S3 June 2002 (MA)

Q1a) $\boxed{\text{Stratified}}$ – data divides naturally into mutually exclusive groups (Deluxe, Premier, Standard)

b) List all Deluxe rooms 1-20
List all Premier rooms 1-40
List all Standard rooms 1-100.

No. of each required : $\left[\frac{\text{room number}}{\text{total number}} \times \text{sample size}\right]$

Standard : $\frac{100}{160} \times 16 = 10$ rooms

Premier : $\frac{40}{160} \times 16 = 4$ rooms

Deluxe : $\frac{20}{160} \times 16 = 2$ rooms

Use random numbers to select (for each list) : 10 Standard rooms
4 Premier rooms
2 Deluxe rooms

Q2a) company 1 / company 2

$\bar{x} = 70.6$ $\quad$ $\bar{x} = 67.2$

$s = 9.1$ $\quad$ $s = 8.4$

$n = 100$ $\quad$ $n = 120$

$$\boxed{\begin{aligned} H_0 &: \mu_1 = \mu_2 \\ H_1 &: \mu_1 \neq \mu_2 \end{aligned}}$$

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

$$= \frac{70.6 - 67.2 - (0)}{\sqrt{\frac{9.1^2}{100} + \frac{8.4^2}{120}}} = \underline{\underline{2.86}}$$

1% level of significance & two tail.
$\therefore$ critical value $= \pm 2.5758$

$2.86 > 2.5758$
$\therefore$ Result is significant, reject H_0.
Evidence suggests that there is a difference in the mean playing times of CDs produced by the two companies.

b) Sample variance = population variance*
$(S^2 = \sigma^2)$

*since the test statistic by definition uses σ^2 not S^2.

Q3a) $M \sim N(80, 2.6^2)$

$n = 10 \therefore \bar{M} \sim N\left(80, \frac{2.6^2}{10}\right)$

b) $P(\bar{M} < 78.5) = P\left(Z < \frac{78.5 - 80}{\sqrt{\frac{2.6^2}{10}}}\right)$

$= P(Z < -1.82) = P(Z > 1.82)$

$= 1 - P(Z < 1.82) = 1 - 0.9656$

$= \boxed{0.0344}$

c) $F \sim N(59, 1.9^2)$

let total weight $= T$,

$T = F_1 + \ldots + F_4 + M_1 + M_2 + \ldots + M_6$

$E(T) = 4(59) + 6(80) = 716.$

$Var(T) = 4\,Var(F) + 6\,Var(M)$
$= 4(1.9^2) + 6(2.6^2) = 55.$

so $T \sim N(716, 55)$

P(max load is exceeded) $= P(T > 730)$

$= P\left(Z > \frac{730 - 716}{\sqrt{55}}\right) = P(Z > 1.89)$

$= 1 - P(Z < 1.89) = \boxed{0.0294}$

1-10 from highest to lowest
1 —— 10

(Q4a)

Athlete	Performance	Dedication	R_P	R_D	d
A	86	72	1	4	3
B	60	69	6	5	1
C	78	59	3	8	5
D	56	68	8	6	2
E	80	80	2	2	0
F	66	84	5	1	4
G	31	65	10	7	3
H	59	55	7	9	2
I	73	79	4	3	1
J	49	53	9	10	1

$$\sum d^2 = 3^2 + 1^2 + 5^2 + 2^2 + 4^2 + 3^2 + 2^2 + 1^2 + 1^2$$
$$= 70$$

so $r_s = 1 - \frac{6(70)}{10(99)} = \boxed{0.576}$

b) $H_0: \rho = 0$ critical value: ± 0.5636
$H_1: \rho \neq 0$ $n=10$, two-tail
r_s

we haven't been told anything about a positive or negative correlation in particular so test for any correlation.

$0.576 > 0.5636$
$\therefore$ Result is significant. Evidence suggests that there is a correlation between performance and dedication.

c) Unlikely that performance & dedication are jointly normally distributed.

