THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE EXAMINATION

NEW MODULAR SCHEME

introduced from the examinations in 2007

MODULE 3

SPECIMEN PAPER B

AND SOLUTIONS

The time for the examination is 1½ hours. The paper contains four questions, of which candidates are to attempt **three**. Each question carries 20 marks. An indicative mark scheme is shown within the questions, by giving an outline of the marks available for each part-question. The pass mark for the paper as a whole is 50%.

The solutions should not be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids. For this reason, they do not carry mark schemes. Please note that in many cases there are valid alternative methods and that, in cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of the questions and solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of the questions or solutions.

1. The following table gives a summary of the numbers of goals scored by the home soccer teams in matches in the English Premier Football League during the 1999–2000 season. It is required to test the assumption that the data follow a Poisson distribution.

Number of goals scored by the home team, r	0	1	2	3	4	≥ 5
Frequency, f	81	112	101	44	28	14

(i) Explain why it might be reasonable to assume that the number of goals, r, scored by the home team would follow a Poisson distribution.

(4)

(ii) The total number of matches was $\Sigma f = 380$, and the total number of goals scored was $\Sigma fr = 634$. Also $\Sigma fr^2 = 1778$. Calculate the mean and variance of the data.

(4)

(iii) Calculate the expected frequencies, on the Poisson hypothesis, for $r = 0$ and $r = 1$. The expected frequencies in the remaining cells of the table are 99.72, 55.46, 23.13 and 10.51. Carry out a χ^2 goodness-of-fit test of the hypothesis that the data follow a Poisson distribution. Explain your conclusions carefully. What problem in carrying out the test would have occurred if the frequencies for values of $r \geq 5$ had not been combined?

(12)

2. The duration of a "normal" pregnancy (i.e. one without medical complications) can be modelled by a Normal distribution with mean 266 days and standard deviation 16 days. In an urban hospital in a relatively deprived area of the USA, a random sample of 60 pregnancies was studied and the duration of each pregnancy determined. The relevant statistics were

$$\sum x = 15568, \quad \sum x^2 = 4054484.$$

(i) Test whether the variance of this sampled population differs from that of "normal" pregnancy durations.

(10)

(ii) Test whether there is evidence that the mean pregnancy duration for this sampled population differs from 266 days. Find an approximate *p*-value for your test, and state your conclusions clearly.

(10)

3. (i) State the assumptions on which an independent (unpaired) two-sample *t* test is based.

(4)

An individual is considering purchasing a two-bedroomed terraced house in one of two adjacent towns in the north of England. To compare prices, he extracts some price data from one week's issue of the local newspaper's "Property Supplement" for all such properties advertised by estate agents with branches in both towns. He wishes to determine whether the mean price in one town differs from that in the other. The data extracted are given in the table below (units are thousands of pounds).

Town 1	77.50	74.95	74.50	60.00	45.00	25.00	25.00
Town 2	72.95	72.95	65.00	62.50	56.95	54.95	52.95
	49.95	46.95	35.00	34.95	30.00	29.95	25.00

(ii) Assuming that all the necessary assumptions hold, perform an independent two-sample *t* test and draw your conclusion.

(12)

(iii) Town 1 has a considerably larger population, and greater numbers of all types of properties, than town 2. Taking note of this information, and of the way in which the data were obtained, discuss critically whether there is a valid basis for the test in part (ii).

(4)

4. (i) A sports equipment company has commissioned an advertising agency to develop an advertising campaign for one of its new products. They can choose between two particular television commercials, *A* and *B*. To aid them in their decision, an experiment is performed in which 200 volunteers are randomly assigned to view one of the two commercials, 100 being assigned to each. After seeing the commercial, each volunteer is asked to state whether they would consider buying the product, with the following results.

		Commercial	
		A	*B*
Purchase product	*No*	70	80
	Yes	30	20

Apply a chi-squared test to these data and comment on your results. What recommendations, if any, would you make to the sports manufacturer concerning the choice of commercial for the proposed advertising campaign?

