

UNIVERSITY OF SWAZILAND

EXAMINATION PAPER 2005/06

TITLE OF PAPER : DISCREPTIVE STATISTICS

COURSE CODE : ST 132

TIME ALLOWED : TWO (2) HOURS

**REQUIREMENTS : CALCULATOR, STATISTICS L TABLES AND
GRAPH PAPER**

**INSTRUCTIONS : ANSWER QUESTION ONE AND ANY OTHER
TWO QUESTIONS**

QUESTION 1

- a) A large organization has been accused of being "ageist", i.e. tending not to employ older people. In response, it publishes the following table showing the age distribution of its current employees.

<i>Age last birthday (years)</i>	<i>Number of employees</i>
15 – 19	240
20 – 24	340
25 – 29	360
30 – 39	420
40 – 49	380
50 – 64	240

- (i) Draw a histogram on graph paper to show the data. (6)
- (ii) State, with reasons, whether the data suggest that the organization is ageist. (2)
- (iii) Explain why the way the data have been presented in the table may be misleading to the casual observer. (2)
- b) Define three different measures of dispersion for a set of data and give one advantage and one disadvantage of each measure. Which of your measures would you recommend in calculating the dispersion of a set of data giving the wages of all employees in a company? Give brief reasons for your choice. (10)

- c) The table shows 100 three-digit numbers x that have been generated using the random number function on an electronic calculator.

Values of x

838	114	017	839	129	298	136	886	058	368
250	877	314	554	200	032	114	415	244	479
522	290	983	522	492	715	160	545	636	642
867	260	202	351	322	134	209	164	025	027
984	319	593	161	035	359	999	243	502	993
830	723	314	574	126	426	601	558	692	867
488	102	834	029	750	425	427	465	681	978
558	385	393	398	592	926	337	683	792	659
078	105	957	150	927	789	904	188	102	299
616	610	877	377	737	610	067	878	472	344

You are given that $\sum_{i=1}^{100} x = 47118$ and $\sum_{i=1}^{100} x^2 = 30710404$

- i) Calculate the mean and the standard deviation of this sample of random numbers. (4)
- ii) Group the data into a frequency distribution with 5 classes intervals. (6)
- iii) Calculate the mean and the standard deviation of this grouped frequency distribution. (6)
- iv) Comment on your results in parts (i) and (iii) and explain why they are not identical. (4)

QUESTION 2

Maccal Trans Co. delivers special farm cars to locations in Swaziland from the Lavumisa border gate. The data below give the prices (in E) for a return fare for a driver and cost for a car on each of 10 routes.

Route	A	B	C	D	E	F	G	H	I	J
Driver cost (x)	20	23	27	33	28	42	38	23	22	19
Car cost (y)	92	107	124	165	105	163	143	85	100	83

You are given that $\sum_{i=1}^{10} x^2 = 8113$, $\sum_{i=1}^{10} y^2 = 144671$ and $\sum_{i=1}^{10} xy = 34046$

- a) Find the correlation coefficient between x and y and comment on its value. (5)
- b) Find the regression line that predicts car cost for a given driver cost and comment on the coefficient values. (6)
- c) Which route gives the cheapest actual car cost compared to predicted cost and which the most expensive? (4)
- d) Calculate the error that one can commit when predicting car cost. (5)

QUESTION 3

- a) The table provides data about electricity consumption in a large university for each quarter from 1996 Qtr 1 to 2000 Qtr 2. The trend has been calculated by the method of moving averages. Calculate the average seasonal component for each type of quarter using an additive model.

<i>Year/Quarter</i>	<i>Units used (000s)</i>
1996/1	207
/2	152
/3	127
/4	176
1997/1	172
/2	186
/3	133
/4	177
1998/1	204
/2	175
/3	131
/4	174
1999/1	183
/2	176
/3	132
/4	175
2000/1	194
/2	164

(12)

- b) The trend for each quarter of 2001 has been predicted as follows:

	<i>Qtr 1</i>	<i>Qtr 2</i>	<i>Qtr 3</i>	<i>Qtr 4</i>
<i>Electricity consumption units (000s)</i>	170	172	174	175

Using the seasonal components you have calculated in (a), predict the actual consumption of electricity in each quarter of 2001.

(4)

NOTE: It is not necessary to draw a graph.

- c) Suggest factors which could affect the accuracy of the forecasts.

(4)

QUESTION 4

An insurance company classifies drivers in the agricultural manufacturing sector according to sex and to whether they are under 25 or 25 years and over. It finds that 60% of its drivers are male; 25% of the male drivers and 30% of the female drivers are under 25.

- a) Find the probabilities that a randomly chosen driver is in each of the four categories
- i) male and under 25,
 - ii) male and 25 or over,
 - iii) female and under 25,
 - iv) female and 25 or over.
- (6)
- b) Hence write down the probabilities of a driver being
- i) under 25,
 - ii) male given that the driver is under 25,
 - iii) male or under 25 (or both),
 - iv) neither male nor under 25.
- (6)
- c) The probability p of having at least one accident in a year is given in the table for the different classes of drivers whose distribution is as above.

Probability p of one or more accidents for different classes of driver

	<i>under 25</i>	<i>25 or over</i>
<i>male</i>	0.09	0.04
<i>female</i>	0.06	0.02

- i) Find the probability that a randomly chosen driver has at least one accident in a year.
- ii) If a driver has at least one accident what is the probability that the driver is male and under 25?

(8)