IAL S3 June 2016 [75]

① (a)

	A	B	C	D	E	F	G	H
distance (x)	20.4	22.2	29.9	37.8	25.5	30.2	35.3	16.5
comission (y)	17.7	24.1	20.3	28.3	34.9	29.3	23.6	26.8
r_x	2	3	5	8	4	6	7	1
r_y	1	4	2	6	8	7	3	5
$d\ (r_x - r_y)$	1	-1	3	2	-4	-1	4	-4
d^2	1	1	9	4	16	1	16	16

$\sum d^2 = 64$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(64)}{8(8^2-1)}$$

$$= 1 - \frac{16}{21}$$

$$= 5/21$$

≈ 0.2381 (4DP) [5]

(b) $H_0: \rho = 0$ (there is no correlation)

$H_1: \rho > 0$ (there is positive correlation)

Critical value $= 0.6429$
(5%; $n = 8$)

$0.2381 < 0.6429 \Rightarrow$ the test is not significant; accept H_0.

There is insufficient evidence to suggest that any positive correlation exists between the distance travelled by car and the amount of comission earned. [4]

② (a) H_0: there is no assosciation between the results of candidates and the test centre.

H_1: there is an assosciation between the results of candidates and the test centre.

	A		B		C	
Observed frequency (O_i)	99	108	110	116	68	119
Expected frequency (E_i)	92.48	114.52	100.97	125.03	83.55	103.45
$\frac{O_i^2}{E_i}$	105.98	101.85	119.84	107.62	55.34	136.89

$$X^2 = \sum \frac{O_i^2}{E_i} - N$$

$$= 627.52 - 620$$

$X^2 = 7.52$ (E_i, $\frac{O_i^2}{E_i}$ and X^2 to 2DP)

$v = (3-1)(2-1) = 2$

$\chi^2_2 (5\%) = 5.991$ (tables)

$7.52 > 5.991 \Rightarrow$ the test is significant; reject H_0.

The researcher's studies show that there is an assosciation between the driving test centre and the results of candidates'. [10]

(b) The researcher should further investigate centre C as it contributes most to the test statistic, X^2. [2]

③ (a) The sample is non-random and not representative of the true population; it will be biased towards those who arrive early to work. [2]

(b) • Divide population into three individual strata: (i) Bristol; (ii) Dudley; (iii) Glasgow.

• Bristol: $\frac{856}{2500} \times 150 = 51.36 \approx 51$

Dudley: $\frac{429}{2500} \times 150 = 25.74 \approx 26$

Glasgow: $\frac{1215}{2500} \times 150 = 72.9 \approx 73$

• Number each strata using the table of random numbers. i.e. 1 to 1815 for Glasgow, 1 to 429 for Dudley, and 1 to 856 for Bristol. Choose a random starting point using table of random numbers and select samples at regular intervals. [3]

(c) Since stratified sampling is random and quota sampling is non-random due to the interviewing process involved, it is less likely for a stratified sample to be biased. [1]

(4)(a) let adults = A and children = C.

$H_0: \mu_A = \mu_C$

$H_1: \mu_A < \mu_C$

$$Z = \frac{59.1 - 61.2 - 0}{\sqrt{\left(\frac{5.9^2}{60} + \frac{5.2^2}{50}\right)}}$$

$Z = -1.9834... < -1.6449 \Rightarrow$ the test is significant; reject H_0.

There is sufficient evidence to suggest that the mean time taken by children to complete the task is greater than that of adults. [6]

(b) Due to the large sample sizes, the Central Limit Theorem allows us to assume that the sample means of both adults and children are normally distributed. [1]

(c) The sample variance, s^2, is the same as the population variance, σ^2. [1]

(5) (a) H_0: a uniform distribution, $U[0, 360]$, is a suitable model.

