S3 June 2013 (MA)

Q1) $\underline{H_0}$: There is no association between cholesterol level and intake of saturated fats.

$\underline{H_1}$: There is an association between cholesterol level and intake of saturated fats.

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

<u>EXPECTED</u> :

S.F \ C.L	High	Low	
High	7.6	12.4	20
Low	30.4	49.6	80
	38	62	$\boxed{100}$

$\underline{O_i}$	$\underline{E_i}$	$\underline{\frac{(O-E)^2}{E}}$
12	7.6	2.5474
8	12.4	1.5613
26	30.4	0.6368
54	49.6	0.3903

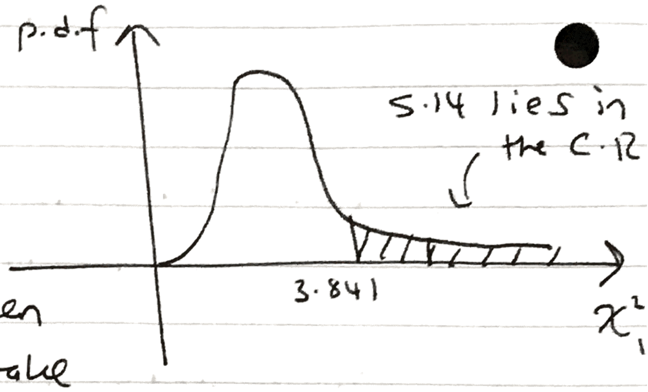
$X^2 = \sum \frac{(O-E)^2}{E} = 5.14$ //

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

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

$5.14 > 3.841$

$\therefore$ Result is significant.
Reject H_0.
Evidence suggests there is an association between cholesterol level and intake of saturated fats.



Q2a)

University	No. students...	Satisfaction	R_N	R_S	d	d^2
A	14.2	4.1	6	5	1	1
B	13.1	4.2	4	6	2	4
C	13.3	3.8	5	2	3	9
D	11.7	4.0	3	4	1	1
E	10.5	3.9	1	3	2	4
F	15.9	4.3	7	7	0	0
G	10.8	3.7	2	1	1	1
						20

$\sum d^2 = 20$

$\therefore r_s = 1 - \frac{6(20)}{7(48)} = \boxed{0.643}$

b) $H_0: \rho = 0$ critical value : ± 0.7857
$H_1: \rho \neq 0$ (5%, 2-tail)

testing for any correlation, not necessarily positive

$0.643 < 0.7857$
$\therefore$ Result is insignificant.
Accept H_0.
Evidence suggests there isn't a correlation between students per staff member and satisfaction score.

Q3ai) ADV : - Fieldwork can be done quickly
- Costs kept to a minimum.
- sampling frame not required.
- suitable for large populations.
- administering the test is easy

any 1

DISADV : - non-random process so can't estimate sampling errors
- non-responses aren't recorded
- Interviewer bias may be introduced.

any 1

ii) ADV : - Representative of entire population
- Can give accurate estimates as it is a random process.

any 1

DISADV : - Sampling frame required
- strata may not be clear - some students may be in more than one course.
- Not suitable for large populations

any 1

b) No. from each course $= \dfrac{\text{Number in course}}{\text{Total number}} \times \text{sample size}$

(Total = 1000)

Leisure/sport $= \dfrac{420}{1000} \times 100 = 42$

$IT = \dfrac{337}{1000} \times 100 = 33.7$

Health /Social care $= \frac{200}{1000} \times 100 = 20$

Media Studies $= \frac{43}{1000} \times 100 = 4.3$

so . . .

Course	Number sampled
Leisure/Sport	42
IT	34
Health/social...	20
Media	4

c) Use the info system to obtain a numbered list for all students on each course. (ie a different list for each course) Then use random numbers to select the appropriate number of students from each course according to (b).

)

Time	Midpoint(y)	Frequency (f)	fy	fy^2
0-3	1.5	8	12	18
3-5	4	12	48	192
5-6	5.5	13	71.5	$\frac{1573}{4}$
6-8	7	9	63	441
8-12	10	8	80	800
			274.5	1844.25

$$\bar{x} = \frac{\sum fy}{n} = \frac{274.5}{50} = \boxed{5.49}$$

$$s_x^2 = \frac{1}{n-1}\left(\sum fy^2 - \frac{(\sum fy)^2}{n}\right)$$

$$= \frac{1}{49}\left(1844.25 - \frac{(274.5)^2}{50}\right) = \boxed{6.88}$$

The method to calculate s_x^2 is slightly different as this is a grouped frequency distribution.

Why: Instead of $\sum x^2$ and $\sum x$ we use $\sum fy^2$ and $\sum fy$ where y is the midpoint of the interval. With this type of data, we don't actually know the exact waiting time of each customer, just the interval they lie in. So we use the midpoint as a way of estimating the waiting-time of each customer. Hence (fx) will give you the total waiting time for an interval and $\sum fx$ gives us the required total waiting time and $\sum fx^2$ the total waiting time squared.

b) from (a), $X \sim N(5.49, 6.88)$

$$P(6 < X < 8) = P\left(\frac{6-5.49}{\sqrt{6.88}} < Z < \frac{8-5.49}{\sqrt{6.88}}\right)$$

$$= P(0.19 < Z < 0.96)$$

$$= P(Z < 0.96) - P(Z < 0.19)$$

$$= 0.8315 - 0.5753$$

50 customers.

$$= 0.2562 \quad \therefore \quad a = 50 \times 0.2562 = \boxed{12.81}$$

so $b = 50 - (\sum E_i) = \boxed{8.34}$

c) H_0: Normal distribution is a good fit

H_1: Normal distribution is not a good fit.

