Paper Reference(s)

6691/01

Edexcel GCE

Statistics S3

Advanced Level

Thursday 21 June 2012 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination	**Items included with question papers**
Mathematical Formulae (Pink)	Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6691), your surname, other name and signature.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
This paper has 7 questions.
The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You must show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

P40472A

1. Interviews for a job are carried out by two managers. Candidates are given a score by each manager and the results for a random sample of 8 candidates are shown in the table below.

Candidate	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Manager X	62	56	87	54	65	15	12	10
Manager Y	54	47	71	50	49	25	30	44

(*a*) Calculate Spearman's rank correlation coefficient for these data.
(5)

(*b*) Test, at the 5% level of significance, whether there is agreement between the rankings awarded by each manager. State your hypotheses clearly.
(5)

Manager Y later discovered he had miscopied his score for candidate D and it should be 54.

(*c*) Without carrying out any further calculations, explain how you would calculate Spearman rank correlation in this case.
(2)

---

2. A lake contains 3 species of fish. There are estimated to be 1400 trout, 600 bass and 450 pike in the lake. A survey of the health of the fish in the lake is carried out and a sample of 30 fish is chosen.

(*a*) Give a reason why stratified random sampling cannot be used.
(1)

(*b*) State an appropriate sampling method for the survey.
(1)

(*c*) Give one advantage and one disadvantage of this sampling method.
(2)

(*d*) Explain how this sampling method could be used to select the sample of 30 fish. You must show your working.
(4)

---

3. (*a*) Explain what you understand by the Central Limit Theorem.

A garage services hire cars on behalf of a hire company. The garage knows that the lifetime of the brake pads has a standard deviation of 5000 miles. The garage records the lifetimes, x miles, of the brake pads it has replaced. The garage takes a random sample of 100 brake pads and finds that $\sum x = 1\ 740\ 000$.

(*b*) Find a 95% confidence interval for the mean lifetime of a brake pad.

(5)

(*c*) Explain the relevance of the Central Limit Theorem in part (*b*).

(2)

Brake pads are made to be changed very 20 000 miles on average. The hire car company complain that the garage is changing the brake pads too soon.

(*d*) Comment on the hire company's complaint. Give a reason for your answer.

(2)

---

4. Two breeds of chicken are surveyed to measure their egg yield. The results are shown in the table below.

Breed \ Egg yield	Low	Medium	High
Leghorn	22	52	26
Cornish	14	32	4

Showing each stage of your working clearly, test, at the 5% level of significance, whether or not there is an association between egg yield and breed of chicken. State your hypotheses clearly.

(10)

5. Mr Allan and Ms Burns are two mathematics teachers teaching mixed ability groups of students in a large college. At the end of the college year all students took the same examination. A random sample of 29 of Mr Allan's students and a random sample of 26 of Ms Burns' students are chosen. The results are summarised in the table below.

	Sample Size, n	Mean, $\bar{x}$	Standard Deviation, s
Mr Allan	29	80	10
Ms Burns	26	74	15

(*a*) Stating your hypothesis clearly, test, at the 10% level of significance, whether there is evidence that there is a difference in the means scores of their students.

(6)

Ms Burns thinks the comparison was unfair as the examination was set by Mr Allan. She looks up a different set of examination marks for these students and, although Mr Allan's sample has a higher mean, she calculates the test statistic for this new set of results to be 1.6.

However, Mr Allan now claims that the mean marks of his students are higher than the mean marks of Ms Burns' students.

(*b*) Test Mr Allan's claim, stating the hypothesis and critical values you would use. Use a 10% level of significance.

(3)

6. A total of 100 random samples of 6 items are selected from a production line n a factory and the number of defective items in each sample is recorded. The results are summarised in the table below.

Number of defective items	0	1	2	3	4	5	6
Number of samples	6	16	20	23	17	10	8

(*a*) Show that the mean number of defective items per sample is 2.91.

(2)

A factory manager suggests that the data can be modelled by a binomial distribution with $n = 6$. He uses the mean from the sample above and calculates expected frequencies as shown in the table below.

Number of defective items	0	1	2	3	4	5	6
Expected frequency	1.87	10.54	24.82	a	22.01	8.29	b

(*b*) Calculate the value of a and the value of b, giving your answers to 2 decimal places.

