THE ROYAL STATISTICAL SOCIETY

2004 EXAMINATIONS – SOLUTIONS

HIGHER CERTIFICATE

PAPER II – STATISTICAL METHODS

The Society provides these solutions to assist candidates preparing for the examinations in future years and for the information of any other persons using the examinations.

The solutions should NOT be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids.

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many 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 these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

<u>Higher Certificate, Paper II, 2004. Question 1</u>

(i) $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$, $i = 1, 2, ..., v$, $j = 1, 2, ..., b$, $\{\varepsilon_{ij}\} \sim \text{ind N}(0, \sigma^2)$.

There are v treatments and b blocks. y_{ij} is the observation (response) on the unit (plot) in block j which receives treatment i. $\Sigma_i \tau_i = 0$, $\Sigma_j \beta_j = 0$ (i.e. fixed effects model). μ is the overall population general mean, τ_i the population mean effect due to treatment i, β_j the population mean effect due to block j. The Normally distributed residual (error) terms ε_{ij} all have variance σ^2 and are uncorrelated (independent). All non-random variation is covered by the τ_i and β_j terms.

(ii) The "blocks" here are subjects 1 to 6. The "treatments" are compounds A to D. In the notation of part (i), $v = 4$ and $b = 6$.

Totals are:

Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
13	24	7	23	8	21

Treatment A	Treatment B	Treatment C	Treatment D
25	23	18	30

The grand total is 96. $\Sigma\Sigma y_{ij}^2 = 486$.

"Correction factor" is $\dfrac{96^2}{24} = 384$.

Therefore total SS = 486 – 384 = 102.

SS for blocks = $\dfrac{13^2}{4} + \dfrac{24^2}{4} + \dfrac{7^2}{4} + \dfrac{23^2}{4} + \dfrac{8^2}{4} + \dfrac{21^2}{4} - 384 \;=\; 457 - 384 \;=\; 73$.

SS for treatments = $\dfrac{25^2}{6} + \dfrac{23^2}{6} + \dfrac{18^2}{6} + \dfrac{30^2}{6} - 384 \;=\; 396.33 - 384 \;=\; 12.33$.

The residual SS is obtained by subtraction.

SOURCE	DF	SS	MS	F value
Blocks	5	73.00	14.600	13.14 Compare $F_{5,15}$
Treatments	3	12.33	4.111	3.70 Compare $F_{3,15}$
Residual	15	16.67	1.111	$= \hat{\sigma}^2$
TOTAL	23	102.00		

The upper 0.1% point of $F_{5,15}$ is 7.57; the blocks effect is very highly significant.

The upper 5% point of $F_{3,15}$ is 3.29; the treatments effect is significant.

Continued on next page

Clearly there are block (subject) differences. Even after removing these, the results are quite variable.

To investigate treatment differences, first calculate the treatment means, which are (in ascending order, for clarity)

$$C:\ 3.00 \qquad B:\ 3.8333 \qquad A:\ 4.1667 \qquad D:\ 5.00$$

The least significant difference between any pair of these means is

$$t_{15}\sqrt{\frac{2\times 1.111}{6}} = 0.6086\,t_{15} \qquad \text{where } t_{15} = \begin{cases} 2.131 & \text{at } 5\% \\ 2.947 & \text{at } 1\% \\ 4.073 & \text{at } 0.1\% \end{cases}$$

so the least significant differences are 1.30 for 5%, 1.79 for 1% and 2.48 for 0.1%. Thus the only apparent difference is between *C* and *D*, significant at the 1% level.

The results must be interpreted with caution.

The data are on a 10-point scale of integers, so obviously cannot have an underlying Normal distribution. However, when the means of 6 replicates are being compared, the (necessarily approximate) results should give a good guide to likely treatment (i.e. compound) differences.

Some of the subjects are considerably more prone to irritation than others. Because of this, the underlying variances might be different in some blocks from others. This would be contrary to an assumption in the modelling, and would thus be a further feature making the results only approximate.

Higher Certificate, Paper II, 2004. Question 2

(i) $n = 16$. $\Sigma x_i = 49.4$, $\Sigma x_i^2 = 157.3$; $\bar{x} = 3.0875$, $s^2 = 0.3185$.

We need to assume that diameters are Normally distributed.

A 95% confidence interval for the true mean of this grower's tomatoes is given by $\bar{x} \pm t\,s/\sqrt{16}$ where t is the double-tailed 5% point of t_{15}, i.e. 2.131. So the interval is $3.0875 \pm 2.131\sqrt{0.3185/16}$, i.e. 3.0875 ± 0.3007, i.e. (2.787, 3.388).

