Paper Reference(s)

6691/01

Edexcel GCE

Statistics S3

Advanced Subsidiary

Thursday 22 May 2014 – Morning

Time: 1 hour 30 minutes

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

Calculators may NOT be used in this examination.

This paper is strictly for students outside the UK.

Instructions to Candidates

Write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6691), your surname, initials 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.
The marks for the parts of questions are shown in round brackets, e.g. **(2)**.
There are 8 questions in this question paper. 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.

1. (*a*) Explain what you understand by a random sample from a finite population.

(1)

(*b*) Give an example of a situation when it is not possible to take a random sample.

(1)

A college lecturer specialising in shoe design wants to change the way in which she organises practical work.

She decides to gather ideas from her 75 students.

She plans to give a questionnaire to a random sample of 8 of these students.

(*c*) (i) Describe the sampling frame that she should use.

(ii) Explain in detail how she should use a table of random numbers to obtain her sample.

(3)

---

2. The weights of pears in an orchard are assumed to have unknown mean μ and unknown standard deviation σ.

A random sample of 20 pears is taken and their weights recorded.

The sample is represented by $X_1, X_2, \ldots, X_{20}$. State whether or not the following are statistics. Give reasons for your answers.

(*a*) (i) $\dfrac{X_1 + 3X_{20}}{2}$

(ii) $\sum_{i=1}^{20} (X_i - \mu)$

(iii) $\sum_{i=1}^{20} \left(\dfrac{X_i - \mu}{\sigma} \right)$

(4)

(*b*) Find the mean and variance of $\dfrac{3X_1 - X_{20}}{2}$.

(3)

---

3. A number of males and females were asked to rate their happiness under the headings "not happy", "fairly happy" and "very happy".

The results are shown in the table below

		Happiness			Total
		Not happy	Fairly happy	Very happy	
Gender	Female	9	43	34	86
	Male	13	25	16	54
Total		22	68	50	140

Stating your hypotheses, test at the 5% level of significance, whether or not there is evidence of an association between happiness and gender. Show your working clearly.

(10)

---

4. The random variable A is defined as

$$A = B + 4C - 3D$$

where B, C and D are independent random variables with

$$B \sim \mathrm{N}(6, 2^2) \qquad C \sim \mathrm{N}(7, 3^2) \qquad D \sim \mathrm{N}(4, 1.5^2)$$

Find $\mathrm{P}(A < 45)$.

(6)

---

5. A research station is doing some work on the germination of a new variety of genetically modified wheat.

They planted 120 rows containing 7 seeds in each row.

The number of seeds germinating in each row was recorded. The results are as follows

Number of seeds germinating in each row	0	1	2	3	4	5	6	7
Observed number of rows	2	6	11	19	25	32	16	9

(*a*) Write down two reasons why a binomial distribution may be a suitable model.

(2)

(*b*) Show that the probability of a randomly selected seed from this sample germinating is 0.6.

(2)

The research station used a binomial distribution with probability 0.6 of a seed germinating. The expected frequencies were calculated to 2 decimal places. The results are as follows:

Number of seeds germinating in each row	0	1	2	3	4	5	6	7
Observed number of rows	0.20	2.06	*s*	23.22	*t*	31.35	15.68	3.36

(*c*) Find the value of *s* and the value of *t*.

(2)

(*d*) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the data can be modelled by a binomial distribution.

(7)

---

6. A random sample $X_1, X_2, \ldots, X_n$ is taken from a population with mean μ.

(*a*) Show that $\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ is an unbiased estimator of the population mean μ.

(1)

A company produces small jars of coffee.

Five jars of coffee were taken at random and weighed.

The weights, in grams, were as follows

197 203 205 201 195

(*b*) Calculate unbiased estimates of the population mean and variance of the weights of the jars produced by the company.

(3)

It is known from previous results that the weights are normally distributed with standard deviation 4.8 g.

The manager is going to take a second random sample. He wishes to ensure that there is at least a 95% probability that the estimate of the population mean is within 1.25 g of its true value.

(*c*) Find the minimum sample size required.

(4)

---

7. A machine fills packets with X grams of powder where X is normally distributed with mean μ. Each packet is supposed to contain 1 kg of powder.