Q5) H_0: no association between gender and use of club facilities

H_1: there is an association between gender and use of facilities.

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

expected:	Male	Female	Total
Pool	50.976	57.024	108
Jacuzzi	27.848	31.152	59
Gym	39.176	43.824	83
Total	118	132	$\boxed{250}$

O_i	E_i	$\frac{(O-E)^2}{E}$
40	50.976	2.3633..
68	57.024	2.1127..
26	27.848	0.12263..
33	31.152	0.1096..
52	39.176	4.1979..
31	43.824	3.7526..
		12.66

$$X^2 = \sum \frac{(O-E)^2}{E} = 12.66$$

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

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

$12.66 > 5.991$ $\therefore$ Result is significant (reject H_0).

Evidence suggests there is an association between gender and use of club facilities.

(Q6a) $X \sim B[8, 0.5]$ for one litter.

$R = 200 \times P(X=3) = 200\left[\binom{8}{3}(0.5^3)(0.5)^5\right]$

for 200 litters

$= \boxed{43.75 = R}$

$S = 200 \times P(X=4) = 200\left[\binom{8}{4}(0.5^4)(0.5^4)\right]$

$= \boxed{54.69 = S}$

$T = 200 - (\Sigma' E_i) = \boxed{43.74}$ mark scheme says 43.75/43.76 but this is fine.

b) $\underline{H_0}$: Binomial $[8, 0.5]$ is a suitable model for these data

$\underline{H_1}$: Binomial $[8, 0.5]$ is not a suitable model for these data.

No. of females	0-1	2	3	4	5	6	7-8
O_i	10	27	46	49	35	26	7
E_i	7.03	21.88	43.75	54.69	43.74	21.88	7.03
$\frac{(O-E)^2}{E}$	1.255	1.198	0.116	0.592	1.746	0.776	1.28×10^{-4}

pool first 2 cells and last 2 cells so $E_i > 5$.

($E_i > 5$) is necessary for the statistic X^2 to be approximated well by the chi-squared distribution(s).

$\sum \frac{(O-E)^2}{E} = X^2 = \underline{\underline{5.69}}$

γ = no. of cells after pooling -1
$= 7-1 = 6$

critical value $= \chi^2_6(5\%) = 12.592$

$5.69 < 12.592$
∴ Result is insignificant.
Evidence suggests that this binomial model $[B(8, \frac{1}{2})]$ is suitable for these data.

c) An extra degree of freedom would be used up. Expected values would differ and as a result so would our test statistic X^2.

(Q7a) $n = 10$
$\bar{x} = 502.2$
$\sum x = 5022$
$\sum x^2 = 2522288$

mean $= \bar{x} = \frac{\sum x}{10} = \frac{5022}{10} = \boxed{502.20}$

$s^2 = \frac{1}{9}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)$

$= \frac{1}{9}\left(2522288 - \frac{(5022)^2}{10}\right)$

$= \boxed{26.62}$

b) 90% C.I: $\left[\bar{x} \pm 1.6449 \frac{\sigma}{\sqrt{n}}\right]$ $\quad [P(Z > 1.6449) = 0.05]$

$= \left[502.2 \pm 1.6449\left(\frac{5}{\sqrt{10}}\right)\right]$

2dp

$= [494, 510] = \boxed{[493.98, 510.42]}$

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

$\Rightarrow \left[502.2 \pm 1.96\left(\frac{5}{\sqrt{10}}\right)\right]$

$\Rightarrow [499, \quad 505]$

d) $H_0: \mu = 500$ $\qquad n = 15$
$H_1: \mu > 500$ $\qquad \bar{x} = 501.9$

Test Stat $= \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{501.9 - 500}{\frac{5}{\sqrt{15}}}$

$= 1.47$

critical value : ± 2.3263

$1.47 < 2.3263$

$\therefore$ Result is insignificant

Evidence suggests that the mean weight is not greater than 500g.