As the specified mean of 3.0 is within this interval, it seems this grower could be accepted.

It is also specified that the true variance σ^2 should not be greater than $(0.5)^2$, which is 0.25. A 95% confidence interval for σ^2 is given by

$$\frac{(n-1)s^2}{\chi^2_{\mathrm{U}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\mathrm{L}}}$$

where χ^2_{L} and χ^2_{U} are the lower and upper 2½% points of χ^2_{n-1}, i.e. of χ^2_{15}, which are 6.262 and 27.488. Thus the interval is $0.1738 < \sigma^2 < 0.7629$, which is equivalent to $0.42 < \sigma < 0.87$. This interval does contain the specified greatest value of 0.5 for σ, but caution is suggested by the fact that the upper limit is well above 0.5; the grower might well not be acceptable on this basis. (Note that the comparatively large value of s^2 has also affected the confidence interval for the mean calculated above – it is, relatively speaking, rather a wide interval.)

The short report should say that although the mean diameter in the sample is near to 3.0, the material is so variable that the specified greatest value of 0.5 for the standard deviation is quite likely to be exceeded, perhaps by a substantial amount. If the directors are still interested, they should examine a larger sample.

The variability could well contain a large between-plant component, so a method which mainly measures within-plant variation is not a good one – however quick and easy it may be.

Higher Certificate, Paper II, 2004. Question 3

(i) Parametric methods need the assumption that the data come from a known distribution, often the Normal distribution for continuous data or the binomial or Poisson for discrete data. When these assumptions are satisfied, parametric tests of hypotheses are the most powerful tests available. However, when these assumptions are not satisfied, any tests or confidence intervals based on them are likely to give wrong conclusions.

Non-parametric methods do not need distributional assumptions (even though the test statistics actually used may have Normal approximations for adequate sizes of sample). They are often based on ordering or ranking, and will serve for skewed data and for ordered (and some categorical) data. They are often surprisingly powerful. Nevertheless, when the conditions for a parametric test are satisfied, at least to a good approximation, it should be used in preference to a non-parametric one to give greater precision.

(ii) (a) Stems of 5 give the following.

BOYS

0	0 2 3
(5)	5 6
10	
(15)	6 8
20	1
(25)	8
30	0

GIRLS

0	0 1 3
(5)	7 8 9 9
10	2
(15)	6
20	1

In each case the data appear to be skewed to the right. The non-Normality is very pronounced, and a t test needs to make the assumption of Normality – so it is not suitable.

(b) A Mann-Whitney U test (or equivalently a Wilcoxon rank sum test) may be used. The data and ranks are as follows, using average ranks for ties.

0	0	1	2	3	3	5	6	7	8	9	9	12	16	16	18	21	21	28	30
1½	1½	3	4	5½	5½	7	8	9	10	11½	11½	13	14½	14½	16	17½	17½	19	20
B	G	G	B	B	G	B	B	G	G	G	G	G	B	G	B	B	G	B	B

$n_1 = 10$, $n_2 = 10$. Total rank for boys $R_B = 113$; total rank for girls $R_G = 97$.

Calculating the Mann-Whitney statistic via the ranks (note: it can also be calculated directly, or the Wilcoxon rank-sum form could be used),

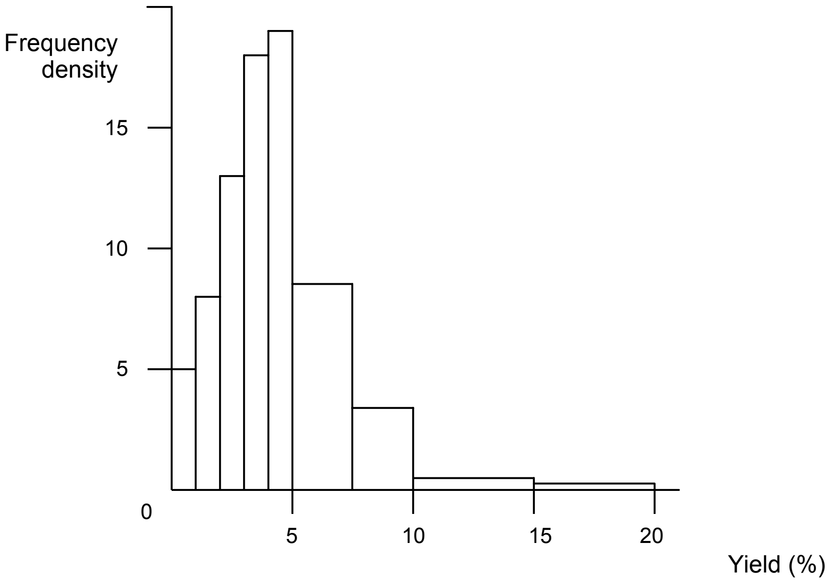
$$U_1 = n_1 n_2 + \tfrac{1}{2} n_1 (n_1 + 1) - R_B = 100 + 55 - 113 = 42.$$
$$U_2 = n_1 n_2 + \tfrac{1}{2} n_2 (n_2 + 1) - R_G = 100 + 55 - 97 = 58.$$

