Paper Reference(s)

6691/01

Edexcel GCE

Statistics S3

Advanced Level

Monday 16 June 2008 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Green)	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 gain no credit.

1. Some biologists were studying a large group of wading birds. A random sample of 36 were measured and the wing length, x mm, of each wading bird was recorded. The results are summarised as follows.

$$\sum x = 6046, \quad \sum x^2 = 1\,016\,338.$$

(*a*) Calculate unbiased estimates of the mean and the variance of the wing lengths of these birds.

(3)

Given that the standard deviation of the wing lengths of this particular type of bird is actually 5.1 mm,

(*b*) find a 99 % confidence interval for the mean wing length of the birds from this group.

(5)

---

2. Students in a mixed sixth form college are classified as taking courses in either Arts, Science or Humanities. A random sample of students from the college gave the following results.

		Course		
		Arts	Science	Humanities
Gender	Boy	30	50	35
	Girl	40	20	42

Showing your working clearly, test, at the 1 % level of significance, whether or not there is an association between gender and the type of course taken. State your hypotheses clearly.

(11)

---

3. The product moment correlation coefficient is denoted by r and Spearman's rank correlation coefficient is denoted by r_s.

(*a*) Sketch separate scatter diagrams, with five points on each diagram, to show

(i) $r = 1$,

(ii) $r_s = -1$ but $r > -1$.

(3)

Two judges rank seven collie dogs in a competition. The collie dogs are labelled *A* to *G* and the rankings are as follows.

Rank	1	2	3	4	5	6	7
Judge 1	*A*	*C*	*D*	*B*	*E*	*F*	*G*
Judge 2	*A*	*B*	*D*	*C*	*E*	*G*	*F*

(*b*) (i) Calculate Spearman's rank correlation coefficient for these data.

(6)

(ii) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the judges are generally in agreement.

(5)

---

4. The weights of adult men are normally distributed with a mean of 84 kg and a standard deviation of 11 kg.

(*a*) Find the probability that the total weight of 4 randomly chosen adult men is less than 350 kg.

(5)

The weights of adult women are normally distributed with a mean of 62 kg and a standard deviation of 10 kg.

(*b*) Find the probability that the weight of a randomly chosen adult man is less than one and a half times the weight of a randomly chosen adult woman.

(6)

---

5. A researcher is hired by a cleaning company to survey the opinions of employees on a proposed pension scheme. The company employs 55 managers and 495 cleaners.

To collect data the researcher decides to give a questionnaire to the first 50 cleaners to leave at the end of the day.

(*a*) Give 2 reasons why this method is likely to produce biased results.

(2)

(*b*) Explain briefly how the researcher could select a sample of 50 employees using

(i) a systematic sample,

(ii) a stratified sample.

(6)

Using the random number tables in the formulae book, and starting with the top left hand corner (8) and working across, 50 random numbers between 1 and 550 inclusive were selected. The first two suitable numbers are 384 and 100.

(*c*) Find the next two suitable numbers.

(2)

---

6. Ten cuttings were taken from each of 100 randomly selected garden plants. The numbers of cuttings that did not grow were recorded.

The results are as follows.

No. of cuttings which did not grow	0	1	2	3	4	5	6	7	8, 9 or 10
Frequency	11	21	30	20	12	3	2	1	0

(*a*) Show that the probability of a randomly selected cutting, from this sample, not growing is 0.223.

(2)

A gardener believes that a binomial distribution might provide a good model for the number of cuttings, out of 10, that do not grow.

He uses a binomial distribution, with the probability 0.2 of a cutting not growing. The calculated expected frequencies are as follows.

No. of cuttings which did not grow	0	1	2	3	4	5 or more
Expected frequency	*r*	26.84	*s*	20.13	8.81	*t*

(*b*) Find the values of *r*, *s* and *t*.

(4)

(*c*) State clearly the hypotheses required to test whether or not this binomial distribution is a suitable model for these data.

(2)

The test statistic for the test is 4.17 and the number of degrees of freedom used is 4.

(*d*) Explain fully why there are 4 degrees of freedom.

(2)

(*e*) Stating clearly the critical value used, carry out the test using a 5% level of significance.

(3)

---

7. A sociologist is studying how much junk food teenagers eat. A random sample of 100 female teenagers and an independent random sample of 200 male teenagers were asked to estimate what their weekly expenditure on junk food was. The results are summarised below.

	n	mean	s.d.
Female teenagers	100	£5.48	£3.62
Male teenagers	200	£6.86	£4.51

(*a*) Using a 5% significance level, test whether or not there is a difference in the mean amounts spent on junk food by male teenagers and female teenagers. State your hypotheses clearly.

