

**UNIVERSITY OF SWAZILND**  
**FACULTY OF HEALTH SCIENCES**  
**FINAL EXAMINATION PAPER, MAY 2012**  
**TITLE OF PAPER: HEALTH STATISTICS**

**COURSE CODE: HSC 404**

**TIME ALLOWED: TWO (2) HOURS**

**MARKS: 75**

**THIS EXAM PAPER HAS TEN (13) PAGES**

**INSTRUCTIONS:**

- 1. THERE ARE THREE (3) QUESTIONS IN THIS PAPER**
- 2. ANSWER ALL THREE QUESTIONS**
- 3. EACH QUESTION IS ALLOCATED 25 MARKS**
- 4. WRITE LEGIBLY**
- 5. ALL FINAL ANSWERS MUST BE TO THE NEAREST 1 DECIMAL, SHOW ALL YOUR CALCULATIONS.**

**THIS PAPER IS NOT TO BE OPENED UNTIL THE INVIGILATOR HAS GRANTED PERMISSION.**

### QUESTION 1

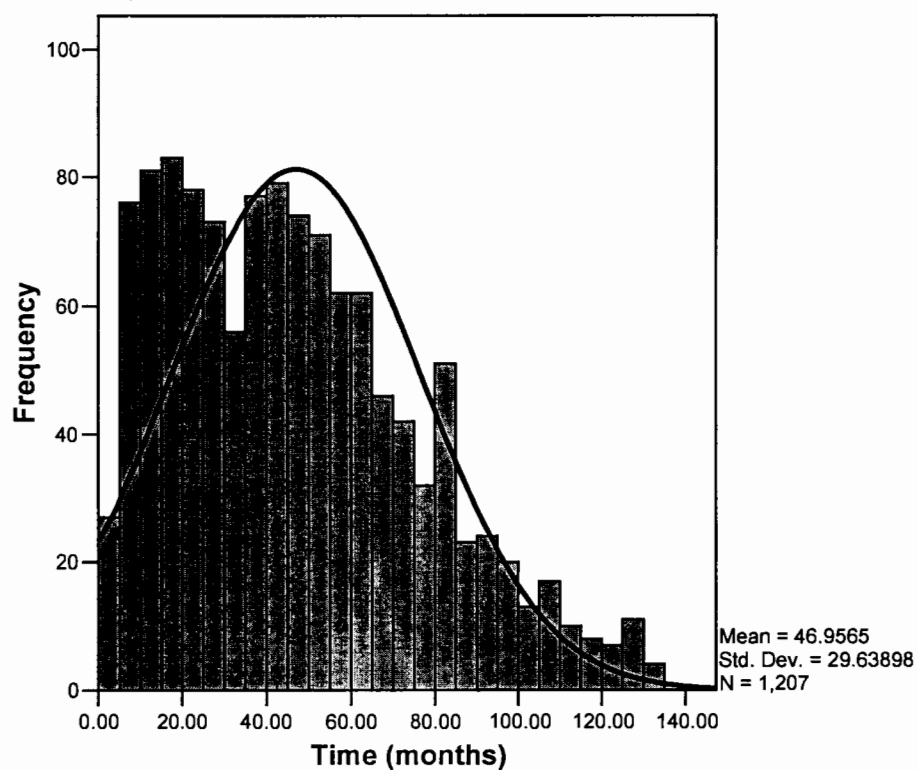
1. Among the variables which the investigator examined among breast cancer survivors, was the period (months) the person had suffered from breast cancer.

#### Statistics

Time (months)

N	Valid	1207
	Missing	0
Mean		46.9565
Median		42.9667
Mode		18.67
Std. Deviation		29.63898
Skewness		.631
Std. Error of Skewness		.070
Minimum		2.63
Maximum		133.80

#### Histogram



A.

- (i) Explain the distribution and justify your explanation with values (scores/figures). (3)
- (ii) What was the range for the number of months that the participant (s) suffered from breast cancer? (2)

B. The investigator determined the relationship between the respondent's age (independent variable) and positive axillary lymph nodes (dependent variable). The correlation findings are presented below.

