EDEXCEL STATISTICS S3 (6685) – JUNE 2003 PROVISIONAL MARK SCHEME

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 $\bar{X} \sim (124, \frac{20^2}{30}$ 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

EDEXCEL STATISTICS S3 (6685) – JUNE 2003 **PROVISIONAL MARK SCHEME**

Question number	**Scheme**		**Marks**
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)	$X\text{-}Y \sim N(10,9)$	Implied	**B1∫**
	P(13<X-Y<16)=P(X-Y<16)-P(X-Y<13)	Subtract	**M1**
	$= P(Z<\frac{16\text{-}10}{3}) - P(Z<\frac{13\text{-}10}{3})$	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 A1∫**

(2)

(e) $E(\bar{X}) = 2\times 0.25 + 6\times 0.2 ..... = 18$ $\sum x P(X = x)$ from table, 18 **M1 A1**

$\mathrm{Var}(\bar{X}) = 2^2\times 0.25 + 6^2\times 0.2 + ... - 18^2 = 224$ $\sum x^2 P(X = x)$ - (theirs)^2, 224**M1A1**

So $E(\bar{X}) = 18 = \mu$ and $\mathrm{Var}(\bar{X}) = 224 = \frac{1}{2}\sigma^2$ as required. **A1**

(5)

17