So $U_{min} = 42$. From tables, the critical value for a U test with $n_1 = n_2 = 10$ at the 5% two-tailed level is 23. As $42 > 23$, we accept the null hypothesis that there is no difference between the distributions.

Higher Certificate, Paper II, 2004. Question 4

Yield (%)	Interval width	Frequency f	Frequency density	Midpoint x	fx	fx^2	Cum freq F
≥ 0 but < 1	1	5	5	0.5	2.5	1.25	5
≥ 1 but < 2	1	8	8	1.5	12.0	18.00	13
≥ 2 but < 3	1	13	13	2.5	32.5	81.25	26
≥ 3 but < 4	1	18	18	3.5	63.0	220.50	44
≥ 4 but < 5	1	19	19	4.5	85.5	384.75	63
≥ 5 but < 7.5	2.5	21	8.4	6.25	131.25	820.3125	84
≥ 7.5 but < 10	2.5	8	3.2	8.75	70.0	612.50	92
≥ 10 but < 15	5	2	0.4	12.5	25.0	312.50	94
≥ 15 but < 20	5	1	0.2	17.5	17.5	306.25	95
		95			439.25	2757.3125	

(i)



(ii) The modal class interval is "≥ 4 but < 5" (based, of course, on frequency density; it would be "≥ 5 but < 7.5" if based only on frequency).

$\bar{x} = 439.25/95 = 4.62$ (%).

For the median, we require the 48th observation from the beginning, which is estimated as being at $4 + (4/19 \times 1) = 4.21$ (%).

$$s^2 = \frac{1}{94}\left(2757.3125 - \frac{439.25^2}{95}\right) = 7.7272. \text{ So } s = 2.78 \text{ (\%)}.$$

(iii) These are only estimates because we have the data grouped into intervals, not the 95 individual values.

(iv) p is estimated by $\hat{p} = 32/95 = 0.337$. The estimated variance of $\hat{p}$ is $(0.337)(0.663)/95 = 0.002352$, so the estimated standard deviation is 0.0485. Thus a 95% confidence interval for p is given by, approximately, $0.337 \pm (1.96)(0.0485)$, i.e. it is $(0.242, 0.432)$.

Higher Certificate, Paper II, 2004. Question 5

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 $X \sim \mathrm{N}(\mu, 15^2)$. The sample size is $n = 9$, so $\bar{X} \sim \mathrm{N}(\mu, 15^2/9)$. Let $Z \sim \mathrm{N}(0, 1)$.

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

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

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

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

(b) Here $\mu = 216$.

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

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

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 $1 - 0.1151 = 0.8849$.)

Higher Certificate, Paper II, 2004. Question 6

Part (i)

(a) If a set of data can be assumed to be a sample from a Normal distribution (with mean μ and variance σ^2), the sample mean ($\bar{x}$) from a sample of size n itself has an underlying Normal distribution (with mean μ and variance σ^2/n). Thus the null hypothesis $\mu = \mu_0$, where μ_0 is a specified value, is tested using the test statistic

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

and referring this to the N(0, 1) distribution.

In practice it is unusual to know σ^2. It might be known from past experience ("historical data"), and this can for example arise in some industrial statistical work, in which case this method can be used.

In large samples from any distribution that is reasonably symmetrical, the method works to a very good level of approximation using the estimated variance s^2 instead of σ^2. As an example, economic data (sales etc) are often so treated. How large the sample needs to be depends on the symmetry of the underlying distribution.

Many "Normal approximations" of other measurements exist. An example is the observed proportion in a binomial situation, such as the proportion of households owning more than one car; the size of sample required for the approximation to be good depends on the value of the proportion.

