THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE EXAMINATION

NEW MODULAR SCHEME

introduced from the examinations in 2007

MODULE 3

SPECIMEN PAPER A

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. An experiment was carried out on random samples of steel to assess whether increasing the tempering temperature from 200^{o}C to 250^{o}C increases the failure stress. The samples were treated and tested and gave the following failure stresses (measured in units of 10 MegaPascals).

	Tempering at 200^{o}C	*Tempering at 250^{o}C*
	66	54
	49	63
	58	43
	77	56
	39	47
	51	97
	46	85
	91	
Sample mean	59.6	63.6
Sample standard deviation	17.4	20.1

(i) Carry out a *t* test, stating carefully your null and alternative hypotheses. (6)

(ii) State the assumptions you have made in carrying out the *t* test and provide any suitable graphical evidence relevant to these assumptions. (4)

(iii) Re-analyse the data using a suitable non-parametric test, once again stating carefully your null and alternative hypotheses. (7)

(iv) Summarise your conclusions and comment on the methods used in your analysis. (3)

2. In the UK, all patients surviving a stroke are supposed to have their cholesterol levels measured soon after their stroke and regularly thereafter. A sample of medical records of men and women who had suffered a stroke was examined to determine whether there was a difference between the sexes in the proportions of stroke survivors who had had a recently recorded cholesterol measurement. The following data were obtained.

		Cholesterol level recorded	
		No	*Yes*
Sex	*Female*	109	22
	Male	97	77

(i) Perform a χ^2 test of the null hypothesis that there is no association between an individual's sex and the chance of he or she having a recently recorded cholesterol measurement.

(9)

(ii) Compute and interpret an approximate 95% confidence interval for the difference between the proportions of females and males having a recently recorded cholesterol measurement.

(8)

(iii) Comment on your conclusions to parts (i) and (ii), relating one to the other.

(3)

3. (i) Explain the meaning of the following terms used in hypothesis tests.

(a) Type I error.

(2)

(b) Type II error.

(2)

(c) Level of significance.

(2)

(d) Power.

(2)

(ii) A manufacturer of coffee uses a machine to fill jars. The machine is calibrated so that the amount of coffee dispensed into each jar is Normally distributed with mean (μ) 200 grams and standard deviation (σ) 15 grams. Each hour, a random sample of 9 jars is taken from the previous hour's output and the sample mean amount ($\bar{x}$) is evaluated. If the sample mean lies in the interval $190 < \bar{x} < 210$, the previous hour's output is accepted, otherwise it is rejected and the machine is recalibrated before continuing.

(a) Calculate the probability of committing a type I error by rejecting the previous hour's output when $\mu = 200$ grams and $\sigma = 15$ grams.

(6)

(b) Calculate the probability that the previous hour's output will be accepted when $\mu = 216$ grams and $\sigma = 15$ grams.

(6)

4. In a plant-breeding programme for maize, there were 1301 plants which could be divided into four types. The experimenter assumed a hypothesis that the types depended on only two genes, A and B, in dominant or recessive forms. The frequencies of the four types AB, Ab, aB, ab should be in the ratio 9:3:3:1. Actual observed frequencies in this experiment were as follows.

AB: 773 Ab: 231 aB: 238 ab: 59

(i) Test whether the assumed hypothesis explains these data satisfactorily, and comment on the result.

(5)

(ii) There is some doubt whether the double recessive ab is as vigorous as the other types. It is possible that not all the plants of type ab will survive. Therefore two further null hypotheses may be examined:

(A) $AB : Ab : aB$ is $9 : 3 : 3$;

(B) (AB, Ab, aB together) : ab is $15 : 1$.

Carry out tests of these hypotheses and explain what can be inferred from all the results.

(9)

(iii) Construct an approximate 95% confidence interval for the proportion of type ab in the population. Comment on the result, and on any further information it gives about the hypotheses being studied.

(6)

SOLUTIONS

Question 1

(i) Null hypothesis: there is no difference between the population mean failure stresses of the 200°C and 250°C tempered steel. Alternative hypothesis: there is a difference between the mean failure stresses, that for 250°C being higher.

$n_1 = 8,\ n_2 = 7;\ \bar{x}_1 = 59.6,\ \bar{x}_2 = 63.6;\ s_1^2 = 17.4^2,\ s_2^2 = 20.1^2.$

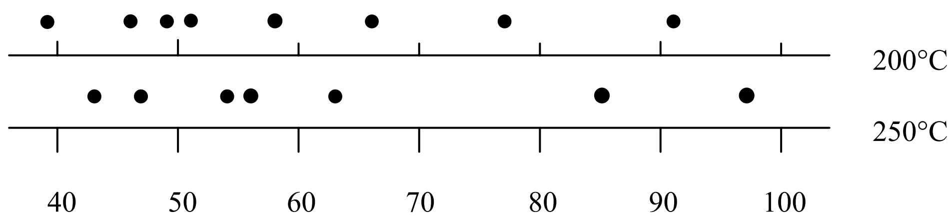
The pooled estimate of the assumed common variance (see part (ii)) is $s^2 = \left((7\times 17.4^2) + (6\times 20.1^2)\right)/13$ = 349.491 (so s = 18.69), with 13 d.f.