(7)

(*b*) Explain briefly the importance of the central limit theorem in this problem.

(1)

TOTAL FOR PAPER: 75 MARKS

END

edexcel

June 2008
6691 Statistics S3
Mark Scheme

Question number	Scheme		Marks
1. (a)	$\bar{x} = \left(\frac{6046}{36} =\right) 167.94...$	awrt **168**	B1
	$s^2 = \frac{1016338 - 36 \times \bar{x}^2}{35}$		M1
	$= 27.0253....$	awrt **27.0** (Accept 27)	A1 (3)
(b)	99% Confidence Interval is: $\bar{x} \pm 2.5758 \times \frac{5.1}{\sqrt{36}}$		M1A1ft
		2.5758	B1
	$= (165.755\ldots, 170.133\ldots)$	awrt (**166,170**)	A1 A1 (5)
			8 marks

(a) M1 for a correct expression for s^2, follow through their mean, beware it is very "sensitive"

$167.94 \rightarrow \frac{999.63..}{35} \rightarrow 28.56...$

$167.9 \rightarrow \frac{1483.24..}{35} \rightarrow 42.37...$

$168 \rightarrow \frac{274}{35} \rightarrow 7.82$

These would all score M1A0

Use of 36 as the divisor (= 26.3…) is M0A0

(b) M1 for substituting their values in $\bar{x} \pm z \times \frac{5.1 \text{ or } s}{\sqrt{36}}$ where z is a recognizable value from tables

1st A1 follow through their mean and their z (to 2dp) in $\bar{x} \pm z \times \frac{5.1}{\sqrt{36}}$

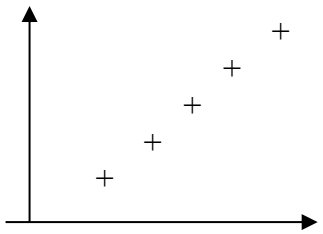
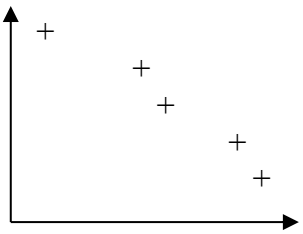
Beware: $167.94 \pm 2.5758 \times \frac{5.1^2}{36} \rightarrow (166.07..,169.8..)$ but scoresB1M0A0A0A0

Correct answer only in (b) scores 0/5

2nd & 3rd A marks depend upon 2.5758 **and** M mark.

Question number	Scheme	Marks
2.	$\frac{115\times 70}{217}=37.0967...$ or $\frac{1150}{31}$ (etc) $\frac{1265}{31},\frac{1020}{31},\frac{1122}{31}$	M1
		A1A1
	H_0 : There is no association between course and gender	
	H_1 : There is some association between course and gender (both)	B1
	$\sum\frac{(O-E)^2}{E}=\frac{(37.1-30)^2}{37.1}+\frac{(32.9-40)^2}{32.9}+...+\frac{(36.2-42)^2}{36.2}$	M1A1ft
	= 1.358 + 4.485 + 0.824 + 1.532 + 5.058 + 0.929 = 14.189… awrt **14.2**	A1
	$\nu=(3-1)(2-1)=2,$ $\chi_2^2(1\%)$ critical value is 9.210 (condone 9.21)	B1, B1ft
	Significant result or reject null hypothesis	M1
	There is evidence of an association between course taken and gender	A1ft (11)
	[Correct answers only score full marks]	**11 marks**
ALT	$\sum\frac{O^2}{E}-N=\frac{30^2}{37.1}+\frac{40^2}{32.9}+...+\frac{42^2}{36.2}-217$	M1A1ft
	1st M1 for some use of the $\frac{\text{row total}\times\text{col total}}{\text{grand total}}$ formula	
	1st A1 for one correct row or one correct column of expected frequencies to nearest integer	
	2nd A1 for all expected frequencies correct to awrt 1 dp (Allow exact fractions)	
	1st B1 for hypotheses. Independence is OK. Must mention courses and gender at least once. Use of ρ or "correlation" is B0 but allow ISW.	
	2nd M1 for an attempt to calculate test statistic. At least one correct expression, ft expected freq.	
	3rd A1 follow through expected frequencies for at least 3 expressions	
	3rd M1 for a correct statement relating their test statistic and their cv (may be implied by comment)	
	5th A1 for a contextualised comment relating their test statistic and their cv. Ignore their H_0 or H_1 or assume that they were correct. Must mention courses and gender	

Expected (Obs)	A	S	H
Boy	37.1 (30)	37.1 (50)	40.8 (35)
Girl	32.9 (40)	32.9 (20)	36.2 (42)

