Version 1.0

General Certificate of Education
June 2010

Statistics SS04

Statistics 4

Mark Scheme

Mark schemes are prepared by the Principal Examiner and considered, together with the relevant questions, by a panel of subject teachers. This mark scheme includes any amendments made at the standardisation meeting attended by all examiners and is the scheme which was used by them in this examination. The standardisation meeting ensures that the mark scheme covers the candidates' responses to questions and that every examiner understands and applies it in the same correct way. As preparation for the standardisation meeting each examiner analyses a number of candidates' scripts: alternative answers not already covered by the mark scheme are discussed at the meeting and legislated for. If, after this meeting, examiners encounter unusual answers which have not been discussed at the meeting they are required to refer these to the Principal Examiner.

It must be stressed that a mark scheme is a working document, in many cases further developed and expanded on the basis of candidates' reactions to a particular paper. Assumptions about future mark schemes on the basis of one year's document should be avoided; whilst the guiding principles of assessment remain constant, details will change, depending on the content of a particular examination paper.

Further copies of this Mark Scheme are available to download from the AQA Website: www.aqa.org.uk

Set and published by the Assessment and Qualifications Alliance.

The Assessment and Qualifications Alliance (AQA) is a company limited by guarantee registered in England and Wales (company number 3644723) and a registered charity (registered charity number 1073334).
Registered address: AQA, Devas Street, Manchester M15 6EX

Key to mark scheme and abbreviations used in marking

M	mark is for method
m or dM	mark is dependent on one or more M marks and is for method
A	mark is dependent on M or m marks and is for accuracy
B	mark is independent of M or m marks and is for method and accuracy
E	mark is for explanation

√or ft or F	follow through from previous incorrect result	MC	mis-copy
CAO	correct answer only	MR	mis-read
CSO	correct solution only	RA	required accuracy
AWFW	anything which falls within	FW	further work
AWRT	anything which rounds to	ISW	ignore subsequent work
ACF	any correct form	FIW	from incorrect work
AG	answer given	BOD	given benefit of doubt
SC	special case	WR	work replaced by candidate
OE	or equivalent	FB	formulae book
A2,1	2 or 1 (or 0) accuracy marks	NOS	not on scheme
$-x$ EE	deduct x marks for each error	G	graph
NMS	no method shown	c	candidate
PI	possibly implied	sf	significant figure(s)
SCA	substantially correct approach	dp	decimal place(s)

No Method Shown

Where the question specifically requires a particular method to be used, we must usually see evidence of use of this method for any marks to be awarded. However, there are situations in some units where part marks would be appropriate, particularly when similar techniques are involved. Your Principal Examiner will alert you to these and details will be provided on the mark scheme.

Where the answer can be reasonably obtained without showing working and it is very unlikely that the correct answer can be obtained by using an incorrect method, we must award **full marks**. However, the obvious penalty to candidates showing no working is that incorrect answers, however close, earn **no marks**.

Where a question asks the candidate to state or write down a result, no method need be shown for full marks.

Where the permitted calculator has functions which reasonably allow the solution of the question directly, the correct answer without working earns **full marks**, unless it is given to less than the degree of accuracy accepted in the mark scheme, when it gains **no marks**.

Otherwise we require evidence of a correct method for any marks to be awarded.

SS04

Q	Solution	Marks	Total	Comments
1(a)	Approximate 99% confidence interval			
	SD = $\sqrt{24}$ used	B1		$\sqrt{24}$ used for s.d.
	$24 \pm Z\text{-value} \times \sqrt{24}$	M1		C1 method, recognisable Z, their s.d.
	Z = 2.5758	B1		Accept 2.58
	24 ± 12.6 or $11.4 \sim 36.6$	A1	4	$24 \pm (12.6 \sim 12.7)$ or $(11.3 \sim 11.4)$ to $(36.6 \sim 36.7)$
(b)	Since 17 lies within the interval	B1		17 lies within CI
	Editors claim incorrect/not justified No evidence of a significant increase in mean number of births	E1√	2	correct conclusion their CI
	Total		**6**	
2(a)(i)	P(Spelling Error) constant/same, or Errors are independent, or Number words per page constant.	E1	1	constant p/errors independent/ n constant (fixed) "Random Sample" E0
(ii)	Poisson approx to B(n = 200, p = 0.0035)	B1		Poisson, mean $200 \times 0.0035 = 0.7$
	$P(X > 2) = 1 - P(X \leqslant 2)$	M1		Attempt, $1 - p(0,1,2)$ or $1 - p(0,1)$ $1 - 0.9659$
	$1 - 0.4966 - 0.3476 - 0.1217$			$1 - 0.8442\ (= 0.156)$
	$1 - 0.9659 = 0.0341$	A1	3	$0.0341\ (0.034 \sim 0.0342)$
(b)(i)	Y=Grammatical errors ~ Poisson(274)	B1	1	Poisson ($np = 365 \times 0.75 = 273.75$), Implied by "normal" & N(274, 274) used in (ii)
(ii)	Y ~ approx N(274, 274)	M1		Normal $\mu = \sigma^2$ = their 365×0.75 $\sigma = \sqrt{273.75} = 16.545$
	P(Y < 300)= P(Z < 299.5 -273.75)/16.545	m1		Standardise 300, ± Z, ignore c.c.
	Z = (±) 1.5563	m1		Correct use of c.c. (± Z)
	P(X < 300) = 0.940	A1	4	0.940 (0.939 ~ 0.941)
	Total		**9**	

