UNIVERSITY OF SWAZILAND

SUPPLEMENTARY EXAMINATION PAPER 2005

TITLE OF PAPER	:	DESCRIPTIVE/INFERENTIAL STATISTICS
COURSE CODE	:	ST230/IDE-ST230-1&2
TIME ALLOWED	:	3 (THREE) HOURS
REQUIRMENTS	:	STATISTICAL TABLES AND CALCULATOR
INSTRUCTIONS	:	ANSWER BOTH QUESTIONS FROM SECTION ONE AND ANSWER ANY THREE (3) QUESTIONS FROM SECTION TWO AND SECTION THREE, EACH. ALL QUESTIONS CARRY MARKS AS GIVEN WITHIN THE PARENTHESIS.

THIS PAPER IS NOT TO BE OPENED UNTIL PERMISSION HAS BEEN GRANTED BY THE INVIGILATOR

SECTION ONE

ANSWER BOTH QUESTIONS:

QUESTION ONE.   [20 marks]

State the most correct answer for each of the following:

1.1   The number 153 has a leaf value of

        a. 15
        b. 150
        c. 153
        d. 3

1.2   Consider the following frequency distribution:

Items	Frequency
10 - 19	4
20 - 29	12
30 - 39	6
40 - 59	4
60 - 79	3

If the height of the rectangle of a histogram representing the second class is 12 units, then the height of the rectangle representing the fourth class would be:

        a. 4 units
        b. 10 units
        c. 2 units
        d. None of the above

1.3   The lower quartile of the data   [11 13 14 15 17 18 20 22] is

        a. 11
        b. 12
        c. 13
        d. 2.5

1.4   The data for this question is per the following time:

Quarter:	1	2	3	4	1	2
Demand:	24	26	32	38	20	24

The trend value can be calculated for the quarter(s):

        a. 2
        b. 3
        c. 4
        d. both (b) and (c).

1.5   Which one of the following statistics is not affected by extreme values?

        a. Interquartile range
        b. Standard deviation
        c. Mean
        d. Both (b) and (c)

1.6   Which of the following statements is true in a Bar Chart?

        a. Height of the bar is proportional to the frequency
        b. Width of the bar is arbitrary
        c. There are equal gaps between bars
        d. All of the above

1.7   Three instructors are comparing scores on their finals; each had 99 students. In class **A**, one student got 1 point, another got 99 points, and the rest got 50 points. In class **B**, 49 students got a score of 1, one student got a score of 50, and 49 students got a score of 99. In class **C**, one student got a score of 1, one student got a score of 2, one student got a score of 3, and so forth, all the way through 99. Which class has the biggest standard deviation?

        a. Class **A**
        b. Class **B**
        c. Class **C**
        d. Not possible to determine without computation

1.8   Suppose you are given a list of scores of ST230 test. The highest score is 98 and the lowest is 41. You are asked to make a frequency table. Which one of the following statements will represent the data best if the frequency distribution includes

        a. the first class interval as 0-9.
        b. the first class interval as 40-49.
        c. 6 class intervals with a length of 10 points.
        d. both (b) and (c).

1.9   Consider these seven observations: 2, 2, 3, 1, 5, 0, 2. Now, 2 is the _____ of those observations.

        a. mean
        b. median
        c. mode
        d. median and mode

1.10   Which of the following is at ordinal level?

        a. Students' major subject
        b. Room numbers in residential halls
        c. Number of school years attended
        d. Annual rainfall in Swaziland

1.11   In the standard normal distribution, the area outside the range $Z = -1.5$ to $Z = +1.5$ is:

      a. 0.8664
      b. 0.1336
      c. 0.9332
      d. 0.0668

1.12   A bag contains 2 red, 5 white and 3 green balls. If two balls are selected without replacement the probability that the first ball is red and the second is green is:

      a. 2/15
      b. 1/15
      c. 4/9
      d. 8/10

1.13   Which of the following statements is not true about binomial distribution:

      a. It is a probability distribution for a discrete random variable.
      b. If the sample size is very large, we can use the normal distribution as an
          approximation to the binomial distribution.
      c. The mean of a binomial distribution is npq.
      d. Both (a) and (b).

