June 2006
6691 Statistics S3
Mark Scheme

Question Number	Scheme	Marks
1 (a)	Advantages: – does not require the existence 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 – non-response not an issue	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 any one – non-random process	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	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. $N_g(90, 0{\cdot}25)$	M1A1
	Application of central limit theorem as (sample large)	B1
		(3)
(b)	$P(\bar{X} \geqslant 91) = 1 - P\left(Z < \frac{91-90}{0{\cdot}5}\right)$ Stand.	M1A1
	$= 1 - P(Z < 2)$	
	$= 1 - 0{\cdot}9772$	
	$= 0{\cdot}0228$ awrt $0{\cdot}0228$	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 $\pm \frac{198-201}{se} = \pm 0{\cdot}4903$ awrt $\pm 0{\cdot}49$	M1A1
	M1A1 probab awrt $0{\cdot}312$	
	B1 probab cv $0{\cdot}025$	
	cv $= (\pm) 1{\cdot}96$	B1
	Insufficient evidence to reject H_0,	
	no significant difference between the mean cholesterol	A1✓
	content of the two samples. (require correct comparison for ft) context required.	
		(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 on price, $\Sigma d^2 = 44$ — M1,A1

(a) $r_s = 1 - \frac{6 \times 286}{10(100-1)} = -0.73$ or $\frac{-11}{15}$ 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 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) — (4)

(–cv from correct table required) (position in context)

TOTAL 9

5. M = wt of male worker $\qquad M \sim N(78\cdot5,\ 12\cdot6^2)$

F = wt of female worker $\qquad F \sim N(62\cdot0,\ 9\cdot8^2)$

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

$E(W) = 7 \times 78\cdot5 + 8 \times 62\cdot0 = 1045\cdot50$ $\qquad$ awrt 1050 $\qquad$ M1 A1

$Var(W) = 7 \times 12\cdot6^2 + 8 \times 9\cdot8^2 = 1879\cdot64$ $\qquad$ 1880 $\qquad$ M1 A1 $\qquad$ (4)

(b) <u>Independent</u> (used in Variance formula) $\qquad$ B1 $\qquad$ (1)

(c) $P(W > 1090) = P\left(Z > \dfrac{1090 - 1045\cdot5}{\sqrt{1879\cdot64}}\right)$ $\qquad$ M1

$= P(Z > 1\cdot03)$ $\qquad$ awrt 1·03 $\qquad$ A1

$= 1 - 0\cdot8485$ $\qquad$ 1 − M1

$= \underline{0\cdot1515}$ $\qquad$ A1 $\qquad$ (4)

AWRT $(0\cdot152)$

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 $\frac{ETXCT}{GT}$, All $\frac{(O-E)^2}{E}$ 3sf or better	M1 A1
	6 7.92 0.4654...	M1 A1
	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	M1 A1
	$\nu = (3-1)(2-1) = 2$, $\chi_2^2 = 5.991$	B1 B1$\checkmark$
	Insufficient evidence to reject H_0.	
	No association between age and colour	A1$\checkmark$ (11)
	(cv from correct H/E for ft)	TOTAL 11
7.(a)	$\bar{x} = \frac{500}{10} = 50$	M1 A1
	$s^2 = \frac{1}{9}\left(25001.74 - \frac{500^2}{10}\right) = 0.193$ — awrt 0.193	M1 A1 A1 (5)
(b)	limits are $50 \pm 1.96\sigma$ (Hairs) — B1 1.96	M1 B1
	$= (49.02, 50.98)$ — awrt 49(0), 51(0)	A1 A1 (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	M1 B1 A1$\checkmark$
	$= (49.59273, 50.40727..)$ — awrt 49.6, awrt 50.4	A1 A1
		(5)
	Use of estimate in (a) in (b) AND (c) assume MISREAD.	TOTAL 14

8 (a) $B_?(5, 0.5)$ — M1 A1 (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

100 P(X=x) for Bin, 1 correct = A1, All correct = A1, 3sf or better — M1 A1 A1

	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. — M1 A1

$\sum \frac{(O-E)^2}{E} = 3.637\dot{3}$ — Σ required, awrt 3.64 — M1 A1

$\nu = 4 - 1 = 3$, $\chi_3^2(0.10) = 6.251$ — 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]~~