Edexcel Maths S3

Mark Scheme Pack

2002-2015

Question Number		Scheme	Marks
1.	(*a*)	Stratified	B1 **(1)**
	(*b*)	Label De-luxe rooms 1 – 20	B1
		Using *random numbers* in range 1 – 20 select *2 rooms*	B1 B1
		Repeat for Premier using 1 – 40 and select 4 rooms Repeat for Standard using 1 – 100 and select 10 rooms	B1 **(4)** **(5 marks)**
2.	(*a*)	$H_0 : \mu_A = \mu_B$ $\quad H_1: \mu_A \neq \mu_B$	B1 B1
		standard error = $\sqrt{\frac{9.1^2}{100}+\frac{8.4^2}{120}} = 1.19$ (awrt)	M1 A1
		$\alpha = 0.01 \Rightarrow$ CR: $z < -2.5758$ or $z > 2.5758$	B1 need both
		$z = \frac{70.6-67.2}{1.19} = 2.86$ (awrt)	M1 A1
		Since 2.86 is in the critical range, H_0 is rejected. There is evidence of a difference in mean playing time.	A1ft **(8)**
	(*b*)	Central Limit Theorem applies to enable normal distribution to be used.	B1 **(1)** **(9 marks)**
3.	(*a*)	$\overline{M} \sim$ N (80, $\frac{2.6^2}{10}$) or N (80, 0.676)	B1 B1 **(2)**
	(*b*)	$P(\overline{M} < 78.5) = P\left(z < \frac{78.5-80}{2.6/\sqrt{10}}\right)$	M1
		$= P(z < -1.82)$	A1
		$= 0.0344$	A1 **(3)**
	(*c*)	Let W = weight of all 10 people	
		$W = M_1 + \ldots + M_6 + F_1 + \ldots + F_4$	
		$E(W) = (6 \times 80) + (4 \times 59) = 716$	B1
		$Var(W) = (6 \times 2.6^2) + (4 \times 1.9^2) = 55$	B1
		$P(W > 730) = P\left(z > \frac{730-716}{\sqrt{55}}\right)$	M1 A1
		$= P(z > 1.89)$	
		$= 0.0294$	A1 **(5)** **(10 marks)**

awrt = "anything which rounds to…"

Question Number	Scheme	Marks
4. (*a*)		
	A B C D E F G H I J Performance 10 5 8 3 9 6 1 4 7 2 Dedication 7 6 3 5 9 10 4 2 8 1	M1
	$\Sigma d^2 = 70$	M1 A1
	$r_s = 1 - \frac{6\times 70}{10\times 99} = 0.576$	M1 A1 **(5)**
(*b*)	$H_0: \rho = 0$; $H_1: \rho \neq 0$	B1 B1
	$n = 10 \Rightarrow$ critical value = 0.5636	B1
	0.576 is in the critical region	M1
	Evidence of correlation between performance and dedication.	A1ft **(5)**
(*c*)	Likely to be an element of judgement in grading. Dedication unlikely to be normally distributed.	B1 **(1)**
		(11 marks)
5.	Expected Frequency Male: 50.98 27.85 39.17 Female: 57.02 31.15 48.83	M1 A1 A1
	H_0: no association between gender and facility	B1
	H_1: Association between gender and facility	B1
	$\sum \frac{(O-E)^2}{E} = \frac{(50.98-40)^2}{50.98} + \frac{(57.02-68)^2}{57.02} + \ldots + \frac{(43.83-31)^2}{43.83}$	M1 A1
	$= 12.7$	A1
	$\alpha = 0.05$, $\underline{\nu = 2} \Rightarrow$ CR: $\chi^2 > \underline{5.991}$	B1 B1
	Evidence of association between gender and facility	A1ft **(11)**
		(11 marks)

ft = follow through mark

Question Number		Scheme	Marks
6.	(*a*)	$R = 43.76$; $S = 54.68$; $T = 43.76$ using tables	M1 A1; B1 B1
		(OR $R = 43.75$; $S = 54.69$; $T = 43.75$ using calculator)	**(4)**
	(*b*)	H_0: Binomial model with $n = 8$, $p = 0.5$ is suitable	
		H_1: Binomial model with $n = 8$, $p = 0.5$ is not suitable	B1 (both)
		Amalgamation of data	M1
		$\sum \frac{(O-E)^2}{E} = 5.69$ (awrt)	M1 A1
		$\alpha = 0.05$, $\underline{\nu = 6} \Rightarrow$ CR: $\chi^2 > \underline{12.592}$	<u>B1</u> <u>B1</u>
		Since 5.69 is not in the critical region there is no evidence to reject H_0. The binomial model with $n = 8$ and $p = 0.5$ is a suitable model.	A1ft **(7)**
	(*c*)	Apart from the expected values and $\sum \frac{(O-E)^2}{E}$ being different, the degrees of freedom would have been reduced by 1 ($\nu = 5$).	B1 **(1)**
			(12 marks)
7.	(*a*)	Cooling by subtracting 500 for each observation gives	
		Mean $= 500 + \frac{22}{10} = 502.2$	M1 A1
		Variance $= \frac{1}{9}\{ 288 - \frac{22^2}{10} \} = 26.622$	M1 A1 A1 **(5)**
	(*b*)	Limits are $502.2 \pm 1.6449 \times 5.0$	M1
		(493.98, 510.42) [accept (494, 510)]	A1 **(2)**
	(*c*)	95 % confidence limits are	
		$502.2 \pm 1.96 \times \frac{5.0}{\sqrt{10}}$	M1 A1ft B1 (for 1.96)
		(499, 505)	A1 A1 **(5)**
	(*d*)	H_0: $\mu = 500$	
		H_1: $\mu > 500$	B1 (both)
		$\alpha = 0.05 \Rightarrow$ CR: $z > 2.3263$	B1
		$z = \frac{503.9 - 500}{5.0/\sqrt{15}} = 1.47$	M1 A1
		1.47 is not in the critical region $\Rightarrow$ no evidence to reject H_0; no evidence to suggest mean is greater than 500g	A1 ft **(5)**
			(17 marks)

Question number	Scheme	Marks
1.		
(a)	Take a (simple) random sample from (mutually exclusive) groups of the population 1g/1h	**B1**
	Sample sizes within strata in strict proportion to numbers in each strata in the population	**B1**
	Advantage:	
	More accurate estimate of variance of population mean	
	Individual estimates for strata available Any one	**B1**
	Disadvantage:	
	Difficult if strata are large	
	Definition of strata problematic (may overlap) Any one	**B1**
		(4)
(b)	Non-random sampling	**B1**
	from groups of the population	**B1**dep
	Advantage:	
	Representative sample can be achieved with small sample size	
	Cheap (costs kept to a minimum)	
	Administration relatively easy Any one (not quick)	**B1**
	Disadvantage	
	Not possible to estimate sampling errors due to lack of randomness	
	Judgment of interviewer can affect choice of sample – bias OK	
	Non-response not recorded	
	Difficulties of defining controls e.g. social class Any one	**B1**
		(4)
		8
2.		
(a)	$X \sim N(124, 20^2)$ or $\overline{X} \sim (124, \frac{20^2}{30}$ or assume σ^2 estimated by s^2 or CLT, vals.	**B1,B1**
	$\bar{x} \pm 2.5758\frac{\sigma}{\sqrt{n}} = 124 \pm 2.5758\frac{20}{\sqrt{30}}$ 2.5758, formula + attempt, all correct&2.58,2.576	**B1M1A1**
	$= 124 \pm 9.405$	
	$= (115, 133)$ 3 sf	**A1**
		(6)
(b)	140 is not in confidence interval	**M1**
	Underweight apples chosen or Sample may not be representative/may be biased Any one	**A1∫**
		(2)
		8

3.

(a) $E(X-Y)=20-10=10$ Require minus, 10 **M1A1** **(2)**

(b) $Var(X-Y)=5+4=9$ Require plus, 9 **M1A1** **(2)**

(c) $X-Y \sim N(10,9)$ Implied **B1∫**

$P(13<X-Y<16)=P(X-Y<16)-P(X-Y<13)$ Subtract **M1**

$= P(Z<\frac{16-10}{3}) - P(Z<\frac{13-10}{3})$ Standardise **M1**

$= P(Z<2) - P(Z<1)$ 2&1 **A1**

$= 0.9772 - 0.8413 = 0.1359$ 0.1359 **A1**

(5)

9

4.

H_0 : Taking drug and catching a cold are independent (not associated) **B1**

H_1 : Taking drug and catching a cold are not independent (associated) (not ditto) both **B1**

	Cold	Not Cold	
Drug	34 (39.5)	66 (60.5)	100
Dummy	45 (39.5)	55 (60.5)	100
	79	121	200

All totals **B1**

$E = \frac{RT \times CT}{GT}$ **M1A1A1**

O	E	$\frac{(O-E)^2}{E}$
34	39.5	0.766
66	60.5	0.5
45	39.5	0.765
55	60.5	0.5

$\sum \frac{(O-E)^2}{E} = 2.53$ (NB with Yates 2.09) attempt & add, awrt 0.766 & 0.5 twice, awrt 2.53 **M1A1A1**

$\nu = 1, \chi_1^2(5\%) = 3.841 > 2.53$ 1, 3.841 **B1,B1**

No reason to believe that the chance of catching a cold is affected by taking the new drug **A1∫**

11

Question number | **Scheme** | **Marks**

5

μ_a and μ_b are mean weight of population after and before closure respectively. **B1**

$H_0: \mu_b = \mu_a$

$H_1: \mu_b > \mu_a$ **B1B1**

$$z = \frac{10-8}{\sqrt{\frac{2.64^2}{100}+\frac{1.94^2}{120}}}$$ Fraction, denom Ok alone **M1A1 / M1A1**

$$z = \frac{2}{\sqrt{0.1011}} = 6.29$$ awrt 6.29 **A1**

Critical region is $z \geq 1.6449,\ 6.29 > 1.6449$ or in critical region or Reject H_0 1.6449 **B1, M1**

(or $P(Z \geq 6.29) = 0, 0 < 0.05$ or z in critical region or Reject H_0 B1M1)

There is evidence that closing the factory has reduced mean river pollution **A1∫**

(11)

11

6 (a)

A	2	5	3	7	8	1	4	6	
B	3	2	6	5	7	4	1	8	
$\lvert d \rvert$	1	3	3	2	1	3	3	2	
d^2	1	9	9	4	1	9	9	4	46

d **M1**

$\sum d^2$ **M1A1**

$$r_s = 1 - \frac{6 \times 46}{8 \times 63}$$ **M1A1∫**

$r_s = 0.452$ 0.452 **A1**

(6)

(b)

$H_0: \rho = 0,\ H_1: \rho \neq 0 (\rho > 0)$ **B1B1**

critical values are ± 0.7381 (0.6429) 0.7381(0.6429) **B1**

0.452<0.7381 (0.452<0.6429) or not sig or Insufficient evidence to reject H_0 **M1**

No agreement between the two judges. Context **A1∫**

(5)

11

Question number	Scheme	Marks
7		
(a)	$\mu = 0.3\times 50 + 0.2\times 10 + 0.5\times 2 = 18$	**M1A1**
	$\sigma^2 = (0.3\times 50^2 + 0.2\times 10^2 + 0.5\times 2^2) - 18^2 = 448$	**M1A1**
		(4)
(b)	(50,50) or (50,50) without ordered pairs (10,2) (10,2) (2,10) (10,10) (10,10) (50,10) (50,10) (2,2) (10,50) (50,2) (2,2) (50,2) (2,50)	either, -1 each missing pair **B2**
		(2)
(c)	$\bar{x}$: 2, 6, 10, 26, 30, 50 P($\bar{X} = \bar{x}$): 0.25, 0.2, 0.04, 0.3, 0.12, 0.09	
		All means, probabs muiltiplied, -1 each error **B1 M1 A2**
		(4)
(d)	$P(2 \le \bar{X} < 7) = 025 + 0.2 = 0.45$	Probabilities of 2 and 6 added, 0.45 **M1 A1ʃ**
		(2)
(e)	$E(\bar{X}) = 2\times 0.25 + 6\times 0.2 ..... = 18$	$\sum xP(X = x)$ from table, 18 **M1 A1**
	$Var(\bar{X}) = 2^2\times 0.25 + 6^2\times 0.2 + ... - 18^2 = 224$	$\sum x^2P(X = x)$ - (theirs)^2 , 224 **M1A1**
	So $E(\bar{X}) = 18 = \mu$ and $Var(\bar{X}) = 224 = \frac{1}{2}\sigma^2$ as required.	**A1**
		(5)
		17

Question Number	Scheme	Marks
1a)	Allocate a number between 1 and N (or equiv) to each pupil.	M1
	Use random number tables, computer or calculator to select 15 different numbers between 1 and 120 (or equiv).	B1
	Pupils corresponding to these numbers become the sample.	B1 (3)
(b)	Allocate numbers 1 – 64 to girls and 1 – 56 to boys. Idea of different sets for boys and girls	M1
	Select $\frac{64}{120}\times 15 = 8$ random numbers between 1 – 64 for girls attempt find no	M1
	Select 7 random numbers between 1 – 56 for boys. Both 7 and 8	A1 (3)
2a)	$H_0: \rho = 0$; $H_1: \rho > 0$ both and ρ	B1 B1
	5% CV – PMCC 0.6215	M1
	0.572 < 0.6215 / not in critical region / not significant	A1
	No evidence of positive correlation	B1
	Spearman 0.6429	B1 (6)
(b)	Evidence of positive correlation	B1
	No evidence to suggest that as Statistics marks increased Geography marks increased. Context and not correlation	B1
	Evidence that students ranked highly in Statistics were also ranked highly in Geography ranked	(2)

Question Number	Scheme	Marks
3a)	$H_0: \mu_A = \mu_B$; $H_1: \mu_B > \mu_A$ both and μ	B1
	$z = \pm \dfrac{249 - 251}{\sqrt{\dfrac{2.5^2}{10} + \dfrac{2.3^2}{15}}}$ 249,251 accept; $\sqrt{\frac{2.5}{10}+\frac{2.3}{15}}$ for M	M1 A1
	$= \pm 2.0227\ldots$ awrt ± 2.02	A1
	$CV = \pm 1.6449$ or $P(Z \geq 2.02) = 0.0212 - 0.0217$, or $P(Z \leqslant 2.02) = 0.9788 - 0.9783$	B1
	$-2.0227 < -1.6449$ or $2.0227 > 1.6449$, or $0.0212 - 0.0217 < 0.05$ comparison and consistency needed or $0.9788 - 0.9783 > 0.95$	M1
	There is evidence that the <u>mean amount of coffee</u> dispensed by B <u>is greater</u> than A. context	A1√ (7)
b)	Machine B amounts are normally distributed.	B1 (1)

Question Number	Scheme		Marks
4a)	$\bar{x} = 75.3$		B1
	$s^2 = \frac{1}{9}\left\{57455 - \frac{753^2}{10}\right\}$		M1
	$= 83.7\dot{8}$, $83\frac{71}{90}$, 83.8	awrt 83.8	A1 (3)
b)	$74.8 \pm 1.96\sqrt{\frac{84.6}{100}}$	1.96 any z value, may use 75.3,83.8 for M	B1 M1 A1√ on z only
	(73.0, 76.6)	awrt 73.0,76.6	A1, A1 (5)
c)	Journey times independent Sample large enough to use central limit theorem Same distribution / population	any 2	B1,B1 (2)

Question Number	Scheme	Marks
5.		M1 convert % to freq
	Never	A1 (26, 91, 30, 132)
	Sometimes	
	Regularly	A1 (143, 78)
	Totals	
	Males	B1
	30	
	132	B1
	78	
	240	
	Females	
	26	
	143	
	91	M1
	260	A1 at least 3sf
	56	
	275	
	169	
	500	B1; B1√
	H_0 : No association (independent) between gender and exercise	
	H1 : association (not independent) between gender and exercise	M1 A1
	Expected Values	A1√
		(12)
	Never	
	Sometimes	
	Regularly	
	Totals	
	Males	
	26.88	
	132	
	81.12	
	240	
	Females	
	29.12	
	143	

87.88
260

56
275
169
500

$\alpha = 0.05 \quad \underline{\nu = 2}$; CV $\chi^2 > \underline{5.991}$

$\Sigma \frac{(O-E)^2}{E}$ *OR* $\Sigma \frac{O^2}{E} - N = 0.9271$ answers in range 0.90 – 0.95

Not in critical region – no evidence of association between gender and exercise

Question Number	Scheme			Marks
6a)	$X \sim B(3,1/6)$		bino 3, 1/6	B1 B1 (2)
b)	X	Prob	Expected freq	prob – must show working and use B(3,p) or may be implied by correct answer expected	M1 M1
	0	$\left(\frac{5}{6}\right)^3$	144.68		
	1	$3\times\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right)$	86.81		
	2	$3\times\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)^2$	17.36	awrt 145,86.8,17.4,1.15/1.16	B2 (-1 ee)
	3	$\left(\frac{1}{6}\right)^3$	1.15 (1.16)		
	H_0 : Binomial model is a good fit H_1 : Binomial model is not a good fit			both, no ditto	B1
	Amalgamate 3 with another group				M1
	$\alpha = 0.01$ $\nu = 2$; CR $\chi^2 > \underline{9.210}$				B1 ; B1√
	$\Sigma\frac{(O-E)^2}{E}$ OR $\Sigma\frac{O^2}{E} - N = 8.6894...$ answers in range 8.67 – 8.70 or				M1 A1
	Evidence that Binomial is a good model.				A1√ (11)

Question Number	Scheme	Marks
6.c)	Estimate p Degrees of freedom reduced by 1	B1 B1 (2)
	<u>Special case</u> Use of B(3,0.192) in part (b)	
	Expected frequencies	M1
	131.8785 94.01242 22.339 1.769	M1 B0
	H_0 : Binomial model is a good fit H_1 : Binomial model is not a good fit	both, no ditto B1
	Amalgamate 3 with another group	M1
	$\alpha = 0.01$ $\nu = 1$; CR $\chi^2 > 6.635$	B1 ; B1√
	$\Sigma\frac{(O-E)^2}{E}$ OR $\Sigma\frac{O^2}{E} - N$ in range 5.45 -5.50	M1 A1
	Evidence that Binomial is a good model.	A1√ (11)

Question Number	Scheme		Marks
7a)	$E(D) = E(A) - 3E(B) + 4E(C)$		M1
	$= 20$		A1
	$Var(D) = Var(A) + 9Var(B) + 16Var(C)$	Use of a^2Var X Adding 3 Var ie 4 +…	M1 M1
	$=341$		A1
	$P(D < 44) = P\left(z < \frac{44-20}{\sqrt{341}}\right)$	standardising their mean and sd	M1, A1√
	$= P(z < 1.30)$	awrt 1.30	A1
b)	$= \underline{0.9032}$		A1 (9)
	$E(X) = 20$		B1
	$Var(X) = Var(A) + 3Var(B) + 16\ Var(C)$	+ and 16 3 Var (B)	M1 M1
	$= 287$		A1
	$P(X > 0) = P\left(z > \frac{-20}{\sqrt{287}}\right)$	standardising their mean and sd	M1
	$= P(z > -1.18)$	awrt -1.18	A1
	$= 0.8810$		A1 (7)

publication

edexcel

June 2005
6685 Statistics S3
Mark Scheme

FINAL

Question Number	Scheme	Marks
1.	(a) Population divided into mutually exclusive (distinct); groups (strata)	B1; B1 (2)
	(b) Advantages – enables fieldwork to be done quickly – costs kept to a minimum – administration is relatively easy Any ONE	B1
	Disadvantages – non-random so not possible to estimate sampling error – subject to possible interviewer bias – non-response not recorded Any ONE	B1 (2)
2.	$X \sim N(10, 3^2) \therefore \bar{X} \sim N(10, 9/5)$ Can be implied: 10; 9/5	B1; B1
	$P(7 \leq \bar{X} \leq 10) = P\left(\frac{7-10}{\sqrt{9/5}} < Z < 0\right)$ Standardising with 10 & their σ	M1 A1
	$= P(-2.236 < Z < 0)$	
	$= \Phi(0) - \{1 - \Phi(2.24)\}$	M1 (p<0.5)
	$= 0.4875$	A1 (6)

Question Number	Scheme	Marks

3.

	No action	Remove diseased branches	Spray with Chemicals	Totals
Tree died within 1 year	10 (7)	5 (7)	6 (7)	21
Survived 1-4 years	5 (7)	9 (7)	7 (7)	21
Survived > 4 years	5 (6)	6 (6)	7 (6)	18
Totals	20	20	20	60

$\frac{RT \times CT}{GT}$ M1

6×7 A1

7×6 A1

H_0: Treatment & Survival are independent (not associated) B1 both

H_1: Treatment & Survival are not independent (associated)

$\alpha = 0.05$

$\nu = (3-1)\times(3-1) = 4$ B1

CR: $\chi^2 > 9.488$ B1✓

$\sum \frac{(O-E)^2}{E} = \frac{9}{7} + \frac{4}{7} + \frac{1}{7} + \frac{4}{7} + \frac{4}{7} + 0 + \frac{1}{6} + 0 + \frac{1}{6}$ Using $\sum \frac{(O-E)^2}{E}$ M1

Any 2 values A1

$= 3.47619....$ AWRT 3.48 A1

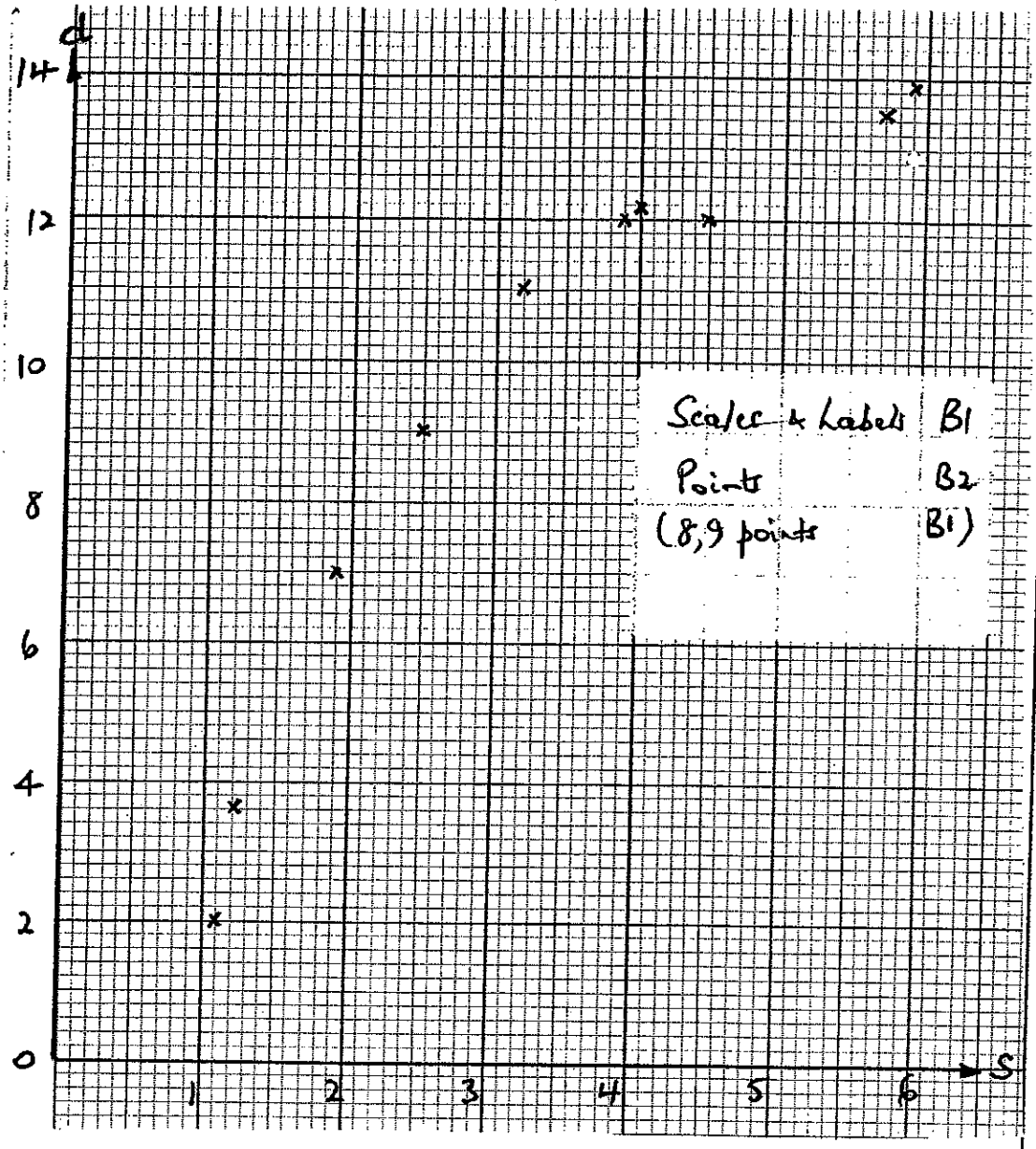
Since 3.47619... is NOT in the critical region (ie < 9.488) there is insufficient evidence to reject H_0.

There is no evidence of association between treatment and length of survival.

Comparison M1

Conclusion A1✓ (11)

4 (a) NB No graph paper ⇒ 0/3



(3)

(b) Linear association between s and d — B1 (1)

(c) $S_{ss} = 141.51 - \frac{33.9^2}{10} = 26.589$; $S_{dd} = 152.444$; $S_{sd} = 59.524$ — B1; B1; B1 (3)

(d) $r = \frac{59.524}{\sqrt{152.444 \times 26.589}}$ — M1

$= 0.93494\ldots$ AWRT 0.935 — A1 (2)

(e) $H_0: \rho = 0$; $H_1: \rho > 0$ — B1

Critical Value at 1% = 0.7155 — B1

Reject H_0; Levels of serum A disease are positively correlated — B1 (3)

(f) Linear correlation significant but scatter diagram looks non-linear. — B1 (1)

5. H_0: Poisson distribution is a suitable model — both B1

H_1: Poisson distribution is not a suitable model

$$\hat{\lambda} = \frac{(0\times 99)+(1\times 65)+\cdots+(4\times 2)}{200} = \frac{153}{200} = 0.765$$ — M1 A1

Using $P(X=x) = \frac{0.765^{x} e^{-0.765}}{x!}$ where X represents the number of restarts gives — $200\times P(X=x)$ M1

X	Observed Frequency	Expected Frequency		
0	99	93.06678....		A1, A1 (-1 e.e.)
1	65	71.19604....	0,1,2	
2	22	27.23250....		
3	12 (14)	6.94428....	8.50468	A1
≥4	2 (14)	1.56040....		