(7)

(ii) A random sample of sportswear manufacturers was surveyed to determine whether they advertised on television and/or the internet. The results are given in the following table.

		Internet	
		No	*Yes*
Television	*No*	3	5
	Yes	15	17

Apply McNemar's test to the above data and comment on your results.

(7)

(iii) Distinguish carefully between chi-squared tests and McNemar's tests, as used to analyse data such as given in parts (i) and (ii) of this question, giving examples of when each would be preferred to the other.

(6)

SOLUTIONS

Question 1

(i) The Poisson distribution to explain numbers of goals might be a reasonable assumption if home team scores can be regarded as random events occurring at a constant average rate throughout the season. If so, the number of home team goals in a match is Poisson with parameter (mean) equal to this constant average rate, μ say.

(ii) $\bar{r} = 634/380 = 1.6684$.

$$s^2 = \frac{1}{\Sigma f - 1}\left\{\Sigma fr^2 - \frac{(\Sigma fr)^2}{\Sigma f}\right\} = \frac{1}{379}\left(1778 - \frac{634^2}{380}\right) = 1.9003.$$

(iii) We take μ as 1.6684. So $P(R=0) = e^{-1.6684} = 0.1885$, and the expected frequency for $r = 0$ is 380 × 0.1885 = 71.65.

Similarly, $P(R=1) = 1.6684e^{-1.6684} = 0.3145$, and the expected frequency for $r = 1$ is 119.51.

Hence we have (taking the remaining expected frequencies from the question paper)

r	0	1	2	3	4	≥5	Total
Observed	81	112	101	44	28	14	380
Expected	71.65	119.51	99.72	55.46	23.13	10.51	379.98

[Note. There is a very small rounding error in the calculations of expected frequencies.]

The test statistic is

$$X^2 = \sum \frac{(O-E)^2}{E} = \frac{(81-71.65)^2}{71.65} + \frac{(112-119.51)^2}{119.51} + \ldots + \frac{(14-10.51)^2}{10.51} = 6.261,$$

which is referred to χ^2_4 (note 4 degrees of freedom because the table has 6 cells and there is one estimated parameter). This is not significant (the 5% point is 9.49); we cannot reject the null hypothesis, i.e. there is no evidence against the Poisson model with these data.

For the test, the expected frequencies need to be not too small (≥5 is often used as a criterion). This would not be the case if frequencies for large r were not combined.

Question 2

(i) The variance of this sample is

$$s^2 = \frac{1}{59}\left(4054484 - \frac{15568^2}{60}\right) = \frac{15106.93}{59} = 256.05\,.$$

The null hypothesis to be tested is "$\sigma^2 = 256$". It seems obvious that this null hypothesis is not likely to be rejected (even if the sample had been of considerably smaller size), but continuing with a formal test we use test statistic

$$\frac{(n-1)s^2}{\sigma^2} = \frac{59\times 256.05}{256} = 59.01$$

which is referred to χ^2_{59}. The upper 5% point is about 78. Clearly we cannot reject the null hypothesis as the data give no evidence for doing so.

(ii) We have $\bar{x} = \frac{15568}{60} = 259.46$ and we wish the test the null hypothesis μ = 266. Taking the value of σ as 16, which seems highly plausible from part (i), we use test statistic

$$\frac{\bar{x}-266}{\frac{16}{\sqrt{60}}} = -3.16$$

and refer to N(0, 1).

[**Alternatively,** we could continue to use the sample variance s^2 (= 256.05) and refer $\frac{\bar{x}-266}{\frac{s}{\sqrt{60}}}$ to t_{59}; this makes hardly any difference in practice in this case.]

This is well beyond the double-tailed 1% point of N(0, 1); there is strong evidence against this null hypothesis. It is reasonable to conclude that this population has a mean different from the "normal" one; it appears to be less.

Using N(0, 1), we have $\Phi(-3.16) = 0.0008$, giving a p-value of 0.0016.