Thus the test statistic for testing that the population means are the same is

$$\frac{\bar{x}_2 - \bar{x}_1 (-0)}{s\sqrt{\frac{1}{8}+\frac{1}{7}}} = \frac{4.0}{9.681} = 0.41,$$

which is referred to t_{13}. This is not significant at the 5% level (upper single-tailed 5% point is 1.771), so there is no evidence to reject the null hypothesis – it seems that the population means are the same.

(ii) The two underlying populations are assumed Normally distributed, with the same variance. A dot plot helps to check these assumptions:



Sample sizes are small but the ranges are similar and it may be reasonable to assume similar variances. However, Normality is in some doubt, as there is no central clustering and some skewness and/or outliers may be present.

Solution continued on next page

(iii) The Wilcoxon rank sum test (or, equivalently, the Mann-Whitney *U* form of this test) is suitable. The null hypothesis is that the population median at 250°C is equal to that at 200°C, and the alternative is that it is greater. We first rank all 15 data items, as follows.

Data	39	43	46	47	49	51	54	56	58	63	66	77	85	91	97
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	A	*B*	*A*	*B*	*A*	*A*	*B*	*B*	*A*	*B*	*A*	*A*	*B*	*A*	*B*

A refers to 200°C, *B* to 250°C.

The rank sum for the smaller sample, which is the sample for *B*, is 2 + 4 + 7 + 8 + 10 + 13 + 15 = 59.

We will again use a 5% significance level.

The required test is one-sided. As the alternative hypothesis is that the population median for *B* is greater than that for *A*, we want to refer the rank sum for *B* to the upper 5% point of the null distribution of the test statistic. However, the Society's tables for use in examinations, like most other sets of published tables, only provide lower-tail points.

One way to overcome this is to re-rank the data in descending order, i.e. ordering from the highest to the lowest. Having done that, it is obviously appropriate to refer to the lower-tail points.

However, the upper-tail points can easily be found from the tables, and it is more usual to proceed in this way. For the general case where the samples are of sizes n_1 and n_2, with that of size n_1 being used, the mean of the null distribution is $n_1(n_1 + n_2 + 1)/2$. This is shown in the tables as the mean of the asymptotic Normal distribution. It is obvious that the null distribution is symmetrical about its mean. So if a lower-tail point is located at *a* below the mean, the corresponding upper-tail point is located at *a* above the mean. Here, the lower-tail 5% point for the (7, 8) case is 41. The mean is $7 \times 16/2 = 56$. So the upper-tail 5% point is 71.

Thus the calculated sample statistic of 59 is (well) outside the upper 5% tail of the null distribution; we have no reason to reject the null hypothesis. It appears that the two populations are the same in this regard.

(iv) Neither test supports the hypothesis of an increase. The non-parametric test is more suitable for these data because there is doubt regarding the assumption of Normality that underlies the *t* test.

Question 2

(i) We have a 2×2 contingency table. The null hypothesis is that there is no association between an individual's sex and the chance of he or she having a recently recorded cholesterol measurement. The contingency table is as follows, with the expected frequencies in brackets in each cell (e.g. 88.48 = 131 × 206 / 305).

		Cholesterol level recorded		
		No	*Yes*	Total
Sex	*Female*	109 (88.48)	22 (42.52)	131
	Male	97 (117.52)	77 (56.48)	174
	Total	206	99	305

All the differences between observed and expected frequencies are ±20.52, becoming ±20.02 if Yates' correction is used. Thus the usual test statistic can be calculated as (using Yates' correction)

$$(20.02)^2 \left\{ \frac{1}{88.48} + \frac{1}{42.52} + \frac{1}{117.52} + \frac{1}{56.48} \right\} = 24.46$$

(or 25.70 if Yates' correction is not used). This is referred to χ_1^2. This is very highly significant (for example, the 1% point is 6.635); we have very strong evidence to reject the null hypothesis and conclude that there is an association.

(ii) $p_f - p_m$ is estimated by $\hat{p}_f - \hat{p}_m = \frac{22}{131} - \frac{77}{174} = 0.168 - 0.443 = -0.275$. The estimated variance of $\hat{p}_f - \hat{p}_m$ is given by

$$\frac{\hat{p}_f\left(1-\hat{p}_f\right)}{n_f} + \frac{\hat{p}_m\left(1-\hat{p}_m\right)}{n_m} = 0.001067 + 0.001418 = 0.002485 .$$

Thus the approximate 95% confidence interval for $p_f - p_m$ is given by –0.275 ± (1.96×√0.002485) i.e. it is (–0.177, –0.373).

(iii) There is clear evidence that the proportions are not the same for men and women. In part (i), this is interpreted via the very strong evidence of an association. In part (ii), the confidence interval does not contain 0, indeed it is a long way from 0, again giving very strong evidence of a real difference.

