S3 January 2006 (MA)

(Q1) Split the data into groups ... one for each of the 15 classes and another for sixth form.

Total students $= (30 \times 15) + 150 = 600$

For each class : $\underbrace{\frac{30}{600} \times 40}_{\frac{\text{number in class}}{\text{total number}} \times \text{sample size}} = 2$ students from each class.

For sixth form : $\underbrace{\frac{150}{600} \times 40}_{\frac{\text{number in } 6^{th} \text{ form}}{\text{total number}} \times \text{sample size}} = 10$ students from sixth form.

So list all boys in each class 1-15 and all girls 1-15. Then use random numbers to select 1 girl and 1 boy in each class.

List all boys in sixth form 1-75 and all girls 1-75 and use random numbers to select 5 boys and 5 girls.

Q2a) $X \sim N(20, 4)$

$Y \sim N(10, 0.84)$

$R = X + Y$

$E(R) = E(X) + E(Y) = 20 + 10 = \boxed{30}$

b) $Var(R) = (4) + (0.84) = \boxed{4.84}$

$(= Var X + Var Y)$

c) $R \sim N(30, 4.84)$

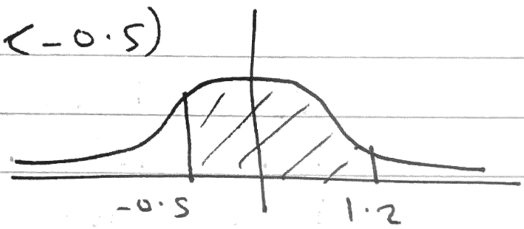
$P(28.9 < R < 32.64)$

$= P\left(\frac{28.9-30}{\sqrt{4.84}} < Z < \frac{32.64-30}{\sqrt{4.84}}\right)$

$= P(-0.5 < Z < 1.2)$

$= P(Z < 1.2) - P(Z < -0.5)$

$= 0.8849 - (1 - 0.6915)$



$= \boxed{0.5764}$

(Q3a) mean $= \frac{\sum x}{n} = \frac{82+98+140+110+90+125+\ldots}{10}$

$= \boxed{110\cdot5}$

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

$= \frac{1}{9}\left(128153 - \frac{(1105)^2}{10}\right) = \boxed{672\cdot28}$

$\left(\begin{array}{l}\sum x^2 = 128153 \\ \sum x = 1105\end{array}\right)$

b) C.I : $\left[\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}\right] \sim \left[95\% \text{ C.V} = 1.96\right]$
2.5% each tail

$\Rightarrow \left[110\cdot5 \pm 1.96\left(\frac{25}{\sqrt{10}}\right)\right] = \left[95, 126\right]$

c) let X represent number of confidence intervals that contain μ . . . .

$X \sim B[15, 0\cdot95]$

mean $= np = 15\times0\cdot95 = \boxed{14\cdot25}$

Q4) H_0: There is no association between gender and acceptance.

H_1: There is an association between gender and acceptance.

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

EXPECTED	Accept	Not Accepted	Total
M	180	100	280
F	270	150	420
Total	450	250	700

O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
170	180	0.5556
110	100	1.0000
280	270	0.3704
140	150	0.6667
		2.59

4 dp.

$$X^2 = \sum \frac{(O-E)^2}{E} = 2.59$$

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

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

$2.59 < 3.841$

$\therefore$ Result is insignificant (accept H_0).

Insufficient evidence to suggest an association between gender & association

$\boxed{\text{error in question: there are 80 girls not 8}}$

Q5a) $H_0: \mu_{boys} = \mu_{girls}$

$H_1: \mu_{boys} \neq \mu_{girls}$

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

$$= \frac{53 - 50 - (0)}{\sqrt{\frac{12^2}{80} + \frac{12^2}{80}}} = 1.58$$

5% level of significance so critical value $= 1.96$.
(two-tail)

$1.58 < 1.96$

$\therefore$ Result is insignificant. (accept H_0)
Evidence suggests there is <u>no difference</u> in the mean scores of boys / girls.

b) $H_0: \mu_{boys} = \mu_{girls}$ critical value: ± 1.6449.
(1 tail)

$H_1: \mu_{boys} < \mu_{girls}$

$$\text{Test Stat} = \frac{62 - 59 - (0)}{\sqrt{\frac{6^2}{80} + \frac{6^2}{80}}} = 3.162$$

$3.162 > 1.6449$

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

Evidence suggests that the mean mark for boys is significantly lower than the mean mark for girls in the 'after 1 year' exam.

c) at the beginning there was no difference in the mean mark. After 1 year the girls are better.

So anything along the lines of:
"Girls have improved much more than the boys"

Q6a) $X \sim Po(2) \rightarrow$ no. of daises in one (1×1) square.

$$P(X=2) = \frac{(e^{-2})(2^2)}{2!} = 0.27067\ldots$$

$$100 \times P(X=2) = \boxed{r = 27.07}$$

for 100 of these (1×1) squares.

$$P(X=3) = \frac{(e^{-2})(2^3)}{3!} = 0.180447\ldots$$

$$100 \times P(X=3) = \boxed{s = 18.04}$$

$$t = 100 - \left(\sum E_i\right) = \boxed{0.12 = t}$$

b) H_0: Poisson ($\lambda=2$) is a suitable distribution to model these data.

H_1: Poisson ($\lambda=2$) is not a suitable model for these data.

Grouping cells... (expected value >5)

No. of daisies	0	1	2	3	4	≥ 5
O_i	8	32	27	18	10	5
E_i	13.53	27.07	27.07	18.04	9.02	5.27
$\frac{(O-E)^2}{E}$	2.260	0.8979	0.00018	8.87×10^{-5}	0.1065	0.01383..

$$X^2 = \sum \frac{(O-E)^2}{E} = 3.28$$

$\nu = 6-1 = 5$ (no. of cells − 1 after grouping)

critical value: $\chi^2_5 (5\%) = 11.070$

$3.28 < 11.070$

$\therefore$ Result is insignificant. (accept H_0).
Evidence suggests that Poisson ($\lambda=2$) is a suitable distribution to model the data.

c) A degree of freedom would be used up and our expected values would be different. As a result, the test statistic X^2 would yield a different value.

Q7a) We are unable to assume an underlying normal distribution.

b)

Age	Pneu...	Lung	R_p	R_L	d	d^2
20-29	12.5	3.7	5	6	1	1
30-39	5.9	9.0	6	5	1	1
40-49	18.5	10.2	4	4	0	0
50-59	19.4	19.0	3	1	2	4
60-69	31.2	13.0	1	3	2	4
≥ 70	31.0	18.0	2	2	0	0
						10

$\sum d^2 = 10$

$$\therefore r_s = 1 - \frac{6(10)}{6(6^2-1)} = \boxed{0.714}$$

c) $H_0: \rho = 0$ critical value: 0.8857
$H_1: \rho \neq 0$
$n = 6$
2.5% each tail

testing for a correlation - could be positive or negative

$r_s = 0.714 < 0.8857$

$\therefore$ Result is insignificant. Accept H_0.

Evidence suggests there isn't a correlation between no. of deaths in different age groups for each disease.