Question number	Scheme	Marks
3. (a)	(i)  (ii) 	(i) B1 (ii) B1B1 (3)
(b)(i)		M1M1

	A	*B*	*C*	*D*	*E*	*F*	*G*
Rank (Judge 1)	1	4	2	3	5	6	7
Rank (Judge 2)	1	2	4	3	5	7	6
d^2	0	4	4	0	0	1	1

	$\sum d^2 = 10$ M1A1
$r_s = 1 - \frac{6\times 10}{7\times(49-1)} = 1 - \frac{5}{28} = \frac{23}{28}$	or awrt **0.821** M1A1 (6)
(ii) $H_0 : \rho = 0$	$H_1 : \rho > 0$ (Allow ρ_S) ($H_1 : \rho \neq 0$ scores B0) B1,B1
r_s 5% one tail critical value is **0.7143**	B1
Significant result or reject null hypothesis	M1
There is evidence of a (positive) correlation between the judges <u>or</u> the judges agree	A1ft (5)
	14 marks

(a) (i) 1st B1 for 5 or more points on a straight line of positive gradient

(ii) 2nd B1 for 4 or more points satisfying -1<*r*< 0

3rd B1 for 5 or more points of decreasing ranks not on a straight line

(b)(i) 1st M1 for attempting to rank one of the judges (at least 2 correct rankings)

2nd M1 for ranking both (may be reversed) (at least 2 correct rankings)

3rd M1 for attempting d^2.

1st A1 for $\sum d^2 = 10$

4th M1 for correct use of the r_s formula

(ii) 3rd B1 for the correct critical value - depends upon their $H_1 : \rho > 0$ needs 0.7143, $\rho \neq 0$, 0.7857

The H_1 may be in words so B0B1 is possible. If no H_1 award for 0.7143 only.

5th M1 for a correct statement relating their r_s and their cv (may be implied by correct comment)

3rd A1ft follow through their r_s and their cv. Comment in context. Must mention judges.

Don't insist on "positive" and condone it if they are using $\rho \neq 0$.

Question number	Scheme		Marks
4. (a)	$X = M_1 + M_2 + M_3 + M_4 \sim \mathrm{N}(336, 22^2)$	$\mu =$ **336**	B1
		$\sigma^2 = 22^2$ or **484**	B1
	$\mathrm{P}(X < 350) = \mathrm{P}(Z < \dfrac{350-336}{22})$		M1
	$= \mathrm{P}(Z < 0.64)$	awrt **0.64**	A1
	$=$	awrt **0.738** or **0.739**	A1 (5)
(b)	$M \sim \mathrm{N}(84, 121)$ and $W \sim \mathrm{N}(62, 100)$ Let $Y = M - 1.5W$		M1
	$\mathrm{E}(Y) = 84 - 1.5 \times 62 = -9$		A1
	$\mathrm{Var}(Y) = \mathrm{Var}(M) + 1.5^2 \mathrm{Var}(W)$		M1
	$= 11^2 + 1.5^2 \times 10^2 = 346$		A1
	$\mathrm{P}(Y < 0)$, $= \mathrm{P}(Z < 0.48\ldots) =$	awrt **0.684 ~ 0.686**	M1, A1 (6)
			11 marks

(a) 2nd B1 for $\sigma = 22$ or $\sigma^2 = 22^2$ or 484
M1 for standardising with their mean and standard deviation (ignore direction of inequality)

(b) 1st M1 for attempting to find Y. Need to see $\pm(M - 1.5W)$ or equiv. May be implied by Var(Y).
1st A1 for a correct value for their E(Y) i.e. usually ± 9. Do not give M1A1 for a "lucky" ± 9.

2nd M1 for attempting Var(Y) e.g. $\ldots + 1.5^2 \times 10^2$ or $11^2 + 1.5^2 \times \ldots$

3rd M1 for attempt to calculate the correct probability. Must be attempting a probability > 0.5.

Must attempt to standardise with a relevant mean and standard deviation

Using $\sigma^2_M = 11$ or $\sigma^2_W = 10$ is not a misread.

Question number	Scheme	Marks
5. (a)	Only cleaners - no managers i.e. not all types. OR Not a random sample 1st 50 may be in same shift/group/share same views. OR Not a random sample (Allow "not a representative sample" in place of "not a random sample")	B1g B1h (2)
(b)(i)	Label employees (1-550) or obtain an ordered list Select first using random numbers (from 1 - 11) Then select every 11th person from the list	B1 B1 B1
(ii)	Label managers (1-55) and cleaners (1-495) Use random numbers to select… …5 managers and 45 cleaners	M1 M1 A1 (6)
(c)	390, 372 (They must be in this order)	B1, B1 (2) **10 marks**

After 1st B1, comments should be in **context**, i.e. mention cleaners, managers, types of worker etc

(a) 1st B1g for one row

2nd B1h for both rows. "Not a random sample" only counts once.