$\nu = 4-1-1 = 2$; CR: $\chi^2_2 > 5.991$ from Poisson — B1; B1√

$\nu = 4-1 = 3$ CR: $\chi^2 > 7.815$ from Poisson (0.765) — M1 Use of $\sum (O-E)^2/E$

$\sum \frac{(O-E)^2}{E} = 5.47368....$ — A1 5.40 – 5.50

5.47 is not in the critical region. — A1√

Number of computer failures per day can be modelled by a Poisson distribution — A1√ (12)

6. (a) Let X represent repair time

$\therefore \sum x = 1435 \quad \therefore \bar{x} = \frac{1435}{5} = 287$ — B1

$\sum x^2 = 442575 \quad \therefore s^2 = \frac{1}{4}\left\{442575 - \frac{1435^2}{5}\right\}$ — M1 A1

$= 7682.5$ — A1 (4)

(b) $P(|\mu - \bar{x}| < 20) = 0.95$ — Use of 1.96, 20 or 400 with their σ & $\sqrt{n}$ — M1

1.96 — B1

$\therefore \frac{20}{\sigma/\sqrt{n}} = 1.96$ — A1

Solving for n — M1

$\therefore n = \frac{1.96^2\sigma^2}{20^2} = \frac{1.96^2 \times 100^2}{400} = 96.04$ — A1

$\therefore$ Sample size ($\geq$)97 required — A1 (6)

7. Let $W = C_1 - C_2$ NB $W = C_1 + C_2 \Rightarrow$ M1 A0 M1 only

(a) $\therefore W \sim N(0, 16)$ — Normal; 0; 16 — M1; A1; A1

$\therefore P(|W| > 6) = 2P(W > 6)$ — M1

NB $W = C - L$ treat as MR, Prob = 0.4346

$= 2 \times P\left(Z > \frac{6-0}{\sqrt{16}}\right)$ — Standardising, their σ — M1

$= 2 \times P(Z > 1.5)$

$= 2 \times (1 - 0.9332) = 0.1336$ — A1 (6)

(b) Let $W = C - L$

$\therefore W \sim N(5, 25)$ — 5; 25 — B1; B1

$P(W > 0) = P\left(Z > \frac{\pm 5}{\sqrt{25}}\right)$ — M1 A1

$= P(Z < 1)$ — M1 ($p > 0.5$)

$= 0.8413$ — A1 (6)

(c) Let $W = C_1 + \cdots + C_{24} + B$

$\therefore E(W) = 24 \times 350 + 100 = \underline{8500}$ B1

$Var(W) = 24 \times 8 + 2^2 = \underline{196}$ B1

$P(8510 \leq W \leq 8520) = P\left(\frac{8510-8500}{\sqrt{196}} \leq Z \leq \frac{8520-8500}{\sqrt{196}}\right)$ M1

$= P(0.71\ldots \leq Z \leq 1.43\ldots)$ AWRT A1√ A1√

$= 0.9236 - 0.7611$

$= \underline{0.1625}$ 0.161 – 0.163 A1 (6)

(d) All random variables are independent. B1 (1)

YE Sharpness
18/06/05

Question Number	Scheme	Marks
1.	Total in School = $(15 \times 30) + 150 = 600$	B1
	random sample of $\frac{30}{600} \times 40$ (Use of $\frac{40}{their 600}$)	M1
	= 2 from each of the 15 classes	A1
	random sample of $\frac{150}{600} \times 40$ Either	
	= 10 from sixth form;	A1
	Label the boys in each class from 1 – 15 and the girls from 1 – 15.	B1
	use random numbers to select 1 girl and 1 boy	B1
	Label the boys in the sixth form from 1 – 75 and the girls from 1 – 75. use random numbers to select 5 different boys and 5 different girls.	B1
		(7)

Question Number	Scheme	Marks
2. (a)	E(R) = 20 + 10 = 30	B1 (1)
(b)	Var(R) = 4 + 0.84, = 4.84	M1, A1 (2)
(c)	R ~ N(30, 4.84) (Use of normal with their (a),(b))	B1ft
	P(28.9 < R < 32.64) = P(R < 32.64) - P(R < 28.9)	
	$= P\left(Z < \frac{32.64-30}{2.2}\right) - P\left(Z < \frac{28.9-30}{2.2}\right)$ Stand their σ and μ	M1
	= P(Z < 1.2) - P(Z < - 0.5)	A1, A1
	= 0.8849 - (1 - 0.6915) Correct area	M1
	= 0.8849 - 0. 3085 = 0.5764 (accept AWRT 0.576)	A1 (6)
		9

3. (a)	$\hat{\mu} = \frac{82+98+140+110+90+125+150+130+70+110}{10}$ = 110.5	M1 A1
	$\hat{\sigma}^2 = \frac{1}{9}\left(128153 - 10\times 110.5^2\right)$ 128153 = 672.28 (AWRT 672)	B1 M1 A1 (5)
(b)	95% confidence limits are (condone use of 5 instead of 25) $110.5 \quad \pm 1.96 \times \frac{25}{\sqrt{10}}$ (for 1.96) 95% conf. lim. = AWRT(95, 126)	M1 B1 A1√ A1 A1 (5)
(c)	Number of intervals = $\frac{95}{100}\times 15$ = 14.25 (Allow 14 or 14.3 if method is clear)	M1 A1 (2)
		12

4.

H_0 : No association between gender and acceptance
H_1 : gender and acceptance are associated

B1

	Accept	Not accept	Total
Males	170 (180)	110 (100)	280
Females	280 (270)	140 (150)	420
Totals	450	250	700

Expected Values

M1 A1

O	E	$\frac{(O-E)^2}{E}$
170	180	0.5556
110	100	1.0000
280	270	0.3704
140	150	0.6667

$\sum \frac{(O-E)^2}{E}$ = 2.59 (Yates' 2.34) (Condone use of Yates')

M1 A1

ν = 1; (5%) = 3.841

B1; B1

3.841 > 2.59. There is insufficient evidence to reject Ho
There is no association between a persons gender and their acceptance (of the offer of a flu jab.)

M1
A1√

(9)

9

5. (a)	μ_b = mean mark of boys, μ_g = mean mark of girls.	
	$H_0 : \mu_b = \mu_g$ both $H_1 : \mu_b \neq \mu_g$	B1
	$z = \dfrac{53-50}{\sqrt{\dfrac{144}{80}+\dfrac{144}{80}}}$	M1 A1
	$= 1.58$	A1
	Critical region $z \geq 1.96$	B1
	$1.58 < 1.96$ insufficient evidence to reject Ho.	M1
	No diff. between mean scores of boys and girls.	A1 (7)
(b)		
	$H_0 : \mu_b = \mu_g$ $H_1 : \mu_b < \mu_g$	B1
	$z = \dfrac{62-59}{\sqrt{\dfrac{36}{80}+\dfrac{36}{80}}}$	M1
	$= 3.16$	A1
	Critical region $z \geq 1.6449$ (accept 1.645)	B1
	$3.16 > 1.6449$ sufficient evidence to reject H_0.	
	the mean mark for boys is less than the mean mark of the girls.	A1 (5)
(c)	Girls have improved more than boys or girls performed better than boys after 1 year	B1 (1)
		13

6. (a)	r = 27.07, s = 18.04, t = 0.11 using tables or 0.12 using totals	M1 A1 B1 B1 ft (4)
(b)	Ho : A Poisson model Po(2) is a suitable model. both H_1 : A Poisson model Po(2) is not a suitable model. Amalgamate data $\sum \frac{(O-E)^2}{E}$ = 3.28 (awrt) ν = 6 –1 =5 $\chi^2_5(5\%)$ = 11.070 (follow through their degrees of freedom) 3.25 < 11.070 There is insufficient evidence to reject H_o, <u>Po(2) is a suitable model.</u>	B1 M1 M1 A1 B1 B1ft A1ft (7)
(c)	The expected values, and hence $\sum \frac{(O-E)^2}{E}$ would be different, and the degrees of freedom would be 1 less.	B1 B1 (2)
		13

7. (a)	The variables cannot be assumed to be normally distributed	B1	(1)
(b)			

	20-29	30-39	40-49	50-59	60-69	70+
Rank x	5	6	4	3	1	2
Rank y	6	5	4	1	3	2
d	1	1	0	2	2	0
d^2	1	1	0	4	4	0

M1 A1

dM1 (depends on ranking attempt)

$\sum d^2 = 10$ (follow through their rankings) A1 ft

$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{60}{210} = 0.714$ $\left(\frac{5}{7} \text{ or awrt } 0.714\right)$ M1 A1 (6)

(c)

$H_o : \rho = 0$ B1

$H_1 : \rho \neq 0$ (or $\rho > 0$) B1

$n = 6 \Rightarrow$ 5% critical value = 0.8857 (or 0.8286) B1✓

0.714 < 0.8857 M1

No evidence to reject H_o; A1

No evidence of correlation between deaths from pneumoconiosis and lung cancer. (5)

12

June 2006
6691 Statistics S3
Mark Scheme

Question Number	Scheme	Marks
1 (a)	**Advantages:** - does not require the existance of a {sampling frame / population list} - field work can be done quickly as representative sample can be achieved with a small sample size - costs kept to a minimum (cheaply) - administration relatively easy - non-response not an issue any one	B1
	Disadvantages: - not possible to estimate sampling errors - interviewer choice and may not be able to judge easily / may lead to bias - non-response not recorded - non-random process any one	B1 (2)
(b)	**Advantages:** - random process so possible to estimate sampling errors - free from bias any one	B1
	Disadvantages: - not suitable when sample size is large - sampling frame required which may not exist or may be difficult to construct for a large population. any one NO REPETITION / OPPOSITES	B1 (2) TOTAL 4

Question Number	Scheme	Marks
2 (a)	$\bar{X} \sim N\left(90, \frac{5^2}{100}\right)$ i.e. $N_g(90, 0.25)$	M1A1 (g)
	Application of central limit theorem as (sample large)	B1
		(3)
(b)	$P(\bar{X} \geqslant 91) = 1 - P\left(Z < \frac{91-90}{0.5}\right)$ Stand.	M1A1
	$= 1 - P(Z < 2)$	
	$= 1 - 0.9772$	
	$= 0.0228$ awrt 0.0228	A1
		(3)
		TOTAL 6
3 (a)	$H_0: \mu_A = \mu_B$, $H_1: \mu_A \neq \mu_B$ $\quad M_1, M_2$ OK both	B1
	$se = \sqrt{\frac{47^2}{70} + \frac{23^2}{90}} \left(= \sqrt{37.43492..}\right)$	M1A1
	Test statistic is $\pm \frac{198-201}{se} = \pm 0.4903$ awrt ± 0.49	M1A1
	M1A1 Probab awrt 0.312	
	B1 Probab cv 0.025	
	cv $= (\pm) 1.96$	B1
	Insufficient evidence to reject H_0, no significant difference between the mean cholesterol content of the two samples. (require correct comparison for ft) context required.	A1√
		(7)
(b)	– require 1 egg from each of 70 chickens of diet A to ensure independence, similarly for diet B.	
	– no chickens in common between the two samples to ensure independence	
	– not same chickens on diet A and diet B because if it were we need to do a paired analysis.	B1, B1
	Any 1	
		(2)
	~~[illegible]~~	
		TOTAL 9

4.

Rank:

Shop	Distance	Price	d	d^2
A	1	9	8	64
B	2	7	5	25
C	3	10	7	49
D	4	6	2	4
E	5	4	1	1
F	6	8	2	4
G	7	2	5	25
H	8	1	7	49
I	9	5	4	16
J	10	3	7	49
				286 Σd^2

ranking M1

Reverse ranking or price, $\Sigma d^2 = 44$ (Hairs) M1, A1

(a) $r_s = 1 - \frac{6 \times 286}{10(100-1)} = -0.73$ or $\frac{-11}{15}$ or $-0.\overset{\text{aurt}}{733}$ M1 A1

or $0.\overset{\text{aurt}}{733}$ for $\Sigma d^2 = 44$ (5)

(b) $H_0 : \rho = 0$ B1

$H_1 : \rho < 0$ ($H_1 : \rho > 0$ if reverse ranking) B1

cv $= -0.5636$ (0.5636) B1

Reject H_0, evidence that is a significant negative correlation between the price of an ice cream and the distance from a tourist attraction. B1

(Ice cream gets cheaper further from the tourist attraction)

(−cv from correct table required) (positive in context) (4)

TOTAL 9

5. M = wt of male worker $\quad M \sim N(78.5, 12.6^2)$

F = wt of female worker $\quad F \sim N(62.0, 9.8^2)$

(a) $W = M_1 + \cdots + M_7 + F_1 + \cdots + F_8$

$E(W) = 7 \times 78.5 + 8 \times 62.0 = 1045.50$ awrt 1050 M1 A1

$\text{Var}(W) = 7 \times 12.6^2 + 8 \times 9.8^2 = 1879.64$ 1880 M1 A1 (4)

(b) Independent (used in Variance formula) B1 (1)

(c) $P(W > 1090) = P\left(z > \dfrac{1090 - 1045.5}{\sqrt{1879.64}}\right)$ M1

$= P(z > 1.03)$ awrt 1.03 A1

$= 1 - 0.8485$ 1 − M1

$= 0.1515$ A1 (4)

Awrt (0.152)

(9)

6.

H_0 : No association between age and colour (independent) — B1

H_1 : Association between age and colour (Not independent) — B1

O	E	$\frac{(O-E)^2}{E}$
12	10.08	0.3657...
6	7.92	0.4654...
10	9.52	0.0242...
7	7.48	0.0308...
6	8.4	0.6857...
9	6.6	0.8727...

at least one $\frac{RTXCT}{GT}$, All $\frac{(O-E)^2}{E}$ 3sf or better — M1 A1 M1 A1

$\sum \frac{(O-E)^2}{E} = 2.4446...$ awrt Σ, 2.44 — M1 A1

$\nu = (3-1)(2-1) = 2$, $\chi^2_2 = 5.991$ — B1 B1√

Insufficient evidence to reject H_0.

No association between age and colour — A1√ (11)

(cv for correct hyp for ft) — TOTAL 11

7.(a)

$\bar{x} = \frac{500}{10} = 50$ — M1 A1

$s^2 = \frac{1}{9}\left(25001.74_{\text{Hairs}} - \frac{500^2}{10}\right) = 0.193$ awrt 0.193 — M1 A1 A1 (5)

(b)

limits are 50 ± 1.966 B1 1.96 — M1 B1

$= (49.02, 50.98)$ awrt 49(.0), 51(.0) — A1 A1 (4)

(c)

Confidence interval is $\left(50_{\text{Hairs}} - 2.5758_{\text{Hairs}} \times \frac{0.5}{\sqrt{10}}, 50 + 2.5758 \times \frac{0.5}{\sqrt{10}}\right)$ B1 2.5758 — M1 B1 A1√

$= (49.59273, 50.40727...)$ awrt 49.6, awrt 50.4 — A1 A1 (5)

Use of estimate in (a) in (b) AND (c) assume MISREAD. — TOTAL 14

8 (a) $B_3(5, 0.5)$ M1A1 (2)

(b) H_0 : B(5, 0.5) is a suitable model (good fit)
H_1 : B(5, 0.5) is not a suitable model (not a good fit) B1 ✓
✓ for $\hat{p} = 0.466$.

No. of heads	0	1	2	3	4	5
Expected	3.125	15.625	31.25	31.25	15.625	3.125
Actual	6	18	29	34	10	3

100P(1=c) for Bin, 1 correct = A1 All correct = A1 3sf or better. M1A1A1

	O	E	$\frac{(O-E)^2}{E}$
0 or 1	24	18.75	1.47
2	29	31.25	0.162
3	34	31.25	0.242
4 or 5	13	18.75	1.763

grouped O and E. All correct 2sf or better. M1A1

$\sum \frac{(O-E)^2}{E} = 3.6373$ Σ required, awrt 3.64 M1A1

$\nu = 4 - 1 = 3$, $\chi^2_3(0.10) = 6.251$ B1 ✓ B1 ✓

Insufficient evidence to reject H_0

B(5, 0.5) is a suitable model.

No evidence that coins are biased A1 ✓

(11)

TOTAL 13

Ungrouped gives awrt 5.44, $\nu = 5$, $\chi^2_5 = 9.236$

edexcel
advancing learning, changing lives

Mark Scheme (Results)
Summer 2007

GCE

GCE Mathematics

Statistics S3 (6691)

Edexcel Limited. Registered in England and Wales No. 4496750
Registered Office: One90 High Holborn, London WC1V 7BH

A PEARSON COMPANY

edexcel

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… 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
	$z = \frac{16.5-18}{3/\sqrt{15}} =, -1.9364...$ AWRT – 1.94	M1, A1
	5% one tail c.v. is z = (-) 1.6449 or probability (AWRT 0.026) (±) 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 ˜ and 18

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

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 ± 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)	$\frac{0\times17+1\times31+...}{17+31+...}=\left(\frac{200}{100}=2\right)$, $\hat{p}=\frac{2}{20}=0.1$ (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 ~ 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
	$= (\underline{+})\ 1.8689\ldots$ AWRT $(\underline{+})\ \underline{1.87}$	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 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 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.	$\bar{x} = \frac{1}{2}(123.5+154.7) = \underline{139.1}$	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 $\underline{(127, 151)}$	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… [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
	$Var(X) = Var(L) + 4^2\,Var(S) = 0.5^2 + 16\times0.2^2$	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.

Mark Scheme (Results)
June 2008

GCE

GCE Mathematics (669101)

Edexcel Limited. Registered in England and Wales No. 4496750
Registered Office: One90 High Holborn, London WC1V 7BH

A PEARSON COMPANY

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…, 170.133…)	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 $\bar{x} \pm z \times \frac{5.1}{\sqrt{36}}$

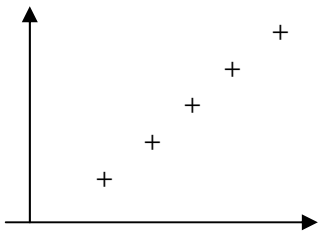
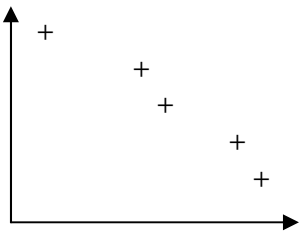
Beware: $167.94 \pm 2.5758 \times \frac{5.1^2}{36} \rightarrow (166.07..,169.8..)$ but scoresB1M0A0A0A0

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

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)	$X = M_1 + M_2 + M_3 + M_4 \sim \text{N}(336, 22^2)$	$\mu =$ **336**	B1	
		$\sigma^2 = 22^2$ or **484**	B1	
	$\text{P}(X < 350) = \text{P}(Z < \dfrac{350-336}{22})$		M1	
	$= \text{P}(Z < 0.64)$	awrt **0.64**	A1	
	$=$	awrt **0.738** or **0.739**	A1	(5)
(b)	$M \sim \text{N}(84, 121)$ and $W \sim \text{N}(62, 100)$ Let $Y = M - 1.5W$		M1	
	$\text{E}(Y) = 84 - 1.5 \times 62 = -9$		A1	
	$\text{Var}(Y) = \text{Var}(M) + 1.5^2 \text{Var}(W)$		M1	
	$= 11^2 + 1.5^2 \times 10^2 = 346$		A1	
	$\text{P}(Y < 0)$, $= \text{P}(Z < 0.48\ldots) =$	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 $\pm$ 9. Do not give M1A1 for a "lucky" $\pm$ 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 <u>types</u>. OR Not a random sample 1st 50 may be in same shift/group/share <u>same views</u>. 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 <u>first</u> using <u>random numbers</u> (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 <u>two</u> groups and labelling <u>both</u> 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 <u>and</u> 45 cleaners. (This mark is dependent upon scoring at least one M)

Question number	Scheme	Marks
6. (a)	$p = \frac{0\times11+1\times21+...}{10\times(11+21+...)}$ or $10\times100, = \frac{223}{1000} = 0.223$ **(*)** (Accept $\frac{223}{1000}$)	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

edexcel
advancing learning, changing lives

Mark Scheme (Results)
Summer 2009

GCE

GCE Mathematics (6691/01)

A PEARSON COMPANY

edexcel

June 2009
6691 Statistics S3
Mark Scheme

Question Number	Scheme	Marks
Q1 (a)	Randomly select a number between 00 and 499 (001 and 500) select every 500th person	B1 B1 (2)
(bi)	Quota **Advantage:** Representative sample can be achieved (with small sample size) Cheap (costs kept to a minimum) **not** "quick" Administration relatively easy **Disadvantage** Not possible to estimate sampling errors (due to lack of randomness) Not a random process Judgment of interviewer can affect choice of sample – bias Non-response not recorded Difficulties of defining controls e.g. social class	B1 B1
(bii)	Systematic **Advantage:** Simple or easy to use **not** "quick" or "cheap" or "efficient" It is suitable for large samples (not populations) **Disadvantage** Only random if the ordered list is (truly) random Requires a list of the population or must assign a number to each member of the pop.	(2) B1 B1 (2) **[6]**
(a)	1st B1 for idea of using random numbers to select the first from1 - 500 (o.e.) 2nd B1 for selecting every 500th (name on the list) If they are clearly trying to carry out stratified sample then score B0B0	
(b)	Score B1 for any one line	
(i)	1st B1 for Quota advantage 2nd B1 for Quota disadvantage	
(ii)	3rd B1 for Systematic Advantage 4th B1 for Systematic Disadvantage	

Question Number	Scheme	Marks
Q2 (a)	Limits are $20.1 \pm 1.96 \times 0.5$	M1 B1
	(19.1, 21.1)	A1cso (3)
(b)	98 % confidence limits are	
	$20.1 \pm 2.3263 \times \frac{0.5}{\sqrt{10}}$	M1 B1
	(19.7, 20.5)	A1A1 (4)
(c)	The growers claim is not correct Since 19.5 does not lie in the interval (19.7, 20.5)	B1 dB1 (2) **[9]**
(a)	M1 for $20.1 \pm z \times 0.5$. Need 20.1 and 0.5 in correct places with no $\sqrt{10}$ B1 for z = 1.96 (or better) A1 for awrt 19.1 and awrt 21.1 **but must have scored both M1 and B1** [Correct answer only scores 3/3]	
(b)	M1 for $20.1 \pm z \times \frac{0.5}{\sqrt{10}}$, need to see 20.1, 0.5 and $\sqrt{10}$ in correct places B1 for z = 2.3263 (or better) 1st A1 for awrt 19.7 2nd A1 for awrt 20.5 [Correct answer only scores M1B0A1A1]	
(c)	1st B1 for rejection of the claim. Accept "unlikely" or "not correct" 2nd dB1 Dependent on scoring 1st B1 in this part for rejecting grower's claim for an argument that supports this. Allow comment on their 98% CI from (b)	

Question Number	Scheme	Marks

Q3 (a)

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
BMI	1	6	3	8	4	5	7	2	9	10
or	10	5	8	3	7	6	4	9	2	1
Finishing position	3	5	1	9	6	4	10	2	7	8
d^2	4	1	4	1	4	1	9	0	4	4

M1

$\sum d^2 = 32$ (298) M1

$r_s = 1 - \frac{6\times 32}{10\times 99}$ M1 A1ft

$= 0.80606\ldots$ (-0.80606) accept $\pm\frac{133}{165}$ **awrt $\pm$ 0.806** A1 (5)

(b)

$H_0 : \rho = 0$, $H_1 : \rho > 0$, B1 B1

Critical value is $(\pm)0.5636$ B1

(0.806 > 0.5636 therefore) in critical region/ reject H_0 M1 A1ft

The lower the BMI the higher the position in the race./ support for doctors belief (5)

(c)

The position is already ranked OR Position is not Normally distributed B1 (1)

[11]

(a)

1st M1 for attempt to rank BMI scores

2nd M1 for attempt at $\sum d^2$ (must be using ranks)

3rd M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required.

1st A1ft for a correct expression. ft their $\sum d^2$ but only if all 3 Ms are scored

2nd A1 awrt $\pm$ 0.806 (but sign must be compatible with their $\sum d^2$)

No ranking can score 3rd M1 only

(b)

2nd B1 for $\rho > 0$ (or <0 but must be one tail and consistent with their ranking)

3rd B1 for critical value that is compatible with their H_1. If one-tail must be $\pm$ 0.5636 if two-tail must be $\pm$ 0.6485 [Condone wrong sign]

M1 for a correct statement relating their r_s with their cv.
e.g. "reject H_0", "in critical region", "significant result"
May be implied by a correct comment

A1ft for correct comment in context. Must mention low/high BMI and race/fitness or doctor's belief. Comment should be one-tailed.
Allow positive correlation between… but NOT …positive relationship…

No H_1 assume one-tail for 3rd B1

(c)

B1 for a correct and relevant comment either based on the fact that the data was originally partially ordered or on the underlying normal assumption
"Quicker" or "easier" score B0

Question Number	Scheme	Marks
Q4	$X \sim \text{N}(55, 3^2)$ therefore $\overline{X} \sim \text{N}(55, \frac{9}{8})$	B1 B1
	$\text{P}(\overline{X} > 57) = \text{P}\left(Z > \frac{57-55}{\sqrt{\frac{9}{8}}}\right) = \text{P}(Z > 1.8856\ldots)$	M1
	$= 1 - 0.9706$	M1
	$= 0.0294$ **<u>0.0294~0.0297</u>**	A1 **[5]**
ALT	1^{st} B1 for $\overline{X}$ ~ normal and $\mu = 55$, may be implied but must be $\overline{X}$ 2^{nd} B1 for Var($\overline{X}$) or st. dev of $\overline{X}$ e.g. $\overline{X} \sim \text{N}(55, \frac{9}{8})$ or $\overline{X} \sim \text{N}\left(55, \left(\frac{3}{\sqrt{8}}\right)^2\right)$ for B1B1 Condone use of X if they clearly mean $\overline{X}$ so $X \sim \text{N}\left(55, \frac{9}{8}\right)$ is OK for B1B1 1^{st} M1 for an attempt to standardize with 57 and mean of 55 and their st. dev. $\neq 3$ 2^{nd} M1 for 1 - tables value. Must be trying to find a probability < 0.5 A1 for answers in the range 0.0294~0.0297 $\sum_1^8 X_i \sim \text{N}\left(8\times 55, 8\times 3^2\right)$ 1^{st} B1 for $\sum X$ ~ normal and mean = 8×55 2^{nd} B1 for variance = 8×3^2 1^{st} M1 for attempt to standardise with 57×8, mean of 55×8 and their st dev $\neq 3$	

Question Number	Scheme	Marks
Q5 (a)	$\lambda = \frac{0\times 40+1\times 33+2\times 14+3\times 8+4\times 5}{100} = 1.05$	M1 A1 (2)
(b)	Using Expected frequency = $100 \times P(X = x) = 100 \times \frac{e^{-1.05}1.05^x}{x!}$ gives	M1
	$r = 36.743$ awrt 36.743 or 36.744	A1
	$s = 19.290$ 19.29 or awrt 19.290	A1 (3)
(c)	H_0 : Poisson distribution is a suitable model H_1 : Poisson distribution is not a suitable model	B1
	Number of goals / Frequency / Expected frequency 0 / 40 / 34.994 1 / 33 / 36.743 2 / 14 / 19.290 3 / 8 / 6.752 ≥ 4 / 5 / 2.221 (8.972443)	M1
	$\upsilon = 4 - 1 - 1 = 2$	B1ft
	CR : $\chi^2_2(0.05) > 5.991$	B1
	$\sum \frac{(O-E)^2}{E} = \frac{(40-34.9937)^2}{34.9937} + \ldots + \frac{(13-8.972443)^2}{8.972443}$	M1
	[=0.7161…+0.3813…+1.4508…+1.80789..]	
	= 4.356. (ans in range 4.2 – 4.4)	A1
	Not in critical region Number of goals scored can follow a Poisson distribution / managers claim is justified	A1 ft (7)
		[12]
(a)	M1 for an attempt to find the mean- at least 2 terms on numerator seen Correct answer only will score both marks	
(b)	M1 for use of correct formula (ft their mean). 1st A1 for r , 2nd A1 for s (19.29 OK)	
(c)	1st B1 Must have both hypotheses and mention Poisson at least once inclusion of their value for mean in hypotheses is B0 but condone in conclusion 1st M1 for an attempt to pool ≥ 4 2nd B1ft for n -1-1 = 2 i.e realising that they must subtract 2 from their n 3rd B1 for 5.991 only 2nd M1 for an attempt at the test statistic, at least 2 correct expressions/values (to 3sf) 1st A1 for answers in the range 4.2~4.4 2nd A1 for correct comment in context based on their test statistic and their cv that mentions goals or manager. Dependent on 2nd M1 Condone mention of Po(1.05) in conclusion Score A0 for inconsistencies e.g. "significant" followed by "manager's claim is justified"	

Question Number	Scheme	Marks
Q6 (a)	μ_U ~ mean length of upper shore limpets, μ_L ~ mean length of lower shore limpets	
	$H_0 : \mu_u = \mu_L$ $H_1 : \mu_u < \mu_L$ both	B1
	s.e. $= \sqrt{\frac{0.42^2}{120} + \frac{0.67^2}{150}}$	M1
	$= 0.0668$	A1
	$z = \frac{5.05 - 4.97}{0.0668} = (\pm)1.1975$ awrt $\pm$ **1.20**	dM1 A1
	Critical region is $z \geq 1.6449$, or probability = awrt (0.115 or 0.116) $z = \pm 1.6449$	B1
	$(1.1975 < 1.6449)$ therefore not in critical region / accept H_0/not significant (or $P(Z \geq 1.1975) = 0.1151$, $0.1151 > 0.05$ or z not in critical region)	M1
	There is no evidence that the limpets on the upper shore are shorter than the limpets on the lower shore.	A1 (8)
(b)	Assume the populations or variables are independent	B1
	Standard deviation of sample = standard deviation of population [Mention of Central Limit Theorem does NOT score the mark]	B1 (2) **[10]**

(a)

1st B1 If μ_1, μ_2 used then it must be clear which refers to upper shore. Accept sensible choice of letters such as *u* and *l*.