		Age (years)	Positive Axillary Lymph Nodes
Age (years)	Pearson Correlation	1	-.120(**)
	Sig. (2-tailed)	.	.000
	N	1207	1207
Positive Axillary Lymph Nodes	Pearson Correlation	-.120(**)	1
	Sig. (2-tailed)	.000	.
	N	1207	1207

\*\* Correlation is significant at the 0.01 level (2-tailed).

- (i) Present the study findings. (3)
- (ii) Describe the meaning of the study findings. (3)

C. A researcher in an alcoholism treatment care, is interested in summarizing the length of stay in the rehabilitation centre for the first-time clients, randomly selects ten (10) charts of clients institutionalized within the previous two years. The length of stay in the rehabilitation centre, is in days, is as follows: 32, 18, 55, 17, 21, 31, 20, 40, 24, and 15. For this data set compute the:

- (i) Percentile rank of 21 days (2)
- (ii) Mean deviation and interpret (3)
- (iii) Standard deviation (3)

- (iv) Skewness and interpret (3)
- (v) Coefficient of variation (3)

**TOTAL = 25 MARKS**

**QUESTION 2**

A.

- (i) At your health facility there were 15 000 individuals who were screened for HIV in 2001. Only 724 of the individuals tested negative of HIV antibodies. What is the probability of being tested HIV positive at your health facility? Show your calculations. (4)
- (ii) In 2011, at the Mbabane Outpatient Department [OPD] 18 523 clients who complained of chest tightness were attended. The probability of diagnosed as asthmatic was 23.5%. How many clients were diagnosed asthma. Show your calculations. (4)

B. The investigator studied the relationship between the independent variable, age, and the dependent variable, general happiness. The findings from regression analysis are presented below.

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.036(a)	.001	.001	17.821

a Predictors: (Constant), General Happiness

**ANOVA(b)**

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	613.974	1	613.974	1.933	.165(a)
	Residual	476364.417	1500	317.576		
	Total	476978.391	1501			

a Predictors: (Constant), General Happiness

b Dependent Variable: Age of Respondent

(i) What conclusion (s) can you draw from the summary model? (3)

(ii) Present and interpret the ANOVA output. (4)

C. Identify the level of measurement for each of the following (also state if data is continuous or discrete):

(i) Years in school

(ii) IQ scores

(iii) Life expectancy

(iv) Hair color

(v) Temperature (5)

D. Identify each of the following variables as qualitative or quantitative (also state if data is continuous or discrete).

- (i) Number of pets in family
- (ii) Number of people you have known with HIV/AIDS (0, 1, 2, 3, 4 or more)
- (iii) Country of birth
- (iv) Distance (in kilometers) commute to school
- (v) Choice of diet (vegetarian, nonvegetarian)

(5)

**TOTAL = 25 MARKS**

**QUESTION 3**

Compute the answer as required **OR** clearly write the letter that corresponds with the most appropriate answer to the statement /question e.g. 2. C.

1. \_\_\_\_\_ results if you accept a false the null hypothesis.
  - A. Type I error
  - B. Type II error
  - C. Type III error
  - D. Type IV
  
2. Which of the following statements is / are TRUE according to the logic of hypothesis testing?
  - A. When the null hypothesis is true, it should be rejected.
  - B. When the null hypothesis is true, it should not be rejected.
  - C. When the null hypothesis is false, it should be rejected.
  - D. When the null hypothesis is false, it should not be rejected.
  - E. Both B and C are true
  
3. A Type I error is also known as \_\_\_\_\_.
  - A. False positive
  - B. False negative
  - C. Double negative
  - D. None of the above
  
4. The standard deviation is:
  - A. The square root of the variance
  - B. A measure of variability

- C. An approximate indicator of how numbers vary from the mean
- D. All of the above
5. Hypothesis testing is an example of descriptive statistics.
- A. True
- B. False
6. The \_\_\_\_\_ is often the preferred measure of central tendency if the data is severely skewed.
- A. Mean
- B. Median
- C. Mode
- D. Range
7. Which of the following is the formula for range?
- A.  $(H / L) + 1$
- B.  $(L \times H) - 1$
- C.  $(L + H) - 1$
- D.  $(H - L) + 1$
8. Which measure of central tendency takes into account the magnitude of scores?
- A. Mean
- B. Median
- C. Mode
- D. Range



9. If a test was generally easy, except for a few students who had very low scores, then the distribution of scores would be \_\_\_\_\_.

