lowest-ranking students (2 marks)
- VI. The grade of the student ranking tenth-highest. (2 marks)
- VII. The number of students who received grades of 75 or higher
- VIII. The number of students who received grades below 85 (2 marks)
- IX. The percentage of students who received grades higher than 85. (2 marks)
- X. The grades that did not appear at all. (2 marks)

B4. (20 marks)

An internationally accepted sports agency published a claim that the average the average height

of an NBA player was 6'6.98=200cm. A sample of 8 NBA athletes are taken aside and their heights taken .The values gotten are as follow: 199, 200, 201, 202, 203, 204, 204, and 206. Make use of the hypothesis testing method and the 95% confidence interval to verify and confirm if the claim published by the agency is acceptable and trust-worthy

B5.

I. What are the requirements for the following :

- a. Binomial distribution? (3 marks)
- b. Poisson distribution? (3 marks)
- c. Normal probability distribution? (3 marks)

II. How the Poisson distribution does differs from the binomial distribution?(4 marks)

III. When an adult is randomly selected, there is a 0.85 probability that this person knows what Twitter is. Suppose we want to find the probability that exactly three of five randomly selected adults know of Twitter.

- a. Does this procedure result in a binomial distribution? (2 marks)
- b. Find the probability (5 marks)