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 any one	B1
	- non-response not an issue	
	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 any one	B1
	- non-random process	(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	B1 (2)
	NO REPETITION / OPPOSITES	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
	Application of <u>central limit theorem</u> as (sample large)	B1
		(3)
(b)	$P(\bar{X} \geq 91) = 1 - P\left(Z < \frac{91-90}{0.5}\right)$ Stand.	M1A1
	$= 1 - P(Z < 2)$	
	$= 1 - 0.9772$	
	$= 0.0228$ awrt 0.0228	A1
		(3)
		TOTAL 6
3 (a)	$H_0: \mu_A = \mu_B$, $H_1: \mu_A \neq \mu_B$ μ_1, μ_2 OK both	B1
	$se = \sqrt{\frac{47^2}{70} + \frac{23^2}{90}} \left(= \sqrt{37.43492..}\right)$	M1A1
	Test statistic is $\pm \frac{198-201}{se} = \mp 0.4903$ 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 <u>independence</u>, similarly for diet B.	
	– no chickens in common between the two samples to ensure <u>independence</u>	
	– not same chickens on diet A and diet B because if it were we need to do a <u>paired analysis</u>.	B1, B1
	Any 1	
	~~[illegible]~~	(2)
		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 on price, $\Sigma d^2 = 44$ (Hairs) — M1, A1

(a) $r_s = 1 - \dfrac{6 \times 286}{10(100-1)} = -0.73$ or $\dfrac{-11}{15}$ or -0.733 (awrt) — M1 A1

or 0.733 (awrt) 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 there 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) — (4)

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

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) $W = M_1 + \dots + M_7 + F_1 + \dots + F_8$

awrt

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

$Var(W) = 7 \times 12.6^2 + 8 \times 9.8^2 = 1879.64$ 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

awrt 1.03

$= P(Z > 1.03)$ A1

1- M1

$= 1 - 0.8485$

$= 0.1515$ 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... at least one $\frac{RTxCT}{GT}$, All	M1A1
	6 7.92 0.4654... $\frac{(O-E)^2}{E}$ 3sf or better	M1A1
	10 9.52 0.0242...	
	7 7.48 0.0308...	
	6 8.4 0.6857...	
	9 6.6 0.8727...	
	$\sum \frac{(O-E)^2}{E} = 2.4446...$ awrt Σ, 2.44	M1A1
	$\nu = (3-1)(2-1) = 2$, $\chi^2_2 = 5.991$	B1 B1√
	Insufficient evidence to reject H_0. No association between age and colour	A1√ (11)
	(cv from correct h/r for ft)	TOTAL 11
7.(a)	$\bar{x} = \frac{500}{10} = 50$	M1A1
	$s^2 = \frac{1}{9}\left(25001.74 - \frac{500^2}{10}\right) = 0.193$ awrt 0.193	M1A1A1 (5)
(b)	limits are 50 ± 1.96 Hairs 5 B1 1.96	M1B1
	$= (49.02, 50.98)$ awrt 49(0), 51(0)	A1A1 (4)
(c)	Confidence interval is $\left(50 - 2.5758 \times \frac{0.5}{\sqrt{10}}, 50 + 2.5758 \times \frac{0.5}{\sqrt{10}}\right)$ Hairs B1 2.5758	M1B1A1√
	$= (49.59273, 50.40727..)$ awrt 49.6, awrt 50.4	A1A1 (5)
	Use of estimate in (a) in (b) AND (c) assume MISREAD.	TOTAL 14

8 (a)	$B_2(5, 0.5)$	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) ✓ for $\hat{p} = 0.466$.	B1 ✓

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

100 P(X = x) 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 3sf or better. — M1A1

$$\sum \frac{(O-E)^2}{E} = 3.637\dot{3}$$

Σ required, awrt 3.64 — M1A1

$\nu = 4 - 1 = 3$, $\chi_3^2(0.10) = 6.251$ — 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, $\nu = 5$, $\chi_5^2 = 9.236$

~~[illegible]~~