Class	0 - 3	3 - 5	5 - 6	6 - 8	> 8
O_i	8	12	13	9	8
E_i	8.56	12.73	7.56	12.81	8.34
$\frac{(O-E)^2}{E}$	0.0366	0.0419	3.9145	1.1332	0.0139

$$X^2 = \sum \frac{(O-E)^2}{E} = 5.14$$

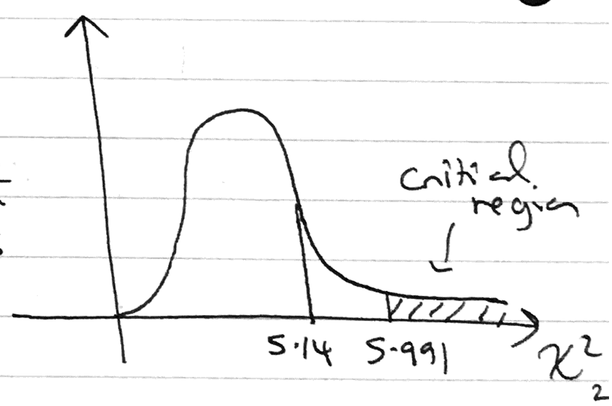
$$\nu = 5 - 1 - 1 - 1 = 2$$

we calculated $\bar{x}$ and s_x^2 so subtract 3.
(ν = no. of cells after pooling -1 $-$ (no. of parameters calculated))

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

$5.14 < 5.991$

$\therefore$ Result is insignificant
Accept H_0.
Evidence suggests that normal distribution is a good fit.



Q5a) $L \sim N(50, 5^2)$

$S \sim N(15, 3^2)$

$P(\text{required}) = P(L < S_1 + S_2 + S_3)$

$= P(L - (S_1 + S_2 + S_3) < 0)$

3 small bottles implies $S_1 + S_2 + S_3$ NOT $3S$.

let $A = L - (S_1 + S_2 + S_3)$,

$E(A) = 50 - (15 + 15 + 15) = 5$

$Var(A) = 5^2 + 3^2 + 3^2 + 3^2 = 52$

$\therefore A \sim N(5, 52)$

$P(\text{required}) = P(A < 0) = P\left(Z < \frac{-5}{\sqrt{52}}\right)$

$= P(Z < -0.69)$

$= 1 - P(Z < 0.69) = \boxed{0.245}$

b) $P(\text{required}) = P(L > 3S) = P(L - 3S > 0)$.

"3 times the amount" so this means $3S$

let $B = L - 3S$,

$E(B) = 50 - 3(15) = 5$

$$\text{Var}(B) = \text{Var}(L) + 3^2\,\text{Var}(S)$$

$$= 5^2 + 3^2(3^2) = 106$$

$$\therefore B \sim N(5, 106)$$

$$P(\text{required}) = P(B>0) = P\left(Z > \frac{-5}{\sqrt{106}}\right)$$

$$= P(Z > -0.49) = P(Z < 0.49)$$

$$= \boxed{0.6879}$$

Q6a) $H_0: \mu_{new} - \mu_{old} = 1$ critical value : ± 1.6449 (5%, 1-tail)

$H_1: \mu_{new} - \mu_{old} > 1$

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

$$= \frac{7 - 5.5 - (1)}{\sqrt{\dfrac{0.5}{60} + \dfrac{0.75}{70}}}$$

$$= 3.62$$

$3.62 > 1.6449$ $\therefore$ Result is significant
Reject H_0.

Evidence suggests mean yield of new variety is more than 1kg greater than old variety.

(Q7a) $\text{mean} = \frac{\sum x}{n} = \frac{33.29}{8} = \boxed{4.16}$

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

$= \frac{1}{7}\left(141.4035 - \frac{(33.29)^2}{8}\right)$

$= \boxed{0.411}$

b) <u>for new sample</u>: $\frac{\sum x_{new}}{32} = 4.55$ (mean)

$\therefore \sum x_{new} = 4.55(32) = 145.6$ //

$S_{new}^2 = \frac{1}{31}\left(\sum x_{new}^2 - \frac{(\sum x_{new})^2}{32}\right) = 0.25$

$\therefore 31(0.25) = \sum x_{new}^2 - \frac{(145.6)^2}{32}$

$\therefore \sum x_{new}^2 = 670.23$ //

<u>for whole sample</u>: $\sum x = 145.6 + 33.29$
$= 178.89$

$\sum x^2 = 141.4035 + 670.23$
$= 811.634$

$\therefore S^2 = \frac{1}{39}\left(811.634 - \frac{(178.89)^2}{40}\right)$

entire sample

$= 0.29725\ldots$

so $S = 0.545$

Standard error $= \frac{\sigma}{\sqrt{n}} = \frac{S}{\sqrt{n}} = \frac{0.545}{\sqrt{40}}$

$\simeq \boxed{0.0862}$

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

$\left[\frac{178.89}{40} \pm 1.96\left(\frac{0.67}{\sqrt{40}}\right)\right]$

$\left[(4.26), (4.68)\right]$