SS04 (cont)

Q	Solution	Marks	Total	Comments
3(a)	H_0: $\mu = 1000$ H_1: $\mu < 1000$	B1		Both hypotheses correct
	$\bar{x}$ = 995.5 & s = 32.725	B1		(995 ~ 996) & (32.7 ~ 32.8) NB $\sigma_n = 31.05$
	Use of SD/ $\sqrt{10}$	M1		Use of their s.d./$\sqrt{10}$
	$t = \pm(995.5 - 1000)/\left(32.725/\sqrt{10}\right)$	m1		Method for ts, - ignore sign use of σ_n gives t = –0.458, m0
	t = – 0.435	A1		-0.435 (-0.43 ~ -0.44)
	ν = 9 df	B1		9 df, may be implied (2.5% = 2.262)
	c.v. t_9 (-)1.833	B1$\surd$		-1.833, their df (5% point)
	Accept H_0 Howard's claim confirmed. There is no significant evidence the mean weight of bags less than 1000g.	A1$\surd$	8	Their –ve ts versus -1.83(3) not inconsistent with their H_0 .
(b)	Type 2 error,	B1		Type II error
	Since Jean suspected that an incorrect null hypothesis had been accepted.	E1dep	2	Correct explanation, false H_0 accepted, dependent on B1
(c)	The sample mean is > 1000g , so the test statistic will be positive.	E1		Sample mean > 1000g, +ve ts
	The critical value is negative so H_0 cannot be rejected.	E1	2	Complete explanation, +ve ts versus -ve CV
(d)	If sample is not random the conclusions are unreliable	E1		Invalid, unreliable- reliability affected
	Manager may select only relatively heavy bags for the sample.	E1	2	Justification, sample may be biased, by selection of heavy bags
	Total		**14**	

SS04 (cont)

Q	Solution	Marks	Total	Comments
4(a)	H_0: p = 0.2 H_1: p < 0.2	B1		Both hypotheses
	B(25, 0.2)	B1		Attempted use of B(25,0.2)
	$P(X \le 1)\{\ P(\ 0,1\ ) = 0.00378, 0.0236\ \}$	M1		$P(X \le 1)$ or P(0), lower tail prob. M0 for Normal approx
	= 0.0274	A1		0.0274 (0.027 ~ 0.0275)
	Reject H_0, since 0.0274 < 0.05	A1√		Conclusion-their p vs 0.05 (5%)
	Simone correct. Less than 20% of members will attend the AGM.	A1	6	Completely correct, conclusion in context
(b)	p = 11/235 = 0.046808	B1		11/235, acf , 0.0468 – 0.047
	Z = (±) 1.6449	B1		1.64 ~ 1.65
	$SE(p) = \sqrt{\{p\times\ (1 - p)/n\}} = 0.01378$	M1		Method for SE(p)
	$CI = p \pm Z\times \sqrt{\{p\times\ (1 - p)/n\}}$	m1		CI method - allow incorrect Z value
	$0.04681 \pm 1.645\times 0.01378$			(0.0468 ~ 0.0470) ± (0.0225 ~ 0.0230) or
	0.0468 ± 0.0227 or 0.024 ~ 0.069	A1	5	(0.024 ~ 0.025) to (0.069~0.070)
(c)	Expected attendance = $7090\times p$ vs 600	E1		7090 × (a probability) or equivalent
	Attendance < 7090 × ('their 0.0694') = 492	B1√		7090 × their upper limit for p
	Room large enough as 492 < 600 Simone correct	B1	3	Completely correct 492 (489 to 496) < 600, room OK
	Equivalent argument based on proportion attending			p = 600/7090 (0.085, 8.5%) E1 600/7090 > their upper limit B1 0.085 > 0.069 ~ 0.070 ⇒ OK B1
			14	

SS04 (cont)