(4)

(*c*) Test, at the 5% level, whether or not the binomial distribution is a suitable model for the number of defective items in samples of 6 items. State your hypotheses clearly.

(8)

---

7. The heights, in cm, of the male employees in a large company follow a normal distribution with mean 177 and standard deviation 5.

The heights, in cm, of the female employees follow a normal distribution with mean 163 and standard deviation 4.

A male employee and a female employee are chosen at random.

(*a*) Find the probability that the male employee is taller than the female employee.

(5)

Six male employees and four female employees are chosen at random.

(*b*) Find the probability that their total height is less than 17 m.

(6)

TOTAL FOR PAPER: 75 MARKS

END

Question Number	Scheme	Marks
1 (a)		

X	Y	Rank X	Rank Y	d	d^2
62	54	3	2	1	1
56	47	4	5	-1	1
87	71	1	1	0	0
54	50	5	3	2	4
65	49	2	4	-2	4
15	25	6	8	-2	4
12	30	7	7	0	0
10	44	8	6	2	4

M1
M1

$\sum d^2 = 18$ — A1

$r_s = 1 - \frac{6 \times 18}{8(64-1)} = 0.7857...$ awrt 0.786 — M1A1

(5)

(b)

$H_0 : \rho = 0$ — B1

$H_0 : \rho > 0$ — B1

Critical region $r_s > 0.6429$ — B1

(0.7857>0.6429 sufficient evidence to) reject H_0 — M1

There is evidence of agreement between the scores awarded by each manager — A1ft

(5)

(c)

(A and D are now) tied ranks (for Manager Y) — B1

Average rank (awarded to A and D) **and** use $r_s = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ — B1

(2)

Total 12

Question Number	Scheme	Marks
2(a)	Sampling frame within each species of fish in the lake impossible to obtain.	B1 **(1)**
(b)	Quota sampling	B1 **(1)**
(c)	Advantages: Sample can be obtained quickly Costs are kept to a minimum Administration of survey is easy Disadvantages: Not possible to estimate sampling errors Process not random Surveyor may not be able to identify species of fish easily	B1 B1 **(2)**
(d)	Species / Quota: Trout: $\frac{1400}{2450}\times 30 = 17.14$ Bass: $\frac{600}{2450}\times 30 = 7.35$ Pike: $\frac{450}{2450}\times 30 = 5.51$ Fish are caught from the lake until the quota of 17 trout, 7 bass and 6 pike are reached. If a fish is caught and the species quota is full, then this is ignored.	B1B1B1 B1 (4)
3(a)	($X_1, X_2, X_3 ...., X_n$ is a random) **sample** of size n , for n is **large**, (from a population with mean μ and variance σ^2) then $\bar{X}$ is (approximately) Normal.	B1 B1 **(2)**
(b)	$\bar{x} = \frac{1740000}{100} = 17400$ $\bar{x} \pm z\frac{\sigma}{\sqrt{n}}, = 17400 \pm 1.96 \times \frac{5000}{\sqrt{100}}$ [16420,18380]	B1 M1, B1 A1A1 **(5)**
(c)	$\bar{X} \sim$ Normal (approx) by CLT, and normal needed to find CI.	B1,B1 **(2)**
3 (d)	20000 **above** upper confidence limit (**not** just outside) Complaint justified.	B1ft dB1ft **(2)**

Question Number	Scheme	Marks
4	H_0 : Egg yield and breed of chicken are independent (not associated) H_1 : Egg yield and breed of chicken are dependent (associated)	B1
	(expected frequencies table below)	M1A1
	(O, E table below)	M1A1
	$\sum\frac{(O-E)^2}{E} = 6.757...$ or $\sum\frac{O^2}{E} - 100 = 6.757...$	A1
	$\nu = 2, \chi_2^2(5\%) = 5.991$	B1B1ft
	(6.757>5.991 so sufficient evidence to) reject H_0	M1
	Egg yield and breed of chicken are dependent (associated)	A1 **(10)**
5(a)	$H_0 : \mu_A = \mu_B$ $H_1 : \mu_A \neq \mu_B$	B1
	$z = \frac{\pm(80-74)}{\sqrt{\frac{100}{29}+\frac{225}{26}}}$	M1A1
	$z = \pm 1.7247...$ awrt ± 1.72	A1
	1.7247>1.6449 o.e. so reject H_0	dM1
	There is evidence of a difference in the (mean) scores of their students.	A1 **(6)**
(b)	(For z=1.6, test above not significant so no evidence of a difference.) For Mr A's claim, $H_0 : \mu_A = \mu_B$, $H_1 : \mu_A > \mu_B$, and critical value is z=1.2816	B1, B1
	(Both z values significant,) Mr Alan's claim supported.	B1 **(3)**