To comply with regulations, the weight of powder in a randomly selected packet should be such that $P(X < \mu - 30) = 0.0005$.

(*a*) Show that this requires the standard deviation to be 9.117 g to 3 decimal places.

(3)

A random sample of 10 packets is selected from the machine. The weight, in grams, of powder in each packet is as follows:

999.8 991.6 1000.3 1006.1 1008.2 997.0 993.2 1000.0 997.1 1002.1

(*b*) Assuming that the standard deviation of the population is 9.117 g, test, at the 1% significance level, whether or not the machine is delivering packets with mean weight of less than 1 kg. State your hypotheses clearly.

(7)

---

8. The heights, in metres, and weights, in kilograms, of a random sample of 9 men are shown in the table below.

Man	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Height (x)	1.68	1.74	1.75	1.76	1.78	1.82	1.84	1.88	1.98
Weight (y)	75	76	100	77	90	95	110	96	120

(a) Given that $S_{xx} = 0.0632$, $S_{yy} = 1957.5556$ and $S_{xy} = 9.3433$ calculate, to 3 decimal places, the product moment correlation coefficient between height and weight for these men.

(2)

(b) Use your value of the product moment correlation coefficient to test whether or not there is evidence of a positive correlation between the height and weight of men. Use a 5 % significance level. State your hypotheses clearly.

(4)

Peter does not know the heights or weights of the 9 men. He is given photographs of them and asked to put them in order of increasing weight. He puts them in the order

A C E B G D I F H

(c) Find, to 3 decimal places, Spearman's rank correlation coefficient between Peter's order and the actual order.

(6)

(d) Use your value of Spearman's rank correlation coefficient to test for evidence of Peter's ability to correctly order men, by their weight, from their photographs. Use a 5% significance level and state your hypotheses clearly.

(4)

TOTAL FOR PAPER: 75 MARKS

END

Question Number	Scheme	Marks
1(a)	(This is a sample where) **every** (possible) **sample** (of size *n*) has an **equal chance** of being chosen.	B1 (1)
(b)	'When it is impossible to provide a **sampling frame**' or a correct example with an indication of sampling frame being impossible.	B1 (1)
(c)(i)	A **list/register** of **all** the students.	B1
(ii)	Number the students (from 0 to 74, 1 to 75 etc.) Using the **random no. table** read off the nos. and **identify or select the students allocated those nos**.	B1 B1 (3) **Total 5**
	Notes	
(a)	Require **all / each / every** etc **sample** and **same/equal** etc **chance / probability** etc for B1	
(b)	Require **impossible / no / doesn't exist** etc and **sampling frame** for B1	
(c)(i)	Require **list/register etc** and **all/every/75 etc students** for B1 List of 8 students is B0	
(ii)	First B1 accept **'in the corresponding position' o.e.** if numbering omitted Second B1 require both for mark.	

Question Number	Scheme	Marks
2a(i)	Only contains **known** data / function of **data only / no population parameters**	B1
	therefore it **is a statistic**	B1d
(ii)(iii)	(ii) and (iii) contain **unknown** parameters / **population parameters** / μ **and / or** σ	B1
	therefore it is **not a statistic**.	B1d
		(4)
(b)	$(E(\frac{3X_1-X_{20}}{2}) = \frac{3\mu-\mu}{2} =) \ \mu$	B1
	$Var(\frac{3X_1-X_{20}}{2}) = \frac{9\sigma^2+\sigma^2}{2^2}$	M1
	$= \frac{5\sigma^2}{2}$	A1
		(3)
		Total 7

Notes		
(a)(i)	First B1 for known / no unknowns o.e. in (i)	
	Second B1 dependent on first B1 for 'Yes' / is a statistic o.e. in (i)	
	Third B1 for unknowns o.e. in **both** (ii) **and** (iii)	
	Fourth B1 dependent on third B1 for 'No' / not a statistic o.e. in **both** (ii) **and** (ii)	
(b)	B1 for μ	
	M1 for some squaring on numerator or denominator **and** must **add** on numerator	
	A1 for $\frac{5\sigma^2}{2}$ o.e.	

