S3 June 2012 (MA)

Q1(a)

Candidate	X	Y	R_x	R_y	d	d^2
A	62	54	6	7	1	1
B	56	47	5	4	1	1
C	87	71	8	8	0	0
D	54	50	4	6	2	4
E	65	49	7	5	2	4
F	15	25	3	1	2	4
G	12	30	2	2	0	0
H	10	44	1	3	2	4
						18

$\sum d^2 = 18$

$\therefore r_s = 1 - \frac{6(18)}{8(63)} = \boxed{0.7857}$

b) $H_0: \rho = 0$

$H_1: \rho > 0$

agreement indicated by a +ve correlation between ranks.

critical value: ± 0.6429 (5%, 1-tail)

$0.7857 > 0.6429$

$\therefore$ Result is significant. Reject H_0.

Evidence suggests that managers are in agreement

c) A and D are now tied ranks for manager Y. Because of this, we cannot use the formula ($r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)}$) to calculate r_s. *

Instead we have to use the 'traditional' PMCC formula ($r_s = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$) with the ranked data. We then rank both A and D as '6.5' since they are tied 6th.

Q2a) Not possible to obtain a sampling frame of each species in the lake.

b) Quota sampling (since you don't need a sampling frame)

c) ADV : - costs kept to a minimum.
- administering the test is easy.
- samples can be obtained quickly

any 1

DISADV : - unable to estimate sampling errors
- not a random process.
- interviewer / surveyor may not be able to identify species of fish easily.

any 1

d) No. of each type needed : $\left(\frac{\text{group size}}{\text{total fish}} \times \text{sample size}\right)$

$$\text{Trout} = \frac{1400}{2450} \times 30 = 17.14$$

$$\text{Bass} = \frac{600}{2450} \times 30 = 7.35$$

$$\text{Pike} = \frac{450}{2450} \times 30 = 5.51$$

so keep catching fish until 17 trout, 7 bass and 6 pike are retrieved. Of course if a fish is caught and enough of that species has already been caught then that fish can be put back into the lake.

Q3a) Suppose X belongs to any distribution with mean μ and variance σ^2 and we take a sample size n $(X_1, X_2, \ldots X_n)$.

Provided n is large, the sample mean $\bar{X}$ will follow a normal distribution (approximately) with mean μ and variance $\frac{\sigma^2}{n}$.

b) $\text{mean} = \frac{\sum x}{n} = \frac{1740000}{100} = 17400$

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

$\Rightarrow \left[17400 \pm 1.96\left(\frac{5000}{\sqrt{100}}\right)\right]$

$\Rightarrow [16420, 18380]$

c) Allows us to assume sample mean $\bar{X}$ is normally distributed. (necessary for C.I).

d) Complaint is a reasonable one since 20000 is above the upper limit of the C.I in (b).

Q4) H_0: There is no association between egg yield and breed of chicken.

H_1: There is an association between egg yield and breed of chicken.

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

EXPECTED:

	Low	Medium	High	total
Leghorn	24	56	20	100
Cornish	12	28	10	50
total	36	84	30	$\boxed{150}$

O_i	E_i	$\frac{(O-E)^2}{E}$
22	24	0.1667
52	56	0.2857
26	20	1.8000
14	12	0.3333
32	28	0.5714
4	10	3.6000
		6.76

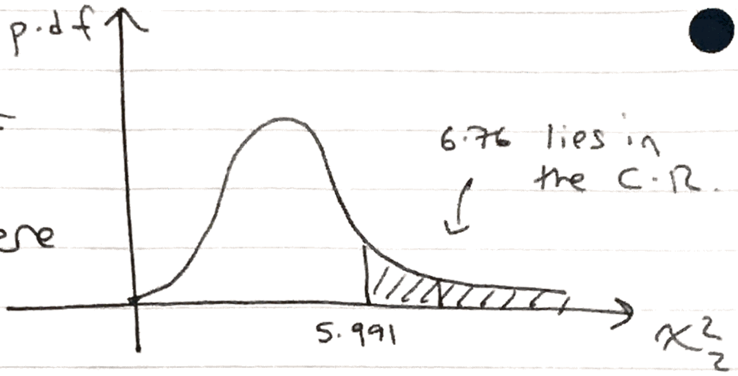
$$X^2 = \sum \frac{(O-E)^2}{E} = 6.76$$

$$\nu = (\text{rows} - 1)(\text{columns} - 1) = (2-1)(3-1) = 2$$

$$\therefore \text{critical value} = \chi^2_2(5\%) = 5.991.$$

$6.76 > 5.991$

$\therefore$ Result is significant
Reject H_0.
Evidence suggests there is in fact an association between egg yield and breed.