Q	Solution	Marks	Total	Comments
5(a)(i)	H_0: P = 0.1 H_1: p $\neq$ 0.1	B1		Both hypotheses
	B(234,0.1)	B1		B(234, 01)identified, implied by use
	$X \sim N(NP, NPQ)$ or $p \sim N(P, PQ/N)$	M1		Attempt normal approx X. their p (gen)
	$\mu = 23.4$, or 0.1 $\sigma^2 = 234(0.1)(0.9)$ or $(0.1)(0.9)/234$	A1		P = 0.1 used for mean, unsimplified var/sd $\mu = 23.4$, $\sigma^2 = 21.06\ (\sigma = 4.589)$ $\mu = 0.1$, $\sigma^2 = 0.01961$, $(\sigma = 0.0003846)$
	Z =(25.5 – 23.4)/4.589 = 0.458 or (26.0 – 23.4)/4.589=0.567 or (0.1111 – 0.1)/0.01961= 0.566	m1		Correct method P= 0.1,ignore CC($\pm$)Z
	Z =0.458 or 0.566 or 0.567	A1		$(0.455 \sim 0.460)$; $(0.565 \sim 0.570)$
	c.v. ± 1.6449	B1		$(\pm)1.6(4 \sim 5)$ or $1 - \Phi(Z)$ vs 0.05 etc
	Accept H_0	A1$\surd$		Conclusion, their Z stat vs correct CV
	Manager correct (do not reject claim 10% of guests require breakfast before 7)	A1	9	Completely correct, conclusion in context $1 - \Phi(z)$: 0.3228 – 0.3246, 0.2843 – 0.2860
	SC: B > P >N:B1 B1 M1 B1 only			SC Normal approx to Poisson approx
(ii)	Guests sharing rooms/work groups may breakfast together etc.	E1		Reason in context for non independence
	Probability may change at weekend, or for reason for stay (business or leisure)	E1	2	Reason in context for non constant p
(b)(i)	$\bar{x}$ = 112.67 and s = 70.61	B1		113(112 ~ 113) & 70.6 (70.5 ~ 70.7)
	$t_{0.025}$ = 2.306	B1		2.306, accept 2.31
	95% CI: $112.67 \pm 2.306 \times 70.61/\sqrt{9}$	M1		CI method, their sd. & t-value
	$\{$ m0 for use of $\sigma_n = 66.57 \div \sqrt{9}$ $\}$	m1		Completely correct method
	112.7 $\pm$ 54.3 or 58.4 ~ 167.0	A1	5	$(112.6 \sim 113.0) \pm 54.3$ $(58.3 \sim 58.7)$ to $(166.8 \sim 167.3)$
(ii)	Times from a normal distribution	B1	1	Sample/data/minutes ~ Normal, disallow “it”, “values”, “normal” alone
(iii)	CI is for (population) mean	E1		CI is for mean
	No reason why 95% of individual times for members should lie within it	E1dep	2	Should not be concerned, not surprising, dependent on previous E
	Total		**19**	

SS04 (cont)

Q	Solution	Marks	Total	Comments
6(a)(i)	Total time, T is normally distributed	B1		Normal implied by use in (ii)
	Mean 12.5 + 9.6 + 19.0 = 41.1	B1		41.1 CAO
	Variance = $1.5^2 + 1.3^2 + 1.9^2 = 7.55$	B1	3	7.55 CAO, sd = 2.75 AWRT
(ii)	Z = ± (42 - 41.1)/2.748 = 0.3275	M1		Standardise 42, their mean, s.d. NO cc
	P(T < 42) = P(Z< 0.3275) = 0.628	A1	2	0.628 (0.625~0.63)
(b)(i)	C < 16.5 required or equivalent	B1		16.5 or conditional $\mu = 42(\text{F})\,\&\,41.5(\text{I})$
	P(C < 16.5) Or P(T < 39$\mid \mu = 42$; 41.5 for F; I) Ian $Z_1 = -2.5/1.9 = -1.315$ Fred $Z_F = -3.0/3.2 = -0.9375$	M1		Method using C or T (at least one variance correct, F or I) $z_1 = (16.5 - 19)/1.9 = -1.315$ or equiv $z_F = (16.5 - 19.5)/3.2 = -0.9375$
	P(T < 39; C < 16.5): Ian = $1 - 0.906 = 0.094$ AND Fred = $1 - 0.825 = 0.175$	A1		$(0.093 \sim 0.095)$ $(0.173 \sim 0.177)$ Both
	(or valid argument using correct Zs)			$(-1.31 \sim -1.32) < (-0.935 \sim -0.940)$
	Fred more likely to beat 39 minutes	B1	4	Fred most likely, ignore method
(ii)	Times for cycling section Fred – Ian			Consider F – I or I – F
	$\text{E}(\text{F}–\text{I}) = 19.5 - 19 = 0.5$	M1		Method for mean & variance (s.d.)
	$\text{V}(\text{F}–1) = 3.2^2 + 1.9^2 = 13.85$			SD = 3.722, equivalently I-F
	P(F > I) = P(Z > – 0.5/3.722)	m1		Standardise 0, allow $\pm Z$
	Z = (±) 0.134	A1		$(\pm)0.130 \sim 0.140$
	P(F > I) = P(Z > – 0.134) = 0.553	A1	4	0.553 (0.55 ~0.56)
	Total		**13**	
	TOTAL		**75**	