S3 June 2006 (MA)

Q1a) ADV : - sampling frame is not required
- field work can be done quickly
(any 1) as a representative sample is easily achieved
- costs kept to a minimum.
- administering the test is easy.

DISADV : - unable to estimate sampling errors.
- interviewer bias may be involved.
(any 1) - non-responses aren't recorded.

b) ADV : - free from bias
- its a random process so you can estimate
(any 1) sampling errors.

DISADV : - not suitable for large sample size
- sampling frame required; may be difficult
(any 1) to obtain.

Q2a) n is large so by C.L.T : $\bar{X} \sim N\left(90, \frac{5^2}{100}\right)$

we can apply the central limit theorem as sample size is large.

b) $P(\text{required}) = P(\bar{X} > 91) = P\left(Z > \frac{91-90}{\frac{5}{\sqrt{100}}}\right)$

$= P(Z > 2.00) = 1 - P(Z < 2)$

$= 1 - 0.9772 = \boxed{0.0228}$

we are testing for a difference so its a two-tailed test.

Q3a) $H_0: \mu_A = \mu_B$
$H_1: \mu_A \neq \mu_B$

A	B
$n = 70$	$n = 90$
$\bar{x} = 198$	$\bar{x} = 201$
$\sigma = 47$	$\sigma = 23$

$$\text{Test Stat} = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x^2} + \frac{\sigma_y^2}{n_y^2}}}$$

$$= \frac{201 - 198 - (0)}{\sqrt{\frac{47^2}{70} + \frac{23^2}{90}}} = 0.4903 //$$

Critical value: ± 1.96 //
(5% two-tail)

$0.4903 < 1.96$
$\therefore$ Result is insignificant. (accept H_0).
Evidence suggests no diference in mean cholesterol content of eggs laid by chickens on diets A and B.

b) each egg selected should be from a different chicken so that all eggs are selected independently.

Also, the chickens on diet A should be different from those on diet B - none should be on both diets as this may affect results.

(Q4a)

Shop	Distance	Price	R_D	R_P	d	d^2
A	50	1.75	1	9	8	64
B	175	1.20	2	7	5	25
C	270	2.00	3	10	7	49
D	375	1.05	4	6	2	4
E	425	0.95	5	4	1	1
F	580	1.25	6	8	2	4
G	710	0.80	7	2	5	25
H	790	0.75	8	1	7	49
I	890	1.00	9	5	4	16
J	980	0.85	10	3	7	49
						286

$\sum d^2 = 286$

$$\therefore r_s = 1 - \frac{6(286)}{10(99)} = \boxed{-0.733}$$

b) $H_0: p = 0$ critical value: ± 0.5636

$H_1: p < 0$ (5%, 1-tail)

$(r_s < 0)$ so use $(p < 0)$ as H_1.

$$-0.733 < -0.5636$$

$\therefore$ result is significant (reject H_0).

Evidence suggests there is a negative correlation between distance and price.

Q5a) $M \sim N(78.5, 12.6^2)$

not 7M and 8F; we are taking samples of M and F essentially.

$F \sim N(62, 9.8^2)$

let $X = M_1 + \dots + M_7 + F_1 + \dots + F_8$

$E(X) = 7(78.5) + 8(62) = \boxed{1045.5}$

$Var(X) = 7(12.6^2) + 8(9.8^2) = \boxed{1879.64}$

$\Rightarrow X \sim N(1045.5, 1879.64)$

b) They are independent (required for variance formulae).

c) $P(\text{max load exceeded}) = P(X > 1090) = P\left(Z > \frac{1090 - 1045.5}{\sqrt{1879.64}}\right)$

$= P(Z > 1.03) = 1 - 0.8485 = \boxed{0.1515}$

Q6) H_0: there is no association between age and colour preference.

H_1: there is an association between age and colour preference.

$$\text{Expected no.} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

$\boxed{\text{expected}}$

Age	Red	Blue
4	10.08	7.92
8	9.52	7.48
12	8.40	6.60

O_i	E_i	$\frac{(O-E)^2}{E}$
12	10.08	0.3657
6	7.92	0.4654
10	9.52	0.0242
7	7.48	0.0308
6	8.40	0.6857
9	6.60	0.8727
		2.44

$$X^2 = \sum \frac{(O-E)^2}{E} = 2.44$$

$$\nu = (3-1)(2-1) = 2$$

$$\text{critical value} = \chi^2_2(5\%) = 5.991$$

$$2.44 < 5.991$$

$\therefore$ Result is insignificant. (accept H_0).
Evidence suggests there is no association between age & colour preference.

Q7a) mean $= \frac{\sum x}{n} = \frac{500}{10} = \boxed{50}$

$\sum x^2 = 25001 \cdot 74.$

$\therefore s^2 = \frac{1}{9}\left(25001 \cdot 74 - \frac{(500)^2}{10}\right) = \boxed{0 \cdot 193}$

b) 95% C.I: $\left[\bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)\right]$

$\left[50 \pm 1.96\left(\frac{0.5}{\sqrt{10}}\right)\right]$

$\left[49.02, \quad 50.98\right]$

c) 99% C.I: $\left[\bar{x} \pm 2.5758\left(\frac{0.5}{\sqrt{10}}\right)\right]$

$\Rightarrow \left[50 \pm 2.5758\left(\frac{0.5}{\sqrt{10}}\right)\right]$

$\Rightarrow \left[49.6, \quad 50.4\right]$

Q8a) $X \sim B[5, 0.5]$ where X = no. of heads.

b) <u>H_0</u>: $B[5, 0.5]$ is a suitable model for these data.

<u>H_1</u>: $B[5, 0.5]$ is not a suitable model for these data.

<u>Expected values</u>

$E_0 = 100 \times P(X=0) = 100\left[\binom{5}{0}(0.5^0)(1-0.5)^{5-0}\right]$

$= 3.125$

similarly,
$E_1 = 100 \times P(X=1) = 15.625$
$E_2 = 100 \times P(X=2) = 31.25$
$E_3 = 100 \times P(X=3) = 31.25$
$E_4 = 100 \times P(X=4) = 15.625$
$E_5 = 100 \times P(X=5) = 3.125$

expected values must all be >5 for the test statistic X^2 to be approximated well by the chi-squared distribution(s). So pool $X=0$ and $X=1$ and $X=4$ and $X=5$.

X	0 - 1	2	3	4 - 5
O_i	24	29	34	13
E_i	18.75	31.25	31.25	18.75
$\frac{(O-E)^2}{E}$	1.47	0.162	0.242	1.763

$X^2 = \sum \frac{(O-E)^2}{E} = 3.64$

$\nu = 4-1=3$ ∴ critical value $= \chi^2_3(10\%) = 6.251$
[(4 cells after pooling)] $3.64 < 6.251$ ∴ accept H_0.
$B[5, \frac{1}{2}]$ is a suitable model.

hence coins are <u>not</u> biased.