1st M1 Condone minor slips e.g. $\frac{0.67^2}{120}$ or $\frac{0.67}{150} + \frac{0.42^2}{120}$ etc i.e. swapped *n* or one sd and one variance but M0 for $\sqrt{\frac{0.67}{150} + \frac{0.42}{120}}$

1st A1 can be scored for a fully correct expression. May be implied by awrt 1.20

2nd dM1 is dependent upon the 1st M1 but can ft their se value if this mark is scored.

2nd A1 for awrt (±) 1.20

3rd M1 for a correct statement based on their *z* value and their cv. No cv is M0A0
If using probability they must compare their *p* (<0.5) with 0.05 (o.e) so can allow 0.884< 0.95 to score this 3rd M1 mark.
May be implied by their contextual statement and M1A0 is possible.

(b)

3rd A1 for a correct comment to accept null hypothesis that mentions length of limpets on the two shores.

1st B1 for one correct statement. Accept "samples are independent"

2nd B1 for both statements

Question Number	Scheme	Marks
Q7 (a)	Estimate of Mean $= \frac{600.9}{5} = 120.18$	M1A1
	Estimate of Variance $= \frac{1}{4}\{ 72216.31 - \frac{600.9^2}{5} \}$ or $\frac{0.148}{4} = 0.037$	M1 A1ft A1 (5)
(b)	$P(-0.05 < \mu - \hat{\mu} < 0.05) = 0.90$ or $P(-0.05 < \overline{X} - \mu < 0.05) = 0.90$ [$\leq$ is OK]	B1
	$\frac{0.05}{\frac{0.2}{\sqrt{n}}} = 1.6449$	M1 A1
	$n = \frac{1.6449^2 \times 0.2^2}{0.05^2}$	dM1
	$n = 43.29\ldots$	A1
	$n = 44$	A1 (6) **[11]**
(a)	1st M1 for an attempt at $\sum x$ (accept 600 to 1sf) 1st A1 for $\frac{600.9}{5}$ = awrt 120 or awrt 120.2. No working give M1A1 for awrt 120.2 2nd M1 for the use of a correct formula including a reasonable attempt at $\sum x^2$ (Accept 70 000 to 1sf) or $\sum \left(x - \bar{x}\right)^2 = 0.15$ (to 2 dp) 2nd A1ft for a correct expression with correct $\sum x^2$ but can ft their mean (for expression - no need to check values if it is incorrect) 3rd A1 for 0.037 Correct answer with no working scores 3/3 for variance	
(b)	B1 for a correct probability statement or "width of 90% CI = $0.05 \times 2 = 0.1$" 1st M1 for $\frac{0.05}{\frac{0.2}{\sqrt{n}}} = z$ value or $2 \times \frac{0.2}{\sqrt{n}} \times z = 0.1$ Condone 0.5 instead of 0.05 or missing 2 or 0.05 for 0.1 for M1 1st A1 for a correct equation including 1.6449 2nd dM1 Dependent upon 1st M1 for rearranging to get n = …Must see "squaring" 2nd A1 for n = awrt 43.3 3rd A1 for rounding up to get n = 44 Using e.g.1.645 instead of 1.6449 can score all the marks except the 1st A1	1st B1 may be implied by 1st A1 scored or correct equation.

Question Number	Scheme	Marks
Q8 (a)	$E(4X-3Y) = 4E(X) - 3E(Y)$ $= 4\times 30 - 3\times 20$ $= 60$	M1 A1 (2)
(b)	$Var(4X-3Y) = 16\,Var(X) + 9\,Var(Y)$ 16 or 9; adding $= 16\times 9 + 9\times 4$ $= 180$	M1; M1 A1 (3)
(c)	$E(B) = 80$ $Var(B) = 16$ $E(B - A) = 20$ $E(B)-E(A)$ $Var(B - A) = 196$ ft on 180 and 16 $P(B - A > 0) = P\left(Z > \frac{-20}{\sqrt{196}}\right) = \left[P(Z > -1.428...)\right]$ stand. using their mean and var $= 0.923\ldots$ awrt 0.923 – 0.924	B1 B1 M1 A1ft dM1 A1 (6) **[11]**

(a) M1 for correct use of $E(aX + bY)$ formula

(b) 1st M1 for $16Var(X)$ or $9Var(Y)$
2nd M1 for adding variances

Key points are the 16, 9 and +. Allow slip e.g using $Var(X)=4$ etc to score Ms

(c) 1st M1 for attempting $B - A$ and $E(B - A)$ or $A - B$ and $E(A - B)$
This mark may be implied by an attempt at a correct probability
e.g. $P\left(Z > \frac{0-(80-60)}{\sqrt{180+16}}\right)$. To be implied we must see the "0"
1st A1ft for $Var(B - A)$ can ft their $Var(A) = 180$ and their $Var(B) = 16$
2nd dM1 Dependent upon the 1st M1 in part (c).
for attempting a correct probability i.e. $P(B-A>0)$ or $P(A-B < 0)$ and
standardising with their mean and variance.
They must standardise properly with the 0 to score this mark
2nd A1 for awrt 0.923 ~ 0.924

edexcel
advancing learning, changing lives

Mark Scheme (Results)
Summer 2010

GCE

GCE Statistics S3 (6691/01)

Edexcel Limited. Registered in England and Wales No. 4496750
Registered Office: One90 High Holborn, London WC1V 7BH

A PEARSON COMPANY

Edexcel is one of the leading examining and awarding bodies in the UK and throughout the world. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers.

Through a network of UK and overseas offices, Edexcel's centres receive the support they need to help them deliver their education and training programmes to learners.

For further information, please call our GCE line on 0844 576 0025, our GCSE team on 0844 576 0027, or visit our website at www.edexcel.com.

If you have any subject specific questions about the content of this Mark Scheme that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

Ask The Expert can be accessed online at the following link:

http://www.edexcel.com/Aboutus/contact-us/

Summer 2010

Publications Code UA024774

Hypothesis Tests (Final M1A1)

For an incorrect comparison (e.g. probability with *z* value) even with a correct statement and/or comment award M0A0

For a correct or no comparison with more than one statement one of which is false
Award M0A0 (This is compatible with the principle above of contradictory statements being penalised)

Apply these rules to all questions

edexcel

June 2010
Statistics S3 6691
Mark Scheme

Question Number	Scheme	Marks
Q1	H_0: $\mu = 80$, H_1: $\mu > 80$	B1,B1
	$z = \dfrac{83-80}{\frac{15}{\sqrt{100}}} = 2$	M1A1
	$2 > 1.6449$ (accept 1.645 or better)	B1
	Reject H_0 or significant result or in the critical region	M1
	Managing director's claim is supported.	A1
		7

1st B1 for H_0. They must use μ not *x, p,* λ or $\bar{x}$ etc
2nd B1 for H_1 (must be > 80). Same rules about μ.
1st M1 for attempt at standardising using 83, 80 and $\frac{15}{\sqrt{100}}$. Can accept $\pm$.
May be implied by $z = \pm 2$
1st A1 for + 2 only

3rd B1 for ± 1.6449 seen (or probability of 0.0228 or better)
2nd M1 for a correct statement about "significance" or rejecting H_0 (or H_1) based on their *z* value and their 1.6449 (provided it is a recognizable critical value from normal tables) or their probability (< 0.5) and significance level of 0.05.
Condone their probability > 0.5 compared with 0.95 for the 2nd M1
2nd A1 for a correct contextualised comment. Must mention "director" and "claim" or "time" and "use of Internet". No follow through.

2nd M1A1

If no comparison or statement is made but a correct contextualised comment is given the M1 can be implied.
If a comparison is made it must be compatible with statement otherwise M0
e.g. comparing 0.0228 with 1.6449 is M0 or comparing probability 0.9772 with 0.05 is M0
comparing -2 with - 1.6449 is OK provided a correct statement accompanies it
condone -2 >-1.6449 provided their statement correctly rejects H_0.

Critical Region

They may find a critical region for $\bar{X}$: $\bar{X} > 80 + \frac{15}{\sqrt{100}} \times 1.6449$ = awrt 82.5

1st M1 for $80 + \frac{15}{\sqrt{100}} \times$ (*z* value)
3rd B1 for 1.645 or better
1st A1 for awrt 82.5
The rest of the marks are as per the scheme.

Question Number	Scheme	Marks
Q2	[$P \sim N(90,9)$ and $J \sim N(91,12)$]	
(a)	$(J-P) \sim N(1,21)$	M1, A1
	$P(J<P) = P(J-P<0)$	
	$= P\left(Z < \frac{0-1}{\sqrt{21}}\right)$	dM1
	$= P(Z < -0.2182...)$	
	$= 1-0.5871 = 0.4129$ awrt (**0.413 ~ 0.414**)	A1
	calculator (0.4136….)	(4)
(b)	$X = (J_1 + J_2 + ... + J_{60}) - (P_1 + P_2 + ... + P_{60})$	M1
	$E(X) = 60\times 91 - 60\times 90 = 60$ [stated as E(X) = 60 or X~N(60, …)]	B1
	$Var(X) = 60\times 9 + 60\times 12 = 1260$	A1
	$P(X>120) = P\left(Z > \frac{120-60}{\sqrt{1260}}\right)$	M1
	$= P(Z > 1.69030...)$	
	$= 1-0.9545 = 0.0455$ awrt (**0.0455**)	A1
		(5)
		9

(a)

1st M1 for attempting J - P and E(J - P) or P - J and E(P - J)

1st A1 for variance of 21 (Accept 9 + 12). Ignore any slip in μ here.

2nd dM1 for attempting the correct probability and standardising with their mean and sd.
This mark is dependent on previous M so if J - P (or P - J) is not being used score M0
If their method is not crystal clear then they must be attempting P(Z< -ve value) or P(Z > +ve value) i.e. their probability <u>after</u> standardisation should lead to a prob. < 0.5 so e.g. P(J - P < 0) leading to 0.5871 is M0A0 unless the M1 is clearly earned.

2nd A1 for awrt 0.413 or 0.414

The first 3 marks may be implied by a correct answer

(b)

1st M1 for a clear attempt to identify a correct form for X. This may be implied by correct variance of 1260

B1 for E(X) = 60. Can be awarded even if they are using X = 60J - 60P. Allow P - J and -60

1st A1 for a correct variance. If 1260 is given the M1 is scored by implication.

2nd M1 for attempting a correct probability and standardising with 120 and their 60 and 1260
If the answer is incorrect a full <u>expression</u> must be seen following through their values for M1 e.g. $P\left(Z > \frac{120 - \text{ their } 60}{\sqrt{\text{their variance}}}\right)$. If using -60, should get $P\left(Z < \frac{-120 - -60}{\sqrt{\text{their variance}}}\right)$

Use of means

Attempt to use $\bar{J} - \bar{P}$ for 1st M1, E($\bar{J} - \bar{P}$) = 1 for B1 and Var($\bar{J} - \bar{P}$) = 0.35 for A1
Then 2nd M1 for standardisation with 2, and their 1 and 0.35

Question Number	Scheme	Marks
Q3 (a)	$E \sim \mathrm{N}(0, 0.5^2)$ or $X \sim \mathrm{N}\left(w, 0.5^2\right)$	
	$\mathrm{P}(\lvert E\rvert < 0.6) = \mathrm{P}\left(\lvert Z\rvert < \frac{0.6}{0.5}\right)$ or $\mathrm{P}\left(\lvert X - w\rvert < 0.6\right) = \mathrm{P}\left(\lvert Z\rvert < \frac{0.6}{0.5}\right)$	M1
	$= \mathrm{P}(\lvert Z\rvert < 1.2)$	
	$= 2 \times 0.8849 - 1 = 0.7698$ awrt **0.770**	A1
		(2)
(b)	$\bar{E} \sim \mathrm{N}\left(0, \frac{1}{64}\right)$ or $\bar{X} \sim \mathrm{N}\left(w, \frac{0.5^2}{16}\right)$	M1
	$\mathrm{P}\left(\lvert \bar{E}\rvert < 0.3\right) = \mathrm{P}\left(\lvert Z\rvert < \frac{0.3}{\frac{1}{8}}\right)$ or $\mathrm{P}\left(\lvert \bar{X} - w\rvert < 0.3\right) = \mathrm{P}\left(\lvert Z\rvert < \frac{0.3}{\frac{1}{8}}\right)$	M1, A1
	$= \mathrm{P}(\lvert Z\rvert < 2.4)$	
	$= 2 \times 0.9918 - 1 = 0.9836$ awrt **0.984**	A1
		(4)
(c)	$35.6 \pm 2.3263 \times \frac{1}{8}$	M1 B1
	(**35.3, 35.9**)	A1,A1
		(4)
		10

(a)
1st M1 for identifying a correct probability (they must have the 0.6) and attempting to standardise. Need | |. This mark can be given for 0.8849 - 0.1151 seen as final answer.
1st A1 for awrt 0.770. NB an answer of 0.3849 or 0.8849 scores M0A0 (since it implies no | |)
M1 may be implied by a correct answer

(b)
1st M1 for a correct attempt to define $\bar{E}$ or $\bar{X}$ but must attempt $\frac{\sigma^2}{n}$. Condone labelling as E or X
This mark may be implied by standardisation in the next line.
2nd M1 for identifying a correct probability statement using $\bar{E}$ or $\bar{X}$. Must have 0.3 and | |
1st A1 for correct standardisation as printed or better
2nd A1 for awrt 0.984
The M marks may be implied by a correct answer.

Sum of 16, not means
1st M1 for correct attempt at suitable sum distribution with correct variance (= $16 \times \frac{1}{4}$)
2nd M1 for identifying a correct probability. Must have 4.8 and | |
1st A1 for correct standardisation i.e. need to see $\frac{4.8}{\sqrt{4}}$ or better

(c)
M1 for $35.6 \pm z \times \frac{0.5}{\sqrt{16}}$
B1 for 2.3263 or better. Use of 2.33 will lose this mark but can still score ¾
1st A1 for awrt 35.3
2nd A1 for awrt 35.9

Question Number	Scheme	Marks
Q4 (a)		

Distance rank	1	2	3	4	5	6	7
Depth rank	1	2	4	3	6	7	5
$\lvert d \rvert$	0	0	1	1	1	1	2
d^2	0	0	1	1	1	1	4

M1

M1

$\sum d^2 = 8$ M1A1

$r_s = 1 - \frac{6\times 8}{7\times 48}$ M1

$= \frac{6}{7} = 0.\dot{8}57142\dot{}$ awrt **0.857** A1

(6)

(b) $H_0 : \rho = 0, H_1 : \rho > 0$ B1

Critical value at 1% level is 0.8929 B1

$r_s < 0.8929$ so not significant evidence to reject H_0, M1

The researcher's claim is not correct (at 1% level). A1ft

or insufficient evidence for researcher's claim

or there is insufficient evidence that water gets deeper further from inner bank.

or no (positive) correlation between depth of water and distance from inner bank

(4)

10

(a) 1st M1 for an attempt to rank the depths against the distances

2nd M1 for attempting d for their ranks. Must be using ranks.

3rd M1 for attempting $\sum d^2$ (must be using ranks)

1st A1 for sum of 8 (or 104 for reverse ranking)

4th M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required.

2nd A1 for awrt (±) 0.857. Sign should correspond to ranking (so use of 104 should get -0.857)

(b) 1st B1 for both hypotheses in terms of ρ, H_1 must be one tail and compatible with their ranking

2nd B1 for cv of 0.8929 (accept ±)

M1 for a correct statement relating their r_s with their cv but cv must be such that |cv|<1

A1ft for a correct contextualised comment. Must mention "researcher" and "claim" or "distance (from bank)" and "depth (of water)"

Follow through their r_s and their cv (provided it is |cv| <1)

Use of "association" is A0

Question Number	Scheme	Marks
Q5		

Finances / Income	Worse	Same	Better	
Under £15 000	10.54	10.54	12.92	34
£15 000 and above	20.46	20.46	25.08	66
	31	31	38	100

M1
A1

H_0 : State of finances and income are independent (not associated)

H_1 : State of finances and income are not independent (associated)

B1

O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$	$\frac{O_i^2}{E_i}$
14	10.54	1.1358….	18.59..
11	10.54	0.0200….	11.48..
9	12.92	1.1893…	6.269..
17	20.46	0.5851…	14.12..
20	20.46	0.0103…	19.55..
29	25.08	0.6126…	33.53..

M1
A1

$\sum \frac{(O_i - E_i)^2}{E_i} = 3.553\ldots$ or $\sum \frac{O_i^2}{E_i} - 100 = 103.553... - 100 = 3.553...$ (awrt **3.55)**

A1

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

B1

cv is 5.991

B1

3.553 < 5.991 so insufficient evidence to reject H_0 or not significant

M1

There is no evidence of association between state of finances and income.

A1

10

1st M1 for some use of $\frac{\text{Row Total} \times \text{Col.Total}}{\text{Grand Total}}$. May be implied by correct E_i

1st A1 for all expected frequencies correct

B1 for both hypotheses. Must mention "state" or "finances" and "income" at least once
Use of "relationship" or "correlation" or "connection"is B0

2nd M1 for at least two correct terms (as in 3rd or 4th column) or correct expressions with their E_i

2nd A1 for all correct terms. May be implied by a correct answer.(2 dp or better-allow eg 1.13…)

3rd M1 for a correct statement linking their test statistic and their cv . Must be χ^2 not normal.

4th A1 for a correct comment in context - must mention "state" or "finances" and "income"
condone "relationship" or "connection" here but **not** "correlation". No follow through.
e.g. "There is no evidence of a relationship between finances and income"

Question Number	Scheme	Marks
Q6		

Distance from centre of site (m)	0-1	1-2	2-4	4-6	6-9	9-12
$b-a$	1	1	2	2	3	3
No of artefacts	22	15	44	37	52	58
$\mathrm{P}(a \le X < b)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{4}$
$228 \times \mathrm{P}(a \le X < b)$	19	19	38	38	57	57

M1

A1

A1

Class	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$	$\frac{O_i^2}{E_i}$
0-1	22	19	$\frac{9}{19} = 0.4736...$	25.57…
1-2	15	19	$\frac{16}{19} = 0.8421...$	11.84…
2-4	44	38	$\frac{36}{38} = 0.9473...$	50.94…
4-6	37	38	$\frac{1}{38} = 0.0263...$	36.02…
6-9	52	57	$\frac{25}{57} = 0.4385...$	47.43…
9-12	58	57	$\frac{1}{57} = 0.0175...$	59.01…

M1

A1

H_0: <u>continuous uniform</u> distribution <u>is</u> a good fit

H_1: <u>continuous uniform</u> distribution <u>is not</u> a good fit

B1

$\sum \frac{(O_i - E_i)^2}{E_i} = \frac{313}{114} = 2.75$ <u>or</u> $\sum \frac{O_i^2}{E_i} - 228 = 230.745... - 228 = \ldots$ (awrt **2.75**)

dM1A1

$\nu = 6 - 1 = 5$

B1

$\chi^2_5(0.05) = 11.070$ (ft their ν i.e. $\chi^2_\nu(0.05)$)

B1ft

2.75<11.070, insufficient evidence to reject H_0

M1

Continuous uniform distribution is a suitable model

A1

12

1st M1 for calculation of at least 3 widths and attempting proportions/probs. <u>or</u> for 1:2:3 ratio seen

1st A1 for correct probabilities

2nd A1 for all correct expected frequencies

2nd M1 for attempting $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$, at least 3 correct expressions or values.

Follow through their E_i provided they are not all = 38

3rd A1 for a correct set of calcs - 3rd or 4th column. (2 dp or better and allow e.g. 0.94…)

3rd dM1 **dependent on 2nd M1** for attempting a correct sum or calculation (must see at least 3 terms and +)

The first three Ms and As can be implied by a test statistic of awrt 2.75

4th M1 for a correct statement based on their test statistic (> 1) and their cv (> 3.8)

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

5th A1 for a correct comment suggesting that continuous uniform model is suitable. No ft

Question Number		Scheme	Marks
Q7	(a)	Label full time staff 1-6000, part time staff 1-4000	M1
		Use random numbers to select	M1
		Simple random sample of 120 full time staff and 80 part time staff	A1 (3)
	(b)	Enables estimation of statistics / errors for each strata <u>or</u> "reduce variability" <u>or</u> "more representative" <u>or</u> "reflects population structure" **NOT** "more accurate"	B1 (1)
	(c)	$H_0: \mu_f = \mu_p$, $H_1: \mu_f \neq \mu_p$ (accept μ_1, μ_2)	B1
		s.e. $= \sqrt{\frac{21}{80}+\frac{19}{80}}$, $z = \frac{52-50}{\sqrt{\frac{21}{80}+\frac{19}{80}}} = \left(2\sqrt{2}\right)$	M1,M1
		$= 2.828\ldots$ (awrt **2.83**)	A1
		Two tailed critical value z = 2.5758 (or prob of awrt 0.002 (<0.005) or 0.004 (<0.01))	B1
		[2.828 > 2.5758 so] significant evidence to reject H_0	dM1
		There is evidence of a difference in policy awareness between full time and part time staff	A1ft (7)
	(d)	Can use mean full time and mean part time	B1
		~ Normal	B1 (2)
	(e)	Have assumed $s^2 = \sigma^2$ or variance of sample = variance of population	B1 (1)
	(f)	2.53 < 2.5758, not significant <u>or</u> do not reject H_0	M1
		So there is insufficient evidence of a difference in mean awareness	A1ft (2)
	(g)	Training course has closed the gap between full time staff and part time staff's mean awareness of company policy.	B1 (1)
			17

	(a)	1st M1 for attempt at labelling full-time and part-time staff. One set of correct numbers. 2nd M1 for mentioning use of random numbers 1st A1 for s.r.s. of 120 full-time and 80 part-time
	(c)	1st M1 for attempt at s.e. - condone one number wrong . NB correct s.e. = $\sqrt{\frac{1}{2}}$ 2nd M1 for using their s.e. in correct formula for test statistic. Must be $\frac{\pm(52-50)}{\sqrt{\frac{p}{q}+\frac{r}{s}}}$ 3rd dM1 **dep. on 2nd M1** for a correct statement based on their normal cv and their test statistic 2nd A1 for correct comment in context. Must mention "scores" or " policy awareness" and types of "staff". Award **A0** for a one-tailed comment. Allow ft
	(d)	1st B1 for mention of mean(s) <u>or</u> use of $\bar{X}$, provided $\bar{X}$ clearly refers to full-time or part-time 2nd B1 for stating that distribution can be assumed normal e.g. "mean score of the test is normally distributed" gets B1B1
	(f)	M1 for correct statement (may be implied by correct contextualised comment) A1 for correct contextualised comment. Accept "no difference in mean scores". Allow ft
	(g)	B1 for correct comment in context that implies training was effective. This must be supported by their (c) and (f). Condone one-tailed comment here.

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467
Fax 01623 450481

Email publications@linneydirect.com

Order Code UA024774 Summer 2010

For more information on Edexcel qualifications, please visit www.edexcel.com/quals

Edexcel Limited. Registered in England and Wales no.4496750
Registered Office: One90 High Holborn, London, WC1V 7BH

edexcel

Mark Scheme (Results)

June 2011

GCE Statistics S3 (6691) Paper 1

Edexcel is one of the leading examining and awarding bodies in the UK and throughout the world. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers.

Through a network of UK and overseas offices, Edexcel's centres receive the support they need to help them deliver their education and training programmes to learners.

For further information, please call our GCE line on 0844 576 0025 or visit our website at www.edexcel.com.

If you have any subject specific questions about the content of this Mark Scheme that require the help of a subject specialist, you may find our **Ask The Expert** email service helpful.

Ask The Expert can be accessed online at the following link:
http://www.edexcel.com/Aboutus/contact-us/

June 2011
Publications Code UA028846

edexcel
advancing learning, changing lives

EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75.

2. The Edexcel Mathematics mark schemes use the following types of marks:

   - M marks: method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
   - A marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
   - B marks are unconditional accuracy marks (independent of M marks)
   - Marks should not be subdivided.

3. Abbreviations

   These are some of the traditional marking abbreviations that will appear in the mark schemes and can be used if you are using the annotation facility on ePEN.

   - bod – benefit of doubt
   - ft – follow through
   - the symbol  will be used for correct ft
   - cao – correct answer only
   - cso - correct solution only. There must be no errors in this part of the question to obtain this mark
   - isw – ignore subsequent working
   - awrt – answers which round to
   - SC: special case
   - oe – or equivalent (and appropriate)
   - dep – dependent
   - indep – independent
   - dp decimal places
   - sf significant figures
   - * The answer is printed on the paper
   -  The second mark is dependent on gaining the first mark

edexcel
advancing learning, changing lives

June 2011
Statistics S3 6691
Mark Scheme

Question Number	Scheme	Marks
1.	$X_1, X_2, ...X_n$ is a random sample of size n , for large n, drawn from a population of any distribution with mean μ and variance σ^2 then $\bar{X}$ is (approximately) $N\left(\mu, \frac{\sigma^2}{n}\right)$	B1 B1 B1 (3) **3**
	1st B for large sample or equivalent 2nd B for 'population of any distribution' or 'any population' 3rd B require mean or symbol and normal (parameters not required)	

Question Number	Scheme	Marks
2. **(a)**	<table><tr><td>Town</td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td><td>F</td><td>G</td></tr><tr><td>h rank</td><td>1</td><td>5</td><td>2</td><td>3</td><td>7</td><td>4</td><td>6</td></tr><tr><td>c rank</td><td>4</td><td>3</td><td>2</td><td>1</td><td>6</td><td>7</td><td>5</td></tr><tr><td>$\lvert d \rvert$</td><td>3</td><td>2</td><td>0</td><td>2</td><td>1</td><td>3</td><td>1</td></tr><tr><td>d^2</td><td>9</td><td>4</td><td>0</td><td>4</td><td>1</td><td>9</td><td>1</td></tr></table> $\sum d^2 = 28$ $r_s = 1 - \frac{6\times 28}{7\times 48}$ $= 0.5$	M1 M1 M1A1 M1 A1 (6)
(b)	$H_0 : \rho = 0, H_1 : \rho \neq 0$ Critical values are $r_s = \pm 0.7857$ 0.5<0.7857 insufficient evidence to reject H_0 Councillor's claim is supported.	B1 B1ft M1 A1ft (4) **10**

Question Number	Scheme	Marks
Notes		
(a)	1st M1 for an attempt to rank the hardship against calls	
	2nd M1 for attempting d for their ranks. Must be using ranks.	
	3rd M1 for attempting $\sum d^2$ (must be using ranks)	
	1st A1 for sum of 28 (or 84)	
	4th M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required.	
	2nd A1 for awrt 0.5 (or -0.5)	
(b)	1st B1 for both hypotheses in terms of ρ, H_1 must be two tail.	
	2nd B1 for cv of $\pm$0.7857 (or 0.7143 to ft from 1-tailed H_1)	
	M1 for a correct statement relating their r_s with their cv but cv must be such that |cv|<1	
	A1ft for a correct contextualised comment. Must mention "Councillor" and "claim" or "hardship" and "number of calls (to the emergency services)"	
	Follow through their r_s and their cv (provided it is |cv| <1	
	Condone use of "association" in conclusion for A1	
	Condone 'positive' in conclusion.	

