UNIVERSITY OF SWAZILAND

MAIN EXAMINATION PAPER 2015

TITLE OF PAPER	:	DESCRIPTIVE STATISTICS
COURSE CODE	:	STA131/IDE-ST 132
TIME ALLOWED	:	TWO (2) HOURS
REQUIREMENTS	:	CALCULATOR
INSTRUCTIONS	:	THIS PAPER HAS FIVE (5) QUESTIONS. ANSWER ANY FOUR (4) QUESTIONS.

Question 1 **[25 marks, 5+10+5+5]**

(a) The price of a kilogram of sugar increased by 5%, 12%, 6%, 4%, 9% and 3% over the past 6 years. Find the *average annual percentage increase* in the price of sugar (per kg).

(b) The management of an office complex in Manzini wants to understand the pattern of energy consumption in the complex. They have assembled quarterly data on energy costs for three years (in SZL10000).

Year	Quarter	Energy Costs	Moving average
2012	**Summer**	2.4	
	Autumn	3.8	
	Winter	4.0	3.35
	Spring	3.1	3.4125
2013	**Summer**	2.6	3.4625
	Autumn	4.1	a
	Winter	4.1	3.5
	Spring	3.2	3.55
2014	**Summer**	b	3.625
	Autumn	4.5	3.6625
	Winter	4.3	
	Spring	3.3	

Compute the *seasonal variates* for the complex's energy costs. Comment on their values.

(c) Find the average price paid per share in an equity portfolio consisting of: 40 shares bought for SZL15 each; 10 shares bought for SZL20 each; 5 shares bought for SZL40 each; and 50 shares bought for SZL10 each.

(d) Which of the following statements are true and which are false. If the *median* mass of 5 parcels for delivery by a courier service in 6.5 kilograms and one further 7 kilogram parcel is added to the consignment, then:

(i) the new median mass will be about 6.6 kilograms

(ii) the median will increase

(iii) it is impossible for the new median mass to be less than it was

(iv) it is impossible for the new median mass to stay exactly at 6.5 kilograms

(v) the median may increase, but that depends on the actual masses of all 6 parcels.

Question 2 **[25 marks, 2+6+8+4+5]**

(a) A bank wanted to find out whether *loan applications received* are influenced by the current loan *interest rate*. The manager selected 11 monthly periods where different interest rates applied and recorded the number of loan applications received. The data are given in the following table:

Loan applications received	
Interest rate %	**Loan applications**
7	18
6.5	22
5.5	30
6	24
8	16
8.5	18
6	28
6.5	27
7.5	20
8	17
6	21

(i) Identify the dependent variable (y) and the independent variable (x).

(ii) Compute the correlation coefficient between the *rate of interest* and number of *loan applications received*. Comment on the strength of the association.

(iii) Derive the least squares regression line between the *rate of interest* and number of *loan applications received*. Interpret the regression coefficients.

(iv) How many loan applications can the bank expect to receive when the interest rate is 6%..

(b) The following measures of location were calculated for a distribution of the number of persons per household in Mdzimba:

mode = 2 persons; mean=4.1 persons; median=3 persons

If there are 9245 households in Mdzimba, which of the following procedures is appropriate to calculate the likely total number of persons living in Mdzimba.

(i) multiply the number of households by 2

(ii) multiply the number of households by 4.1

(iii) multiply the number of households by 3.

Question 3 **[25 marks, 4+4+4+4+4+5]**

(a) Consider the following two-way pivot table of *brand preference* for digital cameras and their *primary usage* (professional and personal).

Usage	Digital Camera Brand Preference		
	Canon	**Nikon**	**Pentax**
Professional	48	15	27
Personal	30	95	65

(i) What is the probability of selecting a *professional* user?

(ii) Find the probability that a user prefers the *Pentax brand* given that their usage is primarily for *personal* use?

(iii) What is the likelihood that a randomly selected user prefers the *Canon brand* and is a *professional* user?

(iv) Find the probability of randomly selecting either a *professional* user or a user who prefers the *Nikon brand* of digital camera?

(b) Two groups of bank trainees wrote a banking exam with the following results:

	Mean	**Variance**	**Sample Size**
Group 1	76	110	34
Group 2	64	88	26

(i) Compute the mean of the combined *exam scores*.

