S3 June 2007 (MA)

Q1a)

Display	P	Q	R_P	R_Q	d	d^2
A	25	20	2	2	0	0
B	19	9	6	8	2	4
C	21	21	4	1	3	9
D	23	13	3	6	3	9
E	28	17	1	3	2	4
F	17	14	7	5	2	4
G	16	11	8	7	1	1
H	20	15	5	4	1	1
						32

highest = 1, lowest = 8
(you can reverse the order, the method will still be valid)

$\sum d^2 = 32$

$\therefore r_s = 1 - \frac{6(32)}{8(63)} = \boxed{0.619}$

b) $H_0: \rho = 0$ critical value: 0.6429
$H_1: \rho > 0$ (5%, 1-tail)

$0.619 < 0.6429$

$\therefore$ Result is insignificant (accept H_0).
Evidence suggests the competitor's claim is correct.

Q2a) $\underline{H_0}$: Student's maths grades are independent of English grades.

$\underline{H_1}$: Maths and english grades are $\underline{\text{not}}$ independent

expected no. = $\dfrac{\text{Row total} \times \text{column total}}{\text{Grand total}}$

<u>EXPECTED</u> :

	A/B	C/D	E/U
A/B	20	27.5	12.5
C-U	20	27.5	12.5

$\underline{O_i}$	$\underline{E_i}$	$\underline{\dfrac{(O-E)^2}{E}}$
25	20	1.2500
25	27.5	0.2273
10	12.5	0.5000
15	20	1.2500
30	27.5	0.2273
15	12.5	0.5000
		3.95 //

$$X^2 = \sum \frac{(O-E)^2}{E} = 3.95 //$$

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

critical value = $\chi^2_2(10\%) = 4.605$

$3.95 < 4.605$

$\therefore$ Result is insignificant (accept H_0)

Evidence suggests that maths grades are independent of english grades.

b) This would likely lead to having some expected frequencies less than 5 (and so we'd need to pool rows and/or columns).

Q3) $H_0: \mu = 18$ critical value : ± 1.6449
$H_1: \mu < 18$ (5%, 1-tail)

$$\text{Test Stat} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{16.5 - 18}{\frac{3}{\sqrt{15}}} = -1.936$$

$-1.936 < -1.6449$

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

Evidence suggests that the mean time Robert takes to complete the puzzle has decreased.

Q4a) $$\hat{p} = \frac{0(17) + 1(31) + 2(19) + 3(14) + 4(9) + 5(7) + 6(3)}{20 \times 100}$$

$$= \frac{20\cancel{0}}{20 \times 1\cancel{0}0} = \frac{2}{20} = \boxed{0.1}$$

b) $X \sim B[20, 0.1]$

$$r = \underset{\text{100 samples}}{100} \times P(X=2) = 100\left[\binom{20}{2}(0.1^2)(0.9^{18})\right]$$

$$= \boxed{28.5}$$

$$s = 100 - \sum E_i = 100 - (12.2 + 27 + 28.5 + 19 + 3.2 + 0.9 + 0.2)$$

$$s = \boxed{9.0}$$

c) H_0: $B[20, 0.1]$ is a suitable model for these data.
H_1: $B[20, 0.1]$ is not a suitable model for these data

X	0	1	2	3	≥ 4
E_i	12.2	27.0	28.5	19.0	13.3
O_i	17	31	19	14	19
$\frac{(O-E)^2}{E}$	1.89	0.59	3.17	1.32	2.44

$$X^2 = \sum \frac{(O-E)^2}{E} = 9.4$$

ν = no. of cells after pooling $- 2 = 5 - 2 = 3$

critical value $= \chi^2_3 (5\%) = 7.815$.

$9.4 > 7.815$

$\therefore$ Result is significant. (reject H_0)
Evidence suggests that Binomial Distribution is not a suitable model.

d) Defective items do not occur with constant probability. (This is an assumption required for a Binomial Distribution).