Question Number	Scheme	Marks
3.		

Defect Type / Shift	D_1	D_2	
First Shift	47.25	15.75	63
Second Shift	56.25	18.75	75
Third Shift	46.5	15.5	62
	150	50	200

M1A1

H_0 : Type of defect is independent of Shift (no association)

H_1 : Type of defect is not independent of Shift (association)

B1

O	E	$\frac{(O-E)^2}{E}$	$\frac{O_i^2}{E_i}$
45	47.25	0.1071…	42.857...
18	15.75	0.3214…	20.571..
55	56.25	0.02777…	53.777...
20	18.75	0.0833…	21.333...
50	46.5	0.2634…	53.763...
12	15.5	0.7903…	9.290...

M1A1

$\frac{(O-E)^2}{E}$ =1.5934.. or $\frac{O_i^2}{E_i}$ -200=201.5934-200=1.5934.. awrt1.59 — A1

$\nu = (3-1)(2-1) = 2$ — B1

$\chi^2_2(0.10) = 4.605$ — B1ft

1.59<4.605 so insufficient evidence to reject H_0 — M1

Insufficient evidence to support manager's belief /claim. — A1

10

Question Number	Scheme	Marks
Notes	1st M1 for some use of $\frac{\text{Row Total} \times \text{Col.Total}}{\text{Grand Total}}$ May be implied by correct E_i 1st A1 for all expected frequencies correct B1 for both hypotheses. Must mention "defect" and "shift" at least once Use of "relationship" or "correlation" or "connection" is B0 2nd M1 for at least two correct terms (as in 3rd or 4th column) or correct expressions with their E_i 2nd A1 for all correct terms. May be implied by a correct answer.(2 dp or better-allow eg 0.10…) 3rd M1 for a correct statement linking their test statistic and their cv . Must be χ^2 not normal. 4th A1 for a correct comment in context - must mention "manager's belief" or "shift" and "defect type" - condone "relationship" or "connection" here but **not** "correlation". No follow through e.g. "There is evidence of a relationship between shift and type of defect"	

Question Number	Scheme	Marks
4. **(a)**	$\bar{x} = \frac{5320}{80} = 66.5$ $s^2 = \frac{392000 - 80 \times (66.5)^2}{79}$ = 483.797… awrt 484	M1,A1 M1A1ft A1 (5)
(b)	$H_0: \mu_m = \mu_{nm}$, $H_1: \mu_m > \mu_{nm}$ (accept μ_1, μ_2 with definition) $z = \frac{69.0 - 66.5}{\sqrt{\frac{483.797}{80} + \frac{446.44}{60}}}$ = 0.6807 awrt 0.681 One tailed cv 1.6449 (Probability is awrt 0.752) 0.6807<1.6449 (or 0.248>0.05) insufficient evidence to reject H_0 Mean money spent is not greater with music playing.	B1B1 M1dM1 A1 B1 dM1 A1ft (8) **13**

Question Number	Scheme	Marks
(b)	**Notes** No definition award B1B0. 1st M1 for attempt at s.e. - condone one number wrong or switched 60 & 80 . 2nd dM1 for using their s.e. in correct formula for test statistic. 3rd dM1 **dep. on 2nd M1** for a correct statement based on their normal cv and their test statistic 2nd A1 for correct comment in context. Must mention "money spent" and "music playing". Allow ft. Critical Region for (b) Standard error x z value for 2nd M1 Standard error x 1.6449= awrt 6.04 for 1st A1 2.5<6.04	

<table>
<tr><th>Question Number</th><th>Scheme</th><th>Marks</th></tr>
<tr><td>5.
(a)</td><td>Hurricanes: occur singly / are independent or occur at random /are a rare event / at a constant rate</td><td>B1B1
(2)</td></tr>
<tr><td>(b)</td><td>From data $\frac{1\times2+2\times5+3\times17+..+7\times12}{80}=4.4875$</td><td>M1A1
(2)</td></tr>
<tr><td>(c)
(d)</td><td>
<table>
<tr><td>No of hurricanes, h</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7+</td></tr>
<tr><td>$80\mathrm{P}(X=h$</td><td>0.9</td><td>4..038</td><td>r=9.06…</td><td>13.55</td><td>s=15.205</td><td>13.647…</td><td>10.206…</td><td>13.388…</td></tr>
<tr><td>Combine to give expected frequencies >5</td><td colspan="3">13.9991</td><td>13.55</td><td>15.205…</td><td>13.647…</td><td>10.206…</td><td>13.388…</td></tr>
<tr><td>Observed</td><td colspan="3">7</td><td>17</td><td>20</td><td>12</td><td>12</td><td>12</td></tr>
<tr><td>$\frac{(O-E)^2}{E}$</td><td colspan="3">3.499…</td><td>0.876…</td><td>1.511…</td><td>0.198…</td><td>0.315…</td><td>0.143…</td></tr>
<tr><td>$\frac{O_i^2}{E_i}$</td><td colspan="3">3.500...</td><td>21.322...</td><td>26.306...</td><td>10.551...</td><td>14.108...</td><td>10.755..</td></tr>
</table>
H_0: Poisson distribution is a good fit o.e.
H_1: Poisson distribution is not a good fit o.e.
$\sum\frac{(O_i-E_i)^2}{E_i}$=6.545.. or $\frac{O_i^2}{E_i}$=86.545-80=6.545.. (awrt 6.55 or 6.54)
$\nu=6-2=4$
cv is 9.488 (ft their ν i.e. $\chi_\nu^2(0.05)$)
6.545<9.488 so insufficient evidence to reject H_0
(Hurricanes) can be modelled by a Poisson distribution</td><td>M1A1A1
(3)
M1

B1
A1
B1
B1ft

A1
(6)
13</td></tr>
</table>

Question Number	Scheme	Marks
	Notes	
(b)	M for at least 2 terms on numerator. 359/80 only award M0A0	
(c)	M for 80xPoisson probability with 4.4875 and either 2 or 4. 1st A1 for awrt 9.06 and 2nd A1 for awrt 15.20 or 15.21	
(d)	1st M1 for some pooling and attempting $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$, at least 3 correct expressions or values. 1st B1 no value for parameter permitted 2nd A1 for a correct comment suggesting that Poisson model is suitable. No ft	

Question Number	Scheme	Marks
6. **(a)**	$L = A_1 + A_2 + ... + A_6$ Mean is $\mathrm{E}(L) = 6 \times 20 = 120$ Standard deviation is $\sqrt{\mathrm{Var}(W)} = \sqrt{6 \times 5^2} = 5\sqrt{6} = 12.247...$ awrt 12.2	B1 B1 (2)
(b)	$P(L > 110) = P\left(Z > \left(\frac{110 - 120}{12.247...}\right)\right)$ $= P(Z < 0.8164...)$ $= 0.7939$ (or 0.7929 using interpolation or 0.79289 by calc)	M1 A1 (2)
(c)	Let $X = 4B - \sum_1^6 A_i$ $\mathrm{E}(X) = 140 - 120 = 20$ $\mathrm{Var}(X) = 16 \times 8^2 + 6 \times 5^2 = 1174$ $\mathrm{P}(X < 0) = \mathrm{P}\left(Z < \frac{-20}{\sqrt{1174}}\right) = \mathrm{P}(Z < -0.583\ldots)$ $= 0.2797$ (or 0.2810 if no interpolation) or 0.27971 by calc.	B1 M1M1A1 M1 A1 (6) **10**

Question Number	Scheme	Marks
	Notes	
(b)	M1 for identifying a correct probability (they must have the 110) and attempting to standardise with their mean and sd. This can be implied by the correct answer. A1 for awrt 0.794 or 0.793	
(c)	Accept ±20 for B mark. Only award for probability statement if 2 terms in var 1st M1 for 1024, 2nd M1 for 150 3rd M for standardising with their mean and 2 term sd and finding probability <0.5 2nd A1 for awrt 0.280 or 0.281	

Question Number	Scheme	Marks
7. **(a)**	$H_0: \mu = 250$, $H_1: \mu < 250$, $z = \dfrac{248-250}{\frac{5.4}{\sqrt{90}}}$ $= -3.513...$ awrt -3.51 Critical value -1.6449 -3.513..<-1.6449 so sufficient evidence to reject H_0 Manager's claim is justified.	B1 M1 A1 B1 A1 (5)
(b)	98% CI for μ is $248 \pm 2.3263 \times \dfrac{5.4}{\sqrt{90}}$ = awrt (247,249) dependent upon z value awrt 2.33	M1B1 A1A1 (4)
(c)	Hypothesis test is significant or CI does not contain stated weight. (Manager should ask the company to investigate if their) stated weight is too high o.e.	B1 B1 (2)
(d)	$P(\lvert\bar{x}-\mu\rvert < 1) = 0.98$ $\dfrac{1}{\frac{3}{\sqrt{n}}} = 2.3263$ $n = (3 \times 2.3263)^2 = 48.7...$ Sample size 49 required.	M1 A1 dM1A1 A1 (5) **16**

Question Number	Scheme	Marks
	Notes	
(a)	1st B1 for H_0 and for H_1 (must be <250) They must use μ not x, p, λ or $\bar{x}$ etc 1st M1 for attempt at standardising using 248, 250 and sd. Can accept $\pm$. Critical region: 250-0.936=249.064 for M1A1 (and compare with 248.) 3rd B1 for $\pm$1.6449 seen (or probability of 0.0002 or better) 2nd A1 for a correct contextualised comment. Must mention "Manager" and "claim" or "weight" and "stated weight". No follow through.	
(b)	2.3263 or better for B mark. Any z value replacing 2.3263 award M.	
(d)	1st M for LHS = z value >1 1st A for RHS awrt 2.33 2nd A1 for answers in the range 48.7-48.9 3rd A1 don't condone $\geq$	

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467
Fax 01623 450481
Email publication.orders@edexcel.com
Order Code UA028846 June 2011

For more information on Edexcel qualifications, please visit
www.edexcel.com/quals

Pearson Education Limited. Registered company number 872828
with its registered office at Edinburgh Gate, Harlow, Essex CM20 2JE

Ofqual

Llywodraeth Cynulliad Cymru
Welsh Assembly Government

Rewarding Learning

edexcel

Mark Scheme (Results)

Summer 2012

GCE Statistics S3 (6691) Paper 1

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please visit our website at www.edexcel.com.

Our website subject pages hold useful resources, support material and live feeds from our subject advisors giving you access to a portal of information. If you have any subject specific questions about this specification that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

www.edexcel.com/contactus

Pearson: helping people progress, everywhere

Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2012
Publications Code UA033146

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the team leader must be consulted.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

Hypothesis Tests (Final M1A1)

For an incorrect comparison (e.g. probability with *z* value) even with a correct statement and/or comment award M0A0

For a correct or no comparison with more than one statement one of which is false
Award M0A0 (This is compatible with the principle above of contradictory statements being penalised)

Apply these rules to all questions

June 2012
6691 Statistics S3
Mark Scheme

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)

1(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)

1(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

Notes

1(a)

1st M1 for an attempt to rank score *X* and score *Y*

2nd M1 for attempting d^2 for their ranks. Must be using ranks.

1st A1 for sum of 18

3rd M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required.

2nd A1 for awrt 0.786

1(b)

1st B1 for null hypotheses in terms of ρ or ρ_s

2nd B1 for alt hyp as given

3rd B1 for cv of +0.6429 (or 0.7381 if two tailed from hyp)

M1 for a correct statement relating their r_s with their cv but cv must be such that |cv|<1

A1ft for a correct contextualised comment. Must mention "scores / rankings" and "manager"

Follow through their r_s and their cv (provided it is |cv| <1

Use of "association" is A0

1(c)

1st B1 Tied ranks can be implied by 2.5, 6.5 or **both** 2 or 6 or description.

2nd B1 Average rank implied by 2.5 or 6.5 or description and 'use of pmcc'.

Question Number	Scheme	Marks
2(a)	Sampling frame within each species of fish in the lake impossible to obtain.	B1 **(1)**
2(b)	Quota sampling	B1 **(1)**
2(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)**
2(d)	<table><tr><th>Species</th><th>Quota</th></tr><tr><td>Trout</td><td>$\frac{1400}{2450}\times 30 = 17.14$</td></tr><tr><td>Bass</td><td>$\frac{600}{2450}\times 30 = 7.35$</td></tr><tr><td>Pike</td><td>$\frac{450}{2450}\times 30 = 5.51$</td></tr></table> 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) **Total 8**
	Notes	
2(a)	'You can't / it's very difficult to number all the fish' or equivalent	
2(c)	Correct answer to (b) required. Some detail required.	
2(d)	1st B1 any one correct calculation seen or implied 2nd B1 all correct to at least 1 dp 3rd B1 for 17,7,6 4th B1 accept equivalent statement. Require comment on what to do with 'extra fish'.	

Question Number	Scheme	Marks
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)**
3 (b)	$\bar{x} = \frac{1740000}{100} = 17400$	B1
	$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}, = 17400 \pm 1.96 \times \frac{5000}{\sqrt{100}}$	M1, B1
	[16420,18380]	A1A1 **(5)**
3(c)	$\bar{X}$: 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)**
		Total 11
3(b)	Notes Recognisable z value required for method. 2nd B1 1.96 or better seen award Final A1s accept 3sf if correct expression seen. 5/5 for [16420,18380]	

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

Egg Yield / Breed	Low	Medium	High	Total
Leghorn	$\frac{100\times36}{150}=24$	$\frac{100\times84}{150}=56$	$\frac{100\times30}{150}=20$	100
Cornish	$\frac{50\times36}{150}=12$	$\frac{50\times84}{150}=28$	$\frac{50\times30}{150}=10$	50
Total	36	84	30	150

M1A1

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

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)

Total 10

Notes

B1 for both hypotheses. Must mention "yield" and "breed" in both but condone ditto marks.

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

1st M1 for some use of $\frac{\text{Row Total}\times\text{Col.Total}}{\text{Grand Total}}$. May be implied by a correct E_i

1st A1 for all expected frequencies correct

2nd M1 for at least two correct terms or correct expressions with their E_i

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

3rd M1 for a correct statement linking their test statistic and their cv . Must be χ^2 not normal.

4th A1 for a correct comment in context - must mention "egg yield" and "breed of chicken" - condone "relationship" or "connection" here but not "correlation". No follow through e.g. "There is no evidence of a relationship between egg yield and breed of chicken" is A0 whatever their test stat or cv.

Question Number	Scheme	Marks
5(a)	$H_0 : \mu_A = \mu_B$ $H_1 : \mu_A \neq \mu_B$	B1
	$z = \dfrac{\pm(80-74)}{\sqrt{\dfrac{100}{29}+\dfrac{225}{26}}}$	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)**
5(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)**
		Total 9
5(a)	Notes 1st M1 for attempt at s.e. (condone one number wrong) and for using their s.e. in correct formula for test statistic. 1st A1 for correct expression for se 2nd dM1 dep. on 1st M1 for a correct statement based on their normal cv and their test statistic 3rd A1 for correct comment in context. Must mention "scores" and "students / groups/classes" Award A0 for a one-tailed comment.	
5(b)	1st B1 Alternative hyp should be clearly defined	

Question Number	Scheme	Marks
6(a)	Mean= $\frac{1\times16+2\times20+...+6\times8}{100}=2.91$ **ag**	M1A1 **(2)**
6(b)	$p=\frac{2.91}{6}=0.485$	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)**
6(c)	H_0 : Binomial is a good fit	
	H_1 : Binomial is a not a good fit	B1

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

M1

$\sum\frac{(O-E)^2}{E}=\frac{(22-12.41)^2}{12.41}+\frac{(20-24.82)^2}{24.82}+...+\frac{(18-9.59)^2}{9.59}=18.998...$ 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

Notes

6(a) 1st M At least 2 correct terms on numerator and 100 for denominator.

6(b) 0.485 can be implied by at least 1 correct answer.

6(c) Accept awrt 2dp for final answers
Clear use of Binomial and x100 required for method.
Parameters in hyps award B0
1st M1 for combining either 0 and 1 or 5 and 6 or both. Require at least 1 value in a combined correct.
2nd M1 for attempting $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$, at least 2 correct expressions or values.

2nd A1 for a correct comment suggesting that Binomial model is not suitable. No ft
Condone parameters here.

Question Number	Scheme	Marks
7(a)	M : N$(177, 25)$, F : N$(163,16)$	
	$\mathrm{E}(M-F)=177-163=14$	B1
	$\mathrm{Var}(M-F)=25+16=41$	M1A1
	$M-F$: 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)	$W=M_1+M_2+...M_6+F_1+F_2+...F_4$	
	$\mathrm{E}(W)=6\times177+4\times163$	
	$=1714$	B1
	$\mathrm{Var}(W)=6\times25+4\times16$	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
7(a)and (b)	Notes Condone reversed sds for method in (b) Accept metres: 2.14 award M1A0 in metres. 2nd M1s for identifying a correct probability and attempting to standardise with their mean and sd. Require explicit sd or accept 1156 for M1A0. This can be implied by the correct answer.	

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467
Fax 01623 450481
Email publication.orders@edexcel.com

Order Code UA033146 Summer 2012

For more information on Edexcel qualifications, please visit our website
www.edexcel.com

Pearson Education Limited. Registered company number 872828
with its registered office at Edinburgh Gate, Harlow, Essex CM20 2JE

Ofqual

Llywodraeth Cynulliad Cymru
Welsh Assembly Government

Rewarding Learning

edexcel

Mark Scheme (Results)

Summer 2013

GCE Statistics 3 (6691/01R)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please visit our website at www.edexcel.com.

Our website subject pages hold useful resources, support material and live feeds from our subject advisors giving you access to a portal of information. If you have any subject specific questions about this specification that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

www.edexcel.com/contactus

Pearson: helping people progress, everywhere

Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2013
Publications Code UA037014

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75.

2. The Edexcel Mathematics mark schemes use the following types of marks:

- **M** marks: method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
- **A** marks: accuracy marks can only be awarded if the relevant method (M) marks have been earned.
- **B** marks are unconditional accuracy marks (independent of M marks)
- Marks should not be subdivided.

3. Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes:

- bod – benefit of doubt
- ft – follow through
- the symbol $\surd$ will be used for correct ft
- cao – correct answer only
- cso - correct solution only. There must be no errors in this part of the question to obtain this mark
- isw – ignore subsequent working
- awrt – answers which round to
- SC: special case
- oe – or equivalent (and appropriate)
- dep – dependent
- indep – independent
- dp decimal places
- sf significant figures
- $*$ The answer is printed on the paper
- ⌐ The second mark is dependent on gaining the first mark ⌙

4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.

6. If a candidate makes more than one attempt at any question:
   - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
   - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.

7. Ignore wrong working or incorrect statements following a correct answer.

8. In some instances, the mark distributions (e.g. M1, B1 and A1) printed on the candidate's response may differ from the final mark scheme.

Question Number	Scheme	Marks
1.	Label females 1 – 100 (or 0 – 99) and males 1 – 300 (or 0 – 299) Using *random numbers* for each group in range 1 – 100 (0 – 99) select 15 females and using 1 – 300 (or 0 – 299) select 45 males	B1 B1 B1 **[Total 3]**
	Notes	
	1st B1 for labelling\numbering\listing females and males 2nd B1 for use of random numbers or "randomly select" in each group (may be implied) 3rd B1 for selecting the correct number of females and males e.g. randomly select 45 males and 15 females scores 2nd and 3rd B marks since randomly selecting and the "each group" is implied, If using systematic sampling within each strata allow 1st B1 and 3rd B1 (if earned) but 2nd B0	

Question Number	Scheme	Marks
2.	$X \sim N(40, 3^2)$ $\quad \overline{X} \sim N\left(40, \frac{9}{n}\right)$ (Condone $Y \sim N\left(40, \frac{9}{n}\right)$)	B1
	$P(\overline{X} > 42) = P\left(Z > \frac{42-40}{\sqrt{\frac{9}{n}}}\right)$	M1
	$\frac{42-40}{\sqrt{\frac{9}{n}}} \geq 1.6449$	B1 dM1
	$n \geq 6.087$ $n = 7$	A1 **[Total 5]**
	1st B1 for stating the correct distribution for $\overline{X}$. May be implied if correctly used in line 2 and no incorrect version seen elsewhere. 1st M1 for an attempt to standardise with 42, 40 and their $\sqrt{\frac{9}{n}}$, must have n. Allow $\pm$ 2nd B1 for using $z = \pm 1.6449$ (or better) 2nd dM1 Dep on 1st M1 for forming an equation in n or $\sqrt{n}$. Allow "=" or "< " i.e. setting their standardised expression = their z value ($\lvert z \rvert > 1.5$) A1 for $n = 7$ only The A1 must follow from correct working so e.g. $n < 6.087$ leading to $n = 7$ is A0	

Question Number	Scheme	Marks
3 (a)		

Town	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Pop	1	2	3	4	5	6	7	8	9	10
Empl	2	1	3	5	4	6	10	8	9	7
$\lvert d \rvert$	1	1	0	1	1	0	3	0	0	3
d^2	1	1	0	1	1	0	9	0	0	9

M1

$\sum d^2 = 22$ M1A1

$r_s = 1 - \dfrac{6 \times 22}{10 \times 99}$ dM1

$= \dfrac{143}{165} = 0.866\dot{}$ awrt **0.867** A1 (5)

(b) $H_0: \rho = 0$; $H_1: \rho > 0$ B1

CV = 0.6485 B1

in critical region / significant/ reject H_0 M1

evidence of positive correlation between population and no. of employees A1 (4)

(c) CV = 0.6319 B1

[not in critical region / not significant/ do not reject H_0]

No evidence of positive correlation B1 (2)

(d) No evidence to suggest that as pop' increased the no. of employees increased linearly. B1

Villages ranked highly for pop' were also ranked highly for the no. of employees. B1

(2)

ALT Alternate for part (d) if different conclusions in part (b) and part (c)

Data probably not (bivariate) normal therefore Spearman's coefficient is more suitable than the product moment correlation coefficient.

[Total 13]

Notes

(a)

1st M1 for an attempt to rank no of employees against the populations

2nd M1 for attempting $\sum d^2$ (must be using ranks) ft their ranks

1st A1 for 22

3rd dM1 dep on 1st M1 for use of the correct formula with their $\sum d^2$.

If ans. is not correct an expr' is required.

(b)

1st B1 for both hypotheses in terms of ρ, H_1 must be one tail and compatible with their ranking

M1 for a correct statement relating their r_s ($|r_s| < 1$)with their cv but cv must be such that |cv|<1

A1 for a correct contextualised comment that is rejecting H_0

Must mention "population" and "no. of employees" and "positive correlation".

Follow through their r_s and their cv (provided it is |cv| <1 Use of "association" is A0

(c)

1st B1 for 0.6319 2nd B1 does not require context just no positive correlation mentioned

(d)

1st B1 for a comment relating to pmcc

(i) no linear relationship or (ii) pmcc requires (joint) normal distribution

2nd B1 for a second comment relating to Spearman's

(i) there is a (non-linear) relationship between ranks..

or (ii) data not (joint) normal so Spearman's is better

Question Number	Scheme	Marks
4 (a)	$\frac{282\times100}{600}$ (Do not accept 282 – 114.2 – 90.2 – 30.6 (o.e.))	B1 **(1)**
(b)	9	B1 **(1)**
(c)	2.5 or better (Do not accept 0.025)	B1 **(1)**
(d)	H_0: hair colour occurs in the ratio 2:6:1:3 H_1: hair colour does not occur in the ratio 2:6:1:3	B1

	black	brown	red	blonde
observed	105	282	48	165
expected	100	300	50	150
$\frac{(O_i-E_i)^2}{E_i}$	0.25	1.08	0.08	1.5
$\frac{O_i^2}{E_i}$	110.25	265.08	46.08	181.5

B1expected
M1
A1

$\sum \frac{(O_i-E_i)^2}{E_i} = 2.91$ or $\sum \frac{O_I^2}{E_I} - 600 = 602.91 - 600 = 2.91$ (awrt **2.91**) A1

$\nu = 3$ B1

cv is 7.815 B1

[2.91< 7.815] so insufficient evidence to reject H_0 or not significant dM1

There is evidence to suggest that hair colour does occur in the given ratio. A1

(9)
[Total 12]

Notes

(d)

1st B1 for both hypotheses. Must mention hair colour and ratio
e.g. "hair colour in the given ratio" Allow use of ditto

2nd B1 for all 4 correct expected frequencies

1st M1 for at least 2 correct calculations from 3rd or 4th row

1st A1 for all correct calculations to at least 3sf if row 4
If awrt 2.91 is seen with no incorrect working award B1M1A1A1

2nd dM1 Dep on 1st M1 for a correct statement linking their test statistic and their cv (cv > 3.5)

3rd A1 for a correct comment in context - must mention "hair colour" and "ratios" or "model"
e.g. "There is evidence of to support the given model" No follow through
If hypotheses are the wrong way round score A0.

Question Number	Scheme	Marks
5 (a)	$\bar{x} = \frac{1}{2}(118.8 + 121.2) = 120$	B1
	1.6449 (or better)	B1
	" their 1.6449" $\frac{\sigma}{\sqrt{n}} = 121.2 - 120$	M1
	2.3263 (or better)	B1
	" their 2.3263" $\frac{\sigma}{\sqrt{n}} = 2.3263 \times \left(\frac{121.2 - 120}{1.6449}\right)$	dM1
	So 98% C.I. = $120 \pm 1.424\ldots$ = (118.3028…, 121.699…)	
	awrt **(118, 122)**	A1
		(6)
(b)	awrt $(118\pi, 122\pi)$ or (371/372, 382/383)	B1ft
		(1)
(c)	P (All) $= (0.98)^3$	M1
	$= 0.941$	A1
		(2)
		[Total 9]

Notes

(a) NB in part (a) only lose one of the B1 marks for not using the percentage points table

1st B1 for $\bar{x} = 120$

2nd B1 for 1.6449 or better in an attempt (could be $1.6449\sigma = k$ or even $1.6449\ \sigma^2 = k$)
Condone strange notation for standard error (E) here if it is used correctly

1st M1 for an attempt to find "width" or "half-width" of a 90% CI ft their z value ($|z| > 1.5$)
e.g. for $z\ E = 121.2 - 120$ (o.e.) N.B. $E = 0.7295...$ Condone missing 2 here.

3rd B1 for 2.3263 or better in an attempt at CI.
If score 2nd B0 for using 1.64 or 1.645 allow 3rd B1 for 2.32 or 2.33 here

2nd dM1 for a correct attempt at "width" or "half-width" of a 98% CI ft their z value ($|z| > 2$)
Dependent on 1st M1 and ft their value or expression for s.e.