(ii) Which group showed greater consistency in *exam scores*? Why?

Question 4 **[25 marks, 4+2+8+3+8]**

The Chamber of Commerce conducted a survey amongst 17 furniture retailers to identify the *percentage of bad debts* in each of the company's debtors' book. The bad debts percentages are as follows:

2.2	4.7	6.3	5.8
5.7	7.2	2.6	2.4
6.1	6.8	2.2	
5.7	3.4	6.6	
1.8	4.4	5.4	

(a) Construct a stem and leaf diagram of the data.

(b) Find the average *% of bad debts* amongst the 17 furniture retailers.

(c) Compute the first quartile and the third quartile of the *% of bad debts* amongst the 17 furniture retailers surveyed. Interpret these quartiles.

(d) The chamber of commerce monitors bad debts levels based on samples of companies. It will advise an industry to take corrective action if the % of bad debts, on average, exceeds 5%. Should the chamber of commerce send out an advisory note to all furniture retailers based on these sample findings? Justify your answer.

(e) Calculate the coefficient of skewness. Interpret this result.

Question 5 **[25 marks, 4+6+6+6+3]**

(a) An economist conducted a study to identify the *percentage of family income allocated to the purchase of groceries.* She surveyed a random sample of 50 families and compiled the following table.

Percent of Income	Number of families
10-under 20%	6
20-under 30%	14
30-under 40%	16
40-under 50%	10
50-under 60%	4

(i) Compute and interpret the mean percentage of family income allocated to grocery purchase.

(ii) What is the maximum percentage of income that is allocated to grocery purchase by the lower 50% of families?

(iii) 25% of families spend more than a specific percentage of their income on groceries. What is that *percentage of income* value?

(b) Two insurance companies (Federal Insurance and Baltic Insurance) both use the Laspeyres index number approach to represent how the number of claims processed annually have either increased or decreased. In 2005 the two companies merged, but decided to maintain separate records of claims processed annually. To make comparisons possible between the two companies, management agreed that their respective index series should have a common base period on 2005. The index series of claims processed for each company is given below:

	2001	2002	2003	2004	2005	2006
Federal Insurance (base=2003)	92.3	95.4	100	102.6	109.4	111.2
Baltic Insurance (base=2004)	93.7	101.1	98.2	100	104.5	107.6

(i) Calculate the revised index for both companies with 2005 as the commom base period. Calculate to one decimal place.

(ii) Which company showed the bigger increase in claims processed between 2003 and 2005?

APPENDIX 2: LIST OF KEY FORMULAE

MEASURES OF CENTRAL LOCATION

Arithmetic mean

Ungrouped data

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad 3.1$$

Grouped data

$$\bar{x} = \frac{\sum_{i=1}^{m} f_i x_i}{n}$$

Mode

Grouped data

$$M_o = O_{mo} + \frac{c(f_m - f_{m-1})}{2f_m - f_{m-1} - f_{m+1}} \qquad 3.3$$

Median

Grouped data

$$M_e = O_{me} + \frac{c\left[\frac{n}{2} - f(<)\right]}{f_{me}} \qquad 3.2$$

Lower quartile

Grouped data

$$Q_1 = O_{q1} + \frac{c\left[\frac{n}{4} - f(<)\right]}{f_{q1}} \qquad 3.7$$

Upper quartile

Grouped data

$$Q_3 = O_{q3} + \frac{c\left[\frac{3n}{4} - f(<)\right]}{f_{q3}} \qquad 3.8$$

Geometric mean

Ungrouped data

$$GM = \sqrt[n]{x_1 \times x_2 \times x_3 \times \ldots \times x_n} \qquad 3.4$$

Weighted arithmetic mean

Grouped data

$$\text{weighted } \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} \qquad 3.5$$

MEASURES OF DISPERSION AND SKEWNESS

Range

$$\text{Range} = \text{Maximum value} - \text{Minimum value} + 1 = x_{max} - x_{min} + 1 \qquad 3.9$$

