Paper Reference(s)

6685/01

Edexcel GCE

Statistics S3

Advanced Level

Wednesday 20 June 2007 – 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 (6685), 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. During a village show, two judges, *P* and *Q*, had to award a mark out of 30 to some flower displays. The marks they awarded to a random sample of 8 displays were as follows:

Display	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Judge *P*	25	19	21	23	28	17	16	20
Judge *Q*	20	9	21	13	17	14	11	15

(*a*) Calculate Spearman's rank correlation coefficient for the marks awarded by the two judges.

(6)

After the show, one competitor complained about the judges. She claimed that there was no positive correlation between their marks.

(*b*) Stating your hypotheses clearly, test whether or not this sample provides support for the competitor's claim. Use a 5% level of significance.

(4)

---

2. The Director of Studies at a large college believed that students' grades in Mathematics were independent of their grades in English. She examined the results of a random group of candidates who had studied both subjects and she recorded the number of candidates in each of the 6 categories shown.

	Maths grade A or B	Maths grade C or D	Maths grade E or U
English grade A or B	25	25	10
English grade C to U	15	30	15

(*a*) Stating your hypotheses clearly, test the Director's belief using a 10% level of significance. You must show each step of your working.

(9)

The Head of English suggested that the Director was losing accuracy by combining the English grades C to U in one row. He suggested that the Director should split the English grades into two rows, grades C or D and grades E or U as for Mathematics.

(*b*) State why this might lead to problems in performing the test.

(1)

---

3. The time, in minutes, it takes Robert to complete the puzzle in his morning newspaper each day is normally distributed with mean 18 and standard deviation 3. After taking a holiday, Robert records the times taken to complete a random sample of 15 puzzles and he finds that the mean time is 16.5 minutes. You may assume that the holiday has not changed the standard deviation of times taken to complete the puzzle.

Stating your hypotheses clearly test, at the 5% level of significance, whether or not there has been a reduction in the mean time Robert takes to complete the puzzle.

(7)

---

4. A quality control manager regularly samples 20 items from a production line and records the number of defective items x. The results of 100 such samples are given in Table 1 below.

x	0	1	2	3	4	5	6	7 or more
Frequency	17	31	19	14	9	7	3	0

Table 1

(*a*) Estimate the proportion of defective items from the production line.

(2)

The manager claimed that the number of defective items in a sample of 20 can be modelled by a binomial distribution. He used the answer in part (*a*) to calculate the expected frequencies given in Table 2.

x	0	1	2	3	4	5	6	7 or more
Expected frequency	12.2	27.0	r	19.0	s	3.2	0.9	0.2

Table 2

(*b*) Find the value of r and the value of s giving your answers to 1 decimal place.

(3)

(*c*) Stating your hypotheses clearly, use a 5% level of significance to test the manager's claim.

(7)

(*d*) Explain what the analysis in part (*c*) tells the manager about the occurrence of defective items from this production line.

(1)

---

5. In a trial of diet *A* a random sample of 80 participants were asked to record their weight loss, *x* kg, after their first week of using the diet. The results are summarised by

$$\sum x = 361.6 \quad \text{and} \quad \sum x^2 = 1753.95.$$

(*a*) Find unbiased estimates for the mean and variance of weight lost after the first week of using diet *A*.

(5)

The designers of diet *A* believe it can achieve a greater mean weight loss after the first week than a standard diet *B*. A random sample of 60 people used diet *B*. After the first week they had achieved a mean weight loss of 4.06 kg, with an unbiased estimate of variance of weight loss of 2.50 kg^2.

(*b*) Test, at the 5% level of significance, whether or not the mean weight loss after the first week using diet *A* is greater than that using diet *B*. State your hypotheses clearly.

(7)

(*c*) Explain the significance of the central limit theorem to the test in part (*b*).

(1)

(*d*) State an assumption you have made in carrying out the test in part (*b*).

(1)

---

6. A random sample of the daily sales (in £s) of a small company is taken and, using tables of the normal distribution, a 99% confidence interval for the mean daily sales is found to be

$$(123.5, 154.7).$$

Find a 95% confidence interval for the mean daily sales of the company.

(6)

---

7. A set of scaffolding poles come in two sizes, long and short. The length L of a long pole has the normal distribution $N(19.7, 0.5^2)$. The length S of a short pole has the normal distribution $N(4.9, 0.2^2)$. The random variables L and S are independent. A long pole and a short pole are selected at random.

(a) Find the probability that the length of the long pole is more than 4 times the length of the short pole.

(7)

Four short poles are selected at random and placed end to end in a row. The random variable T represents the length of the row.

(b) Find the distribution of T.