Question Number	Scheme	Marks
3		

		Happiness		
		Not happy	Fairly happy	Very happy
Gender	Female	13.51	41.77	30.71
	Male	8.49	26.23	19.29

M1
A1

H_0 : Happiness and gender are independent/ not associated
H_1 : Happiness and gender are not independent/ associated

B1

O	E	$\frac{(O-E)^2}{E}$	$\frac{O^2}{E}$
9	13.51	1.508	5.996
43	41.77	0.0361	44.264
34	30.71	0.351	37.637
13	8.49	2.402	19.915
25	26.23	0.0575	23.829
16	19.29	0.560	13.274

dM1
A1

$\sum \frac{(O-E)^2}{E} = 4.91$ or $\sum \frac{O^2}{E} - N = 144.91 - 140 = 4.91$

A1

$\nu = (3-2)(2-1) = 2$

B1

$\sum \frac{(O-E)^2}{E} < 5.991$

B1ft

4.91 < 5.991 so 'insufficient evidence to reject H0 ' or 'Accept H0'

M1

No association between gender and happiness.

A1

(10)

Total 10

Notes

1st M1 for some use of $\frac{Row\ Total \times Column\ Total}{Grand\ Total}$. May be implied by at least 1 correct *Ei*

1st A1 awrt 13.5, 41.8, 30.7, 8.5, 26.2 and 19.3 Allow M1A0 for *Ei* rounded to integers

1st B1 for both hypotheses. Must mention "happiness" and "gender" at least once.

Use of "relationship" or "correlation" or "connection" is B0

2nd dM1 for at least 2 correct terms (in 3rdor 4thcolumns) or correct expressions with their *Ei*

Dependent on 1st M1. Accept 2sf accuracy for the M mark.

2nd A1 for all correct terms (2sf or better). May be implied by a correct ans

Allow truncation e.g. 44.2…

3rd A1 awrt 4.91 . Condone 4.915

2nd B1 for correct degrees of freedom (may be implied by a cv of 5.991)

3rd B1ft for cv that follows from their degrees of freedom

3rd M1 for a correct statement linking their test statistic and their cv

Contradictory statements score M0 e.g. "significant, do not reject H0"

Condone "reject H1 "

4th A1 for a correct comment in context - must mention "gender" and "happiness"

Condone "relationship" or "connection" here but **not** "correlation".

e.g. "There is no evidence of a relationship between gender and happiness"

No follow through. If e.g hypotheses are the wrong way around A0 here.

SC Use of calculator with **no working** may get M0A0B1M1A0A1B1B1M1A1

Question Number	Scheme	Marks
4	$E(A) = E(B) + 4E(C) - 3E(D)$	M1
	$= 22$	A1
	$Var(A) = Var(B) + 16Var(C) + 9Var(D)$	M1
	$= 168.25$	A1
	$P(A < 45) = P\left(Z < \frac{45-22}{\sqrt{168.25}}\right)$	M1
	$= P(Z < 1.773)$	
	$= 0.9616$ awrt 0.962	A1
		(6)
		Total 6
Notes		
	1st M1 for $E(4C) = 4E(C)$ **and** $-E(3D) = -3E(D)$	
	1st A1 for 22 cao	
	2nd M1 for use of $Var(aX) = a^2 Var\ X$ and + their '9Var(D)'	
	2nd A1 for 168.25 cao	
	3rd M1 for standardising using their mean and their sd	
	3rd A1 for awrt 0.962. NB Calculator gives 0.961899….	

Question Number	Scheme	Marks
5(a)	The seeds are **independent /** There are a **fixed number** of seeds in a row / There are only **two outcomes** to the seed germinating – either it germinates or it does not / The **probability** of a seed germinating is **constant**	B1 B1 (2)
(b)	$\frac{(0\times2)+(1\times6)+(2\times11)+(3\times19)+(4\times25)+(5\times32)+(6\times16)+(7\times9)}{120\times7}=\frac{504}{840}$ = 0.6 **	M1 A1cso (2)
(c)	$p = 0.6 \quad q = 0.4$ $s = 120\times 21q^5p^2 = 120\times 21 \text{ x } 0.4^5 \text{ x } 0.6^2 = 9.29$ $t = 120\times 35q^3p^4 = 120\times 35 \text{ x } 0.4^3 \text{ x } 0.6^4 = 34.84$	B1 B1 (2)
(d)	H_0 : A binomial distribution is a suitable model. H_1: A binomial distribution is not a suitable model.	B1
		M1