(Q5a) mean $= \dfrac{\sum x}{n} = \dfrac{361.6}{80} = \boxed{4.52}$

$$s^2 = \frac{1}{79}\left(\sum x^2 - \frac{(\sum x)^2}{80}\right)$$

$$= \frac{1}{79}\left(1753.95 - \frac{(361.6)^2}{80}\right)$$

$$= \boxed{1.51}$$

b) let mean weight loss with diet A = μ_A,
and mean weight loss with diet B = μ_B,

$H_0 : \mu_A = \mu_B$
$H_1 : \mu_A > \mu_B$

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

from (a)

$$= \frac{4.52 - 4.06 - (0)}{\sqrt{\frac{2.50}{60} + \frac{1.51}{80}}} = 1.8689$$

critical value : ± 1.6449
(1-tail, 5%)

So $1.8689 > 1.6449$
$\therefore$ Result is significant (reject H_0)
Evidence suggests that mean weight loss is greater with diet A.

c) Allows us to assume sample means for diet A and B are normally distributed.

d) $\sigma_A^2 = S_A^2$ and $\sigma_B^2 = S_B^2$.

(By definition the test statistic uses population variance).

Q6) 99% C.I: $\left[\bar{x} \pm 2.5758 \frac{\sigma}{\sqrt{n}}\right]$ //

so this means the width of the interval will be

$$2 \times 2.5758 \frac{\sigma}{\sqrt{n}}.$$

width of interval $= 154.7 - 123.5 = \frac{2(2.5758)\sigma}{\sqrt{n}}$

$$\Rightarrow \frac{\sigma}{\sqrt{n}} = \frac{31.2}{2(2.5758)} = 6.056\ldots$$

and so $\bar{x} + 2.5758 \frac{\sigma}{\sqrt{n}} = 154.7$

$$\bar{x} = 154.7 - 2.5758\left(\frac{31.2}{2(2.5758)}\right)$$

$$\bar{x} = 139.1 \text{ //}$$

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

$$\Rightarrow \left[139.1 \pm 1.96\left(\frac{31.2}{2(2.5758)}\right)\right]$$

$$\Rightarrow [127.22, 150.97] \text{ //}$$

Q7a) $L \sim N(19.7, 0.5^2)$

$S \sim N(4.9, 0.2^2)$

$P(L > 4S) = P(L - 4S > 0)$

let $A = L - 4S$,

$E(A) = 19.7 - 4(4.9) = 0.1$

$Var(A) = 0.5^2 + 16(0.2^2) = 0.89$

$\therefore A \sim N(0.1, 0.89)$

so $P(L - 4S > 0) = P(Z > \frac{-0.1}{\sqrt{0.89}})$

$= P(Z < 0.11) = \boxed{0.5438}$

b) $T = S_1 + S_2 + S_3 + S_4$.

We are taking a sample (size 4) of S, this is not the same as $4S$. S_1, S_2, S_3, S_4 are all different. (independent)

$\therefore E(T) = 4E(S) = 19.6$
$Var(T) = 4\,Var(S) = 0.16$

$\Rightarrow \boxed{T \sim N(19.6, 0.16)}$

c) $P(|L - T| < 0.1) = P(-0.1 < L - T < 0.1)$

$E(L - T) = E(L) - E(T) = 0.1$
$Var(L - T) = Var(L) + Var(T) = 0.41$

$\therefore L - T \sim N(0.1, 0.41)$

$P(\text{required}) = P(-0.1 < L-T < 0.1)$

$= P\left(\frac{-0.1-0.1}{\sqrt{0.41}} < Z < \frac{0.1-0.1}{\sqrt{0.41}}\right)$

$= P(-0.31 < Z < 0)$

$= P(Z < 0) - P(Z < -0.31)$

$= 0.5 - [1 - P(Z < 0.31)]$

$= 0.5 - [1 - 0.6217]$

$= \boxed{0.1217}$ from tables.