Question 3

Part (i)

(a) A type I error is to reject the null hypothesis, in favour of the alternative hypothesis, when in fact the null hypothesis is true.

(b) A type II error is to fail to reject the null hypothesis when in fact the alternative hypothesis is true.

(c) The level of significance of a test is the probability of rejecting the null hypothesis when in fact it is true, i.e. it is the probability of making a type I error. It is conventionally denoted by α.

(d) The power of a test is the probability of rejecting the null hypothesis, expressed as a function of the parameter (or equivalently, if it is not a test for a single parameter) being investigated. So it is given by $1 - \beta$, where β is the probability of making a type II error similarly expressed as a function.

Part (ii)

Let X represent the amount of coffee in a jar. We have $X \sim N(\mu , 15^2)$. The sample size is $n = 9$, so $\bar{X} \sim N(\mu, 15^2/9)$. Let $Z \sim N(0, 1)$.

(a) We have $\mu = 200$.

$$P\left(\bar{X} < 190\right) = P\left(Z < \frac{190-200}{15/3} = -2.0\right) = 0.02275.$$

$$P\left(\bar{X} > 210\right) = P\left(Z > \frac{210-200}{15/3} = 2.0\right) = 0.02275.$$

So the probability of committing a type I error is 0.02275 + 0.02275 = 0.0455.

(b) Here $\mu = 216$.

$$P\left(\bar{X} < 190\right) = P\left(Z < \frac{190-216}{15/3} = -5.2\right) = \text{ ZERO to several decimal places.}$$

$$P\left(\bar{X} < 210\right) = P\left(Z < \frac{210-216}{15/3} = -1.2\right) = 0.1151 .$$

So the total probability of accepting the output is 0.1151. (This is the probability of a Type II error for this procedure, i.e. the value of β, for $\mu = 216$. Thus the power of the procedure when in fact $\mu = 216$ is $1 - 0.1151 = 0.8849$.)

Question 4

(i) The observed (o) and expected (e) frequencies on the assumed (i.e. null) hypothesis are as given in the following table.

	AB	Ab	aB	ab	
o	773	231	238	59	(Total 1301)
e	731.81	243.94	243.94	81.31	(Total 1301)

The test statistic is

$$X^2 = \sum \frac{(o-e)^2}{e} = \frac{(773-731.81)^2}{731.81} + \dots + \frac{(59-81.31)^2}{81.31}$$

$$= 2.318 + 0.686 + 0.145 + 6.121 = 9.27,$$

which is referred to χ_3^2 (note there are no estimated parameters here). This is significant at the 5% level (critical point is 7.815). The null hypothesis is rejected at the 5% level; there is evidence that the ratio 9:3:3:1 does not apply here. We note that the main contribution to the test statistic is from the *ab* cell, so this is the principal course of discrepancy from the assumed ratio; fewer *ab* plants are observed than would be expected.

(b) First we consider only the first three cells of the table, and the observed and expected frequencies on the 9:3:3 null hypothesis are as follows.

	AB	Ab	aB	
o	773	231	238	(Total 1242)
e	745.2	248.4	248.4	(Total 1242)

The test statistic is

$$X^2 = \sum \frac{(o-e)^2}{e} = \frac{(773-745.2)^2}{745.2} + \frac{(231-248.4)^2}{248.4} + \frac{(238-248.4)^2}{248.4}$$

$$= 1.037 + 1.219 + 0.435 = 2.69,$$

which is referred to χ_2^2. This is not significant (the 5% point is 5.991). There is no evidence to reject the 9:3:3 (i.e. 3:1:1) null hypothesis for *AB*, *Ab* and *aB*.

Solution continued on next page

Now we combine the first three cells and reintroduce the *ab* cell, as follows.

	Combined	*ab*	
o	1242	59	(Total 1301)
e	1219.69	81.31	(Total 1301)

The test statistic is

$$X^2 = \sum \frac{(o-e)^2}{e} = \frac{(1242-1219.69)^2}{1219.69} + \frac{(59-81.31)^2}{81.31} = 0.408 + 6.121 = 6.53 ,$$

which is referred to χ_1^2. This is significant at the 5% level (critical point is 3.841). The null hypothesis is rejected at the 5% level; there is evidence that the ratio 15:1 does not apply here.

Overall, the results suggest that the *ab* type is not occurring with sufficient frequency, since it is mainly responsible for the discrepancies in both (i) and (ii), in both cases with lower *o* than *e*. Hence the suggestion of lower survival rate is the one to follow up.

(iii) The proportion for the *ab* type, *p*, is estimated by $\hat{p} = \frac{59}{1301} = 0.0453$. The estimated variance of $\hat{p}$ is given by

$$\frac{\hat{p}(1-\hat{p})}{n} = 0.00003324 .$$

Thus the approximate 95% confidence interval for *p* is given by 0.0453 ± (1.96×√0.00003324) i.e. it is (0.0340, 0.0566).

We note that 1/16 (= 0.0625) is not contained in this interval, the interval being entirely below that value. Thus there is evidence that the true proportion of type *ab* is less than 1/16.