Observed number of rows	19	19	25	32	25
Expected number of rows	11.55	23.22	34.84	31.35	19.04
$\frac{(O-E)^2}{E}$	4.81	0.77	2.78	0.013	1.87
$\frac{O^2}{E}$	31.26	15.55	17.94	32.66	32.83

$\nu = 5 - 2 = 3$	B1ft
Critical value for $\chi^2 = 11.345$	B1ft
$\sum \frac{(O-E)^2}{E} = 10.23$ or $\sum \frac{O^2}{E} - N = 130.23 - 120 = 10.23$	M1A1
10.23 < 11.345 therefore do not reject H_0 A binomial is a suitable model.	A1 (7)
	Total 13

Notes

(a) Any two and at least one must have context. 2 correct, no context B1B0. Do not award B0B1.

(b) M1 require at least two correct terms in numerator **and** /(120x7) or /120 then /7
A1 cso as given answer

(c) Cao for each B1

(d) 1st B1 for both hypotheses. B0 if they include 0.6 Condone X~B(n,p) etc

1st M1 for using some combined columns (<8)

2nd B1ft follows from 'their no of columns' -2

3rd B1ft follows from the degrees of freedom

2nd M1 for attempting $\frac{(O-E)^2}{E}$ **or** $\frac{O^2}{E}$ with at least 2nd (3 seeds) and 4th (5 seeds) accurate to 2sf

Contradictory statements score M0 e.g. "significant" do not reject H0

1st A1 for awrt 10.2

2nd A1 dependent on 2nd M for a correct comment suggesting that binomial model is suitable.
No follow through .
Condone mention of 0.6 here. Hypotheses wrong way round scores A0

Question Number	Scheme	Marks
6(a)	$\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$ $E(\bar{X}) = \frac{1}{n}E(X_1 + \ldots + X_n)$ $= \frac{1}{n}(E(X_1) + \ldots + E(X_n))$ $= \frac{1}{n}(\mu + \ldots + \mu)$ $= \frac{n\mu}{n} = \mu$	B1cso (1)
(b)	$\bar{x} = \frac{1}{5}(197 + 203 + 205 + 201 + 195)$ $= 200.2$(g)	B1
	$s^2 = \frac{1}{n-1}(\sum x^2 - n\bar{x}^2)$ or $\frac{n}{n-1}\text{Var}\,x$ $= \frac{1}{5-1}(200469 - 5(200.2^2))$ $= 17.2$	M1 A1 (3)
(c)	We require $2 \times 1.25 \geq$ Width of confidence interval	
	$2.5 \geq \frac{2\times1.96\times4.8}{\sqrt{n}}$ or $1.25 \geq \frac{1.96\times4.8}{\sqrt{n}}$ or $\frac{1.25}{\frac{4.8}{\sqrt{n}}} \geq 1.96$	M1B1
	$\sqrt{n} \geq \frac{2\times1.96\times4.8}{2.5} = 7.5264$ $n \geq 56.6(5)$	A1
	Minimum sample size is 57	A1 (4)
		Total 8

Notes

(a)	B1 cso: require $E(\bar{X}) = \mu$ with at least 1 correct intermediate step and no incorrect working.
(b)	B1 for 200.2 or $\frac{1001}{5}$
	M1 for use of correct formula. Accept $\frac{1}{4}S_{xx} = \frac{1}{4}\times 68.8$
	A1 for awrt 17.2
(c)	M1 for use of any equivalent expression. Accept equality. Accept their *s* instead of 4.8
	B1 for 1.96 seen with s.e.
	1st A1 for 56.6(5)
	2nd A1 for 57. Must follow from correct working e.g. $\sqrt{n} \leq 7.5264$ resulting in $n = 57$ award A0