(3)

(c) Find $P(|L - T| < 0.1)$.

(5)

TOTAL FOR PAPER: 75 MARKS

END

June 2007
6691 Statistics S3
Mark Scheme

Question number	Scheme	Marks
1. (a)		M1A1

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*
P Rank	2	6	4	3	1	7	8	5
Q Rank	2	8	1	6	3	5	7	4
d^2	0	4	9	9	4	4	1	1

$\sum d^2 = 32$ M1A1

$r_S = 1 - \frac{6\times 32}{8\times(8^2-1)}$ M1

$= \frac{13}{21}$ or AWRT 0.619 A1 (6)

(b) $H_0 : \rho = 0$ $H_1 : \rho > 0$ (ρ_s is OK) both B1

r_S 1 tail 5% critical value is 0.6429 (Independent of their H_1) B1 (± is OK)

0.619 < 0.6429 or not significant M1

So insufficient evidence of a positive correlation between judges competitor is justified

Or A1f.t. (4)

10

(a) 1st M1 for attempting to rank both *P* and *Q*.
1st A1 for both correct (could be reversed)
2nd M1 for attempting d^2
2nd A1 for $\sum d^2 = 32$.
3rd M1 for correct use of formula for r_S

(b) M1 for a correct comparison or statement about significance (o.e.)
Follow through their r_s provided $0 < r_s < 1$

A1f.t. for a conclusion in context. Must mention judges or marks or competitor.
If they use correlation they must say it is positive.
Follow through their positive r_s with their positive c.v. and ignore hypotheses.
So r_s =0.667 they could say competitor's claim is not justified etc.

S.C. No ranking Typical answer (-3.82) can get mark for use of r_s formula and hypotheses in (b) only
(a) M0A0M0A0M1A0 (b) B1B1M0A0

Question number	Scheme	Marks
2. (a)	H_0 : Maths grades are independent of English grades <u>or</u> No association …	
	H_1 : Maths and English grades are dependent <u>or</u> There is an association …	B1
	Expected Frequencies e.g. $\frac{60\times 40}{120}=20$ 20 | 27.5 | 12.5 20 | 27.5 | 12.5	M1 A1
	$\sum\frac{(O-E)^2}{E}=2\times\left(\frac{5^2}{20}+\frac{2.5^2}{27.5}+\frac{2.5^2}{12.5}\right)$, $=3.9545\ldots$ AWRT <u>3.95</u> or <u>3.955</u>	M1, A1
	$\nu=(3-1)(2-1)=2$; $\chi_2^2(10\%)$ c.v. $=4.605$	B1; B1
	$3.95<4.605$ or not significant or do not reject H_0 (allow reject H_1)	M1
	Insufficient evidence of an association between English and maths grades <u>or</u> there is support for the Director's belief <u>or</u> Student's grades in maths and English are independent	A1 (9)
(b)	May have some expected frequencies <5 (and hence need to pool rows/cols)	B1 (1)
		10

(a) 1st B1 for both hypotheses in terms of independence or association and in context.
Must mention Maths and English in at least one of the hypotheses.
"relationship" or "correlation" or "connection" or "link" is B0

1st M1 for some correct calculation seen

1st A1 for all expected frequencies correct. Accept answers without formula seen.

2nd M1 for some evidence seen of attempt to calculate test statistic.
At least one correct term seen. Follow through their expected frequencies.

2nd A1 for AWRT 3.95. Answers only please escalate!

3rd M1 for correct comparison or statement – may be implied by correct conclusion.

3rd A1 for conclusion in context using "association" or "independence" in connection with grades.
Don't insist on seeing English or maths mentioned here.
Use ISW for comments if a false statement and correct statement are seen.

(b) B1 If they just say expected frequencies are "small" they must go onto mention need to pool.

Question number	Scheme	Marks
3.	$H_0 : \mu = 18, \quad H_1 : \mu < 18$	B1, B1
	$z = \dfrac{16.5-18}{3/\sqrt{15}} =, -1.9364...$ AWRT – 1.94	M1, A1
	5% one tail c.v. is z = (-) 1.6449 or probability (AWRT 0.026) ($\pm$) 1.6449	B1
	- 1.94 < -1.6449 or significant or reject H_0 or in critical region	M1
	There is evidence that the (mean) time to complete the puzzles has reduced	
	Or Robert is getting faster (at doing the puzzles)	A1f.t.
		7

1st & 2nd B1 must see μ and 18

1st M1 for attempting test statistic, allow $\pm$. Or attempt at critical value for $\bar{X}$: $\mu - z \times \dfrac{3}{\sqrt{15}}$

1st A1 for AWRT – 1.94. Allow use of $|z|$ = +1.94 to score M1A1. Or critical value = AWRT 16.7.

