publication

edexcel

FINAL

June 2005
6685 Statistics S3
Mark Scheme

Question Number	Scheme	Marks
1.	(a) Population divides into mutually exclusive; groups (distinct) (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 errors – subject to possible interviewer bias Any ONE – non-response not recorded	B1 (2)
2.	$X \sim N(10, 3^2) \quad \therefore \quad \bar{X} \sim N(10, \frac{9}{5})$ Can be implied 10; 9/5	B1; B1
	$P(7 \le \bar{X} \le 10) = P\left(\frac{7-10}{\sqrt{9/5}} < Z < 0\right)$ Standardising with 10 & their σ	M1 A1
	$= P(-2.236 < Z < 0)$	
	$= \Phi(0) - \{1 - \Phi(2.24)\}$	M1 ($p<0.5$)
	$= \underline{0.4875}$	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

3×6 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

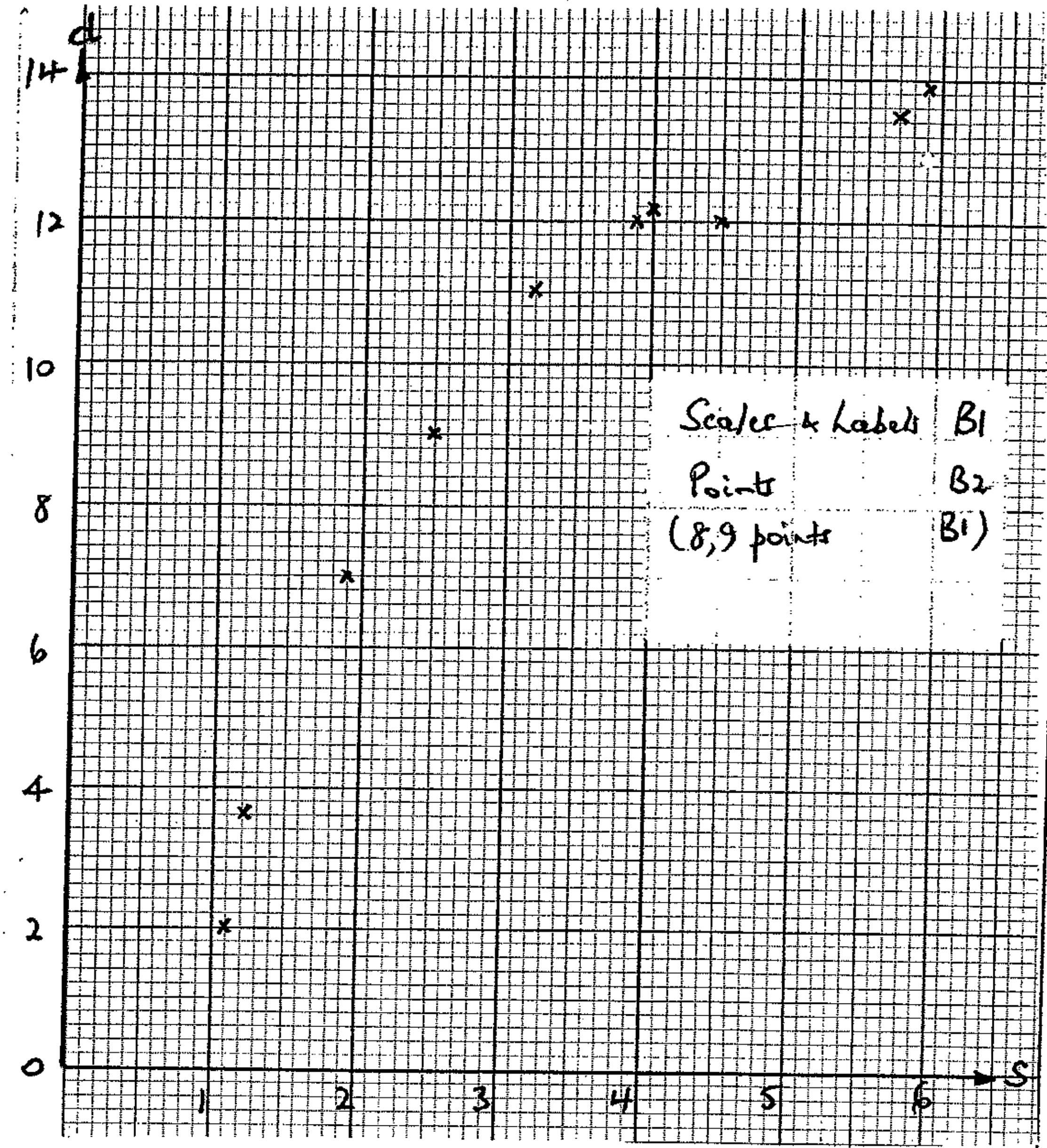
$= 3.47619....$ AWRT 3.48 A1

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	
	$= 0.93494....$ 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 & 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 H_1: Poisson distribution is not a suitable model (both)	B1
	$\hat{\lambda} = \frac{(0\times99)+(1\times65)+\dots+(4\times2)}{200} = \frac{153}{200} = 0.765$	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
	(table below)	A1, A1 (0,1,2) (-1 e.e.) ; A1
	$\nu = 4-1-1 = 2$; CR: $\chi^2_2 > 5.991$ from Poisson $\nu = 4-1 = 3$ CR: $\chi^2 > 7.815$ from Poisson (0.765)	B1; B1√
	$\sum\frac{(O-E)^2}{E} = 5.47368\ldots$ Use of $\sum(O-E)^2/E$	M1
	5.40 – 5.50	A1
	5.47 is not in the critical region. Number of computer failures per day can be modelled by a Poisson distribution	A1√ (12)

X	Observed Frequency	Expected Frequency
0	99	93.06678....
1	65	71.19604....
2	22	27.23250...
3	12 } 14	6.94428... } 8.50468
≥4	2	1.56040...

Question	Scheme	Marks
6.	(a) Let X represent repair time	
	$\therefore \sum x = 1435 \quad \therefore \bar{x} = \frac{1435}{5} = \underline{287}$	B1
	$\sum x^2 = 442575 \quad \therefore s^2 = \frac{1}{4}\left\{442575 - \frac{1435^2}{5}\right\}$	M1 A1
	$= \underline{7682.5}$	A1 (4)
	(b) $P(\lvert \mu - \hat{\mu} \rvert) < 20 = 0.95$ — Use of 19.20 or 40 with their σ & $\sqrt{n}$	[redacted] M1
	1.96	B1
	$\therefore \frac{20}{\sigma/\sqrt{n}} = 1.96$	[redacted] A1
	Solving for n	M1
	$\therefore n = \frac{1.96^2 \sigma^2}{20^2} = \frac{1.96^2 \times 100^2}{400} = \underline{96.04}$	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	M1
	0; 16	A1; A1
	$\therefore P(\lvert W \rvert > 6) = 2P(W > 6)$	M1
	$= 2 \times P\left(Z > \frac{6-0}{\sqrt{16}}\right)$ — Standardising, their σ	M1
	$= 2 \times P(Z > 1.5)$	
	$= 2 \times (1 - 0.9332) = \underline{0.1336}$	A1 (6)
	NB $W = C - L$ treat as MR; Prob $= 0.4346$	
	(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 $(p > 0.5)$
	$= \underline{0.8413}$	A1 (6)

	(c) Let $W = C_1 + \dots + C_{24} + B$	
	$\therefore E(W) = 24 \times 350 + 100 = \underline{8500}$	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)

13/06/05