

Statistics Olympiad 2018 (Senior Level)

Time: Two hours

Instructions to the candidates:

- This paper consists of two parts (Part I and Part II).
- Part I consists of **Thirty (30)** multiple choice questions. Underline the most appropriate answer. Each correct answer is worth 5 marks.
- Part II consists of **Two (02)** questions. Write down the answer in the separate paper provided.
- Calculators are not allowed.

PART I

1. A numerical value used as a summary measure for a sample, such as sample mean, is known as a
 - (a) population parameter
 - (b) sample parameter
 - (c) sample statistic
 - (d) None of the above answers is correct
2. One use of a regression line is
 - (a) to determine if any x-values are outliers.
 - (b) to determine if any y-values are outliers.
 - (c) to determine if a change in x causes a change in y.
 - (d) to estimate the change in mean of y for a one-unit change in x.
3. Tickets numbered 1 to 20 are mixed up and then a ticket is drawn at random. What is the probability that a ticket drawn has a number which is a multiple of 3 or 5?
 - (a) $1/2$
 - (b) $2/5$
 - (c) $8/15$
 - (d) $9/20$

4. In a five number summary, which of the following is not used for data summarization?

- (a) the smallest value
- (b) the median
- (c) the 25th percentile
- (d) the mean

5. Three unbiased coins are tossed. What is the probability of getting at most two heads?

(a) $\frac{3}{4}$

(b) $\frac{1}{4}$

(c) $\frac{3}{8}$

(d) $\frac{7}{8}$

6. A list of 5 pulse rates is: 70, 64, 80, 74, 92. What is the median for this list?

- (a) 74
- (b) 76
- (c) 77
- (d) 80

7. Since the mode is the most frequently occurring data value,

- (a) it can never be larger than the mean
- (b) it is always larger than the mean
- (c) it can never be larger than the median
- (d) None of the above answers is correct

8. What is one of the distinctions between a population parameter and a sample statistic?
- (a) A population parameter is only based on conceptual measurements, but a sample statistic is based on a combination of real and conceptual measurements.
 - (b) A sample statistic changes each time you try to measure it, but a population parameter remains fixed.
 - (c) A population parameter changes each time you try to measure it, but a sample statistic remains fixed across samples.
 - (d) The true value of a sample statistic can never be known but the true value of a population parameter can be known.
9. Which one of these statistics is unaffected by outliers?
- (a) Mean
 - (b) Interquartile range
 - (c) Standard deviation
 - (d) Range
10. At a high school with 200 students, 32 play soccer, 18 play basketball and 8 play both sports. If a student is selected at random, find the probability that a student plays soccer or basketball.
- (a) $1/100$
 - (b) $1/4$
 - (c) $4/25$
 - (d) $1/5$
11. Since the population size is always larger than the sample size, then the sample statistic
- (a) can never be larger than the population parameter
 - (b) can never be equal to the population parameter
 - (c) can never be zero
 - (d) None of the above answers is correct.
12. The probability that a family visits City Museum is 0.36 and the probability that a family rides on the ferry is 0.47. The probability that a family does both is 0.22. Find the probability that a family visits the museum or rides the ferry:
- (a) 0.83
 - (b) 0.61
 - (c) 0.58
 - (d) 0.69

13. The value of a correlation coefficient is reported by a researcher to be $r = -0.5$. Which of the following statements is correct?

- (a) The x-variable explains 25% of the variability in the y-variable.
- (b) The x-variable explains -25% of the variability in the y-variable.
- (c) The x-variable explains 50% of the variability in the y-variable.
- (d) The x-variable explains -50% of the variability in the y-variable.

14. Past data has shown that the regression line relating the final exam score and the midterm exam score for students who take statistics from a certain professor is:

$$\text{final exam score} = 50 + 0.5 \times \text{midterm score}$$

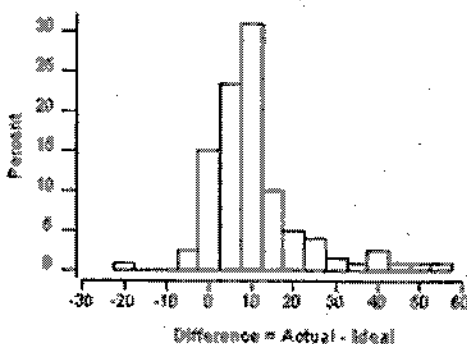
One interpretation of the slope is

- (a) a student who scored 0 on the midterm would be predicted to score 50 on the final exam.
- (b) a student who scored 0 on the final exam would be predicted to score 50 on the midterm exam.
- (c) a student who scored 10 points higher than another student on the midterm would be predicted to score 5 points higher than the other student on the final exam.
- (d) students only receive half as much credit (.5) for a correct answer on the final exam compared to a correct answer on the midterm exam.