3rd B1 for AWRT 0.026 (i.e. correct probability only) or $\pm$ 1.6449. (May be seen in cv formula)

2nd M1 for correct comparison or statement relating their test statistic and 1.6449 or their probability and 0.05. Ignore their hypotheses if any or assume they were correct.

2nd A1f.t. for conclusion in context which refers to "speed" or "time". Depends only on previous M

Question number	Scheme	Marks
4. (a)	$\frac{0\times17+1\times31+\ldots}{17+31+\ldots}=\left(\frac{200}{100}=2\right)$, $\hat{p}=\frac{2}{20}=0.1$ (Accept $\frac{2}{20}$ or 2 per 20)	M1, A1 (2)
(b)	e.g. $r=100\times\binom{20}{2}(0.1)^2(0.9)^{18}$	M1
	$r=28.5$, $s=$ AWRT 9	A1, A1 (3)
(c)	x: 0 | 1 | 2 | 3 | ≥ 4 O_i: 17 | 31 | 19 | 14 | 19 E_i: 12.2 | 27.0 | 28.5 | 19.0 | 13.3 $\frac{(O-E)^2}{E}$: 1.89 | 0.59 | 3.17 | 1.32 | 2.44 Pooling	M1
	$\sum\frac{(O-E)^2}{E}=$ AWRT 9.4	M1A1c.a.o.
	$v=5-2=3$, $\chi_3^2(5\%)=7.815$	B1ft, B1ft
	H_0 : Binomial distribution is a good/suitable model/fit [Condone: B(20, 0.1) is…]	
	H_1 : Binomial distribution is not a suitable model — both	B1
	(Significant result) Binomial distribution is not a suitable model	A1cao (7)
(d)	defective items do not occur independently **or** not with constant probability	B1ft (1)
		13

(a) M1 for attempt to find mean or $\hat{p}$ (as printed or better). The 0.1 must be seen in part (a).

(b) M1 for correct expression for r or s using the binomial distribution. Follow through their $\hat{p}$.

(c) 1st M1 for some pooling (accept $x \geq 5$, obs.freq. …14, 9, 10 and exp.freq. 19.0, s, 4.3)

2nd M1 for calculation of test statistic (N.B. $x \geq 5$ gives 14.5). One correct term seen.

1st B1ft for number of classes – 2 (N.B. $x \geq 5$ will have 6 – 2 = 4)

2nd B1ft for the appropriate tables value, ft their degrees of freedom. (NB $\chi_4^2(5\%)=9.488$)

3rd B1 (for hypotheses) allow just "$X \sim$ B(20, 0.1)" for null etc.

2nd A1 for correctly rejecting Binomial model. No ft and depends on 2nd M1.

(d) B1ft for independence or constant probability – must mention defective items or defectives
Follow through their conclusion in (c). So if they do not reject they may say "defectives occur with probability 0.1". Stating the value implies constant probability.

Question number	Scheme	Marks
5. (a)	$\hat{\mu} = \bar{x} = \frac{361.6}{80}, = \underline{4.52}$	M1, A1
	$\hat{\sigma}^2 = s^2 = \frac{1753.95 - 80 \times \bar{x}^2}{79} = (1.51288...)$	M1A1ft
	AWRT $\underline{1.51}$	A1 (5)
(b)	$H_0 : \mu_A = \mu_B$ $\quad H_1 : \mu_A > \mu_B$	B1 B1
	Denominator	M1
	$z = \frac{4.52 - 4.06}{\sqrt{\frac{1.51...}{80} + \frac{2.50}{60}}} = \left(\frac{0.46}{\sqrt{0.0605...}}\right)$ $\quad z$	dM1
	$= (\pm)\ 1.8689...$ $\quad$ AWRT $(\pm)\ \underline{1.87}$	A1
	One tail c.v. is $z = 1.6449$ (AWRT 1.645 or probability AWRT 0.0307 or 0.0308)	B1
	(significant) there is evidence that diet A is better than diet B $\underline{\text{or}}$	
	evidence that (mean) weight lost in first week using diet A is more than with B	A1ft (7)
(c)	CLT enables you to assume that $\bar{A}$ and $\bar{B}$ are normally distributed	B1 (1)
(d)	Assumed ${\sigma_A}^2 = {s_A}^2$ and ${\sigma_B}^2 = {s_B}^2$ (either)	B1 (1)
		14

(a) 2nd M1 for a correct attempt at s or s^2, A1ft for correct expression for s^2, ft their mean.

