EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

ORDINARY CERTIFICATE IN STATISTICS, 2010

Paper II

Time allowed: Three Hours

*Candidates may attempt **all** the questions.*

The number of marks allotted to each question or part-question is shown in brackets.

The total for the whole paper is 100.

A pass may be obtained by scoring at least 50 marks.

Graph paper and Official tables are provided.

*Candidates may use calculators in accordance with the regulations published in
the Society's "Guide to Examinations" (document Ex1).*

1. In words, without using any formulae, define the following three measures of location: the *mode*, the *median* and the (arithmetic) *mean*.

(3)

Three children are carrying out a survey on fruit at school.

Hannah is asking children to name their favourite fruit.

Joshua is asking children to rate how much they like eating pears on an integer scale from 1 to 5 where 1 represents "I hate eating pears" and 5 represents "I love eating pears".

Sarah is asking children to record how many bananas they ate in a week.

For each of these three surveys, choose the two words from the list below which best describe the type of data that the children will obtain.

(3)

Categorical Continuous Counting Discrete Measurement Ordered Unordered

For each of the surveys, write down which one or more of the measures of location you have defined would be appropriate to use with the data.

(3)

Turn over

2. Some UK supermarkets have recently introduced a "traffic light system" of food labelling. As part of this system, foodstuffs are labelled according to whether they are high, medium or low in fat and sugar. Those which are high are given a red (R) label, those which are medium an orange (O) label and those which are low a green (G) label. On a recent week's shopping my purchases had the labels given in the following table.

Foodstuff	Fat	Sugar	Foodstuff	Fat	Sugar	Foodstuff	Fat	Sugar
Bacon	R	G	Biscuits	R	O	Bread	G	G
Cheesecake	O	R	Cola	G	R	Cream	R	G
Crispbake	O	G	Curry	R	R	Custard	R	R
Custard Tart	O	O	Diet Cola	G	G	Eclairs	R	O
Eggs	O	G	Gateau	O	R	Honey	G	R
Hot Pot	R	G	Jam	G	R	Lamb	O	G
Marmalade	G	R	Mascarpone	R	G	Meringue	O	R
Pâté	O	G	Pies	R	G	Pizza	G	O
Prawns	G	G	Pudding	O	R	Sandwich	R	G
Sauce	O	G	Soup	G	G	Treacle Tart	O	R

Construct a 3 × 3 table in which the three columns are the levels of sugar content and the three rows are the levels of fat content. Fill in the row and column totals and the grand total.

(7)

Use your table to answer the following questions. Where the answer is a proportion, give your answer as a fraction in its simplest form.

(i) What proportion of my purchases has a red fat content?

(ii) What proportion of my purchases has both a green fat content and a green sugar content?

(iii) What is the modal category with respect to sugar content?

(iv) What is the modal category overall, with respect to both sugar and fat content?

(v) What proportion of my purchases with a green sugar content has an orange fat content?

(5)

3. The ONS Centre for Demography lists the male/female splits of the UK population at particular ages in 2007 as follows.

	Thousands	
Age	Males	Females
0	387.5	368.3
10	339.3	324.5
20	423.2	397.0
30	402.3	398.6
40	434.1	445.7
50	441.3	455.4
60	353.5	363.5
70	242.8	268.8
80	138.1	195.3

Draw a bar chart showing the numbers of males and of females at each of these ages.

(6)

Construct a table showing the ratio, correct to 2 decimal places, of females to males at each age.

(4)

Comment on the results.

(2)

4. A seed merchant took delivery of a sack of beans. Looking into the sack, he thought the beans appeared much more variable in size than his usual stock. The lengths of beans in his usual stock have the following characteristics.

Mean = 1.69 cm   Range = 0.3 cm   Inter-quartile range = 0.2 cm

Standard deviation = 0.12 cm   Coefficient of variation = 6.9%

He took a sample of 15 beans from the sack and measured their lengths. His results were as follows, in cm.

2.4 2.7 3.0 2.5 3.2 2.8 2.8 2.3 3.0 2.8 2.8 2.4 2.9 2.8 3.1

Calculate the following measures of variability for these 15 beans.

Range   Inter-quartile range   Standard deviation   Coefficient of variation

(13)

Comment on whether there is evidence to support the seed merchant's views on the variability in size of the beans.

(2)

Turn over

5.   The monthly values of two different measures of the cost of living in the UK, the Retail Prices Index (RPI) and Two-person Pensioner Households Index (PPI), are given for July 2007 – June 2008, taken from the UK *Monthly Digest of Statistics*. Both index numbers exclude Housing.

UK Retail Price Index Numbers

Month	RPI (Jan 1987 = 100)	PPI (Jan 1987 = 100)
2007 Jul	182.2	179.8
2007 Aug	182.9	180.3
2007 Sep	183.5	181.4
2007 Oct	184.3	182.2
2007 Nov	185.1	183.4
2007 Dec	186.3	184.5
2008 Jan	185.2	183.7
2008 Feb	187.0	186.3
2008 Mar	188.2	187.3
2008 Apr	189.6	189.0
2008 May	191.2	191.0
2008 Jun	193.2	193.3

Explain what is meant by *(Jan 1987 = 100)*.