(b) If data are sampled from a Normal distribution where σ^2 is not known (so that s^2 is used in its place) and where the sample size is small – typically smaller than 30 – the t distribution must be used instead of N(0, 1) as above. Most sets of biological and agricultural data are like this; the samples are not large, and the underlying variability needs to be estimated from the data in each new experiment or study.

Part (ii)

$n_S = 10;\ \bar{x}_S = 2159,\ s_S{}^2 = 22360.44.\qquad n_N = 10;\ \bar{x}_N = 2830,\ s_N{}^2 = 19310.22.$

The "pooled estimate" of variance is $s^2 = 20835.33$.

The test statistic for testing the null hypothesis $\mu_N - \mu_S = 600$ is

$$\frac{\bar{x}_N - \bar{x}_S - 600}{s\sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{71}{64.55} = 1.10,$$

which is referred to t_{18}. This is not significant, so the null hypothesis cannot be rejected. (The alternative hypothesis would be "difference > 600". Presumably the result of the test would lead to the change not being made, on economic grounds.)

The time to recharging in each case is assumed to be Normally distributed, and the variances of these Normal distributions are assumed to be equal.

Higher Certificate, Paper II, 2004. Question 7

Part (i)

(a) Any scores occurring for more than one subject lead to an average rank being given. This is shown for some subjects in part (b).

(b)

Subject	1	2	3	4	5	6	7	8	9	10	11	12
IT rank	5	12	1	10	11	6	8	3	4	2	9	7
LT rank	2	12	6½	8½	10½	1	4½	4½	6½	3	10½	8½
Difference d_i	3	0	–5½	1½	½	5	3½	–1½	–2½	–1	–1½	–1½

$\Sigma d_i^2 = 9 + 0 + \ldots + 2.25 = 93.$

Spearman's coefficient $r_S = 1 - \dfrac{6\Sigma d^2}{n\left(n^2-1\right)} = 1 - \dfrac{558}{1716} = 0.675$.

From tables, this is significantly different from zero at the 5% level (two-sided).

[Note. Use of this formula for r_s leads to slight inaccuracy where there are tied ranks – but unlikely to make much difference here.]

(c) There is evidence to reject the null hypothesis of no correlation, so it seems that there is some association between intelligence and the ability to think laterally; but it does not appear that the relationship is a very strong one.

Part (ii)

Because of the pairing, McNemar's test should be used. It gives the test statistic

$$\frac{(15-9)^2}{15+9} = \frac{36}{24} = 1.5\,,$$

which is referred to χ_1^2. This is not significant, so we do not reject the null hypothesis that the two sexes have the same abilities to think laterally.

Higher Certificate, Paper II, 2004. Question 8

(a) The total number of nests is 240. On the null hypothesis that each of the 8 directional categories has the same probability of being used, the expected number will be 30 for each. Thus the observed and expected frequencies (O and E) are as follows.

Position	N	NE	E	SE	S	SW	W	NW
O	27	24	35	33	38	33	24	26
E	30	30	30	30	30	30	30	30

The test statistic is

$$X^2 = \sum \frac{(O-E)^2}{E} = \frac{3^2}{30} + \frac{6^2}{30} + \ldots + \frac{4^2}{30} = \frac{204}{30} = 6.80,$$

which is referred to χ_7^2 (note 7 degrees of freedom because the table has 8 cells and there are no estimated parameters here). This is not significant (the 5% point is 14.07); we cannot reject the null hypothesis.

Splitting the data table into 8 positions, each with a fairly small expected frequency for a chi-squared test, limits the power.

A Kolmogorov-Smirnov test uses an empirical cumulative distribution function, taking a starting point which in this case would be arbitrary, say N (North), and following in order round the positions. Thus it is not testing a relevant null hypothesis for this problem.

(b) We have a 2×3 contingency table. The null hypothesis is that males and females have the same ratio of preferences for designs A, B and C. The contingency table is as follows, with the expected frequencies in brackets in each cell.

	A	*B*	*C*	Total
Males	129 (120)	24 (32)	47 (48)	200
Females	126 (135)	44 (36)	55 (54)	225
Total	255	68	102	425

The test statistic is

$$X^2 = \sum \frac{(O-E)^2}{E} = \frac{9^2}{120} + \frac{8^2}{32} + \frac{1^2}{48} + \frac{9^2}{135} + \frac{8^2}{36} + \frac{1^2}{54} = 5.09,$$

which is referred to χ_2^2. This is not significant (the 5% point is 5.99); we cannot reject the null hypothesis.