A1 for lower limit in range [118, 118.35) and upper limit in range (121.65, 122]

Answer only of awrt (118, 122) with no incorrect working seen scores 6/6/ if 1.6449 and 2.3263 are seen and 5/6 (B1B1M1B0M1A1) otherwise.

(c) M1 for a correct expression i.e. $(0.98)^3$

A1 for awrt 0.941

Question Number	Scheme	Marks
6 (a)	Var $(X) = \frac{(a+5-a+1)^2}{12}$ [=3]	M1
	$\overline{X} \sim \text{N}\left(a+2, \frac{3}{50}\right)$	A1, A1ft
		(3)
(b)	$17.2 - 1.96\times\sqrt{\frac{3}{50}} < \mu < 17.2 + 1.96\times\sqrt{\frac{3}{50}}$	B1
		M1
	$17.2 - 1.96\times\sqrt{\frac{3}{50}} < a+2 < 17.2 + 1.96\times\sqrt{\frac{3}{50}}$	B1
	$14.7 < a < 15.7$	A1
		(4)
		[Total 7]
	Notes	
(a)	M1 for a correct expression for Var(X) in terms of a or Var(X) = 3	
	1st A1 for normal and correct mean must be $a + 2$	
	NB N(17.2, ...) is A0 and N(17.2, $\frac{3}{50}$) is M1A0A1	
	2nd A1ft for correct Var($\overline{X}$), i.e. (their "3")/50	
(b)	1st B1 for correct use of $z = 1.96$ in an attempt e.g. $\overline{x} \pm z\sigma$ or $\overline{x} \pm z\sigma^2$	
	M1 for $17.2 \pm z \times \sqrt{\frac{"3"}{50}}$ where $\lvert z\rvert > 1.5$ accept just + or just −	
	Answer of (16.7, 17.7) scores B1M1B0A0	
	2nd B1 for either of the inequalities with $a + 2$ and any z ($\lvert z\rvert > 1.5$) or $a = 15.2 \pm z \times \sqrt{\frac{"3"}{50}}$	
	A1 for awrt 14.7 and 15.7	

Question Number	Scheme	Marks
7 **(a)**	$H_0: \mu_a = \mu_b$, $H_1: \mu_a < \mu_b$	B1
	s.e. $= \sqrt{\frac{25^2}{100} + \frac{10^2}{150}}$, $z = \frac{67-60}{\sqrt{\frac{25^2}{100} + \frac{10^2}{150}}}$ CR $= 1.6449 \times \sqrt{\frac{25^2}{100} + \frac{10^2}{150}}$	M1,dM1
	$z = \pm 2.6616....$ $= \pm 4.326...$ (awrt **2.66/4.33**)	A1
	One tailed critical value $z = 1.6449$ (or prob of awrt 0.004 (<0.05))	B1
	[Condone 0.996 if compared correctly with 0.95 for the B1]	
	2.6616 > 1.6449 so] significant evidence to reject H_0	dM1
	There is evidence that the amount of lead present in the soil has decreased.	A1ft
		(7)
(b)	CLT enables you to assume that means are normally distributed	B1
		(1)
(c)	Have assumed $s^2 = \sigma^2$ or variance of sample = variance of population	B1
		(1)
		[Total 9]
	Notes	
(a)	1st B1 for both hypotheses in terms of μ not words.	
	Accept μ_1, μ_2 etc if there is some indication of which is which e.g $X \sim N(67, 25^2)$ implies X is "before".	
	1st M1 for attempt at s.e. - condone one number wrong or mis-matched variances	
	i.e. $\sqrt{\frac{p}{q} + \frac{r}{s}}$ (3 of p,q,r& s correct) or $\sqrt{\frac{10^2}{100} + \frac{25^2}{150}}$	
	2nd dM1 Dep on 1st M1 for using their s.e. in correct formula for test statistic. Num of $\pm (67 - 60)$	
	or for correct expression for CR	
	3rd dM1 dep. on 2nd M1 for a correct statement based on their normal cv (|cv| > 1.5) and their test statistic	
	2nd A1ft for correct comment in context. Must mention "lead" or "soil" and "factory". Allow ft	
	If hypotheses are the wrong way round score A0	
	If hypotheses are not for a difference between 2 means award A0	
(b)	B1 must mention mean and normal. In words or symbols e.g. $\bar{X} \sim N(...$	

Question Number	Scheme	Marks
8 (a)	Let $W = D_1 - D_2$	M1
	$W \sim$ N(0, 2.88)	A1, A1
	P($\lvert W \rvert$>3) = 2 × P(W >3)	M1
	$= 2 \times \text{P}\left(Z > \frac{3-0}{\sqrt{2.88}} \right)$	dM1
	= 2 × P(Z > 1.76776….)	
	= 2 × (1 – 0.9616)	
	= 0.0768 awrt **0.077**	A1
		(6)
(b)	Let $T = 5C - 4D$ or $4D - 5C$ or $C - \frac{4}{5}D$ or $\frac{4}{5}D - C$	M1
	$T \sim$ N($\pm$ 4, 39.04) or N($\pm$ 0.8, 1.5616)	A1 A1
	$\text{P}(T < 0) = \text{P}\left(Z < \frac{0-4}{\sqrt{39.04}} \right)$ or $\text{P}\left(Z < \frac{0-0.8}{\sqrt{1.5616}} \right)$	M1
	= P(Z < -0.64018….)	
	= (1 – 0.7389)	
	= 0.2611 awrt **0.261**	A1
		(5)
(c)	Let $P = D + D + D + D + D + D + B$	M1
	Let $Q = C + C + C + C + C + C + B$	
	$P \sim$ N(352, 13.64) and $Q \sim$ N(292, 8.84)	A1, A1
	[Let $R = P - Q$] $R \sim$ N($\pm$ 60, 22.48)	M1
	$\text{P}(R > 50) = \text{P}\left(Z > \frac{50-60}{\sqrt{22.48}} \right)$	dM1
	= P($Z > -2.10$….)	
	= 0.9821 awrt **0.982 ~ 0.983**	A1
		(6)
		[Total 17]

Notes

Award full marks in each part for a correct answer with no incorrect working seen.

(a)
1st M1 for explicitly defining a suitable W and attempt to find the distribution of W.
May be implied by sight of N(0, 2.88)
1st A1 for normal and mean of 0, 2nd A1 for variance of 2.88. Award M1A1A1 for N(0, 2.88) seen.
2nd M1 for realising need $2 \times \text{P}(W > 3)$
3rd dM1 Dep on 1st M1 for standardising with 3, 0 and their s.d. Must lead to P(Z > +ve) (o.e.)

(b)
1st M1 for explicitly defining a suitable T but may be implied by sight of one of these normals
1st A1 for normal and correct mean, 2nd A1 for correct variance. Accept awrt 3sf i.e. 39.0, 1.56
2nd M1 for standardising with 0 and their mean and their s.d. Must lead to P(Z < -ve) (o.e.)

(c)
1st M1 for explicitly defining a correct P or Q. May be implied by a correct distribution for P or Q
1st A1 for a correct distribution for P 2nd A1 for a correct distribution for Q
2nd M1 for attempting R and obtaining its distribution- ft their P and Q means and variances
3rd dM1 for attempting P(R > 50) and standardising with 50 and their E(R) and their $\sqrt{\text{Var}(R)}$
Dependent on 2nd M1. Must lead to a P(Z > -ve) (o.e.)

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467
Fax 01623 450481
Email publication.orders@edexcel.com

Order Code UA037014 Summer 2013

For more information on Edexcel qualifications, please visit our website
www.edexcel.com

Pearson Education Limited. Registered company number 872828
with its registered office at Edinburgh Gate, Harlow, Essex CM20 2JE

Ofqual

Llywodraeth Cynulliad Cymru
Welsh Assembly Government

Rewarding Learning

edexcel

Mark Scheme (Results)

Summer 2013

GCE Statistics 3 (6691/01)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please visit our website at www.edexcel.com.

Our website subject pages hold useful resources, support material and live feeds from our subject advisors giving you access to a portal of information. If you have any subject specific questions about this specification that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

www.edexcel.com/contactus

Pearson: helping people progress, everywhere

Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2013
Publications Code UA037011

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75.

2. The Edexcel Mathematics mark schemes use the following types of marks:

- **M** marks: method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
- **A** marks: accuracy marks can only be awarded if the relevant method (M) marks have been earned.
- **B** marks are unconditional accuracy marks (independent of M marks)
- Marks should not be subdivided.

3. Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes:

- bod – benefit of doubt
- ft – follow through
- the symbol $\surd$ will be used for correct ft
- cao – correct answer only
- cso - correct solution only. There must be no errors in this part of the question to obtain this mark
- isw – ignore subsequent working
- awrt – answers which round to
- SC: special case
- oe – or equivalent (and appropriate)
- dep – dependent
- indep – independent
- dp decimal places
- sf significant figures
- $*$ The answer is printed on the paper
- ⌐ The second mark is dependent on gaining the first mark

4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.

6. If a candidate makes more than one attempt at any question:
   - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
   - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.

7. Ignore wrong working or incorrect statements following a correct answer.

8. In some instances, the mark distributions (e.g. M1, B1 and A1) printed on the candidate's response may differ from the final mark scheme.

Question Number	Scheme	Marks
1.		

Cholesterol Level	High	Low	
High	7.6	12.4	20
Low	30.4	49.6	80
	38	62	100

M1A1 (2)

H_0 : Cholesterol level is independent of intake of saturated fats(no association) B1

H_1 : Cholesterol level is not independent of intake of saturated fats (association) (1)

O	E	$\frac{(O-E)^2}{E}$	$\frac{O^2}{E}$
12	7.6	2.547… or $\frac{242}{95}$	18.947… or $\frac{360}{19}$
8	12.4	1.56129… or $\frac{242}{155}$	5.161… or $\frac{160}{31}$
26	30.4	0.6368… or $\frac{121}{190}$	22.236… or $\frac{845}{38}$
54	49.6	0.3903… or $\frac{121}{310}$	58.790… or $\frac{3645}{62}$

dM1 A1

$\sum\frac{(O-E)^2}{E}$=5.1358234.. or $\frac{1.2^2}{7.6}+\frac{8^2}{12.4}+\frac{26^2}{30.4}+\frac{54^2}{49.6}-100=5.14$ (awrt **5.14**) A1 (3)

$\nu=(2-1)(2-1)=1$ B1

$\chi^2_1(0.05)=3.841$ B1 (2)

5.14 > 3.841 so sufficient evidence to reject H_0 [Condone "accept H_1"] M1

Association between cholesterol level and saturated fat intake A1 (2)

Total 10

Notes

Minimum working use part marks: E_i (2), Hyp (1), 5.14 (3), 3.841 (2), Conclusion (2)

1st M1 for some use of $\frac{\text{Row Total}\times\text{Col.Total}}{\text{Grand Total}}$. May be implied by correct E_i

1st A1 for all expected frequencies correct. Allow M1A0 for E_i rounded to integers

1st B1 for both hypotheses. Must mention "cholesterol" and "fats" at least once
Use of "relationship" or "correlation" or "connection" is B0

2nd dM1 for at least 2 correct terms (as in 3rd or 4th column) or correct expressions with their E_i
Dependent on 1st M1 Accept 2sf accuracy for the M mark

2nd A1 for all correct terms. May be implied by a correct ans.(2 dp or better)
Allow truncation eg 2.54… 3rd A1 for awrt 5.14

2nd B1 for correct degrees of freedom (may be implied by a cv of 3.841)

3rd M1 for a correct statement linking their test statistic and their cv(cv could be 2.705 or > 3.5)
Contradictory statements score M0 e.g. "significant, do not reject H_0"

4th A1 for a correct comment in context - must mention "cholestorol" and "fats"
condone "relationship" or "connection" here but **not** "correlation".
e.g. "There is evidence of a relationship between cholesterol level and fat intake"
No follow through. If e.g hypotheses are the wrong way round A0 here.

Question Number	Scheme	Marks
2(a)	<table><tr><td>Uni</td><td>*A*</td><td>*B*</td><td>*C*</td><td>*D*</td><td>*E*</td><td>*F*</td><td>*G*</td><td></td></tr><tr><td>Staff-Stu</td><td>2</td><td>4</td><td>3</td><td>5</td><td>7</td><td>1</td><td>6</td><td></td></tr><tr><td>Satisfaction</td><td>3</td><td>2</td><td>6</td><td>4</td><td>5</td><td>1</td><td>7</td><td></td></tr><tr><td>[*d*]</td><td>–1</td><td>2</td><td>–3</td><td>1</td><td>2</td><td>0</td><td>–1</td><td></td></tr><tr><td>d^2</td><td>1</td><td>4</td><td>9</td><td>1</td><td>4</td><td>0</td><td>1</td><td>20</td></tr></table>	M1A1A1
	$r_s = 1 - \frac{6\times 20}{7(49-1)} = 0.642857....$ (accept $\frac{9}{14}$) (awrt **0.643)**	dM1A1
		(5)
(b)	H_0: $\rho = 0$	
	$H_{1:}$ $\rho \neq 0$ ($\rho > 0$)	B1
	Critical value is ± 0.7857(± 0.7143 for a one tailed test)	B1
	0.643<cv so insufficient evidence to reject H_0	
	There is insufficient evidence to suggest a (positive) correlation between staff-student ratio and satisfaction.	B1ft
		(3)
		Total 8

	Notes
(a)	1st M1 for an attempt to rank the staff-students ratio or satisfaction (at least 4 correct) 1st A1 for correct rankings for both (one or both may be reversed) 2nd A1 for $\sum d^2$ =20 or correct d^2 row (NB $\sum d^2 = 92$ for one set of reversed ranks) 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. 3rd A1 If $\sum d^2$ = 20 for awrt 0.643 or if $\sum d^2$ =92 for awrt – 0.643 (accept $\pm\frac{9}{14}$)
(b)	1st B1 for both hypotheses in terms of ρ, 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 0.7857 or 0.7143 for one-tailed test (accept +) Their cv must be compatible with their H_1 which may be in words If hypotheses are the wrong way around this must be B0 but 3rd B1 is possible. 3rd B1ft for a correct contextualised comment. Must mention "ratio" or "no. of students per member of staff" and "satisfaction" Follow through their r_s and their cv (provided it is |cv| <1) Don't insist on the word "positive" for a one-tailed test Use of "association" is B0 Independent of 1st B1 so if $\|r_s\|$ > |cv| must say there is sufficient evidence of …….(o.e.) and if $\|r_s\|$ < |cv| must say insufficient evidence of … (o.e.) regardless of their hypotheses Contradictory statements score B0 (This mark is just testing interpretation of comparison of their r_s and their cv)

Question Number	Scheme	Marks
3(a)i e.g e.g	**Quota Sampling:** Advantages: Fieldwork can be done **quick**ly, or administering the test is **easy**, or costs are kept to a minimum (**cheap**), or gives estimates for each course. or OK for large populations or sampling frame not required (o.e.) Disadvantages: **Non-random** process or not possible to estimate the sampling errors, or non response not recorded, or interviewer can introduce **bias** in sample choice. (o.e.)	B1
3(a)ii e.g. e.g.	**Stratified Sampling:** Advantages: Can give accurate estimates as it is a **random** process, or gives estimates for each course or **representative** of [BUT not "proportional" to] the whole population. (o.e.) Disadvantages: Sampling frame required, or strata may not be clear as some students overlap courses or not suitable for large populations. (o.e.)	B1 (2)
3(b)	Total enrolments=1000 (may be implied by calculations) Leisure and Sport= $\frac{420}{1000}\times 100 = 42$ Information Technology= $\frac{337}{1000}\times 100 = 33.7 = 34$ Health and Social Care= $\frac{200}{1000}\times 100 = 20$ Media Studies= $\frac{43}{1000}\times 100 = 4.3 = 4$	B1 M1 A1 (3)
3(c)	The college's information system would be used to identify each student and which course they are enrolled on. i.e. idea of **sampling frame** or **list** for **each course**. Use of **random numbers** to select required number of students **from each course**	B1 B1 (2) **Total 7**
	Notes	
(a)	Do not penalise for lack of context in part (a) 1st B1 for an advantage and a disadvantage for quota sampling (must be 1st or labelled (i)) 2nd B1 for an advantage and a disadvantage for stratified sampling (2nd or labelled (ii)) Do not allow opposite pairs e.g. "quicker/easier" for quota sampling and "takes a long time/more difficult" for stratified or quota "easy to use" but strat. "hard for large populations" Do not allow same reason for both e.g. "gives estimates for each course"	
(b)	M1 for one correct calculation, ft their "1000" A1 for 42, 34, 20 and 4 only	
(c)	1st B1 for some mention of a suitable sampling frame. Need not give the specific term but a suitable source of list is required for all students in each course. 2nd B1 for mentioning use of random numbers or some random selection process for each course. If they are describing systematic sampling score B0 here	

Question Number	Scheme	Marks
4 (a)	$\bar{x} = \dfrac{8\times1.5+12\times4+13\times5.5+9\times7+8\times10}{50} = \dfrac{274.5}{50} = 5.49$ **(*)**	B1cso
	$s^2 = \dfrac{8\times1.5^2+12\times4^2+13\times5.5^2+9\times7^2+8\times10^2}{49} - \dfrac{50}{49}5.49^2, = 6.88$ **(*)**	M1,A1cso
		(3)
(b)	$a = \quad 50\times \mathrm{P}(6 < X < 8) = 50\times \mathrm{P}(0.194.. < Z < 0.956..)$	M1
	a = 12.81 (tables) <u>or</u> 12.68 (calc)	A1
	$b = 50 - (28.85 + a)$ = 8.34 (tables) <u>or</u> 8.47 (calc)	A1ft
		(3)
(c)	H_0: Normal distribution is a good fit H_1: Normal distribution is not a good fit	B1

Class	O	E	$\dfrac{O^2}{E}$	$\dfrac{(O-E)^2}{E}$
0-3	8	8.56	7.4766…	0.0366…
3-5	12	12.73	11.31186….	0.0418…
5-6	13	7.56	22.354497…	3.9144…
6-8	9	12.68 or (12.81)	(6.32) ~ 6.38801..	1.0680...~ (1.13)
8-12	8	(8.34) or 8.47	7.556080…~ (7.67)	(0.013) ~ 0.0260..

M1 A1

$\sum \dfrac{O^2}{E} - N = 5.087\ldots \sim 5.1400\ldots$ awrt **(5.09 ~ 5.14)**	A1
$\nu = 5-3 = 2$ (for 5 – 3 or 2 can be implied by 5.991 seen)	B1
$\chi^2_2(0.05) = 5.991$	B1
5.09<5.991 so insufficient evidence to reject H_0	M1
Normal distribution is a good fit.	A1 (8)
	Total 14

Notes

(a)

B1cso for denominator of 50 and at least 3 products on num <u>or</u> 274.5 on num

M1 for a correct expression with at least 3 correct products on num or $\dfrac{1844.25}{49} - \dfrac{1507.005}{49}$

<u>or</u> $\dfrac{337.245}{49}$ <u>or</u> $\left(\dfrac{7377}{200} - 5.49^2\right)\times\dfrac{50}{49}$ etc Allow 3sf accuracy

A1cso for 6.88 with M1 scored and no incorrect working seen

(b)

M1 a full method for a or b using the normal dist. Correct use of (6), 8, 5.49 and $\sqrt{6.88}$ seen

1st A1 for a in range 12.68 ~ 12.81 or b in range 8.34~ 8.47 or awrt these values

2nd A1ft for 50 – 28.85 – their a (or b) (but requires M1). Allow awrt 3sf. Must add up to 50

(c)

1st B1 for both hypotheses. B0 if they include 5.49 or 6.88. Condone X~N(μ, σ^2) etc

1st M1 for attempting $\dfrac{(O-E)^2}{E}$ or $\dfrac{O^2}{E}$, at least 3 correct expressions or values.

1st A1 for at least 4 correct calcs - 3rd or 4th column. (2 dp or better and allow e.g. 7.47) Allow any value in the ranges for the last two rows.

2nd A1 for a test statistic that is awrt 5.09 ~ 5.14. Award M1A1A1 if this is obtained.

2nd M1 for a correct statement based on their test statistic (> 1) and their cv (> 3.8) Contradictory statements score M0 e.g. "significant" do not reject H_0.

3rd A1 for a correct comment suggesting that normal model is suitable <u>or</u> manager's belief is correct. **No f t** . Condone mention of 5.49 or 6.88 here. Hypotheses wrong way round scores A0

Question Number	Scheme	Marks
5 (a)	Let $L \sim \mathrm{N}(50, 25)$ and $S \sim \mathrm{N}(15, 9)$	
	Let $X = L - (S_1 + S_2 + S_3)$	B1
	$\mathrm{E}(X) = 50 - 3\times 15 = 5$	B1
	$\mathrm{Var}(X) = 25 + 3\times 9 = 52$	M1A1
	$\mathrm{P}(X < 0) = \mathrm{P}\left(Z < \frac{-5}{\sqrt{52}}\right)$	dM1
	$= \mathrm{P}(Z < -0.693..)$	
	=0.244 or 0.2451 (tables) (awrt **0.244 ~ 0.245)**	A1
		(6)
(b)	Let $Y = L - 3S$	B1
	$\mathrm{E}(Y) = 50 - 3\times 15 = 5$	B1
	$\mathrm{Var}(Y) = 25 + 3^2 \times 9 = 106$	M1A1
	$\mathrm{P}(Y > 0) = \mathrm{P}\left(Z > \frac{-5}{\sqrt{106}}\right)$	dM1
	$= \mathrm{P}(Z > -0.4856..)$	
	=0.686 or 0.6879 (tables) (awrt **0.686 ~ 0.688)**	A1
		(6)
		Total 12

	Notes
(a)	1st B1 for forming a suitable variable X explicitly seen. Do not give for $L - 3S$ but allow $L - (S + S + S)$ 2nd B1 for E(X) = 5 (or – 5 if their X is defined the other way around) 1st M1 for an attempt at Var(X) = Var(L) + 3Var(S). Do not condone 5 for "25" or 3 for "9" 1st A1 for 52 2nd dM1 for attempting the correct probability and standardising with their mean and sd. This mark is dependent on 1st M1 so if X is not being used or wrong variance score M0 If their method is not crystal clear then they must be attempting P(Z < -ve value) or P(Z > +ve value) i.e. their probability after standardisation should lead to a prob. < 0.5 2nd A1 for awrt 0.244 ~ 0.245 Correct ans. only scores 5/6 (or 6/6 if 1st B1) but must be clearly labelled as (a) or the first answer.
(b)	1st B1 for defining a new variable [Y =]± (L – 3S). May be implied by a correct variance. 2nd B1 for E(Y) = 5 (or – 5 if their Y is defined as $Y = 3S - L$) 1st M1 for an attempt at Var(Y) = Var(L) + 3^2 Var(S). Do not condone 5 for "25" or 3 for "9" 1st A1 for 106 only 2nd dM1 for attempting the correct probability and standardising with their mean and sd. This mark is dependent on 1st M1 so if Y is not being used or wrong variance score M0 If their method is not crystal clear then they must be attempting P(Z > -ve value) or P(Z < +ve value) i.e. their probability after standardisation should lead to a prob. > 0.5 2nd A1 for an awrt 0.686 ~ 0.688 Correct answer only scores 6/6 but must be clearly labelled as (b) or the second answer.

Question Number	Scheme	Marks
6 (a)	$H_0 : \mu_{new} - \mu_{old} = 1$	B1
	$H_1 : \mu_{new} - \mu_{old} > 1$	B1
	$z = \dfrac{7 - 5.5 - 1}{\sqrt{\dfrac{0.5}{60} + \dfrac{0.75}{70}}} = 3.62254...$ (awrt **3.62**)	M1 A1A1 A1
	Critical value $z = 1.6449$ (allow $\pm$)	B1
	[3.62 >1.6449] so sufficient evidence to reject H_0	dM1
	Evidence that the mean yield of new variety is more than 1 kg greater than the old variety.	A1
		(9)
(b)	Mean yield is normally distributed	B1
	Sample size is large. Must state or imply that **in this case** sample size is large	B1
		(2)
		Total 11

	Notes
(a)	1st & 2nd B1 for hypotheses. Accept μ_1, μ_2 or μ_A, μ_B etc if there is some indication of which is which e.g. $A \sim N(\mu_A, 0.5)$
	1st M1 for an attempt at se. Condone switching 0.5 and 0.75 $\sqrt{\dfrac{0.5 \text{ or } 0.75}{60} + \dfrac{0.75 \text{ or } 0.5}{70}}$
	1st A1 for a correct expression for denominator of test statistic or 0.138… or $\sqrt{0.0190...}$
	2nd A1 for a correct numerator of test statistic (must have the -1)
	3rd A1 for awrt 3.62
	[Allow – 3.62 from numerator of 5.5 – 7 − − 1 and compatible H_1]
	3rd B1 for $\pm$ 1.6449 seen or
	probability of 0.0002 (tables) or 0.000145…(calc) [allow 0.0001]
	2nd dM1 dep. on 1st M1 for a correct statement based on their normal cv and their test statistic
	2nd A1 for correct comment in context. Must mention "yield" and "varieties" or "old" and "new" and "1"
	If second B mark is B0 award A0 here
ALT	**Pooled estimate:** If they calculate $s_p = \sqrt{0.41845...} = 0.64688...$ allow 1st M1, 1st A1 for expression (or awrt 0.114) and 2nd A1 if numerator correct but A0 for test statistic (4.39)
(b)	1st B1 for mention of mean (yield) and normal (distribution)
	2nd B1 for mention of sample (size) being large in this case

Question Number		Scheme	Marks
7	**(a)**	$\hat{\mu} = \bar{x} = \frac{33.29}{8} = 4.16125$ (awrt **4.16**)	B1
		$\hat{\sigma}^2 = s^2 = \frac{4.12^2 + 5.12^2 + ... - 8 \times \bar{x}^2}{7}$	M1
		$\hat{\sigma}^2 = s^2 = \frac{141.4035 - 138.528013}{7} = 0.41078...$ (awrt **0.411**)	A1
			(3)
	(b)	$\sum x = 33.29 + 32 \times 4.55 = 178.89,$ (awrt **179**)	B1
		$\sum x^2 = "141.4035" + 31 \times 0.25 + 32 \times 4.55^2 (= 811.6335)$ (awrt **812**)	M1A1
		Combined sample: $s^2 = \frac{811.6335 - \frac{178.89^2}{40}}{39} = 0.29724865...$ (awrt **0.297**)	M1A1
		$\frac{s}{\sqrt{n}} = \frac{\sqrt{0.297....}}{\sqrt{40}} = 0.0862$ (awrt **0.0862)**	M1A1
			(7)
	(c)	$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = \frac{178.89}{40} \pm 1.96\frac{0.67}{\sqrt{40}}$	M1B1
		= (4.2646…, 4.67988…) awrt (**4.26**[or 4.265]**, 4.68**)	A1
			(3)
			Total 13

	Notes
(a)	M1 for an attempt at s^2: correct denom, clear attempt at $\sum x^2$ and ft their $\bar{x}$ Ans only 2/2
(b)	B1 for correct sum or mean or fully correct expression (accept mean = awrt 4.47) **May be in (c)** 1st M1 for their 141.4035 + $31 \times 0.25 + 32 \times 4.55^2$ or "141.4035" + 7.75+ 662.48 (accept 3sf) **Beware:** 32(0.25 + 4.55^2) + "141.4035" = awrt 812 but scores M0A0. 1st A1 for a fully correct expression (all to 3sf or better) or answer only = awrt 812 2nd M1 for a correct expression using their values 3rd M1 dependent on using a changed s^2 (not their 0.411 or 0.25) for $\frac{\sqrt{"0.297"}}{\sqrt{40}}$ This s^2 must be based on a combination of their 0.411 and 0.25 e.g. 0.661
(c)	M1 for $\bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}$ for any z (> 1.5) and ft their $\bar{x}$ based on combining their 4.16 and 4.55, do not award for simply using 4.55 or their 4.16. Condone $\sigma = \sqrt{\text{their } 0.297}$ or their (b) B1 for z = 1.96 used in an attempt at a CI, may for example miss $\sqrt{n}$ A1 for both limits awrt 3sf. Allow lower limit of 4.265

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467
Fax 01623 450481
Email publication.orders@edexcel.com
Order Code UA037011 Summer 2013

For more information on Edexcel qualifications, please visit our website
www.edexcel.com

Ofqual

Llywodraeth Cynulliad Cymru
Welsh Assembly Government

Rewarding Learning

Mark Scheme (Results)

Summer 2014

Pearson Edexcel International A Level in Statistics 3 (WST03/01)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please visit our website at www.edexcel.com.