<u>Question 3</u>

(i) The two samples should be from independent Normal distributions with the same variance but possibly different means (the null hypothesis is usually that the two means are equal). The samples are random samples and are independent of each other.

(ii)

$n_1 = 7;\ \ \bar{x}_1 = 54.56,\ s_1^{\,2} = 534.6956.$ $\qquad n_2 = 14;\ \ \bar{x}_2 = 49.29,\ s_2^{\,2} = 261.3001.$

The "pooled estimate" of variance is $s^2 = \dfrac{(n_1 - 1)s_1^{\,2} + (n_2 - 1)s_2^{\,2}}{n_1 - 1 + n_2 - 1} = 347.6355.$

The test statistic for testing the null hypothesis $\mu_1 - \mu_2 = 0$, where μ_1 and μ_2 are the respective population mean prices, is

$$\frac{\bar{x}_1 - \bar{x}_2 - 0}{s\sqrt{\frac{1}{7} + \frac{1}{14}}} = \frac{5.27}{8.631} = 0.611\,,$$

which is referred to t_{19}. This is not significant, so the null hypothesis cannot be rejected. There is no evidence that the true means in the two towns differ.

(iii) Town 1 has greater population and greater numbers of all types of properties, yet only half the sample size was used compared with town 2. The probabilities of selection in the two towns are thus very different. Also, the samples were restricted to the "Property Supplement" and to agents dealing in both towns. The assumption of randomness is doubtful, even whether we have representative samples. The test based on these data must be suspect for practical reasons, even if Normality and constant variance are acceptable. [The usual $F_{6,13}$ test for equality of population variances gives a test statistic value of 2.05 which is not significant.]

Question 4

(i) The chi-squared test will examine the null hypothesis that there is no relation between "Commercial" and "Purchase", against the alternative hypothesis that there is a relation. Observed frequencies and (in brackets) expected frequencies on the null hypothesis are as follows.

		Commercial		
		A	*B*	*Total*
Purchase	*No*	70 (75)	80 (75)	150
	Yes	30 (25)	20 (25)	50
	Total	100	100	200

$$\text{Test statistic} = \frac{(70-75)^2}{75}+\frac{(80-75)^2}{75}+\frac{(30-25)^2}{25}+\frac{(20-25)^2}{25} = 2.667$$

[or 2.16 if calculated with Yates' correction so that "$(O-E)^2$" becomes $(4.5)^2$].

Refer to χ_1^2: not significant. There is no evidence of a relation between "Commercial" and "Purchase". Hence a decision could be made on non-statistical grounds, such as cost of the commercial or the potential size of the audience.

(ii) McNemar's test for paired data tests similar hypotheses on association or otherwise of the two classifications, in this case which advertising medium each manufacturer uses. It does not use either "No–No" or "Yes–Yes" manufacturers.

Test statistic $= \frac{(5-15)^2}{5+15} = \frac{100}{20} = 5.00$, refer to χ_1^2, significant at 5%. There is evidence against a null hypothesis of no preference of advertising medium.

(iii) In part (i), two different random samples of responses are obtained, and the problem is that of comparing the proportions giving a particular response in the two samples. This is often the case in a chi-squared test, for example in opinion surveys where the two samples are drawn from males and females, or "young" and "old" age-groups, in otherwise similar populations. It also applies in medical trials where different groups of patients (e.g. smokers and non-smokers) are classified as having or not having a particular disease.

However, in part (ii) there are not two independent samples, and this may occur more often in medical trials; for instance, suppose two drugs are used to treat a chronic illness, each for a short period of time, on the same patients (or at the least on patients who have been closely paired for age, sex and general medical condition). No information is gained from patients (or pairs) where both drugs worked, or both failed. The McNemar test examines whether, in cases where only one worked, drug A was more successful than drug B or not. It compares the proportions of preferences in this part of the data only. The example in part (ii) uses the same manufacturers, so a McNemar test is valid.