15. A coin is tossed six times. What is the probability of getting six tails?

- (a) $1/2$
- (b) $1/4$
- (c) $1/12$
- (d) $1/64$

16. The following histogram shows the distribution of the difference between the actual and "ideal" weights for 119 female students. Notice that percent is given on the vertical axis. Ideal weights are responses to the question "What is your ideal weight"? The difference = actual - ideal.



What is the approximate shape of the distribution?

- (a) Nearly symmetric.
- (b) Skewed to the left.
- (c) Skewed to the right.
- (d) Bimodal (has only two peaks).

17. The mean of a distribution is 23, the median is 24 and the mode is 25.5. It is most likely that this distribution is

- (a) Positively skewed
- (b) Symmetrical
- (c) Asymptotic
- (d) Negatively Skewed

18. Which of the following describe the middle of a group of numbers?

- (a) Measure of Variability
- (b) Measure of Central Tendency
- (c) Measure of Association
- (d) Measure of shape

19. According to the empirical rule, approximately what percent of the data should lie within $\pm\sigma$?

- (a) 75%
- (b) 68%
- (c) 99.7%
- (d) 95%

20. The sum of the deviation about the mean is always:

- (a) Range
- (b) Zero
- (c) Total standard Deviation
- (d) Positive

21. Which of these represent qualitative data

- (a) Height of a student
- (b) Liking or disliking of a product
- (c) Income of government employee
- (d) Yield from wheat plot

22. Given $P(A) = 0.4$, $P(B) = 0.5$ and $P(A \cup B) = 0.9$, then:

- (a) A and B are not mutually exclusive events
- (b) A and B are equally likely events
- (c) A and B are independent events
- (d) A and B are mutually exclusive events

23. Two events A and B are called mutually exclusive if:

- (a) $A \cup B = \Phi$
- (b) $A \cap B = \Phi$
- (c) $A \cap B = S$
- (d) $A \cap B = 1$

24. When A and B are two non-empty and mutually exclusive events, then:

- (a) $P(A \cup B) = P(A) \cdot P(B)$
- (b) $P(A \cup B) = P(A) + P(B)$
- (c) $P(A \cap B) = P(A) \cdot P(B)$
- (d) $P(A \cap B) = P(A) + P(B)$

25. Two dice are rolled. Probability of getting total less than 4 or total more than 10 is given by:

- (a) $10/36$
- (b) $4/36$
- (c) $1/36$
- (d) $14/36$

26. If A and B are any two events, then $P(\bar{A} | B)$ is equal to:

- (a) $P(A/B)$
- (b) $1 - P(A/B)$
- (c) $1 + P(A/B)$
- (d) $P(\bar{A} \cap B)$

27. If A and B are two non-independent events, then the probability that both A and B will happen together is:

- (a) $P(A \cap B) = P(A)P(B/A)$
- (b) $P(A \cap B) = P(A)P(B)$
- (c) $P(A \cap B) = P(A) + P(B)$
- (d) $P(A \cap B) = P(A)$

28. The digits 1, 2, 3, 4, 5 are the roll numbers of 5 students. These roll numbers are written on the paper slips and two paper slips are selected at random without replacement. The possible combinations are:
- (a) 5
 - (b) 2
 - (c) 25
 - (d) 10
29. Five cards are selected at random from a pack of 52 cards with replacement. The possible combinations are:
- (a) 52
 - (b) $(52)^5$
 - (c) 52×52
 - (d) $(5)^{52}$
30. If first quartile and third quartile are 32 and 35 respectively with median of 20, then the distribution is skewed to
- (a) lower tail.
 - (b) upper tail.
 - (c) close end tail.
 - (d) open end tail.

Part II

1. (a) Let A and B be two random events. Define the conditional probability $P(A|B)$ when $P(B) > 0$.
- (b) Show that

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2)$$

for three random events A_1, A_2, A_3 .

- (c) According to the Report of the Health Ministry of a certain country, 7.0% of the population has lung cancer. Of those having lung cancer, 90.0% are smokers; of those not having lung cancer, 25.3% are smokers.
- (i). Find the probability that the person selected randomly is a smoker,
- (ii). Determine the probability that a randomly selected smoker has a lung cancer.
2. Shown here is a frequency distribution for the number of inches of rain received in year 2000 in 25 selected cities.

Number of inches	Frequency
05.5 – 20.5	2
20.5 - 35.5	3
35.5 -50.5	8
50.5 – 65.5	6
65.5 – 80.5	3
80.5 – 95.5	3

- (i). Construct the histogram and the frequency polygon for the above distribution.
- (ii). Using a suitable coding method, find the mean, the mode, and the standard deviation of this distribution.
- (iii) Later, it was found that each rain fall amount in the above distribution was 2 more than the actual amount. Find the mean and the standard deviation of the actual distribution.