Question 4 expected frequencies:

Egg Yield / Breed	Low	Medium	High	Total
Leghorn	$\frac{100\times 36}{150} = 24$	$\frac{100\times 84}{150} = 56$	$\frac{100\times 30}{150} = 20$	100
Cornish	$\frac{50\times 36}{150} = 12$	$\frac{50\times 84}{150} = 28$	$\frac{50\times 30}{150} = 10$	50
Total	36	84	30	150

O	E	$\sum\frac{(O-E)^2}{E}$	$\sum\frac{O^2}{E}$
22	24	0.166667	20.2
52	56	0.285714	48.3
26	20	1.8	33.8
14	12	0.333333	16.3
32	28	0.571429	36.6
4	10	3.6	1.6

		Total 9

Question Number	Scheme	Marks
6(a)	Mean$=\dfrac{1\times16+2\times20+...+6\times8}{100}=2.91$ **ag**	M1A1
		(2)
(b)	$p=\dfrac{2.91}{6}=0.485$	B1
	$a=100\times C_3^6\times0.485^3\times0.515^3=31.17$	M1A1
	$b=100\times0.485^6=1.3(0)$	A1
		(4)
(c)	H_0 : Binomial is a good fit H_1 : Binomial is a not a good fit	B1
	(table below)	M1
	$\sum\dfrac{(O-E)^2}{E}=\dfrac{(22-12.41)^2}{12.41}+\dfrac{(20-24.82)^2}{24.82}+...+\dfrac{(18-9.59)^2}{9.59}=18.998...$ awrt 19.0	M1A1
	ν =5-2=3 degrees of freedom	B1
	$\chi_3^2(5\%)=7.815$	B1ft
	18.998...>7.815 so reject H_0	M1
	Binomial is a not a good fit (and is not a good model for the number of defective items in samples of size 6)	A1
		(8)
		Total 14

Number of defective items	0 or 1	2	3	4	5 or 6
O	22	20	23	17	18
E	12.41	24.82	31.17	22.01	9.59

Question Number	Scheme	Marks
7(a)	$M \sim \mathrm{N}(177, 25)$, $F \sim \mathrm{N}(163,16)$	
	$\mathrm{E}(M - F) = 177 - 163 = 14$	B1
	$\mathrm{Var}(M - F) = 25 + 16 = 41$	M1A1
	$M - F \sim \mathrm{N}(14, 41)$	
	$\mathrm{P}(M - F > 0) = \mathrm{P}\left(Z > \frac{-14}{\sqrt{41}}\right)$ or $\mathrm{P}\left(Z < \frac{14}{\sqrt{41}}\right)$	M1
	$= \mathrm{P}(Z < 2.186...)$	
	$= 0.9854$ or 0.9856 by calculator awrt 0.985 or 0.986	A1
		(5)
7(b)	$W = M_1 + M_2 + ...M_6 + F_1 + F_2 + ...F_4$	
	$\mathrm{E}(W) = 6 \times 177 + 4 \times 163$	
	$= 1714$	B1
	$\mathrm{Var}(W) = 6 \times 25 + 4 \times 16$	M1
	$= 214$	A1
	$\mathrm{P}(W < 1700) = \mathrm{P}\left(Z < \frac{1700 - 1714}{\sqrt{214}}\right)$ or $\mathrm{P}\left(Z > \frac{1714 - 1700}{\sqrt{214}}\right)$	M1
	$= \mathrm{P}(Z < -0.957..)$ awrt $Z < -0.96$ or $Z > 0.96$	A1
	$= 1 - 0.8315$	
	$= 0.1685$ awrt 0.169	A1
	(0.1693 by calculator)	**(6)**
		Total 11