S3 June 2004 (MA)

Q1a) Number each pupil 1 to 120. Then use random numbers to select 15 pupils. The numbers selected should be between 1 and 120 and will correspond to the numbers given to each pupil.

b) Population divides into two groups: girls and boys.

No. required from each group = $\frac{\text{group size}}{\text{population size}} \times \text{sample size}$

No. of girls = $\frac{64}{120} \times 15 = 8$

No. of boys = $\frac{56}{120} \times 15 = 7$

so List all girls 1 - 64 and all boys 1 - 56 seperately and use random numbers to select 8 girls and 7 boys.

Q2a) $H_0: \rho = 0$ PMCC critical value: 0.6215
$H_1: \rho > 0$ $n = 8$, 1-tail

$0.572 < 0.6215$ ∴ Result is insignificant.
No evidence of a positive correlation.

Spearman critical value: 0.6429

$0.655 > 0.6429$ ∴ Result is significant
Evidence suggests there is a positive correlation.

b) PMCC test yielded no correlation (positive) so conclude that there is no evidence that as Stats marks increased, so did Geography marks.

Spearmans test yielded a positive correlation so conclude that students ranked highly in Stats were also ranked highly in Geography.

Q3a) $H_0: \mu_A = \mu_B$

$H_1: \mu_A < \mu_B$

critical value : ± 1.6449

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

$$= \frac{251 - 249 - (0)}{\sqrt{\frac{2.5^2}{10} + \frac{2.3^2}{15}}} = 2.0227$$

$2.02 > 1.6449$

Result is significant

Evidence suggests population mean weight dispensed by B is greater than that of A.

b) Machine B weight dispensed is also normally distributed.

Q4a) mean $= \frac{\sum x}{n} = \frac{753}{10} = \boxed{75.3}$

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

$$= \frac{1}{9}\left(57455 - \frac{(753)^2}{10}\right) = \boxed{83.8}$$

b) 95% CI : $\left[\bar{x} \pm 1.96\sqrt{\frac{\sigma^2}{n}}\right]$

$$\left[74.8 \pm 1.96\sqrt{\frac{84.6}{100}}\right]$$

$$\left[73.0, 76.6\right]$$

c) - sample size large enough to apply CLT (required for confidence interval)
- Journey times are independent of each other.

(Q5a) H_0: there is no association between gender and amount of exercise.

H_1: there is an association between gender and amount of exercise.

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

OBSERVED:

	Never	Sometimes	Regularly	
Male	30	132	78	240
Female	26	143	91	260
	56	275	169	$\boxed{500}$

EXPECTED:

	Never	Sometimes	Regularly	
Male	26.88	132	81.12	240
Female	29.12	143	87.88	260
	56	275	169	$\boxed{500}$

O_i	E_i	$\frac{(O-E)^2}{E}$
30	26.88	0.3621
132	132	0
78	81.12	0.12
26	29.12	0.3343
143	143	0
91	87.88	0.1108
		0.927 //

$$X^2 = \sum \frac{(O-E)^2}{E} = 0.927$$

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

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

$0.927 < 5.991$

$\therefore$ Result is insignificant
Evidence suggests there is no association between gender and amount of exercise.

Q6a) $X \sim B[3, \frac{1}{6}]$ where X = no. of sixes.

b) $\underline{H_0}$: Binomial $B[3, \frac{1}{6}]$ is a suitable model for these data

$\underline{H_1}$: Binomial $B[3, \frac{1}{6}]$ is not a suitable model for these data.

No. of sixes	0	1	2 - 3
E_i	144.68	86.81	18.52
O_i	125	109	16
$\frac{(O-E)^2}{E}$	2.677	5.672	0.343

working

$250 \times P(X=0) = E_0 = 144.68$

$250 \times P(X=1) = E_1 = 86.81$

$250 \times P(X=2) = E_2 = 17.36$

$250 \times P(X=3) = E_3 = 1.16$

Pool cells so expected >5 value

$X^2 = \sum \frac{(O-E)^2}{E} = 8.69$ //

3 cells $\therefore \nu = 2$ $\therefore$ critical value $= \chi^2_2(1\%) = 9.210$

$8.69 < 9.210$ $\therefore$ Result is insignificant.
evidence suggests Binomial $B[3, \frac{1}{6}]$ is a suitable model.

(Q7a) $D = A - 3B + 4C$

$E(D) = E(A) - 3E(B) + 4E(C)$

$= 5 - 3(7) + 4(9) = 20$

$Var(D) = Var(A) + 9\,Var(B) + 16\,Var(C)$

$= 4 + 9(9) + 16(16) = 341$

$\therefore D \sim N[20, 341]$

$P(D < 44) = P\left(Z < \frac{44-20}{\sqrt{341}}\right) = P(Z < 1.30)$

$= \boxed{0.9032}$

b) $X = A - (B_1 + B_2 + B_3) + 4C$

$E(X) = 5 - (7+7+7) + 4(9) = 20$

$Var(X) = Var(A) + 3\,Var(B) + 16\,Var(C)$

$= 4 + 3(9) + 16(16) = 287$

we dont square the 3 as we are taking 3 samples of B (ie $B_1 + B_2 + B_3$). This is not equivalent to 3B

$\Rightarrow X \sim N(20, 287)$

$P(X > 0) = P\left(Z > \frac{-20}{\sqrt{287}}\right)$

$= P(Z > -1.18)$
$= P(Z < 1.18)$

$= \boxed{0.8810}$