S3 June 2003 (MA)

Q1a) Divide the population into mutually exclusive groups, with each group having a sample size representative of the entire population.

[Each group should have sample size u, where

$$u = \frac{\text{number in group}}{\text{total number (population)}} \times \text{sample size}$$
]

Then use simple random sampling to select u elements from each strata.

ADV: Can analyse data about specific subgroups.

DISADV: Not good for large n (population).

b) Quota sampling involves an 'interviewer' selecting the sample units to be used. A sample reflective of the population can be achieved – it is vital to keep in mind this process is non-random and some bias can exist with who is selected.

ADV: Administration of the entire process is easy.
OR: - costs kept to a minimum.
- representative sample can be achieved with small sample size

DISADV: unable to estimate sampling errors.
OR: Non-responses aren't recorded
Interviewer bias may exist with who is selected

critical value : 2.5758

Q2a) 99% C.I : $\left[\bar{x} \pm 2.5758 \frac{\sigma}{\sqrt{n}}\right]$

$\Rightarrow \left[124 \pm 2.5758\left(\frac{20}{\sqrt{30}}\right)\right]$

$\Rightarrow [115, 133]$

b) 140 is outside of our confidence interval, so the sample selected may be biased since it does not seem to accurately represent the population.

Q3a) $X \sim N(20, 5)$

$Y \sim N(10, 4)$

$E(X-Y) = E(X) - E(Y) = 20 - 10 = \boxed{10}$

b) $Var(X-Y) = Var(X) + Var(Y) = 5 + 4 = \boxed{9}$

c) $P(13 < X-Y < 16)$

$(X-Y) \sim N(10, 9)$ from a/b.

$\therefore P\left(\frac{13-10}{3} < Z < \frac{16-10}{3}\right)$

$= P(1 < Z < 2)$

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

$= 0.9772 - 0.8413 = \boxed{0.1359}$

Q4) H_0: No association between taking the drug and catching a cold.

H_1: There is an association between taking the drug and catching a cold.

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

$\boxed{\text{expected}}$

	Cold	No cold	
Drug	39.5	60.5	100
Dummy pill	39.5	60.5	100
	79	121	$\boxed{200}$

O_i	E_i	$\frac{(O-E)^2}{E}$
34	39.5	0.7658
66	60.5	0.5000
45	39.5	0.7658
55	60.5	0.5000
		2.5316

$$X^2 = \sum \frac{(O-E)^2}{E} = 2.5316$$

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

critical value $= \chi^2_1 (5\%) = 3.841$

$2.5316 < 3.841$ $\therefore$ result is insignificant.

Evidence suggests no association between catching a cold and taking the drug.

(Q5) Before

$\bar{x} = 10$

$s = 2.64$

$n = 100$

After

$\bar{x} = 8$

$s = 1.94$

$n = 120$

$H_0: \mu_{old} = \mu_{new}$

$H_1: \mu_{old} > \mu_{new}$

critical value : ± 1.6449

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

$$= \frac{10 - 8 - (0)}{\sqrt{\frac{2.64^2}{100} + \frac{1.94^2}{100}}} = \underline{\underline{6.29}}$$

$6.29 > 1.6449$

$\therefore$ Result is significant (reject H_0)

Evidence suggests that the mean river pollution fell after the factory closed down.

unnecessary as data is already ranked

Q6a)

Skater	A	B	~~r_A~~	~~r_B~~	d	d^2
1	2	3			1	1
2	5	2			3	9
3	3	6			3	9
4	7	5			2	4
5	8	7			1	1
6	1	4			3	9
7	4	1			3	9
8	6	8			2	4

$\sum d^2 = 46$

$$\therefore r_s = 1 - \frac{6(46)}{8(63)} = \boxed{0.452}$$

b) $H_0: \rho = 0$ critical value: 0.6429

$H_1: \rho > 0$ $n=8$, 1-tail

(or $\rho \neq 0$)

$r_s = 0.452 < 0.6429$

$\therefore$ result is insignificant.

no evidence to suggest a correlation.

If you used ($\rho \neq 0$) as H_1 then your critical value will be different but the conclusion will be the same.

Q7a)

x	2	10	50
$P(X=x)$	0.5	0.2	0.3

mean $= E(X) = \sum x P(X=x) = 2(0.5) + 10(0.2) + 50(0.3)$
$= \boxed{18}$

variance $= E(X^2) - [E(X)]^2$

$E(X^2) = 2^2(0.5) + 10^2(0.2) + 50^2(0.3) = 772$

$\therefore$ Variance $= \sigma^2 = 772 - 18^2 = \boxed{448}$

b)

50, 50
10, 10
2, 2
50, 10
50, 2
10, 50
10, 2
2, 50
2, 10

c)

$\bar{x}$	2	6	10	26	30	50
$P(\bar{X}=\bar{x})$	0.25	0.20	0.04	0.30	0.12	0.09

$P(\bar{X}=2) = 0.5^2 = \frac{1}{4}$

$P(\bar{X}=50) = 0.3^2 = 0.09$

$P(\bar{X}=6) = (0.5)(0.2) \times 2$

$P(\bar{X}=30) = (0.2)(0.3) \times 2$
$= 0.12$

$P(\bar{X}=10) = 0.2^2 = 0.04$

$P(\bar{X}=26) = (0.3)(0.5) \times 2 = 0.30$

e) $E(\bar{X}) = \sum x P(X = x) = 2(\frac{1}{4}) + 6(0.2) + 10(0.04)$
$+ 26(0.3) + 30(0.12) + 50(0.09) = 18$ //

$E(\bar{X}^2) = 2^2(0.25) + 6^2(0.2) + 10^2(0.04) + 26^2(0.3)$
$+ 30^2(0.12) + 50^2(0.09) = 548$ //

$\therefore Var(\bar{X}^2) = 548 - (18)^2 = 224$ //

$\frac{1}{2}\sigma^2 = \frac{1}{2} \times 448 = 224$