1.14   Samples of 64 items are drawn from a population with $\mu = 144$ and $\sigma = 24$. The value of the standard error of the mean is

      a. 144
      b. 64
      c. 24
      d. 3

1.15   In the large sample case, the constant used in a 99% confidence interval is:

      a. 1.645
      b. 1.960
      c. 2.326
      d. 2.576

1.16   When a 95% confidence interval is calculated instead of a 99% confidence interval with *n* being the same, the maximum error of estimate will be

      a. smaller
      b. larger
      c. the same
      d. difficult to determine without additional information.

1.17 Two different investigators are working on a growth study. The first measures the heights of 100 children in inches. The second prefers the metric system, and changes the results to centimeters, using the conversion factor 2.54 cm/inch. If no mistakes are made in the conversion, what is the correlation between the two sets of measurements?

     a. $0.5 < r < 1.0$
     b. $r > 1.0$
     c. $r = 1.0$
     d. $r = 0$

1.18 Which is not a property of the normal distribution?

     a. It is symmetric about the mean.
     b. It is uniform.
     c. It is bell-shaped.
     d. It is unimodal.

1.19 The sample space for tossing two coins consists of how many outcomes?

     a. 2
     b. 4
     c. 6
     d. 8

1.20 The alternative hypothesis for the Chi-square test of independence is that the variables are

     a. dependent.
     b. not related.
     c. independent.
     d. none of the above.

QUESTION TWO.   [20 marks]

State which of the following statements are **TRUE** and which are **FALSE**?

2.1   If the class "20, but under 30" is used in a frequency distribution, then by implication the data is continuous.

2.2   The rectangles of a histogram are always of uniform width.

2.3   The sum of the deviations from the mean is always a non-zero number.

2.4   The median cannot be affected by extreme values in a data set.

2.5   The mode will always have a unique value.

2.6   The lower quartile is the value such that 25% of items in a set have this value or less.

2.7   The inclusion of additional observations in a data set will always change the value of the interquartile range.

2.8   If we know the standard deviation, we can always calculate the coefficient of variation.

2.9   The adjusted seasonal variates must add up to zero.

2.10   If the value of an index number is 100, it must be the index number for the base year.

2.11   Binomial tables cannot be used to find the probability that the random variable has a particular value.

2.12   In any Poisson distribution, the mean is equal to the standard deviation.

2.13   A confidence interval converts a point estimate into an interval estimate.

2.14   If the largest possible samples are drawn, then the standard error of the sample statistics will have its maximum value.

2.15   Increasing the level of confidence involves narrowing the range for sample results.

2.16   The standard normal distribution has a mean of one.

2.17   Increasing the constant in the interval estimate increases the size of the sampling error.

2.18   The higher the level of significance, the more often the null hypothesis would be accepted.

2.19   The scatter diagram is a useful illustration of any relationship that may exist between variables.

2.20   If the null hypothesis is true, the value of the Chi-square Test Statistics is zero.

SECTION TWO

ANSWER ANY THREE QUESTIONS: (You must show all your works in order to obtain full marks.)

QUESTION THREE.   [10 marks]

Consider the following table (use 2002 as base year):

Commodity	2000		2003	
	Prices (E per kg)	Quantity Purchased	Prices (E per kg)	Quantity Purchased
Milk	10.90	10	12.75	18
Butter	26.30	8	28.90	10
Cheese	35.40	5	45.80	8

Compute Fisher's price index.

QUESTION FOUR.   [6 + 4 marks]

The following table shows the quarterly sales of a company in thousands of tons for a period of four years:

Sales:			Quarter 1	2	3	4
	Year	1	70	41	52	83
		2	78	44	48	85
		3	83	54	51	96

(a)   Find the trend using four-quarterly moving average

(b)   Find the average seasonal variation.

QUESTION FIVE.   [2 + 4 + 4 marks]

The number of telephone calls received by a firm on 30 randomly chosen mornings:

43	54	67	53	61	71	46	61	55	56
58	60	66	54	68	52	36	64	51	52
57	52	63	59	69	70	42	66	58	53

(a)   Construct a Stem and Leaf diagram for the above data.
(b)   Find the median and the mode.
(d)   Find the lower and upper quartiles.