Our website subject pages hold useful resources, support material and live feeds from our subject advisors giving you access to a portal of information. If you have any subject specific questions about this specification that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

www.edexcel.com/contactus

Pearson: helping people progress, everywhere

Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2014
Publications Code IA040147

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

EDEXCEL IAL MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75.

2. The Edexcel Mathematics mark schemes use the following types of marks:

- **M** marks: method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
- **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
- **B** marks are unconditional accuracy marks (independent of M marks)
- Marks should not be subdivided.

3. Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes.

- bod – benefit of doubt
- ft – follow through
- the symbol $\surd$ will be used for correct ft
- cao – correct answer only
- cso - correct solution only. There must be no errors in this part of the question to obtain this mark
- isw – ignore subsequent working
- awrt – answers which round to
- SC: special case
- oe – or equivalent (and appropriate)
- dep – dependent
- indep – independent
- dp decimal places
- sf significant figures
- $*$ The answer is printed on the paper
- $\lfloor$ The second mark is dependent on gaining the first mark

4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.

6. Ignore wrong working or incorrect statements following a correct answer.

Question Number	Scheme	Marks
1. (a)	165, 8	B1 **[1]**
(b)	Select <u>every 6th person</u> {having chosen the first person by} Selecting a random number between 1 and 6 **or** selecting a random number and then loop back to start when you reach the end.	B1 dB1
(b)(ii)	The <u>list</u> is alphabetical and <u>has not been **sorted** by gender</u>.	B1 **[3]**
(c)	Label male members 1- 180, female members 1 – 120 <u>Use random numbers</u> to select a ... Simple random sample of <u>30 male</u> members and <u>20 female</u> members	M1 M1 A1 **[3]**
(d)	**Any one of** • It (a stratified sample) is <u>not biased</u> as the members are chosen randomly. • You <u>can estimate</u> the <u>sampling errors</u> (for a stratified sample) • It (a stratrified sample) gives <u>more accurate estimates</u> as it is a random process. • A quota sample may <u>be biased</u> (whereas a stratrified sample is not). • It's <u>not possible</u> to <u>estimate/find</u> the <u>sampling errors</u> for a <u>quota sample</u> (whereas you can for a stratified sample)	B1 **[1]** **8**

	Notes	
(a)	**B1**	165 followed by 8 or 008.
(b)(i)	**1st B1**	For selecting every 6th (name on the list)
	2nd dB1	***is dependent on the first B1 mark being awarded.*** For idea of using random numbers to select first from 1 to 6 or 0 to 5 (o.e.) **or** selecting a random number between 1 and 300 and then looping back when the end of the list has been reached.
(b)(ii)	**B1**	A comment that implies <u>the list</u> (or sampling frame) has **not been sorted** by gender.
	Note	B0 for "the ordered list is not truly random"
	Note	B0 for "sample does not divide the members into gender."
(c)	**1st M1**	For suitable labelling of all 180 males <u>and</u> all 120 females. E.g. Allow labelling female members 181 – 300. Also allow labelling male members 0 – 179 and female members either 0 to 119 or 180 to 299.
	2nd M1	For use of random numbers to select males and females.
	A1	For 30 males <u>and</u> 20 females **(dependent on 2nd M1 only)**
	Note	A simple random sample of 30 males and 20 females scores 2nd M1 and A1.
(d)	**Note**	B0 for "a stratified sample can reflect the population structure." B0 for "estimates obtained from each of the strata."

Question Number	Scheme	Marks
2.	X follows a continuous unform distribution over $[\alpha - 3,\ 2\alpha + 3]$	
(a)	$\left\{\mathrm{E}(\bar{X}) = \mu =\right\}\ \dfrac{2\alpha + 3 + \alpha - 3}{2}$	M1
	$= \dfrac{3\alpha}{2}$. So $\bar{X}$ is a biased estimator.	A1
	bias $\left\{= \dfrac{3\alpha}{2} - \alpha\right\} = \pm\dfrac{\alpha}{2}$ bias $= \pm\dfrac{\alpha}{2}$	B1
		[3]
(b)	$k = \dfrac{2}{3}$ $\dfrac{2}{3}$	B1
		[1]
(c)	$\alpha = \dfrac{2}{3}\bar{X} = \dfrac{2}{3}(8)$ "their k" $\times 8$	M1
	Max value $= 2\left(\dfrac{16}{3}\right) + 3$ $2 \times$ "their α" $+ 3$ See notes	M1
	$= \dfrac{41}{3}$ $\dfrac{41}{3}$ or $13\dfrac{2}{3}$ or awrt 13.7	A1
		[3]
		7

Notes

(a)

M1 Using the formula $\left(\dfrac{a+b}{2}\right)$ or getting $\dfrac{3\alpha}{2}$

A1 $\dfrac{3\alpha}{2}$ and concluding. Allow A1 for $\dfrac{3\alpha}{2} \neq \alpha$.

Note Also allow A1 for bias $= \pm\dfrac{\alpha}{2} \neq 0$

(c)

1st M1 An attempt to use the sample data given to find $\bar{x}$ and multiply by their k. Allow full expression for $\bar{x}$ or $\dfrac{\sum x}{n}$.

Note 1st M1 can be implied by a correct recovery leading to $\alpha = \dfrac{16}{3}$

2nd M1 $2 \times$ "their α" $+ 3$ where their α is a function of the sample mean - which found by applying $\dfrac{\sum x}{n}$ from the data values given in the question.

Note $2(13) + 3 = 39$ is M0M0A0

Question Number	Scheme	Marks
3. (a)	$H_0 : \mu_A = \mu_B \quad H_1 : \mu_A > \mu_B$	B1
	s.e. $= \sqrt{\frac{35^2}{80} + \frac{28^2}{100}}$ $\{= 4.81170448...\}$	M1 A1
	$z = \frac{532 - 520}{"4.8117..."}$; $= 2.4939...$ $\frac{\pm(532-520)}{"4.8117..."}$	dM1;
	awrt 2.49	A1
	One tailed c.v. $Z = 2.3263$ or CR: $Z \geqslant 2.3263$ or p-value = awrt $0.006 < 0.01$ or $"0.994" > 0.99$ Critical value of 2.3263 Or a correct probability comparison.	B1
	[in the CR/significant/Reject H_0 / $"0.006" < 0.01$ / $"0.994" > 0.99$] Conclude either • that the <u>mean weight</u> of <u>grapefruit</u> from <u>farm *A*</u> is <u>greater</u> than that of <u>farm *B*</u>. • that the <u>average weight</u> of <u>grapefruit</u> from <u>farm *A*</u> is <u>greater</u> than that of <u>farm *B*</u>. • that the <u>grocer</u>'s belief is <u>correct</u>. A correct conclusion in context which is based on ***their*** *z*-value and ***their*** critical value, where $\lvert c.v. \rvert > 1$.	A1
		[7] **7**

Notes

B1 If μ_1, μ_2 used then it must be clear which refers to farm *A* and to farm *B*.

1st M1 Condone minor slips e.g. $\sqrt{\frac{35^2}{100} + \frac{28^2}{80}}$ or $\sqrt{\frac{35}{80} + \frac{28^2}{100}}$ etc.
i.e. swapped *n* or one s.d. and one variance.

1st A1 s.e. $= \sqrt{\frac{35^2}{80} + \frac{28^2}{100}}$. Or can be implied by s.e. = awrt 4.81

2nd dM1 ***is dependent upon the 1st M1.***
You can follow through their s.e. if 1st M1 mark has been awarded.

Note M1A1dM1 is scored for writing $z = \pm \frac{(532 - 520)}{\sqrt{\frac{35^2}{80} + \frac{28^2}{100}}}$

Special Case **SC: M1A0M0A0 for** s.e. $= \sqrt{\frac{35}{80} + \frac{28}{100}}$ $\{= 0.847...\}$

Final A1 Dependent on the first two method marks being scored.
For a contextualised comment which is rejecting H_0.
Contradictory statements score final A0. E.g. "significant, do not reject H_0".

<u>Alternative method for 2nd "M1A1B1" marks:</u> Let $D = \bar{x}_A - \bar{x}_B$

$2.3263 = \frac{D - 0}{4.8117...}$	dM1: dependent upon the 1st M1 for $\frac{D}{\text{their } "4.8117..."} = 2.3263 / 2.32 / 2.33$
	A1: D = awrt 11.2
So, $D = 11.193$	B1: 2.3263

Question Number	Scheme	Marks
4. (a)	**Man** | ***A*** | ***B*** | ***C*** | ***D*** | ***E*** | ***F*** | ***G*** | ***H*** | ***I*** | ***J*** Rank x | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 Rank w | 2 | 7 | 4 | 3 | 1 | 9 | 6 | 5 | 8 | 10 or **Man** | ***A*** | ***B*** | ***C*** | ***D*** | ***E*** | ***F*** | ***G*** | ***H*** | ***I*** | ***J*** Rank x | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 Rank w | 9 | 4 | 7 | 8 | 10 | 2 | 5 | 6 | 3 | 1 Attempt to rank both for x and for w. (at least four correct).	M1
	$\sum d^2 = 1 + 25 + 1 + 1 + 16 + 9 + 1 + 9 + 1 + 0; = 64$ For finding the difference between each of the ranks and evaluating $\sum d^2$.	M1
	$\sum d^2 = 64$	A1
	$r_s = 1 - \frac{6(64)}{10(99)}; = 0.6121212...$ Using $1 - \frac{6\sum d^2}{10(99)}$ with their $\sum d^2$	dM1;
	$\frac{101}{165}$ or awrt 0.612	A1
		[5]
(b)	$H_0 : \rho = 0$, $H_1 : \rho > 0$ Both hypotheses stated correctly	B1
	Critical Value $r_s = 0.5636$ or CR: $r_s \geqslant 0.5636$ Critical value of 0.5636	B1
	Either • Since $r_s = 0.6121...$ lies in the CR • Result is significant • Reject H_0 (condone H_1) see notes	M1
	conclude that there is a positive correlation between systolic blood pressure and weight. Conclusion in context	A1
		[4]
(c)	**Both** either • Critical Value $r = 0.5494$ • CR: $r \geqslant 0.5494$ **and** either • Since $r = 0.5114$ does not lie in the CR • Result is not significant • Do not reject H_0 (or accept H_0)	M1
	Conclude that there is no positive correlation Context not required here.	A1
		[2]
(d)	Either • A comment that conveys both the ideas "as x increases, w increases" **and** "the relationship is non-linear" • "There is a positive correlation" **and** "the relationship is non-linear" • Data is not (bi-variate) normal Any one of these or equivalent.	B1
		[1]
		12

	Notes	
4. (a)		
	3rd dM1	***is dependent on 1st M1*** for use of $1-\frac{6\sum d^2}{10(99)}$ with their $\sum d^2$
	Note	If a candidate finds $\sum d^2 = 266$, leading to r_s = awrt -0.612 then award M1M1A1M1A1.
(b)	**1st B1**	Both hypotheses stated in terms of ρ.
	M1	For a correct statement relating their r_s $\left(\lvert r_s\rvert < 1\right)$ with their c.v. where $\lvert$their c.v.$\rvert < 1$
	A1	For a contextualised comment which is rejecting H_0, which must mention "positive correlation", "blood pressure" and "weight". (Use of "association" is A0.) Follow through their r_s with their c.v. (provided $\lvert$their c.v.$\rvert < 1$)
	Two-tailed test	Applying a two-tailed test scores a maximum of B0B1M1A0 **So Award SC B0B1** for $H_0: \rho = 0$, $H_1: \rho \neq 0$ followed by critical value r_s = $(\pm)$ 0.6485 and allow access to the M1 mark only.

Question Number	Scheme	Marks
5. (a)	H_0 : There is no association between type of drink and gender (independent) H_1 : There is an association between type of drink and gender (dependent) — Correct hypotheses	B1
	Expected / **Tea** / **Coffee** / **Hot Chocolate** / **Total** **Male** 46.53 34.31 13.16 94 **Female** 52.47 38.69 14.84 106 **Total** 99 73 28 200 Some attempt at $\frac{(\text{Row Total})(\text{Column Total})}{(\text{Grand Total})}$ Can be implied by at least one correct E_i to 1d.p.	M1
	All expected frequencies are correct. Condone exact fractions.	A1
	Observed / Expected / $\frac{(O-E)^2}{E}$ / $\frac{O^2}{E}$ 57 46.53 2.3559... 69.8259... 26 34.31 2.0127... 19.7027... 11 13.16 0.3545... 9.1945... 42 52.47 2.0892... 33.6192... 47 38.69 1.7849... 57.0949... 17 14.84 0.3144... 19.4744... Totals 8.9116... 208.9116... At least 2 correct terms for $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ or correct expressions with their E_i. Accept 2 sf accuracy for the dM1 mark.	dM1
	At least 5 correct $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ terms to either 2 dp or better. Allow truncation.	A1
	$X^2 = \sum \frac{(O-E)^2}{E}$ or $\sum \frac{O^2}{E} - 200$;= 8.9116... For applying either $\sum \frac{(O-E)^2}{E}$ or $\sum \frac{O^2}{E} - 200$	dM1
	8.9 or awrt (8.88 – 8.91)	A1
	$\nu = (2-1)(3-1) = 2$ — $\nu = 2$	B1
	$\chi^2_2(0.05) = 5.991 \Rightarrow$ CR: $X^2 \geqslant 5.991$ — 5.991 or ft $\chi^2_{\text{their } \nu}(0.05)$	B1ft
	[in the CR/significant/Reject H_0]	
	conclude that there is an association between type of drink preferred and gender. (or they are not independent.) A correct conclusion in context which is based on ***their*** X^2-value and ***their*** χ^2-critical value.	A1
		[10]
(b)	$\chi^2_2(0.005) = 10.597 \Rightarrow$ CR: $X^2 \geqslant 10.597$ — Critical value of 10.597	B1
	[not in the CR/not significant/do not reject H_0]	
	Either • Conclude there is no association between type of drink preferred and gender (or they are independent). • The conclusion would change (if a correct H_0 has been rejected in part (a)). Any one of these.	B1
		[2]
		12

		Notes
5. (a)	**1st B1**	For both hypotheses. Must mention "drink" and "gender" or "sex" at least once. Use of "relationship" or "correlation" or "connection" is B0.
	2nd dM1	Dependent on the first method mark. At least 2 correct terms (as in 3rd or 4th column) or ***correct expressions*** with their E_i
	2nd A1	All correct terms to either 2 d.p. or better. Allow truncated answers.
	3rd dM1	***Dependent on the second method mark.*** For applying either $\sum \frac{(O-E)^2}{E}$ or $\sum \frac{O^2}{E} - 200$
	3rd A1	8.9 or awrt (8.88 – 8.91)
	2nd B1	$v = 2$ This mark can be implied by a correct critical value of 5.991
	Note	If 8.9 or awrt (8.88 – 8.91) is seen (from a calculator) **without** the expected frequencies stated then award **special case** M0A0M1A1M1A1.
	Final A1	***Dependent on the third method mark.*** A correct contextualised conclusion which is rejecting H_0. Must mention "drink" and "gender" or "sex". No follow through. If e.g. hypotheses are the wrong way round A0 here.
	Note	Contradictory statements score A0. E.g. "significant, do not reject H_0".
	Note	Condone "relationship" or "connection" here but **not** "correlation". e.g. "There is evidence of a relationship between grades and gender"
	Note	**Full accuracy gives** $X^2 = 8.911619...$ and p-value 0.0116 to 0.0117

Question Number	Scheme	Marks
6. (a)	$\hat{p} = \dfrac{0(2)+1(21)+2(45)+3(42)+4(12)+5(3)}{8(2+21+45+42+12+3) \text{ or } 8(125)} \left\{= \dfrac{300}{1000}\right\} = 0.3$ (*) Answer is given. See notes.	M1 A1cso **[2]**
(b)	$r = 125 \times {}^{8}C_{3}(0.3)^{3}(0.7)^{5}$ $\{= 31.76523...\}$ (formula) or $r = 125 \times (0.8059 - 0.5518)$ $\{= 31.7625\}$ (tables) $s = 125 - (7.21 + 24.71 + 37.06 + \text{their } r + 17.02 + 5.83)$ $\{= 1.40477... \text{ or } 1.4075\}$ or $s = 125 \times (1 - 0.9887)$ $\{= 1.4125\}$	M1
	$r = 31.76523$ or 31.7625 or 31.7575 r = awrt 31.77 or r = awrt 31.76	A1
	$s = 1.40477$ or 1.4075 or 1.4125 $s = 1.4$ or awrt1.40 or s = awrt1.41	A1 **[3]**
(c)	See table below	

# failed tasks	O_i	E_i	Comb O_i	Comb E_i	$\dfrac{(O-E)^2}{E}$	$\dfrac{O^2}{E}$
0	2	7.21	2	7.21	3.7648...	0.5548...
1	21	24.71	21	24.71	0.5570...	17.8470...
2	45	37.06	45	37.06	1.7011...	54.6411...
3	42	31.77 (31.76)	42	31.77 (31.76)	3.2941... (3.3016...)	55.5241... (55.5416...)
4	12	17.02	12	17.02	1.4806...	8.4606...
5	3	5.83	3	7.23 (7.24) {7.25}	2.4748… (2.4831…)	1.2448… (1.2431…)
$\geqslant 6$	0	1.40 (1.41)				
				Totals	13.2724... (13.2882...)	138.2724... (138.2882...)

M1 M1

For applying either $\mathrm{X}^2 = \sum \dfrac{(O-E)^2}{E}$ or $\sum \dfrac{O^2}{E} - 125$; = awrt 13.3 $\sum \dfrac{(O-E)^2}{E}$ or $\sum \dfrac{O^2}{E} - 125$ — dM1

awrt 13.3 — A1

$v = 6 - 1 - 1 = 4$ see notes — B1 ft

$\chi^2_4(0.05) = 9.488 \Rightarrow$ CR: $\mathrm{X}^2 \geqslant 9.488$ a correct ft for their $\chi^2_k(0.05)$, where $k = n - 1 - 1$ from their n. — B1

H_0 : Binomial distribution is a good(or suitable) model (or fit).
H_1 : Binomial distribution is not a suitable model.
[in the CR/significant/Reject H_0] Correct hypotheses — B1

Binomial distribution is not a suitable model. A correct conclusion (context not required here) which is based on ***their*** X^2-value and ***their*** χ^2-critical value. — A1 **[8]**

(d)	Following from a correct conclusion in part (c), a comment conveying **either** • p is not constant • employer's belief is not justified.	B1 **[1]**
		14

	Notes	
6. (a)	**M1**	Must show clearly how to get either 300 or 1000.
	A1 cso	Showing how to get <u>both</u> 300 and 1000 and reaching $p = 0.3$
(b)	**M1**	For any correct method (or a correct expression) for finding either r or s.
	A1	r = awrt 31.77 or r = awrt 31.76
	A1	$s = 1.4$ or awrt1.40 or s = awrt1.41
(c)	**1st M1**	For an attempt to pool 5 failed tasks and $\geqslant 6$ failed tasks ONLY.
	Note	Give 1st M0 for pooling 0 failed tasks and 1 failed task.
	2nd M1	For an attempt at the test statistic, at least 2 correct expressions/values (to awrt 2 d.p. or truncated 2 d.p.)
	3rd dM1	***Dependent on the second method mark.*** For applying either $\sum \frac{(O-E)^2}{E}$ or $\sum \frac{O^2}{E} - 125$
	1st A1	awrt 13.3
	1st B1ft	For their evaluated $n-1-1$. i.e. realising that they must subtract 2 from their n.
	2nd B1	For a correct ft for their $\chi^2_k(0.05)$, where $k = n-1-1$ from their n.
	3rd B1	Must have both hypotheses and mention Binomial at least once. Inclusion of 0.3 for p in hypotheses is B0 but condone in conclusion.
	Final A1	***Dependent on the 2nd and 3rd Method marks only.*** A correct conclusion (context not required) which is rejecting H_0.
	Note	No follow through on their hypotheses if they are stated the wrong way round.
	Note	Contradictory statements score A0. E.g. "significant, do not reject H_0".
	Note	Condone mentioning of Bin(8, 0.3) in conclusion
	Note	Full accuracy gives a combined expected frequency of 7.245956..., $\frac{(O-E)^2}{E} = 2.4880$, $\frac{O^2}{E} = 1.2421$, $X^2 = 13.28333...$
	Note	p-value for the test is 0.0099 to 0.0100
	Note	No combining gives $X^2 = 13.58...$
	Note	Combining 0/1 and 4/5/$\geqslant$6 gives $X^2 = 11.02$

Question Number	Scheme	Marks
7. (a)	$X = 4Y - 3W$, $Y \sim N(40, 3^2)$, $W \sim N(50, 2^2)$; Y, W are independent.	
	$\{E(X) = 4E(Y) - 3E(W) = 4(40) - 3(50)\} \Rightarrow E(X) = 10$ $\quad E(X) = 10$ (seen or implied)	B1
	$\text{Var}(X) = 16\,\text{Var}(Y) + 9\,\text{Var}(W)$ $\quad$ Either $(4^2)\text{Var}(Y)$ or $+(3^2)\text{Var}(W)$	M1
	For adding the variances	M1
	$\{\text{Var}(X) = 16(9) + 9(4)\} \Rightarrow \text{Var}(X) = 180$ $\quad \text{Var}(X) = 180$	A1
	{So $X \sim N(10, 180)$}	
	$\{P(X > 25) =\}\ P\left(Z > \frac{25 - 10}{\sqrt{180}}\right)$ $\quad$ Standardising $(\pm)$ with their mean and their standard deviation	M1
	$= P(Z > 1.11803...)$ $\quad$ awrt ± 1.12	A1
	$= 1 - 0.8686$	
	$= 0.1314$ (or 0.131777...) $\quad$ awrt 0.131 or awrt 0.132	A1
		[7]
(b)	$A = \sum_{i=1}^{3} Y_i$, $C \sim N(115, \sigma^2)$; $P(A - C < 0) = 0.2$; A, C are independent.	
	$\{E(A - C) = 3E(Y) - E(C) = 3(40) - (115)\} \Rightarrow E(A - C) = 5$ $\quad E(A - C) = 5$	B1
	$\text{Var}(A - C) = 3\text{Var}(Y) + \text{Var}(C)$ $\quad 3\text{Var}(Y)$ and a + ...	M1
	$\{\text{Var}(A - C) = 3(9) + \sigma^2\} \Rightarrow \text{Var}(A - C) = 27 + \sigma^2$ $\quad \text{Var}(A - C) = 27 + \sigma^2$	A1
	{So $A - C \sim N(5, 27 + \sigma^2)$}	
	$\{P(A - C < 0) = 0.2\} \Rightarrow P\left(Z < \frac{-5}{\sqrt{27 + \sigma^2}}\right) = 0.2$	
	$\frac{-5}{\sqrt{27 + \sigma^2}} = k\ (= -0.8416)$ $\quad$ Standardising $(\pm)$ with their mean and their standard deviation which is in terms of σ^2 and setting the result equal to k, where $\lvert k \rvert$ is in the interval [0.84, 0.85].	M1
	± 0.8416 or awrt ± 0.8416	B1
	Correct equation. See notes	A1
	$\sigma^2 = \left(\frac{-5}{-0.8416}\right)^2 - 27 \Rightarrow \sigma^2 = ...$ $\quad$ Squaring and rearranging leading to a positive value for σ^2.	dM1
	$\sigma^2 = 8.2962...$ (= 8.4308... from using -0.84) $\quad$ awrt 8.3 or awrt 8.4	A1 **cso**
	(= 8.2945... from calculator, so need awrt 8.29 for full marks if no prior working is shown.)	**[8]**
		15

(a)	**Note**	Condone applying reversed variances, e.g. $16(4) + 9(9)$ for the first 2 method marks.
	Note	$\text{Var}(X) = 180$ with no working gets M1M1A1
	Note	$\text{Var}(X) = 48$ with no working gets M0M1A0
	Note	$\text{Var}(X) = 108$ with no working gets M1M0A0
	Note	$\text{Var}(X) = 24$ with no working gets M0M0A0
(b)	**2nd M1**	Allow $\frac{\pm \text{their } E(A - C)}{\sqrt{\text{their Var}(A - C)}} = k$, where $\lvert k \rvert$ is in the interval (0.84, 0.85).
	2nd B1	For either -0.8416 or 0.8416
	2nd A1	E.g. Allow $\frac{-5}{\sqrt{27 + \sigma^2}} = [-0.85, -0.84]$ or $\frac{5}{\sqrt{27 + \sigma^2}} = [0.84, 0.85]$
	3rd M1	***Dependent on the 2nd M1 mark being awarded.***

Pearson Education Limited. Registered company number 872828
with its registered office at Edinburgh Gate, Harlow, Essex CM20 2JE

Mark Scheme (Results)

Summer 2014

Pearson Edexcel GCE in Statistics 3R (6691/01R)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please visit our website at www.edexcel.com.

Our website subject pages hold useful resources, support material and live feeds from our subject advisors giving you access to a portal of information. If you have any subject specific questions about this specification that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

www.edexcel.com/contactus

Pearson: helping people progress, everywhere

Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2014
Publications Code UA040138

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

PEARSON EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75

2. The Edexcel Mathematics mark schemes use the following types of marks:

- **M** marks: Method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
- **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
- **B** marks are unconditional accuracy marks (independent of M marks)
- Marks should not be subdivided.

3. Abbreviations

   These are some of the traditional marking abbreviations that will appear in the mark schemes.

   - bod – benefit of doubt
   - ft – follow through
   - the symbol $\surd$ will be used for correct ft
   - cao – correct answer only
   - cso - correct solution only. There must be no errors in this part of the question to obtain this mark
   - isw – ignore subsequent working
   - awrt – answers which round to
   - SC: special case
   - oe – or equivalent (and appropriate)
   - d… or dep – dependent
   - indep – independent
   - dp decimal places
   - sf significant figures
   - $*$ The answer is printed on the paper or ag- answer given
   - $\lfloor$ or d… The second mark is dependent on gaining the first mark

4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.

6. If a candidate makes more than one attempt at any question:
   - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
   - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.