Variance

Mathematical – ungrouped data

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{(n-1)} \qquad 3.10$$

Computational – ungrouped data

$$s^2 = \frac{\Sigma x_i^2 - n\bar{x}^2}{(n-1)} \qquad 3.11$$

Standard deviation

$$s = \sqrt{s^2} \qquad 3.12$$

Coefficient of variation

$$CV = \frac{s}{\bar{x}} \times 100\% \qquad 3.13$$

Pearson's coefficient of skewness

$$sk_p = \frac{n\Sigma(x_i - \bar{x})^3}{(n-1)(n-2)s^3} \qquad 3.14$$

$$sk_p = \frac{3\,(\text{Mean} - \text{Median})}{\text{Standard deviation}} \quad \text{(approximation)} \qquad 3.15$$

PROBABILITY CONCEPTS

Conditional probability

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \qquad 4.2$$

Addition rule

Non-mutually exclusive events

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad 4.3$$

Mutually exclusive events

$$P(A \cup B) = P(A) + P(B) \qquad 4.4$$

Multiplication rule

Statistically dependent events

$$P(A \cap B) = P(A/B) \times P(B) \quad 4.5$$

Statistically independent events

$$P(A \cap B) = P(A) \times P(B) \quad 4.6$$

$n!$ = n factorial

$$n \times (n-1) \times (n-2) \times (n-3) \times \ldots \times 3 \times 2 \times 1 \quad 4.8$$

Permutations

$$_nP_r = \frac{n!}{(n-r)!} \quad 4.10$$

Combinations

$$_nC_r = \frac{n!}{r!\,(n-r)!} \quad 4.11$$

PROBABILITY DISTRIBUTIONS

Binomial distribution

$$P(x) = {}_nC_x\, p^x(1-p)^{(n-x)} \quad \text{for } x = 0, 1, 2, 3, \ldots, n \quad 5.1$$

$$P(x \text{ successes}) = \frac{n!}{x!\,(n-x)!}\, p^x(1-p)^{(n-x)} \quad \text{for } x = 0, 1, 2, 3, \ldots, n$$

Binomial descriptive measures

Mean $\mu = np$

Standard deviation $\sigma = \sqrt{np(1-p)}$ 5.2

Poisson distribution

$$P(x) = \frac{e^{-a}a^x}{x!} \quad \text{for } x = 0, 1, 2, 3 \ldots \quad 5.3$$

Poisson descriptive measures

Mean $\mu = a$

Standard deviation $\sigma = \sqrt{a}$ 5.4

Standard normal probability

$$z = \frac{x-\mu}{\sigma} \quad 5.6$$

INDEX NUMBERS

Price relative

$$\text{Price relative} = \frac{p_1}{p_0} \times 100\% \quad 13.2$$

Laspeyres price index

Weighted aggregates method

$$\text{Laspeyres price index} = \frac{\Sigma(p_1 \times q_0)}{\Sigma(p_0 \times q_0)} \times 100\% \quad 13.5$$

Laspeyres price index

Weighted average of relatives method

$$\text{Laspeyres price index} = \frac{\Sigma\left[\left(\frac{p_1}{p_0}\right) \times 100 \times (p_0 \times q_0)\right]}{\Sigma(p_0 \times q_0)} \quad 13.9$$

Paasche price index

Weighted aggregates method

$$= \frac{\Sigma(p_1 \times q_1)}{\Sigma(p_0 \times q_1)} \times 100\% \quad 13.8$$

Paasche price index

Weighted average of relatives method

$$= \frac{\Sigma\left[\left(\frac{p_1}{p_0}\right) \times 100 \times (p_0 \times q_1)\right]}{\Sigma(p_0 \times q_1)} \quad 13.10$$

Quantity relative

$$\text{Quantity relative} = \frac{q_1}{q_0} \times 100\% \quad 13.11$$

Laspeyres quantity index

Weighted aggregates method

$$\text{Laspeyres quantity index} = \frac{\Sigma(p_0 \times q_1)}{\Sigma(p_0 \times q_0)} \times 100\% \quad 13.12$$

Laspeyres quantity index

Weighted average of relatives method

$$\text{Laspeyres quantity index} = \frac{\Sigma\left[\left(\frac{q_1}{q_0}\right) \times 100 \times (p_0 \times q_0)\right]}{\Sigma(p_0 \times q_0)} \quad 13.14$$