- A. Positively skewed
- B. Negatively skewed
- C. Not skewed at all
- D. Normal

10. When a set of numbers is heterogenous, you can place more trust in the measure of central tendency as representing the typical person or unit.

- A. True
- B. False

11. What is the mean of the following numbers? 12.5, 13, 11, 10.3.

Mean = \_\_\_\_\_

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

- A. Negatively skewed
- B. Positively skewed
- C. Symmetrical
- D. Normal

13. What is the median of the following numbers? 83, 54, 48, 60.

Median = \_\_\_\_\_

14. The middle value of an ordered array of numbers is the:

- A. Mean
- B. Median
- C. Mode
- D. Midpoint

15. Consider the following ungrouped data. What is the mode? 16, 35, 15, 20.

Mode = \_\_\_\_\_

16. The sum of deviations about the mean are always:

- A. The range
- B. The standard deviation total
- C. Zero
- D. Positive

17. Which of the following statements is NOT CORRECT?

- A. Nominal variables consist of labels or names used to identify a property of an element in a data set.
- B. An ordinal scale refers to nominal data that is ordered or ranked in a meaningful way.
- C. Qualitative data consist of nominal variables
- D. Age and monthly earnings of workers are qualitative data.

18. In general, which of the following statements is FALSE?
- A. The sample mean is more sensitive to extreme values than the median
  - B. The sample range is more sensitive to extreme values than the standard deviation
  - C. The sample standard deviation is a measure of spread around the sample mean
  - D. The sample standard deviation is a measure of central tendency around the median
19. The term test scores of 15 students enrolled in a Health Statistics class were recorded in ascending order as follows: 4, 7, 7, 9, 10, 11, 13, 15, 15, 15, 17, 17, 19, 19, 20.
- After calculating the mean, median, and mode, an error is discovered: one of the 15's is actually a 17. The measures of central tendency which will change are:
- A. The mean only
  - B. The mode only
  - C. The median only
  - D. The mean and mode
  - E. All the three measures
20. The heights in centimeters (cm) of 5 students are: 165, 175, 176, 159, and 170.
- The sample median and sample mean are respectively:
- A. 170, 169
  - B. 170, 170
  - C. 169, 170
  - D. 176, 169

21. Which of the following statements is NOT true?
- A. In a symmetric distribution, the mean and the median are equal
  - B. The first quartile is equal to the 25th percentile
  - C. The median is always greater than the mean
  - D. In a symmetric distribution, the median is halfway between the first and the third quartiles.
22. A researcher randomly assigned 50 students to take a test in a hot room and 50 students to take the same test in a cold room. The researcher recorded the percent correct on the test. The researcher gave each group 50 minutes to take the test. Which of the following is a **CORRECT** statement?
- A. The independent variable is the temperature of the room -- hot vs. cold.
  - B. The independent variable is the percent correct on the test.
  - C. The independent variable is that each group had 50 minutes to take the test.
  - D. There are no independent variables in this study. There are only subject variables.
23. If we want to generalize what we know about a sample to a population, we must ensure that
- A. The sample is sufficiently large.
  - B. The sample is selected in a random manner from the population.
  - C. The sample consists of the entire population.
  - D. Answers A and B.

24. The median is another name for

- A. The difference of the first and third quartiles.
- B. The 50<sup>th</sup> percentile.
- C. The mean.
- D. The square root of variance

25. What sample statistic is used to estimate a population value?

- A. Parameter
- B. Sampling error
- C. Point estimate
- D. Interval estimate
- E. None of the above

**TOTAL = 25 MARKS**