N.B. $\sigma^2_n = 1.49...$ so $\frac{80}{79} \times 1.49...$ is M1A1ft

(b) 1st B1 can be given for $\mu_1 = \mu_2$, but 2nd B1 must specify which is A or B.
1st M1 for the denominator, follow through their 1.51.

Must have square root can condone 2.50^2 but $\sqrt{\frac{1.51^2}{80} + \frac{2.50^2}{60}}$ is M0.

Allow $\sqrt{\frac{1.51}{79} + \frac{2.50}{59}}$ leading to AWRT 1.85 to score M1M1A0 in (b) and can score in (d).

2nd dM1 for attempting the correct test statistic, dependent on denominator mark
1st A1 for AWRT $\pm$ 1.87, may be implied by a correct probability.
2nd A1ft ft their test statistic vs their cv **only if** H_1 is correct and both Ms are scored

(c) B1 for stating $\underline{\text{either}}$ $\bar{A}$ or $\bar{B}$ (but not A or B) are normally distributed

(d) B1 for either, can be stated in words in terms of variances or standard deviations.

Question number	Scheme	Marks
6.	$\bar{x} = \frac{1}{2}(123.5+154.7) = 139.1$	B1
	2.5758	B1
	"their 2.5758" $\frac{\sigma}{\sqrt{n}} = 154.7 - 139.1 = 15.6$	M1
	AWRT 1.96	B1
	"their 1.96" $\frac{\sigma}{\sqrt{n}} = \frac{15.6 \times 1.96}{2.5758} = (11.87...)$	M1
	So 95% C.I. = 139.1 $\pm$ 11.87…= (127.22…, 150.97…) AWRT (127, 151)	A1
		6

1st B1 for mean = 139.1 only

1st M1 for UL – mean or mean – LL set equal to z value times standard error or some equivalent expression for standard error. Follow through their 2.5758 provided a z value.

May be implied by $\frac{\sigma}{\sqrt{n}} = 6.056\ldots$ [N.B. $\frac{15.6}{2.3263} = 6.705...$]

Condone poor notation for standard error if it is being used correctly to find CI.

2nd M1 for full method for semi-width (or width) of 95% interval

Follow through their z values for both M marks

N.B. Use of 2.60 instead of 2.5758 should just lose 2nd B1 since it leads to AWRT (127, 151)

Question number	Scheme	Marks
7. (a)	Let $X = L - 4S$ then $E(X) = 19.7 - 4\times4.9$,$= 0.1$	M1, A1
	$Var(X) = Var(L) + 4^2 Var(S) = 0.5^2 + 16\times 0.2^2$	M1, M1
	$= 0.89$	A1
	$P(X > 0) = [P(Z > -0.10599\ldots)]$	M1
	= AWRT (0.542 – 0.544)	A1 (7)
(b)	$T = S_1 + S_2 + S_3 + S_4$ (May be implied by 0.16)	M1
	$T \neg N(19.6, 0.16)$ $E(T) = 19.6$	B1
	$Var(T) = 0.16$ or 0.4^2	A1 (3)
(c)	Let $Y = L - T$ $E(Y) = E(L) - E(T) = [0.1]$	M1
	$Var(Y) = Var(L) + Var(T) = [0.41]$	M1
	Require $P(-0.1 < Y < 0.1)$	M1
	$= P(Z < 0) - P(Z < -0.31..)$ or $0.5 - P(Z < -0.31..)$ or $P(Z < 0.31..) - P(Z < 0)$	M1
	= 0.1217 (tables) or 0.1226.. (calc) AWRT (0.122 – 0.123)	A1 (5)
		15

(a)

1st M1 for defining X and attempting $E(X)$

1st A1 for 0.1. Answer only will score both marks.

2nd M1 for $Var(L) + \ldots..$

3rd M1 for $\ldots. 4^2 Var(S)$. For those who don't attempt $L - 4S$ this will be their only mark in (a).

2nd A1 for 0.89

4th M1 for attempting a correct probability, correct expression and attempt to find, which should involve some standardisation: ft their $\sqrt{0.89}$ and their 0.1.

If 0.1 is used for $E(X)$ answer should be > 0.5, otherwise M0.

(c)

1st M1 for a correct method for $E(Y)$, ft their $E(T)$.

2nd M1 for a correct method for $Var(Y)$, ft their $Var(T)$. Must have +.

3rd M1 for dealing with the modulus and a correct probability statement. Must be modulus free.

May be implied by e.g. $P(Z < \frac{0.2}{\sqrt{\text{their } 0.41}}) - 0.5$, or seeing both 0.378… (or 0.622…) and 0.5

4th M1 for correct expression for the correct probability, as printed or better. E.g. 0.5 + 0.378.. is M0

A1 for AWRT in range.