Score B1B0 or B1B1 or B0B0 on EPEN

(b)(i) 1st B1 for idea of labelling or getting an ordered list. No need to see 1-550.

2nd B1 selecting first member of sample using random numbers (1-11 need not be mentioned)

3rd B1 selecting every *n*th where $n = 11$.

(ii) 1st M1 for idea of two groups and labelling both groups. (Actual numbers used not required)

2nd M1 for use of random numbers within each strata. Don't give for SRS from all 550.

"Assign random numbers to managers and cleaners" scores M0M1

A1 for 5 managers and 45 cleaners. (This mark is dependent upon scoring at least one M)

Question number	Scheme	Marks
6. (a)	$p = \dfrac{0\times11+1\times21+...}{10\times(11+21+...)}$ or $10\times100, = \dfrac{223}{1000} = 0.223$ **(*)** (Accept $\dfrac{223}{1000}$)	M1, A1cso (2)
(b)	$r = (0.8)^{10}\times100 = 10.7374$ awrt **10.74**	M1A1
	$s = \binom{10}{2}(0.8)^8\times(0.2)^2\times100 = 30.198\ldots$ awrt **30.2**	A1
	t = 100 - [r + s + 26.84 + 20.13 + 8.81] = awrt **3.28**	A1cao (4)
(c)	H_0 : Binomial ([n =10], p = 0.2) is a suitable model for these data	B1
	H_1 : Binomial ([n =10], p = 0.2) is NOT a suitable model for these data	B1 (2)
(d)	Since t < 5 , the last two groups are combined	M1
	and $\nu = 4$ = 5 - 1	A1 (2)
(e)	Critical value $\chi_4^2(5\%)$ = 9.488	B1
	Not significant or do not reject null hypothesis	M1
	The binomial distribution with p = 0.2 is a suitable model for the number of cuttings that do not grow	A1 (3)
		13 marks

(a) M1 Must show clearly how to get either 223 or 1000. As printed or better.
A1cso for showing how to get <u>both</u> 223 and 1000 and reaching p = 0.223

(b) M1 for any correct method (a correct expression) seen for r or s.
1st A1 for correct value for r awrt 10.74
2nd A1 for s = awrt 30.2
3rd A1 for t = 3.28 only

(c) B1 for each. The value of p must be mentioned at least once. Accept B(10, 0.2)
If hypotheses are correct but with no value of p then score B0B1
Minimum is X~B(10, 0.2). If just B(10, 0.2) and not B(10, 0.2) award B1B0

(d) M1 for combining groups (must be stated or implied by a new table with combined cell seen)
A1 for the calculation 4 = 5 - 1

(e) M1 for a correct statement based on 4.17 and their cv(context not required) (may be implied)
Use of 4.17 as a critical value scores B0M0A0
A1 for a correct interpretation in context and p =0.2 and cuttings mentioned.

Question number	Scheme	Marks
7. (a)	$H_0 : \mu_F = \mu_M$ $\quad H_1 : \mu_F \neq \mu_M$ (Allow μ_1 and μ_2)	B1
	$z = \dfrac{6.86 - 5.48}{\sqrt{\dfrac{4.51^2}{200} + \dfrac{3.62^2}{100}}}$	M1 A1
	$= 2.860\ldots$ awrt (±)**2.86**	A1
	2 tail 5% critical value (±) 1.96 (or probability awrt 0.0021~0.0022)	B1
	Significant result or reject the null hypothesis (o.e.)	M1
	There is evidence of a difference in the (mean) amount spent on junk food by male and female teenagers	A1ft (7)
(b)	CLT enables us to assume $\bar{F}$ and $\bar{M}$ are normally distributed	B1 (1)
		8 marks

(a) 1st M1 for an attempt at $\dfrac{a-b}{\sqrt{\dfrac{c}{100 \text{ or } 200} + \dfrac{d}{100 \text{ or } 200}}}$ with 3 of *a, b, c* or *d* correct

1st A1 for a fully correct expression

2nd B1 for ± 1.96 but only if their H_1 is two-tail (it may be in words so B0B1 is OK)

If H_1 is one-tail this is automatically B0 too.

2nd M1 for a correct statement based on comparison of their *z* with their cv. May be implied

3rd A1 for a correct conclusion in context based on their *z* and 1.96.

Must mention junk food or money and male vs female.

(b) B1 for $\bar{F}$ or $\bar{M}$ mentioned. Allow "mean (amount spent on junk food) is normally distributed"

Read the whole statement e.g. " original distribution is normal so mean is…" scores B0