UNIVERSITY OF SWAZILAND

EXAMINATION PAPER 2006

TITLE OF PAPER   : RESEARCH METHODS AND STATISTICS

COURSE CODE   : HSC 301

TIME ALLOWED   : TWO (2) HOURS

REQUIREMENTS   : CALCULATOR

INSTRUCTIONS   : ANSWER QUESTION ONE AND ANY OTHER TWO QUESTIONS

                                  (EACH QUESTION CARRIES 20 MARKS)

Question 1

a) In the year 2000, the National Center for Health Statistics estimated that the numbers (in millions) of older people in the USA who suffered from Alzheimer's Disease, by age group, were;

Age 65–74	Age 75–84	Age 85+
0.3	2.4	1.8

The Center's projections for 2050 (in millions) are;

Age 65–74	Age 75–84	Age 85+
0.4	4.8	8.0

   i) Draw separate bar charts, using the same scale, to represent the data. (4)

   ii) Comment on what these charts show. (2)

   iii) Give one advantage and one disadvantage of pie charts as opposed to bar charts in representing data. (2)

b) The charge in a town centre car park depends on how long a car is parked, rounded up to the next hour. Penalty charges are levied for stays of more than six hours. For a random sample of 60 cars, the parking times were recorded to the nearest minute. A stem-and-leaf diagram of these times is shown below.

```
Times (minutes) in car park

hundreds | tens

       0 | 1124
       0 | 5555556
       1 | 0123344
       1 | 5566667777799
       2 | 022233334
       2 | 556799
       3 | 00123
       3 | 5555666
       4 |
       4 | 9
       5 | 2
```

   i) Explain clearly what the entries in the diagram signify. By reference to the final entry 5 | 2, identify the "stem" and the "leaf". (3)

   ii) Comment on the distribution of times. What does this distribution suggest about the pricing policy at the car park? (3)

   iii) Assuming that each of the times recorded had a units digit of 5, find the median and quartiles of the distribution of times. (3)

   iv) Comment on the skewness of the distribution. (3)

Question 2

a) A restaurant in a certain area currently serves quiches that contain four fresh eggs. Research in the area has shown that one in four locally produced eggs contains the salmonella bacterium. A health inspector has therefore suggested that in order to avoid food poisoning with salmonella the restaurant should use no more than three eggs in its quiches.

   i) Comment on the logic behind the inspector's suggestion.   (3)

Assuming that any egg has, independently, the same chance as any other of containing salmonella bacterium, find
   ii) the probability that a three-egg quiche is salmonella free,   (3)
   iii) the probability that a four-egg quiche is salmonella free,   (3)
   iv) Comment on the assumption of independence between the eggs in a given quiche.
                                                    (3)

b) A disease, which can only be diagnosed with certainty after death, exists in a proportion p_0 of the population. A clinical test is known such that;

   P(test is positive given disease is present) $= p_1$
   and
   P(test is negative given disease is absent) $= p_2$.

   i) Find, in terms of p_0, p_1 and p_2, the probability that a randomly chosen individual who tests positive actually has the disease.   (4)
   ii) Calculate the answer to (i) in the case $p_0 = 0.005$, $p_1 = 0.95$, $p_2 = 0.95$, and comment on your result.   (4)

Question 3

The breaking stresses in newtons per square metre of standard samples of pine, measured at varying angles $x°$ to the grain of the wood, are tabulated below.

Breaking Stress y (N/m^2)×10^{-10}	0.987	1.064	1.337	1.912	2.740	5.771	11.494
Angle $x°$	0	15	30	45	60	75	90

a) Plot these data, use the information given to calculate the Pearsons's correlation coefficient between x and y, and comment briefly. (8)

b) Conversion of the data using Hankinson's formula for y in terms of x;

$$y = \left(\frac{1-\sin^2 x}{a} + \frac{\sin^2 x}{b} \right)^{-1}$$

the data are tabulated below.

$X = \sin^2 x$	0	0.067	0.250	0.500	0.750	0.933	1
$Y = 1/y$	1.013	0.940	0.748	0.523	0.365	0.173	0.087

Plot Y against X, comment on the suitability of this relationship for regression analysis, estimate a and b coefficients. Also compute corr(X, Y) and compare this with the correlation computed in part (a). (12)

Question 4

a) Explain what is wrong with the conclusions in **bold type** in each of the following statements.

    i) In Town M, the correlation coefficient between the ages of men and women at marriage is 0.94. **Hence a man is likely to marry a woman of his own age.**

    ii) Company N is a company employing 500 people, a few of whom are highly paid specialists. The mean annual salary of employees in the company is £15,355. **Hence half the employees earn less than £15,355 per year.**

    iii) "Migraines affect approximately 14% of women and 6% of men, **that is one fifth of the population**, assuming that there are equal numbers of men and women in the population."

    iv) "The medical insurance premium is reduced by 70% for five years without claims, a further reduction of 10% of the reduced premium is given for buying on the internet, **that's a total reduction of 80%.**"

(12)

b) Explain what the terms *quantitative* and *qualitative* mean when referring to variables in a set of data. (2)

c) Suggest two categories into which qualitative variables can be sub-divided and give an example of a variable of each category. (4)

d) Quantitative variables may be divided into the categories *discrete* and *continuous*. Give one example of each of these categories. (2)

Question five

Certain factors in the Environment Health profession necessitate the conducting of research for academic purposes. These factors could be professional and non-professional related.

In your own understanding describe five factors that necessitate the conducting of academic research in the environmental health profession. Five marks for each factor.

(20 marks)

Question six

In research, evaluation could be done through analyzing the nature of the compatibility or congruence ensured by the research while developing the research proposal. That a, number of elements are considered when evaluating the research.

a) In your own understanding what are the between content and face validity?

(6 marks)

b) How would you ensure content validity when developing a research proposal?

(8 marks)

c) How are variables derived during proposal development?

(6 marks)