7. Ignore wrong working or incorrect statements following a correct answer.

Question Number	Scheme	Marks
1.		
a)		M1

Car model	A	B	C	D	E	F	G	H
Sales rank	8	6	1	5	4	7	2	3
Fuel efficiency rank	8	1	5	6	2	7	4	3
d^2	0	25	16	1	4	0	4	0

	$\sum d^2 = 50$	M1A1
	$r_s = 1 - \frac{6\sum d^2}{8(64-1)} = 1 - \frac{6 \times 50}{8 \times 63}$	M1
	$r_s = \frac{204}{504} = 0.40476\ldots$ awrt 0.405	A1
		(5)
b)		
	$H_0: \rho_s = 0$ $H_1: \rho_s > 0$ (accept ρ_s or ρ)	B1
	1 tail critical value $\rho = 0.6429$	B1
	Test value is not in critical region so insufficient evidence to reject H_0	M1A1ft
	No significant evidence at 5% level to support journalist's belief	
		(4)
c)	Underlying (bivariate) **Normal** distribution	B1
		(1)
d)		
	Evidence does not support Normal distribution since mean< median or (negative) skew, oe	B1
		(1)
		(11 marks)

Notes	
a)	M1 for attempting to rank at least one set of data A1 for at least one set of data ranked correctly(NB this mark comes after 2nd M1) M1 for attempting Σd^2 M1 for correct use of formula for r_s
b)	B1 for H_0 **and** H_1 correct (condone $\leq$ for H_0) 2nd B1 allow 0.7381 if their H_1:$\rho_s \neq 0$ M1 for correct statement relating their test statistic and critical value A1ft their test statistic, H_1 and critical value but must be in context.
c)	B1 require **Normal** distribution, ignore additional assumptions
d)	B1 require not Normal **and** valid reason

Question Number	Scheme	Marks
2) **(a)** **(i)** **(ii)** **(b)**	Expected value = $\frac{50\times74}{200} = 18.5$ χ^2 contribution = $\frac{(27-18.5)^2}{18.5} = 3.905405405 = 3.91$ to 3sfs H_0: users age and main mobile phone use are independent/ no association between users age and main mobile phone use H_1: users age and main mobile phone use are not independent/ some association between users age and main mobile phone use $\nu = 4$ Critical value χ^2 = 9.488 Test statistic is in critical region therefore significant evidence to reject H_0 and accept H_1. Evidence at 5% level that age and main phone use are not independent.	B1 cso B1 cso (2) B1 B1 B1ft M1 A1ft (5) (7 marks)
Notes		
(b)	3rd B1 ft on their value of ν M1 for attempt to compare test statistic and their critical value A1 ft on test statistic and critical value but must be comment in context. (A0 if hypotheses are the wrong way around)	

Question Number	Scheme	Marks
3) **(a)**	$P(S > 2C) = P(S - 2C > 0)$	B1
	$E[S - 2C] = 4.9 - 2 \times 2.5 = -0.1$	M1A1
	$Var(S - 2C) = 0.64 + 4 \times 0.16 = 1.28$	M1, M1
	$P(S - 2C > 0), = P(Z > \frac{0--0.1}{\sqrt{1.28}})$	
	$= P(Z > 0.08838\ldots)$	A1
	$=0.4641$ (tables), or $0.4648\ldots$ (calculator) accept awrt 0.464 or 0.465	(6)
(b)	Let $T = S_1 + S_2 + \ldots + S_{100}$	M1A1
	$E[T] = 100 \times 4.9 = 490$	A1
	$Var(T) = 100 \times 0.64 = 64$	M1
	$P(T < 500) = P(Z < \frac{500-490}{\sqrt{64}})$	
	$= P(Z < 1.25)$	A1
	$= 0.8944$	(5)
		(11 marks)

Notes

(a)

1st M1 for $\ldots \pm 4Var(C)$
2nd M1 for $P(S - 2C > 0)$
3rd M1 ft their expectation and variance but not if $Var(S - 2C)$ is negative. (Should lead to $P(Z >$ +ve)

(b)

1st M1 for attempt to find mean or variance of total

1st A1 either correct

2nd A1 both correct
2nd M1 for standardising using 500, their mean and their sd leading to $P(Z <$ +ve) o.e.

Sample mean, $\bar{x} = \frac{660+\alpha}{5} = 132 + \frac{\alpha}{5}$

Question Number	Scheme	Marks
4)	Test statistic, $z = \frac{132+\frac{\alpha}{5}-160}{\frac{6}{\sqrt{5}}}$	M1A1ft
	Critical z values is 1.6449	B1
	Therefore the test statistic is significant if $\frac{132+\frac{\alpha}{5}-160}{\frac{6}{\sqrt{5}}} > 1.6449$	M1
	Therefore $132+\frac{\alpha}{5}-160 > 1.6449\times\frac{6}{\sqrt{5}}$ $\alpha > 5\left(1.6449\times\frac{6}{\sqrt{5}}+28\right)$ $\alpha > 162.0686493\ldots$	A1
	Accept awrt 162.1	(6) (6 marks)
Notes		
	1st A1 ft on their $\bar{x}$ 1st B1 given for 1.6449 seen (condone sign) 3rd M1 inequality using their test statistic, accept incorrect signs for M1	

Question Number	Scheme	Marks
5)	$S_E{}^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) = \frac{1}{119}\left(956909 - \frac{10650^2}{120}\right)$	M1
(a)	$= \frac{11721.5}{119} = 98.5$	A1 (2)
(b)	H_0: $\mu_F = \mu_E$, H_1: $\mu_F \neq \mu_E$,	B1
	$\bar{x}_E = \frac{10650}{120} = 88.75$ and $\bar{x}_F = \frac{6510}{70} = 93$	M1
	Test statistic, $z = \frac{93-88.75-0}{\sqrt{\frac{151}{70}+\frac{98.5}{120}}} = 2.4627$ ...	M1A1
	Critical values, $z = (\pm)2.5758$	B1ft
	Test stat is not in critical region	M1
	Insufficient evidence to reject H_0 at 1% level	A1ft
	No significant evidence of a difference in mean lengths of English and French films	(7)
(c)	By CLT we can assume that the mean of a large sample has a Normal distribution	B1 (1)
(d)	On a list, label English films 1 – 724 and French films 1-473 (oe)	B1
	Use random number table/generator to select	
	$\frac{724}{724+473} \times 190 = 115$ English films and	M1A1
	$\frac{473}{1197} \times 190 = 75$ French films	(3)
		(13 marks)

Notes		
	Alternative	
(a)	$S_E{}^2 = \frac{n}{n-1}\left(\frac{\sum x^2}{n} - \bar{x}^2\right) = \frac{120}{119}\left(\frac{956909}{120} - 88.75^2\right) = 98.5$	
(b)	1st B1 needs both H_0 and H_1, can be in words 2nd B1ft on their H_1 1st M1 for attempt @ both means ($\bar{x}_E$ may be in (a)) 2nd M1 for attempt at correct test statistic, ft their values 3rd M1 for attempt to compare their test stat and critical values A1 ft on their test and critical values but must include comment in context	
(c)	Require mention of mean of *E* or *F* and normal distribution	
(d)	M1 requires use of <u>random numbers</u> and attempt to find correct sample sizes A1 both 115 and 75 found.	

Question Number	Scheme	Marks
6)	Independence of each occurrence (of a fake coin)	B1
(a)	Constant probability for each occurrence (of a fake)	B1
		(2)
(b)	$r = 150 \times P(X = 2) = 150 \times \binom{20}{2} \times 0.05^2 \times 0.95^{18}$	M1
	r = 28.3015… awrt 28.3	A1
	s = 150 – (53.8 + 56.6 + 28.3 + 8.9) = 2.4	A1ft
		(3)
(c)	H_0: Bin(20, 0.05) is a suitable model H_1: Bin(20, 0.05) is not a suitable model	B1
	Combining last two groups	
	≥ 3 / **Observed frequency** 19 / **Expected frequency** 11.3	M1
	$\nu = 4 - 1 = 3$	B1
	Critical value, χ^2 (0.05) = 7.815 (accept 9.488 if their $\nu = 4$)	B1ft
	Test statistic, $\sum \frac{(O-E)^2}{E} = \frac{(43-53.8)^2}{53.8} + \frac{(62-56.6)^2}{56.6} + \cdots$	M1
	= 2.168…+ 0.515… + 0.186… + 5.246…	
	= 8.117 (accept 10.16 if groups not combined)	A1ft
	In critical region, sufficient evidence to reject H_0, accept H_1	
	Significant evidence at 5% level to reject the manager's model	A1ft
		(7)

Question Number	Scheme	Marks
(d)	$\nu = 4 - 2 = 2$ 4 classes due to pooling 2 restrictions (equal total and mean/proportion)	 B1 B1 (2)
(e)	H_0: Binomial distribution is a good model H_1: Binomial distribution is not a good model Critical value, χ^2 (0.05) = 5.991 Test statistic is not in critical region, insufficient evidence to reject H_0 Accept the assistant manager's model for the number of fake coins per bag.	B1 B1 B1 (3) (17 marks)
	Notes	
(b)	M1A1 for one of *r* or *s* correct A1ft for other value if using 150 - … and answer must be >0	
(c)	1st B1 can be in words but must include p = 0.05 3rd B1 ft on their ν Test statistic alternative method Test stat = $\sum \frac{O^2}{E} - 150 = \frac{43^2}{53.8} + \frac{62^2}{56.6} + \cdots - 150 = 8.117$ ... 1st A1 ft if their groups not combined 2nd A1 ft their test and critical values but must be comment in context e.g. mention of "manager's model" <u>or</u> "fake coins"	
(d)	1st B1 evidence that pooling is required 2nd B1 must have correct reasons for restrictions.	

Question Number	Scheme	Marks
7) **(a)** **(i)**	$\bar{x} = \frac{10.01+9.97+9.93+\cdots}{8} = 9.9775$	M1
	95% CI $\bar{x} \pm 1.96 \times \frac{0.08}{\sqrt{8}}$	B1M1
	95% CI for μ (9.92, 10.03)	A1 (4)
(ii)	10.00 is within confidence interval so accept that pump may be performing correctly (although sample mean is low).	B1 (1)
(b)	Upper limit of CI is $9.96 + 1.6449 \times \frac{0.08}{\sqrt{n}} < 10.00$	B1, M1A1ft
	$\frac{1.6449 \times 0.08}{\sqrt{n}} < 0.04$	
	$\sqrt{n} > \frac{1.6449 \times 0.08}{0.04}$	M1
	$n > 10.82\ldots$ therefore minimum $n = 11$	A1 cao (5) (10 marks)

Notes		
(a) **(i)**	1st M1 attempt to find sample mean B1 for correct *z* value A1 limits correct to 2 decimal places (or more) B1 for correct z value	
(b)	1st M1A1, ft their z value	

Pearson Education Limited. Registered company number 872828
with its registered office at Edinburgh Gate, Harlow, Essex CM20 2JE

Mark Scheme (Results)

Summer 2014

Pearson Edexcel GCE in Statistics 3 (6691/01)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2014
Publications Code UA040135

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

PEARSON EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75

2. The Edexcel Mathematics mark schemes use the following types of marks:

- **M** marks: Method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
- **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
- **B** marks are unconditional accuracy marks (independent of M marks)
- Marks should not be subdivided.

3. Abbreviations

   These are some of the traditional marking abbreviations that will appear in the mark schemes.

   - bod – benefit of doubt
   - ft – follow through
   - the symbol $\surd$ will be used for correct ft
   - cao – correct answer only
   - cso - correct solution only. There must be no errors in this part of the question to obtain this mark
   - isw – ignore subsequent working
   - awrt – answers which round to
   - SC: special case
   - oe – or equivalent (and appropriate)
   - d… or dep – dependent
   - indep – independent
   - dp decimal places
   - sf significant figures
   - * The answer is printed on the paper or ag- answer given
   - $\sqsubset$ or d… The second mark is dependent on gaining the first mark

4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.

6. If a candidate makes more than one attempt at any question:
   - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
   - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.

7. Ignore wrong working or incorrect statements following a correct answer.

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

A1

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

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

B1

B1ft

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

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 3rd or 4th columns) 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)	$\frac{(0\times 2)+(1\times 6)+(2\times 11)+(3\times 19)+(4\times 25)+(5\times 32)+(6\times 16)+(7\times 9)}{120\times 7} = \frac{504}{840}$ = 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)	$\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$ $\mathrm{E}\left(\bar{X}\right) = \frac{1}{n}\mathrm{E}(X_1 + \ldots + X_n)$ $= \frac{1}{n}(\mathrm{E}(X_1) + \ldots + \mathrm{E}(X_n))$ $= \frac{1}{n}(\mu + \ldots + \mu)$ $= \frac{n\mu}{n} = \mu$	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 $\frac{n}{n-1}\mathrm{Var}\,x$ $= \frac{1}{5-1}(200469 - 5(200.2^2))$ $= 17.2$	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 $\frac{1.25}{\frac{4.8}{\sqrt{n}}} \geq 1.96$ $\sqrt{n} \geq \frac{2\times1.96\times4.8}{2.5} = 7.5264$ $n \geq 56.6(5)$ Minimum sample size is 57	M1B1 A1 A1 (4) **Total 8**

Notes

(a)	B1 cso: require $\mathrm{E}\left(\bar{X}\right) = \mu$ 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 $\|r_s\| < 1$)	

Pearson Education Limited. Registered company number 872828
with its registered office at Edinburgh Gate, Harlow, Essex CM20 2JE

edexcel

Mark Scheme (Final)

Summer 2015

Pearson Edexcel International A Level in Statistics 3 (WST03/01)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2015
Publications Code IA042726

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

PEARSON EDEXCEL IAL MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75

2. The Edexcel Mathematics mark schemes use the following types of marks:

- **M** marks: Method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
- **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
- **B** marks are unconditional accuracy marks (independent of M marks)
- Marks should not be subdivided.

3. Abbreviations

   These are some of the traditional marking abbreviations that will appear in the mark schemes.

   - bod – benefit of doubt
   - ft – follow through
   - the symbol $\surd$ will be used for correct ft
   - cao – correct answer only
   - cso - correct solution only. There must be no errors in this part of the question to obtain this mark
   - isw – ignore subsequent working
   - awrt – answers which round to
   - SC: special case
   - oe – or equivalent (and appropriate)
   - d… or dep – dependent
   - indep – independent
   - dp decimal places
   - sf significant figures
   - $*$ The answer is printed on the paper or ag- answer given
   - $\sqsubset$ or d… The second mark is dependent on gaining the first mark

4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.

6. If a candidate makes more than one attempt at any question:
    - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
    - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.

7. Ignore wrong working or incorrect statements following a correct answer.

June 2015
WST03 Statistics 3 Mark Scheme

Question Number	Scheme	Marks
1. (a)	$\{w\} = 018$ or 18 **018 or 18**	B1 **[1]**
(b)	$\{x\} = 18$ **18**	B1 **[1]**
(c)	$\{\text{prob} =\}\ 0$ **0**	B1 **[1]**
(d)	**Advantage:** Any one of: • Simple or easy to use also allow "quick" or "efficient" (o.e.) • It is suitable for large samples (or populations) • Gives a good spread of the data **Disadvantage:** Any one of: • The alphabetical list is (probably) not random • Biased since the list is not (truly) random • Some combinations of names are not possible	B1 B1 **[2]** **(Total 5)**

	Notes
(d)	If no labels are given treat the 1st reason as an advantage and the 2nd as a disadvantage 1st B1: for advantage 2nd B1: for disadvantage "it requires a sampling frame" is 2nd B0 since the alphabetical list is given. Note: Do not score both B1 marks for opposing advantages and disadvantages.

Question Number	Scheme	Marks
2. (a)		
	A B C L N R S T Y	
	Judge 1 6 3 4 9 2 8 1 5 7	
	Judge 2 8 4 5 7 3 9 1 2 6	
	or	M1
	S N B C T A Y R L	
	Judge 1 1 2 3 4 5 6 7 8 9	
	Judge 2 1 3 4 5 2 8 6 9 7	
	$\sum d^2 = 4+1+1+4+1+1+0+9+1$	M1
	or $0+1+1+1+9+4+1+1+4 = 22$ $\sum d^2 = 22$	A1
	$r_s = 1-\frac{6(22)}{9(80)}; = 0.8166666...$ $\frac{49}{60}$ or awrt **0.817**	M1; A1
		[5]
(b)	$H_0: \rho = 0$, $H_1: \rho > 0$	B1
	Critical Value = 0.7833 or CR: $r_s \geqslant 0.7833$ **0.7833**	B1
	Since $r_s = 0.8166...$ it lies in the CR , or reject H_0 (o.e.)	M1
	The two judges (or "they") are in agreement or there is a positive correlation between the ranks of the two judges.	A1ft
		[4]
		(Total 9)

	Notes
(a)	1st M1 for an attempt to rank at least one row (at least 4 correct) 2nd M1 for an attempt at d^2 row (may be implied by sight of $\sum d^2 = 22$ or 221 for reverse ranks) 1st A1 for $\sum d^2 = 22$ (or 221 if reverse ranking is used) Can be implied by correct answer. 3rd M1 for use of the correct formula with their $\sum d^2$ (if it is clearly stated) If the answer is not correct then a correct expression is required
False Ranking	e.g Alphabetic ranking: Gives Judge 1: 7 5 2 3 8 1 9 6 4 Judge 2: 7 8 5 2 3 9 4 1 6 $\sum d^2 = 162$ and $r_s = -0.35$ **Scores:** M0(for ranking), M1(for attempt at d^2 row), A0, M1 (for use of their $\sum d^2$), A0 i.e. 2 out of 5 Can follow through their r_s in (b)
(b)	1st B1 for both hypotheses stated correctly in terms of ρ (allow ρ_s) H_1 must be compatible with ranking 2nd B1 for cv = 0.7833 (independent of their H_1 (no 2-tail value in tables) but compatible sign with their r_s) M1**:** for a correct statement (in words) relating their r_s with their critical value. e.g. "reject H_0 ","in critical region", "significant", "positive correlation" May be implied by a correct contextual comment.
|cv|>1	If their cv is |cv| > 1 (often from using normal tables) award M0A0 If |their r_s | > |their cv| then "significant" (o.e.) for M1 and "judges are in agreement" (o.e.) for A1ft If |their r_s | < |their cv| then "not significant" (o.e.) for M1 and "judges don't agree" (o.e.) for A1ft
	A1ft**:** for a correct follow through conclusion in context. "positive correlation" alone scores M1 A0 For reverse ranking should still say " judges are in agreement"

Question Number	Scheme	Marks
3. (a)	$\widehat{\lambda} = \dfrac{0(47)+1(57)+2(46)+3(35)+4(9)+5(6)}{200} = \dfrac{320}{200} = 1.6$ * Full exp' or at least 2 products and 320/200 seen	B1 * **[1]**
(b)	$r = 200 \times \dfrac{e^{-1.6}(1.6)^2}{2!}$ $\{= 51.68550861...\}$ Using $r = 200 \times \dfrac{e^{-1.6}(1.6)^2}{2!}$	M1
	$s = 200 - (40.38 + 64.61 + \text{their } r + 27.57 + 11.03)$ $\{= 4.72449139...\}$ or their $r + s = 56.41$	M1
	$r = 51.68550861...$ and $s = 4.72449139...$ r = awrt **51.69** and s = awrt **4.72**	A1 **[3]**
(c)	H_0 : Poisson (distribution) is a suitable/ sensible (model)	B1
	H_1 : Poisson (distribution)is not a suitable/ sensible (model).	

Number of accidents	Observed	Expected	Combined Observed	Combined Expected	$\frac{(O-E)^2}{E}$	$\frac{O^2}{E}$
0	47	40.38	47	40.38	1.0853	54.7053
1	57	64.61	57	64.61	0.8963	50.2863
2	46	51.69	46	51.69	0.6264	40.9364
3	35	27.57	35	27.57	2.0024	44.4324
4	9	11.03	15	15.75	0.0357	14.2857
$\geqslant 5$	6	4.72				
				Totals	4.6461	204.6461

	$X^2 = \sum \dfrac{(O-E)^2}{E}$ or $\sum \dfrac{O^2}{E} - 200 ;= 4.6461$	M1;
		awrt **4.65** A1
	$v = 5 - 1 - 1 = 3$	**3** B1 ft
	$\chi^2_3(0.10) = 6.251 \Rightarrow$ CR: $X^2 \geqslant 6.251$	**6.251** B1 ft
	[Since $X^2 = 4.6461$ does not lie in the CR, then there is insufficient evidence to reject H_0]	
	The number of ***accidents*** per day can be modelled by a Poisson distribution or the ***supervisor's*** belief is correct.	A1 ft **[7]** **(Total 11)**

Notes

(b) Note: Allow A1 for s = awrt 4.74 (found as a result of using expected values to full accuracy.)

(c) 1st B1: for both hypotheses and mentioning Poisson at least once.
Allow Poisson is a "good fit/model" but not "good method"
Inclusion of 1.6 for mean in hypotheses is B0 but condone in conclusion.

1st M1: For an attempt to pool 4 accidents and $\geqslant 5$ accidents or pool when $E_i < 5$ No pooling is M0

2nd M1: For an attempt at the test statistic, at least 2 correct expressions/values (to awrt 2 d.p.)

1st A1: For awrt 4.65 (score M1M1A1 if awrt 4.65 seen)

No pooling If no pooling can allow 2nd M1 if $X^2 = 5.33$ is seen

2nd B1ft: For $n-1-1$ i.e. subtracting 2 from their n.

3rd B1ft: For a correct ft for their $\chi^2_k(0.10)$, where $k = n - 1 - 1$ from their n.

B1B1 may be implied by 6.251 (if pooling) or 7.779 for no pooling

2nd A1ft**: (*Dep. on the* 2nd M1)** For correct comment in context based on their test statistic and their critical value that mentions ***accidents*** or ***supervisor***. Condone mention of Po(1.6) in conclusion
Score A0 for inconsistencies e.g. "significant" followed by "supervisor's belief is justified"

Note: Full accuracy gives a combined expected frequency of 15.76, $\dfrac{(O-E)^2}{E} = 0.0366$, $\dfrac{O^2}{E} = 14.2766$,
$X^2 = 4.64855...$ and p-value 0.199

Question Number	Scheme		Marks
4. (a)	Let X = weight of a sack of potatoes, $X \sim \mathrm{N}(25.6\ ,\ 0.24^2)$		
	So $D = X_1 - X_2 \sim \mathrm{N}(0\ ,\ 2(0.24)^2)$ or $D \sim \mathrm{N}(0\ ,\ 0.1152)$	Attempt at D and $D \sim \mathrm{N}(0, ..)$ $(0.24)^2 + (0.24)^2$; **0.1152**	M1 A1; A1
	$\left\{\mathrm{P}\left(\lvert D\rvert > 0.5\right) =\right\} 2\,\mathrm{P}(D > 0.5)$	$2 \times \mathrm{P}(D > 0.5)$ can be implied	dM1
	$= 2 \times \mathrm{P}\left(Z > \dfrac{0.5}{\sqrt{0.1152}}\right)$		dM1
	$= 2 \times \mathrm{P}\left(Z > 1.4731...\right)$ <u>or</u> $= 2(1 - 0.9292)$		
	$= 0.1416$	**awrt 0.141** or **awrt 0.142**	A1
			[6]
(b)	Let Y = weight of an empty pallet, $Y \sim \mathrm{N}(20.0\ ,\ 0.32^2)$		
	So $T = X_1 + X_2 + \ ... \ + X_{30} + Y$		
	$T \sim \mathrm{N}(30(25.6) + 20,\ 30(0.24)^2 + 0.32^2)$	$30(25.6) + 20$ <u>or</u> **788** $30(0.24)^2 + 0.32^2$	B1 M1
	$T \sim \mathrm{N}(788, 1.8304)$	N and 1.8304 or awrt **1.83**	A1
	$\left\{\mathrm{P}(T > 785) =\right\}\ \mathrm{P}\left(Z > \dfrac{785 - 788}{\sqrt{1.8304}}\right)$		M1
	$= \mathrm{P}(Z > -2.2174...)$		
	$= 0.9868$	**awrt 0.987**	A1
			[5]
			(Total 11)

Notes

(a)

1st M1: For clear definition of D and normal distribution with mean of 0 (Can be implied by 3rd M1)
1st A1: for correct use of Var($X_1 - X_2$) formula
2nd A1: for 0.1152
2nd dM1: For realising need $2 \times \mathrm{P}(D > 0.5)$ (Dependent on 1st M1 i.e. must be using suitable D)
3rd dM1: Dep on 1st M1 for standardising with 0.5, 0 and their s.d.($\neq 0.24$)Must lead to $\mathrm{P}(Z > +\mathrm{ve})$ (o.e.)
$\mathrm{P}(Z > 1.47)$ implies 1st M1 1st A1 2nd A1 and 3rd M1
Correct answer only will score 6 out of 6

(b)

B1: For a mean of $30(25.6) + 20$. Can be implied by 788.
1st M1: For $30(0.24)^2 + 0.32^2$. Can be implied by 1.8304 or awrt 1.83
Allow M1 for swapping error i.e. $30 \times 0.32^2 + 0.24^2$ if the expression is seen
1st A1: For normal and correct variance of 1.8304 or awrt 1.83.
Normality may be implied by standardisation
2nd M1: For standardising with 785 with their mean and st. dev..($\neq 0.24$) Must lead to $\mathrm{P}(Z > -\mathrm{ve})$ oe.
2nd A1: awrt 0.987
Correct answer only will score 5 out of 5

Note: Calculator answers are (a) 0.14071... , (b) 0.98670...

Question Number	Scheme	Marks
5.	H_0 : Grades and gender are independent (or not associated) H_1 : Grades and gender are dependent (or associated) "grades" and "gender" mentioned at least once.	B1 (1)
	Observed / Male / Female Distinction 37 44 Merit 127 96 Unsatisfactory 36 20 An attempt to convert percentages to observed frequencies.	M1
	All observed frequencies are correct.	A1
	Expected / Male / Female / Totals Distinction 45 36 81 Merit 123.889 99.111 223 Unsatisfactory 31.111 24.889 56 Totals 200 160 360 Some attempt at $\frac{(\text{Row Total})(\text{Column Total})}{(\text{Grand Total})}$ Can be implied by a correct E_i	M1
	All expected frequencies are correct to nearest integer.	A1
	Observed / Expected / $\frac{(O-E)^2}{E}$ / $\frac{O^2}{E}$ 37 45 1.422 30.422 44 36 1.778 53.778 127 123.889 0.078 130.189 96 99.111 0.098 92.987 36 31.111 0.768 41.657 20 24.889 0.960 16.071 Totals 5.104 365.104 At least 2 correct terms for $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ or correct expressions with their E_i. Accept 2 sf accuracy for the M1 mark.	M1
	All correct $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ terms to either 2 dp or better. Allow truncation. ($\Rightarrow$ by awrt 5.1 if 3rd M1 seen)	A1
	$X^2 = \sum \frac{(O-E)^2}{E}$ or $\sum \frac{O^2}{E} - 360$;= awrt 5.1 awrt **5.1**	A1 (7)
	$\nu = (3-1)(2-1) = 2$ ($\nu =$) **2** (Can be implied by 5.991)	B1
	$\chi^2_2(0.05) = 5.991 \Rightarrow$ CR: $X^2 \geqslant 5.991$ For **5.991** only	B1
	Since $X^2 = 5.1$ does not lie in the CR, then there is insufficient evidence to reject H_0	M1
	Business Studies grades and gender are independent or There is no association between Business Studies grades and gender. Or Head of department's (belief) is correct	A1ft (4)
		[12] **(Total 12)**

	Notes
	Final M1: For a correct statement linking their test statistic and their critical value (> 3.8) Note: Contradictory statements score M0. E.g. "significant, do not reject H_0 ". Final A1ft: For a correct ft statement in context – must mention "grades" and "gender" or "sex" or "head of department" Condone "relationship" or "connection" here but **not** "correlation". e.g. "There is no evidence of a relationship between grades and gender"
5.10 only	Just seeing 5.10... only can imply 1st 3 Ms but loses 1st 3 As so can score 4 out of 7 (Qu says "show..")
	Note: Full accuracy gives $X^2 = 5.104356...$ and p-value 0.0779

Question Number	Scheme	Marks
5.	**Mark Scheme for candidates who use percentages instead of observed values.**	
	H_0 : Grades and gender are independent (or not associated) H_1 : Grades and gender are dependent (or associated) — "grades" and "gender" mentioned at least once.	B1
	Observed / Male / Female Distinction 18.5 27.5 Merit 63.5 60.0 Unsatisfactory 18.0 12.5 — These marks cannot be obtained.	M0 A0
	Expected / Male / Female / Totals Distinction 23 23 46 Merit 61.75 61.75 123.5 Unsatisfactory 15.25 15.25 30.5 Totals 100 100 200 — Some attempt at $\frac{(\text{Row Total})(\text{Column Total})}{(\text{Grand Total})}$ Can be implied by one of these E_i's	M1
	Expected frequencies are not correct.	A0
	Observed / Expected / $\frac{(O-E)^2}{E}$ / $\frac{O^2}{E}$ 18.5 23 0.8804 14.8804 27.5 23 0.8804 32.8804 63.5 61.75 0.0496 65.2996 60.0 61.75 0.0496 58.2996 18.0 15.25 0.4959 21.2459 12.5 15.25 0.4959 10.2459 Totals 2.8518 202.8518 — At least 2 "correct" terms for $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ or correct expressions with their E_i. Accept 2 sf accuracy for the M1 mark.	M1
	This mark cannot be obtained.	A0
	$X^2 = \sum \frac{(O-E)^2}{E}$ or $\sum \frac{O^2}{E} - 360 ;= 2.8518$ — This mark cannot be obtained.	A0
	$\nu = (3-1)(2-1) = 2$ — $(\nu =)$ **2** (Can be implied by 5.991)	B1
	$\chi^2_2(0.05) = 5.991 \Rightarrow$ CR: $X^2 \geqslant 5.991$ — For **5.991** only	B1
	Since $X^2 = 2.86$ does not lie in the CR, then there is insufficient evidence to reject H_0	M1
	Not available since comes from incorrect working.	A0
		[12] **(Total 12)**
	Notes	
	If a candidate uses percentages rather than observed values then they can obtain a maximum of **6 marks**. They can get B1 M0A0 M1A0 M1A0A0 B1B1M1A0.	