Q5a) $H_0 : \mu_{Alan} = \mu_{BURNS}$

$H_1 : \mu_{Alan} \neq \mu_{Burns}$

critical value : ± 1.6449.
(10%, 2-tail)

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

$$= \frac{80 - 74 - (0)}{\sqrt{\frac{100}{29} + \frac{225}{26}}}$$

$$= 1.7247\ldots$$

$1.72 > 1.6449$

$\therefore$ Result is significant
Reject H_0.
Evidence suggests there is a difference in the mean scores of students from both classes.

b) $H_0: \mu_{alan} = \mu_{burns}$ critical value $= \pm 1.2816$ (10%, 1-tail)

$H_1: \mu_{alan} > \mu_{burns}$.

$1.6 > 1.2816$

$\therefore$ Result is significant.

Reject H_0.

Mr. Alan's claim is supported.

Q6a) $$\text{mean} = \frac{\text{total defective items}}{\text{total no. of samples}}$$

$$= \frac{1(16) + 2(20) + 3(23) + 4(17) + 5(10) + 6(8)}{100}$$

$$= \boxed{2.91}$$

b) $X \sim B[6, \frac{2.91}{6}]$ where X = no. of defective items in a sample of 6.

$$a = 100 \times P(X=3) = 100\left[\binom{6}{3}\left(\frac{2.91}{6}\right)^3\left(1 - \frac{2.91}{6}\right)^3\right]$$

for 100 samples $= \boxed{31.17}$

$$b = 100 - \sum E_i = \boxed{1.30}$$

c) $H_0: B[6, \frac{2.91}{6}]$ is a good fit

$H_1: B[6, \frac{2.91}{6}]$ is not a good fit.

No. defective	0 - 1	2	3	4	5 - 6
O_i	22	20	23	17	18
E_i	12.41	24.82	31.17	22.01	9.59
$\frac{(O-E)^2}{E}$	7.4108	0.9630	2.1414	1.1404	7.3752

Remember, expected frequencies must be greater than 5 for X^2 to be approximated well by the chi-squared distribution(s). So I have pooled the first two and last two cells.

$$X^2 = \sum \frac{(O-E)^2}{E} = 19.03$$

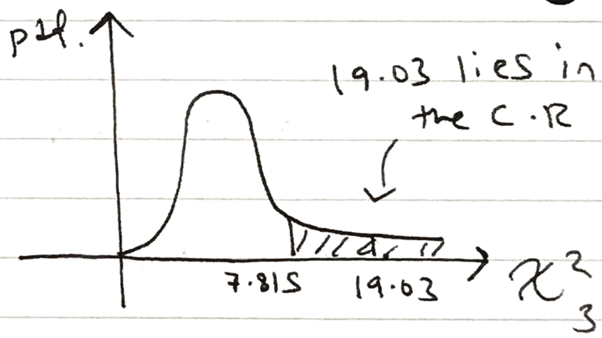
$$\nu = 5 - 1 - 1 = 3$$

(we calculated p so subtract 2)

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

$$19.03 > 7.815$$

∴ Result is significant.
Reject H_0
Evidence suggests that Binomial is not a good fit.



Q7a) $M \sim N(177, 5^2)$

$F \sim N(163, 4^2)$

$P(M > F) = P(M - F > 0)$

let $A = M - F$,

$E(A) = 177 - 163 = \boxed{14}$

$Var(A) = 4^2 + 5^2 = \boxed{41}$

$\therefore \quad A \sim N(14, 41)$

$\Rightarrow \quad P(M - F > 0) = P(Z > \frac{-14}{\sqrt{41}})$

$\Rightarrow \quad P(Z < 2.186..) \simeq P(Z < 2.18)$

$\simeq \boxed{0.9854}$

b) $P(\text{required}) = P(M_1 + \ldots + M_6 + F_1 + \ldots + F_4 < 1700)$

let $B = M_1 + \ldots + M_6 + F_1 + \ldots + F_4$,

$M_1 + \ldots + M_6 \neq 6M$

$E(B) = 6(177) + 4(163) = 1714$

$Var(B) = 4(4^2) + 6(5^2) = 214$

$B \sim N(1714, 214)$

$P(B < 1700) = P(Z < \frac{1700 - 1714}{\sqrt{214}})$

$= P(Z < -0.96) = 1 - P(Z < 0.96)$

$= \boxed{0.1685}$