QUESTION SIX.   [4 + 4 + 2 marks]

Consider the following grouped frequency distribution:

Class limits	Frequency
Less than 5	7
Less than 8	22
Less than 11	44
Less than 14	58
Less than 17	60

(a)   Compute the mean.
(b)   Compute the standard deviation.
(d)   Compute the coefficient of variation.

QUESTION SEVEN.   [4 + 6 marks]

The following are daily numbers of automobiles rented by an automobile rental company in 100 business days.

Automobile rentals	Number of days
30, but under 35	3
35, but under 40	20
40, but under 45	24
45, but under 50	25
50, but under 55	14
55, but under 60	9
60, but under 65	5

(a)   Draw a Histogram and a Frequency Polygon, considering equal classes.
(b)   Find the median and mode number of automobile rented.

SECTION THREE

ANSWER ANY THREE QUESTIONS: (You must show all your works in order to obtain full marks.)

QUESTION EIGHT.   [1 + 3 + 3 + 3 marks]

In a sample of 50 people, 21 had type O blood, 22 had type A blood, 5 had B blood and 2 had type AB blood. Find the probability that:

(a) A person has type O blood.
(b) A person has type A or type B blood.
(c) A person has neither type A nor type O blood.
(d) A person does not have type AB blood

QUESTION NINE.   [4 + 6 marks]

(a) A company manufactures steel rods of 500 cm length, each one having two flaws on average. The process ultimately divides these rods into 100 cm segments. What is the probability that a particular 100 cm segment will have no flaws?

(b) A buying department is considering an acceptance sampling scheme for incoming lots of a manufactured item that can be classified as either good or defective. The plan calls for a random sample of 50 items from each lot. If there is one or less defective items in the sample than the lot is accepted, otherwise it is rejected. Using the binomial distribution, find the probability of rejecting a lot that is 1% defective.

QUESTION TEN.   [4 + 6 marks]

A motor car manufacturer purchases gear assemblies from a sub-contractor who undertakes to ensure that not more than 5% of his supplies will be defective. In order to provide a check on the quality of incoming supplies a random sample of 200 assemblies is selected of which 17 are found to be defective.

(a)   Construct an interval estimate the proportion of all the sub-contractor's supplies that are defective. Use a confidence level of 95% .

(b)   Does the sample evidence indicate that the sub-contractor is not maintaining the quality of his supplies at the agreed level? Use the significance level of 0.01 .

QUESTION ELEVEN.   [6 + 4 marks]

Consider the following results from a sample of size 7:

$$\Sigma X = 57 \qquad \Sigma Y = 511 \qquad \Sigma X^2 = 579 \qquad \Sigma XY = 3745 \qquad \Sigma Y^2 = 38993$$

(a)   Estimate the linear regression line.
(b)   Compute the coefficient of correlation (r) and interpret the answer.

## QUESTION TWELVE.   [10 marks]

In an experiment to study the dependence of hypertension on smoking habits, the following data were taken on 180 individuals:

Hypertension	Smoking	Habits	
	Heavy	Moderate	Nonsmokers
Yes	30	36	21
No	19	26	48

Test the hypothesis that the presence of hypertension is independent of smoking habits. Use a 0.05 level of significance.

Normal curve areas
Standard normal probability in right-hand
tail (for negative values of z areas are found
by symmetry)
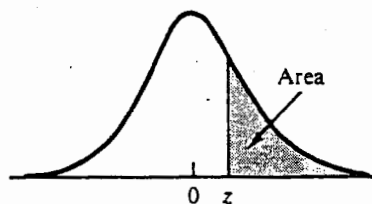
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
						Second decimal place of z				
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014

z	area		
3.0	.00135		
3.5	.000 233		
4.0	.000 031 7		
4.5	.000 003 40		
5.0	.000 000 287		

Selected χ^2-values

$\chi^2_{.05}$	df	$\chi^2_{.05}$	df
3.84	1	11.07	5
5.99	2	12.59	6
7.81	3	14.07	7
9.49	4	15.51	8

From R. E. Walpole, *Introduction to Statistics* (New York: Macmillan, 1968).