Question Number	Scheme		Marks
6. (a)	$\left\{\hat{\mu} = \dfrac{\sum x}{n} = \dfrac{1570}{50} =\right\}\ \bar{x} = 31.4$	$\bar{x} =$ **31.4**	B1 cao
	$\left\{\hat{\sigma}^2 = \dfrac{\sum x^2 - n\bar{x}^2}{n-1} =\right\}\ s_x^2 = \dfrac{49467.58 - 50(31.4)^2}{50-1}$		M1 A1ft
	$= 3.460816...$	awrt **3.46**	A1
			[4]
(b)	[Let Y = time taken to complete obstacle course in the afternoon.]		
	$H_0 : \mu_x = \mu_y$, $H_1 : \mu_x > \mu_y$		B1
	$(z =) \dfrac{"31.4" - 30.9}{\sqrt{\dfrac{"3.46"}{50} + \dfrac{3.03}{50}}}$		M1 A1ft
	$= 1.38781...$	awrt **1.39**	A1
	CR: $Z \geqslant 1.6449$ or probability = awrt 0.082 or awrt 0.083	**1.6449** or better	B1
	Since $z = 1.38781...$ does not lie in the CR, then there is insufficient evidence to reject H_0		M1
	Conclude that the <u>mean time</u> to complete the obstacle course is the same for the early <u>morning</u> and late <u>afternoon</u>.		A1
			[7]
(c)	$\bar{X}$ <u>and</u> $\bar{Y}$ are both approx. <u>normal</u>ly distributed <u>or</u> $\bar{X} - \bar{Y}$ normal (Condone $\bar{x}$ and $\bar{y}$)		B1
			[1]
(d)	Have assumed $s^2 \simeq \sigma^2$ or variance of sample $\simeq$ variance of population		B1
			[1]
			(Total 13)

	Notes
(a)	B1: 31.4 cao Allow 31 minutes, 24 seconds. 1st M1: A correct expression for either s or s^2 (ignore label) 1st A1ft: A correct expression for s^2 with their ft $\bar{x}$. 3rd A1: awrt 3.46 (Correct answer scores 3 out of 3)
(b)	1st B1: Both hypotheses stated correctly, with some indication of which μ is which. Eg: μ_M , μ_A 1st M1: For an attempt at $\dfrac{a-b}{\sqrt{\dfrac{c}{50} + \dfrac{d}{50}}}$ with at least 3 of *a, b, c* or *d* correct. Allow $\pm$ 1st A1ft: for $\pm \dfrac{\text{their } 31.4 - 30.9}{\sqrt{\dfrac{\text{their } 3.46}{50} + \dfrac{3.03}{50}}}$ Allow $D = \bar{x} - \bar{y}$ $1.64 \sim 1.65 = \dfrac{D - 0}{\sqrt{\dfrac{"3.46"}{50} + \dfrac{3.03}{50}}}$ [SE = 0.360277..] 2nd A1: for awrt 1.39 (possibly $\pm$)(Allow for CV D = awrt 0.593) (NB d = 0.5) Correct answer scores M1A1ftA1 <u>but</u> $0 - (31.4 - 30.9) \rightarrow -1.39$ loses this 2nd A mark 2nd B1: Critical value of 1.6449 or better (seen). Allow for probability = awrt 0.082 or awrt 0.083 Note: p-values are 0.0823 (tables) and 0.0826 (calculator). 2nd M1: For a correct statement linking their test statistic and their critical value. Note: Contradictory statements score M0. E.g. "significant, do not reject H_0". 3rd A1: For a correct statement in context that accepts H_0 (no ft) Condone "no difference in mean times" Must mention "<u>mean time</u>", "<u>morning</u>" and "<u>afternoon</u>" or "<u>both</u> <u>times</u> <u>of day</u>"
(c)	B1 E.g. $\bar{X} \sim N(...)$ need both. Allow in words e.g "sample means are normally distributed"
(d)	B1 condone only mentioning "*x*" or "*y*" <u>but</u> watch out for $s_x = s_y$ or $\sigma_x = \sigma_y$ which scores B0

Question Number	Scheme	Marks
7.	Let X = score on a die	
(a)	$\mathrm{E}(S) = 3.5$, $\mathrm{Var}(S) = \frac{35}{12}$	$\mathrm{E}(S) =$ **3.5** B1 $\mathrm{Var}(S) = \frac{35}{12}$ or awrt **2.92** B1 **[2]**
(b)	So, $\overline{S} \sim \mathrm{N}\left("3.5", \frac{"\left(\frac{35}{12}\right)"}{40}\right)$ or $\overline{S} \sim \mathrm{N}\left("3.5", \frac{7}{96}\right)$	B1ft
	$\mathrm{P}\left(\overline{S} < 3\right) = \mathrm{P}\left(Z < \frac{3 - "3.5"}{\sqrt{\frac{7}{96}}}\right)$ $\{= \mathrm{P}(Z < -1.85164...)\}$	M1
	$\{= 1 - 0.9678\} = 0.0322$ **0.032 to 0.0322**	A1 **[3]** **(Total 5)**
	Notes	
(a)	2nd B1 allow awrt 2.92	
(b)	B1ft for $\overline{S} \sim \mathrm{N}\left("3.5", \frac{"\left(\frac{35}{12}\right)"}{40}\right)$ seen or implied. Follow through their E(S) and their Var(S) NB $\frac{7}{96} = 0.072916\dot{}$ accept awrt 0.0729 M1 for an attempt to standardise with 3, their mean (>3) and $\sqrt{\frac{\text{their Var}(S)}{40}}$. Must lead to P($Z <$ −ve) A1 for 0.032 ~ 0.0322	
ALT ΣS	B1ft for $\sum S \sim \mathrm{N}\left(140, \frac{350}{3}\right)$ where 140 is 40 × their E(S) and variance is 40 × their Var(S) M1 for $\mathrm{P}\left(Z < \frac{120 - "140"}{\sqrt{\frac{350}{3}}}\right)$ or $\mathrm{P}\left(Z < \frac{119.5 - "140"}{\sqrt{\frac{350}{3}}}\right)$ {= P($Z <$ −1.8979...)} A1 for 0.032~0.0322 or (with continuity correction) 0.0287 (tables) or 0.0289 (calculator).	

Question Number	Scheme	Marks
8. (a)	$\left\{\bar{x} = \dfrac{29.74 + 31.86}{2}\right\} \Rightarrow \bar{x} = 30.8$ $\bar{x} = 30.8$ This can be implied. See note.	B1
	$"1.96"\left(\dfrac{\sigma}{\sqrt{n}}\right) = 31.86 - 30.8$ or $2("1.96")\left(\dfrac{\sigma}{\sqrt{n}}\right) = 31.86 - 29.74$	M1
	$SE_{\bar{x}} = \dfrac{31.86 - 30.8}{1.96} = 0.540816... = 0.54\ (2\text{dp})$ awrt **0.54**	A1
		[3]
(b)	A 90% CI for μ is $\bar{x} \pm 1.6449\left(\dfrac{\sigma}{\sqrt{n}}\right)$	B1
	$= 30.8 \pm 1.6449(0.54)$ $(\text{their } \bar{x}) \pm (\text{their } z)(\text{their } SE_{\bar{x}} \text{ from (a)})$	M1
	$= (29.91,\ 31.69)$ (awrt **29.9**, awrt **31.7**)	A1
		[3]
(c)	Let X = number of confidence intervals containing μ	
	or Y = number of confidence intervals not containing μ	
	So $X \sim \text{Bin}(4, 0.9)$ or $Y \sim \text{Bin}(4, 0.1)$	M1
	$P(X \geqslant 3)$ or $P(Y \leqslant 1) = {}^4C_3(0.9)^3(0.1) + (0.9)^4$ ${}^4C_3(0.9)^3(0.1) + (0.9)^4$	A1 oe
	$= 0.2916 + 0.6561 = 0.9477$ **0.9477** or **0.948**	A1
		[3]
		(Total 9)

Notes

(a)

B1: $\bar{x} = 30.8$ may be implied by $1.96\left(\dfrac{\sigma}{\sqrt{n}}\right) = [31.86 - 30.8] = 1.06$ or $2(1.96)\left(\dfrac{\sigma}{\sqrt{n}}\right) = 31.86 - 29.74$

M1: A correct equation for either a width or a half-width involving a z-value $1.5 \leqslant z \leqslant 2$

Eg: $"\text{their } z"\left(\dfrac{\sigma}{\sqrt{n}}\right) = 31.86 - "30.8"$ ft their $\bar{x}$ or $2("\text{their } z")\left(\dfrac{\sigma}{\sqrt{n}}\right) = 31.86 - 29.74$

or $"\text{their } z"(SE_{\bar{x}}) = 31.86 - "30.8"$ or $2("\text{their } z")(SE_{\bar{x}}) = 31.86 - 29.74$ are fine for M1.

A1: 0.54 or awrt 0.54 Must be seen as final answer to (a) NB $\dfrac{53}{98}$ as final answer is A0

Condone $\bar{x} \pm 1.96\sigma = ...$ for B1 and M1 but A0 even if they say "σ = standard error = 0.54"
Otherwise answer only of 0.54 scores 3 out of 3

(b)

B1 for use of 1.6449 or better in an attempt at a CI formula. Need at least 1.6449 (their SE)
M1 for attempt at CI ft their values and provided $1 \leqslant z \leqslant 1.7$

(c)

M1: States or applies either $X \sim \text{Bin}(4, 0.9)$ or $Y \sim \text{Bin}(4, 0.1)$
Condone M1 for $0.9^4 + 0.9^3 \times 0.1$ (o.e.)

1st A1: ${}^4C_3(0.9)^3(0.1) + (0.9)^4$ or $(0.9)^4 + {}^4C_1(0.1)(0.9)^3$ oe

2nd A1: 0.9477 or 0.948

G. B. Attwood 30/05/15

Pearson Education Limited. Registered company number 872828
with its registered office at 80 Strand, London, WC2R 0RL, United Kingdom

Mark Scheme (Results)

Summer 2015

Pearson Edexcel GCE in Statistics 3 (6691/01)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2015
Publications Code UA042717

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

PEARSON EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75

2. The Edexcel Mathematics mark schemes use the following types of marks:

- **M** marks: Method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
- **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
- **B** marks are unconditional accuracy marks (independent of M marks)
- Marks should not be subdivided.

3. Abbreviations

   These are some of the traditional marking abbreviations that will appear in the mark schemes.

   - bod – benefit of doubt
   - ft – follow through
   - the symbol $\surd$ will be used for correct ft
   - cao – correct answer only
   - cso - correct solution only. There must be no errors in this part of the question to obtain this mark
   - isw – ignore subsequent working
   - awrt – answers which round to
   - SC: special case
   - oe – or equivalent (and appropriate)
   - d… or dep – dependent
   - indep – independent
   - dp decimal places
   - sf significant figures
   - * The answer is printed on the paper or ag- answer given
   - $\lfloor$ or d… The second mark is dependent on gaining the first mark

4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.

6. If a candidate makes more than one attempt at any question:
    - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
    - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.

7. Ignore wrong working or incorrect statements following a correct answer.

June 2015
6691 S3
Mark Scheme

Question Number	Scheme	Marks
1. (a)	Label all the books from 1 – 160 (o.e.) Use random numbers to select the 10 books	B1 B1 (2)
(b)	<table><tr><td>Book</td><td>*A*</td><td>*B*</td><td>*C*</td><td>*D*</td><td>*E*</td><td>*F*</td><td>*G*</td><td>*H*</td><td>*I*</td><td>*J*</td></tr><tr><td>Borrow rank</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr><tr><td>Page rank</td><td>1</td><td>6</td><td>4</td><td>2</td><td>8</td><td>3</td><td>10</td><td>7</td><td>5</td><td>9</td></tr><tr><td>d^2</td><td>0</td><td>16</td><td>1</td><td>4</td><td>9</td><td>9</td><td>9</td><td>1</td><td>16</td><td>1</td></tr></table> M1 M1 $r_s = 1 - \frac{6\times 66}{10(100-1)}, \left[= 1 - 0.4\right]$ = 0.6 **0.6** M1,A1 (4)
(c)	H_0: $\rho = 0$ H_1: $\rho > 0$ Critical value is 0.5636 0.6 > cv so significant result and sufficient evidence to reject H_0 There is support for the librarian's belief or there is evidence of a correlation between the number of pages in a book and the number of times it is borrowed.	B1 B1 B1ft (3) **Total 9**

Notes

(a)

1st B1 for labelling\numbering\listing\using sampling frame of all 160 books

2nd B1 for use of random numbers\selection and mentioning the number 10

(b)

1st M1 for an attempt to rank the number of pages (at least 4 correct) Allow reverse ranks

2nd M1 for attempt at d^2 row (may be implied by sight of $\sum d^2 = 66$ or 264 for reverse ranks)

3rd M1 for use of the correct formula, follow through their $\sum d^2$ if clearly stated

If answer is not correct, a correct expression is required.

A1 for 0.6 (or -0.6 for reverse ranks)

(c)

1st B1 for both hypotheses in terms of ρ, one tail H_1 (compatible with ranks) Allow use of ρ_s

Hypotheses just in words e.g. "no correlation" score B0.

2nd B1 for cv of 0.5636 [If they have a two tail H_1 then allow 0.6485]

Allow $\pm$ for reverse ranking but must be same sign as r_s

If hypotheses are the wrong way around this must be B0 but 3rd B1 is possible.

3rd B1ft for a correct contextualised comment. Must mention "librarian" (or he)

or "number of pages" and "borrowing"

Follow through their r_s and their cv (provided it is |cv| <1)

Don't insist on the word "positive" or "negative" for a one-tailed test

Use of "association" is B0

Independent of 1st B1 so if $|r_s|$ > |cv| must say there is sufficient evidence of …….(o.e.)

and if $|r_s|$ < |cv| must say insufficient evidence of … (o.e.) regardless of their hypotheses

Question Number	Scheme	Marks
2. (a)	$H_0 : \mu_g - \mu_s = 1.5$ [g = in a group, s = on their own]	B1
	$H_1 : \mu_g - \mu_s > 1.5$	B1
	s.e. = $\sqrt{\frac{2.1^2}{80} + \frac{1.4^2}{65}} = \left[\sqrt{0.08527...}\right]$ = [0.292]	M1
	$z = \frac{8.7 - 6.6 - 1.5}{"\sqrt{\frac{2.1^2}{80} + \frac{1.4^2}{65}}"}$	dM1
	= 2.0546…. awrt 2.05(5)	A1
	cv 1% one tailed = 2.3263	B1
	Not significant, accept H_0	dM1
	Insufficient evidence that using plan as part of a group leads to weight loss of more than 1.5 kg than using plan on one's own or researcher's belief not supported	A1ft
		(8)
(b)	Since sample is large Central Limit Theorem (CLT) applies	B1
	No need to assume normal distribution	dB1
		(2)
		Total 10
	Notes	
(a)	1st & 2nd B1 for hypotheses. Accept μ_1, μ_2 or μ_A, μ_B etc if there is some indication of which is which e.g. $G \sim N(\mu_g, 8.7)$	
	1st M1 for an attempt at se with 3 out of 4 values correct. Condone switching 2.1 and 1.4 $\sqrt{\frac{2.1^2 \text{ or } 1.4^2}{80} + \frac{1.4^2 \text{ or } 2.1^2}{65}}$	
	2nd dM1 dependent on 1st M1 for a correct numerator(must have -1.5) and ft their se.	
	1st A1 for awrt 2.05	
	3rd B1 for ± 2.3263 or better seen or probability of awrt 0.02	
	3rd dM1 dep. on 1st M1 for a correct statement based on their normal cv and their test statistic	
	2nd A1ft for correct comment in context. Must mention "plan" and "group or individual" and "1.5" or "researcher" and "belief or claim"	
	NB Use of cv for difference in means D will have D = 1.5 + 2.3263 × s.e. = awrt 2.18 and requires sight of d = 2.1 with a comment for the 3rd M1	
(b)	1st B1 for mentioning "large samples" and "CLT"	
	2nd dB1 dependent on 1st B1 for stating **no need to** assume normality (since CLT assures it)	

Question Number	Scheme	Marks
3. (a)	Label staff (from 1 – 16) and children (from 1 – 40) Use random numbers to select 4 staff and 10 children	B1 B1 B1 (3)
(b)	$\bar{x} = \hat{\mu} =$ 31.2142… **awrt** **<u>31.2</u>** $s^2 = \frac{26983 - 14 \times "31.2..."^2}{13}$ = 1026.33… **awrt** **<u>1030</u>**	B1 M1 A1ft A1 (4)
(c)	$\frac{"\sqrt{1026.33...}"}{\sqrt{14}}$, = 8.562.. **awrt** **<u>8.56</u>**	M1, A1 (2)
(d)	The variation within each stratum is quite small (o.e.) **The difference in the means will be quite large**, (so variations from the overall mean will be large giving a larger overall s.e.)	B1 B1 (2) **Total 11**

Notes

(a)

1st B1 for labelling\numbering\listing staff <u>and</u> children

2nd B1 for use of random numbers or "randomly select" in <u>each group</u> (may be implied)

3rd B1 for selecting the correct number of staff <u>and</u> children

e.g. randomly select 4 staff and 10 children scores 2nd and 3rd B marks since randomly selecting and the "each group" is implied,

(b)

B1 for awrt 31.2

M1 for a correct expression ft their $\bar{x}$ and allow transcription error in $\sum x^2$ e.g. 29683

1st A1ft for a fully correct expression ft their $\bar{x}$ only

2nd A1 for awrt 1030

(c)

M1 for attempting $\frac{"\text{their } s"}{\sqrt{14}}$ (must have 14)

A1 for awrt 8.56

(d)

1st B1 for a suitable comment about **variation (se)** suggesting that **variation (se) within strata is less than that overall**

2nd B1 for a suitable **reason** about **means**, pointing out that the individuals' weights will vary a lot from the overall mean and so overall s.e. will be higher.

Question Number	Scheme	Marks
4. (a)	$H_0 : \mu = 0.5$ $H_1 : \mu \neq 0.5$	B1
	(Significance level =)10%	dB1
	(0.5 is in the interval so not significant, accept H_0, can accept) $\mu = 0.5$	B1
		(3)
(b)	$1.6449 \times \frac{\sigma}{\sqrt{100}} = 0.0247$	M1
		B1
	$\sigma = 0.15016$ or $\frac{10 \times 0.0247}{1.6449}$ (awrt 0.15)	A1
	$0.479 \pm 1.96 \times \frac{"\sigma"}{\sqrt{150}}$	M1
		B1
	awrt **(0.455, 0.503)**	A1
		(6)
		Total 9
	Notes	
(a)	1st B1 for both hypotheses in terms of μ.	
	2nd dB1 for 10% but accept 5% if they have a one-tail test as H_1	
	3rd B1 for a correct comment leading to accepting H_0	
	Ignore any 'further calculations'.	
(b)	1st M1 for $z\frac{\sigma}{\sqrt{100}} = k$, using $n = 100$ and where $\lvert z \rvert > 1.5$ and $0.02 < k < 0.03$	
	1st B1 for 1.6449 or better in an attempt (could be $1.6449\sigma = k$ or even $1.6449\ \sigma^2 = k$)	
	1st A1 for a correct expression for σ e.g. awrt 0.15	
	2nd M1 for $\bar{x} \pm z \times \frac{\sigma}{\sqrt{150}}$ for any z (> 1) and ft their σ and allow $\bar{x} \in (0.4633, 0.5127)$	
	Allow use of letter σ without a value.	
	2nd B1 for 1.96 or better in an attempt (could be 1.96σ or even $1.96\ \sigma^2$)	
	2nd A1 for awrt 0.455 and awrt 0.503	

Question Number	Scheme	Marks
5 (i)	Let $R = B_1 + B_2 + B_3 + B_4 + B_5 - 5H$ so E(R) = -25 (o.e.)	B1
	Var(R) = $5\times 6^2 + 5^2 \times 4^2$ $\quad R$~N(-25, $\sqrt{580}^2$)	M1A1
	P(R > 0) = P(Z > $\frac{0--25}{\sqrt{580}}$) = P(Z > 1.04), = 0.149619…(calc) <u>or</u> 0.1492 (tables)	dM1 A1 (5)
(ii)(a)	$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{5}\right)$	B1
	Var(D) = $\sigma^2 + "\frac{\sigma^2}{5}"\left[=\frac{6\sigma^2}{5}\right]$, so $\quad D \sim \mathrm{N}\left(0, \frac{6\sigma^2}{5}\right)$	M1, A1 (3)
(b)	P($Y_1 > \bar{X} + \sigma$) = P($D > \sigma$) = $\mathrm{P}\left(Z > \frac{\sigma}{\sqrt{\frac{6}{5}}\sigma}\right)$	M1
	= P(Z > 0.912…) = 0.181(3 dp) **(*)**	A1cso (2)
(c)	Since U_1 and $\bar{U}$ are not independent (so variance formula cannot be used)	B1
	Can be implied e.g. U_1 used to calculate $\bar{U}$, U_1 and $\bar{U}$ from same sample o.e.	(1)
(d)	Let $F = U_1 - \bar{U} = U_1 - \frac{(U_1+U_2+U_3+U_4+U_5)}{5}, = \frac{4U_1 - (U_2+U_3+U_4+U_5)}{5}$	M1, A1
	Var(F) = $\frac{4^2\sigma^2 + 4\sigma^2}{5^2}$ = $0.8\sigma^2$,so $\quad F$ ~ N(0, $0.8\sigma^2$)	dM1, A1
	P($F > \sigma$) = $\mathrm{P}\left(Z > \frac{\sigma}{\sigma\sqrt{0.8}}\right)$ = P(Z > 1.118..)	M1
	= 0.1314 (tables) or 0.131776…(calc) **awrt 0.131~0.132**	A1cso (6)
		Total 17

	Notes
(i)	1st B1 for E(R) = – 25 (or 25 if their R is defined the other way around) 1st M1 for an attempt at Var(R) = 5Var(B) + 25Var(H). Condone swapping of 6^2 and 4^2 1st A1 for normal and correct variance (ft their mean) 2nd dM1 for attempting the correct probability and standardising with their mean and sd. This mark is dependent on 1st M1 so if R is not being used or M0 for variance score M0 If their method is not crystal clear then they must be attempting P(Z > +ve value) o.e 2nd A1 for answer in the range [0.149, 0.150]
(ii)(a)	B1 for correct distribution of $\bar{X}$ (may be implied for a correct answer for D) M1 for correct attempt at Var(D) (ft their Var($\bar{X}$)) [A1 needs must be fully correct]
(ii)(b)	M1 for expressing probability in terms of D and standardising A1cso for seeing P(Z> 0.912..) or prob of 1 – 0.8186 (tables) or 0.180655…(calc)
(c)	B1 correct statement that should mention U_1 and $\bar{U}$
(d)	1st M1 for forming an expression in terms of U_1...U_5 only 1st A1 for collecting U_1 terms and getting in a form where Var($aX \pm bY$) can be used. 2nd dM1 for a correct expression for Var(their F). Dependent on 1st M1. 2nd A1 for a correct distribution for F 3rd M1 attempting a correct prob and standardising using their Var(F), σ must cancel 3rd A1cso for awrt 0.131 or 0.132

Question Number		Scheme	Marks
6.	**(a)**	H_0 : U[0, 10] is a suitable model H_1: U[0, 10] is not a suitable model	B1
		Values of D Expected Freq	B1 M1A1
		4^{th} or 5^{th} col	M1
		$\chi^2 = 13.65$	A1
		$\nu = 3, \quad \chi_3^{\,2}(1\%) = 11.345$	B1, B1
		[Reject H_0 ,] the uniform distribution over [0, 10] is not a suitable model	A1
			(9)
	(b)	Area $\propto \pi R^2$ so r = 81, – 49 = **32**	M1, A1
		s = 100 – "32" – 49 or 100 – 81 = **19**	B1ft
			(3)
	(c)	Not significant, Henry's model is suitable	M1, A1
			(2)
	(d)	H_0 : The colour\region chosen for the points is independent of gender(or no assoc')	B1
		H_1 : The colour\region chosen for the points is dependent on gender(or assoc')	
			(1)
	(e)	$\frac{39 \times 65}{100}$	B1
			(1)
	(f)	Expected frequency for Yellow and Boys is 4.9 < 5 so col. must be pooled/combined.	B1
		[This gives a 2×3 table so $\nu = (2-1) \times (3-1) = 2$]	
			(1)
	(g)	cv = 4.605	B1
		[Not significant] so the data do not support Phoebe's belief oe	B1
			(2)
			Total 19

	Notes
(a)	2^{nd} B1 for the correct values for D (can be implied by 40, 30, 20, and 10.) 1^{st} M1 for at least 2 expected frequencies or clear use of a correct formula e.g. $0.4N$ 1^{st} A1 for all the correct E_i 2^{nd} M1 for at least 2 correct calculations from 4^{th} or 5^{th} column 2^{nd} A1 for a test statistic of 13.65 (accept 13.7 to 3 sf) Awrt 13.7 only scores 2^{nd} B1M1A1M1A1 3^{rd} A1 for a correct conclusion rejecting the uniform model. Award provided their test statistic > 11.345
(b)	M1 for some attempt to use πR^2 to find r
(c)	M1 for a correct statement that it is not significant A1 for correctly stating that Henry's model is suitable o.e.
(d)	B1 Independence or association mentioned at least once if ditto marks used. Allow connection but not correlation.
(f)	B1 for recognising there is an $Ei < 5$ **and** need for pooling/combining oe
(g)	2^{nd}B1 for correctly stating that Phoebe's belief is not supported by the data oe (depends on their cv being > 1.411)

Pearson Education Limited. Registered company number 872828
with its registered office at 80 Strand, London, WC2R 0RL, United Kingdom