(1)

What is the percentage rise in the RPI from July 2007 to June 2008?

(1)

What is the percentage rise in the PPI from July 2007 to June 2008?

(1)

By how many percentage points has the RPI risen between July 2007 and June 2008?

(1)

By how many percentage points has the PPI risen between July 2007 and June 2008?

(1)

Rebase each series to July 2007 and display your results in a table.

(5)

Using all your results, comment on the differences in price rises faced by the pensioner households as compared to general households over this period.

(2)

6. Meteorologists researching climate change have been studying the records for Bradford UK between 1990 and 2007. The numbers of days each year with air frost are shown in the table.

Days with air frost in Bradford UK

Year	Number of days
1990	15
1991	53
1992	43
1993	46
1994	34
1995	47
1996	61
1997	32
1998	28
1999	28
2000	24
2001	57
2002	18
2003	52
2004	29
2005	42
2006	32
2007	28

Plot a time chart of the data and comment on what it shows.

(6)

State one benefit and one drawback of using the moving average method to estimate the trend of a time series.

(2)

Calculate a 7-year moving average of the data and plot the results on the chart.

(6)

Comment briefly on whether you think a 7-year moving average is an appropriate method of estimating the trend for this particular data series.

(1)

Turn over

7.   Define the probability of an event *A* given an event *B*. You may assume that the probability of *B* is greater than zero.

(1)

A serious disease occurs in 1% of a certain population. A diagnostic test is available to help in screening for the disease; a positive test result is taken as an indication that the disease may be present and further different tests are necessary, a negative result suggests that no further tests are necessary. It has been found that 95% of those with the disease have a positive test result and 5% of those who do not have the disease have a positive test result.

Copy into your answer book the following probability tree for an individual chosen at random from the population, and write the three pairs of probabilities along the branches.

(4)
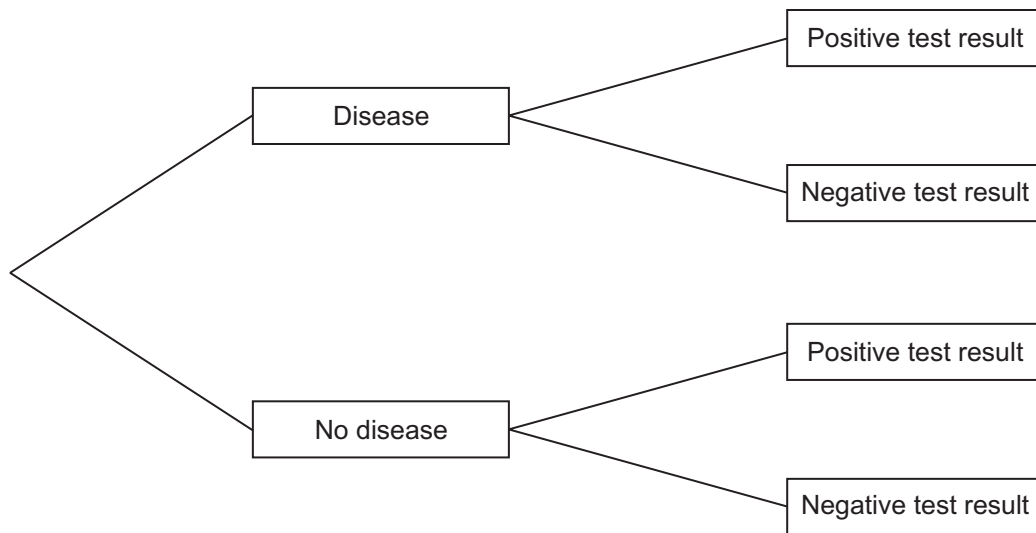
Hence calculate the following.

(i)   The probability that a randomly chosen member of the population has a positive test.

(ii)   The probability that a person has the disease given that this person's test result is positive.

(iii) The probability that a person has the disease given that this person's test result is negative.

(5)

Comment briefly on how well the test is working given your answers to parts (ii) and (iii).

(2)

8.   An experiment is being conducted into the physical fitness of older people. A physical fitness test is given to each of 10 subjects (a higher score represents greater fitness). Each subject's fitness score for the test is recorded as value y and age as value x (ages are between 40 and 80 years). The following totals are given.

$$\Sigma x = 581 \quad \Sigma x^2 = 37\,193 \quad \Sigma y = 607 \quad \Sigma y^2 = 38\,795 \quad \Sigma xy = 33\,426$$

(i)   Calculate $\bar{x}$, $\bar{y}$, $\Sigma(x-\bar{x})^2$, $\Sigma(y-\bar{y})^2$, $\Sigma(x-\bar{x})(y-\bar{y})$ and hence the product-moment correlation coefficient.

(6)

(ii)   Comment on both the sign and magnitude of the coefficient.

(2)

(iii)   By fitting the appropriate straight line, estimate the average fitness score of a person who is aged 45.

(3)

(iv)   Explain briefly why your value is an estimate and why it is the average fitness score that is being estimated.

(2)