Question Number	Scheme	Marks
7(a)	$z = \pm\ 3.2905$	B1
	$\sigma = \frac{30}{3.2905}$	M1
	$\sigma = 9.117$ **	A1cso
		(3)
(b)	$H_0: \mu = 1000$ $H_1: \mu < 1000$	B1
	mean weight = 999.54	B1
	$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{(999.54-1000)}{\frac{9.117}{\sqrt{10}}} = -0.160$ or $\frac{c-1000}{\sqrt{83.12/10}} = -2.3263 \therefore \text{CR } c < 993.29$	M1A1
	1% critical value = $-\ 2.3263$	B1
	$-\ 2.3263 < -0.160$	
	Accept H_0 / not in critical region	dM1
	There is no evidence that that the machine is delivering packets of mean weight less than 1 kg	A1ft
		(7)
		Total 10

Notes

(a)

M1 for 30/'their $|z|$ ' >1

A1 cso as given answer

(b)

1st B1 both hypotheses correct.

Accept 1kg in hypotheses if consistent units used in working usually either kg or g.

2nd B1 999.54 (g) or 0.99954 (kg)

1st M1 for standardising using their mean allow $\pm$, 1000 and $\frac{9.117}{\sqrt{10}}$ o.e. in kg

1st A1 awrt -0.160 unless clearly using $|z|$ (stated) then accept 0.160 or CR awrt 993

Condone -0.16 if fully correct expression seen.

3rd B1 $\pm$ 2.3263 sign consistent with test statistic or $p = 0.4364 > 0.01$ NB $p = 0.5636 < 0.99$

2nd dM1 dependent upon 1st M for a correct statement linking their test statistic and their cv

Contradictory statements score M0 e.g. "significant, do not reject H_0"

2nd A1 for correct conclusion in context. Must mention 'machine' and 'packets'.

Question Number	Scheme	Marks
8(a)	$r = \frac{9.3433}{\sqrt{0.0632 \times 1957.5556}}$	M1
	$= 0.840$	A1
		(2)
(b)	$H_0 : \rho = 0$ $H_1 : \rho > 0$	B1
	Critical value = 0.5822	B1
	$0.840 > 0.5822$ There is evidence to reject H_0.	M1
	There is evidence of a positive correlation between a man's height and his weight.	A1ft
		(4)
(c)		

Man	A	B	C	D	E	F	G	H	I
Actual weight	1	2	7	3	4	5	8	6	9
Peter's order	1	4	2	6	3	8	5	9	7
d^2	0	4	25	9	1	9	9	9	4

B1 B1

$\sum d^2 = 70$ M1A1

$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$ dM1

$= 1 - \frac{6 \times 70}{9(81-1)}$

$= 0.417$ A1

(6)

(d)	$H_0 : \rho = 0$ $H_1 : \rho > 0$	B1
	Critical value 0.600	B1
	$(0.417 < 0.600)$ There is insufficient evidence to reject H_0.	M1
	Peter does not have the ability to correctly order men, by weight, from their photograph.	A1
		(4)
		Total 16

Notes		
(a)	M1 Clear use of $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ A1 0.840 cao	
(b)	1st B1 for both hypotheses in terms of ρ , one tail H_1 must be compatible with their r Hypotheses just in words e.g. "no correlation" score B0 2nd B1 for 0.5822 cao M1 for a statement comparing 'their r' with 'their cv' A1 for a correct contextualised comment. Must mention positive correlation, be carrying out a 1-tailed test and mention height and weight. Follow through their r and their cv (provided their |cv| <1 and their $\|r\|$ <1)	
(c)	1st B1 for attempt to rank actual weight / Peter's order with at least 4 correct 2nd B1 for correct rankings for both (one or both may be reversed) 1st M1 for use of $\sum d^2$ with at least 4 values correct and attempt to add 1st A1 for 70 or 170 with reversed rankings 2nd dM1 for use of the correct formula, follow through their $\sum d^2$.Dependent on 1st M1 If answer is not correct, a correct expression is required. 2nd A1 for awrt 0.417 or $\frac{5}{12}$	
(d)	1st B1 for both hypotheses in terms of ρ or ρ_s One tail H_1 must be compatible with their ranking Hypotheses just in words e.g. "no correlation" score B0 2nd B1 for cv of 0.6(00) cao Their cv must be compatible with their H_1 which may be in words M1 for statement comparing 'their r' with 'their cv' A1 for a correct contextualised comment. Must mention Peter and Men. Follow through their r and their cv (provided their |cv| <1and their $\|r_s\| < 1$)	