H_1: a uniform distribution, $U[0, 360]$, is not a suitable model.

direction of flight	0 – 72	72 – 140	140 – 190	190 – 260	260 – 360
observed (O_i)	78	69	51	108	144
expected (E_i)	90	85	62.5	87.5	125
$\frac{O_i^2}{E_i}$	67.6	$\frac{4761}{85}$	41.616	$\frac{23328}{175}$	165.888

$$X^2 = \sum \frac{O_i^2}{E_i} - N$$

$= 464.4186... - 450$

$= 14.41862...$

$\therefore X^2 \approx 14.419$ (3DP)

$\nu = 5-1 = 4$

$\chi^2_4 (1\%) = 13.277$

$13.277 < 14.419 \Rightarrow$ the test is significant; reject H_0.

There is evidence to suggest that a uniform distribution is not a suitable model

Kylie's belief is incorrect. [9]

⑥(a) $X \sim N(21, 2^2)$

$Y \sim N(8.5, \sigma^2)$

$W = 3X - 4Y$

$\Rightarrow W \sim N(29, (36+16\sigma^2))$

$P(W \leq 44) = 0.9$

$\Rightarrow P\left(Z < \frac{44-29}{\sqrt{36+16\sigma^2}}\right) = 0.9$

$\Rightarrow P\left(Z > 15/\sqrt{(36+16\sigma^2)}\right) = 0.10$

$$\Rightarrow \frac{15}{\sqrt{36+16\sigma^2}} = 1.2816$$

$$\sqrt{36+16\sigma^2} = \frac{15}{1.2816}$$

$$36+16\sigma^2 = \left(\frac{15}{1.2816}\right)^2$$

$$16\sigma^2 = \left(\frac{15}{1.2816}\right)^2 - 36$$

$$\sigma^2 = \frac{1}{16}\left[\left(\frac{15}{1.2816}\right)^2 - 36\right]$$

$\sigma^2 = 6.31165\ldots$

$\sigma = \sqrt{(6.31165\ldots)}$

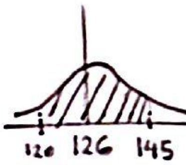
$\sigma = 2.5123\ldots \approx 2.51$ (2DP) [8]

(b) $B = 2X + \sum_{i=1}^{3} A_i$

$B = 2X + A_1 + A_2 + A_3$

$\Rightarrow B \sim N(126, 91)$

$P(B \leq 145 \mid B > 120)$



$$= \frac{P(B \leq 145 \cap B > 120)}{P(B > 120)}$$

$$= \frac{P(120 < B < 145)}{1 - P(B < 120)}$$

$$= \frac{P\left(\frac{120-126}{\sqrt{91}} < Z < \frac{145-126}{\sqrt{91}}\right)}{1 - P\left(Z < \frac{120-126}{\sqrt{91}}\right)}$$

$$= \frac{P(-0.63 < Z < 1.99)}{1 - P(Z < -0.63)}$$

$$= \frac{P(Z < 1.99) - P(Z > 0.63)}{1 - P(Z > 0.63)}$$

$$= \frac{\Phi(1.99) - 1 + \Phi(0.63)}{\Phi(0.63)}$$

$$= \frac{0.9767 - 1 + 0.7357}{0.7357}$$

$= 0.9683$ (4DP) [7]

⑦ (a) $\sum x = 1152$

$\sum x^2 = 167218$

$n = 8$

$$\hat{\mu} = \frac{\sum x}{n}$$

$$= \frac{1152}{8}$$

$\therefore \hat{\mu} = 144$ g

$$s^2 = \frac{\sum x^2 - n\hat{\mu}^2}{n-1}$$

$$= \frac{167218 - 8(144)^2}{8-1}$$

$\therefore s^2 = 190$ g^2 [4]

(b) $\sum_{i=1}^{8} (x_i - \mu)^2$ contains an unknown population parameter, μ, and hence, cannot be a statistic. [1]

(c) $$Y = \frac{1}{8}\left(\sum_{i=1}^{8} X_i^2 - 8\bar{X}^2\right)$$

$$E(Y) = E\left[\frac{1}{8}\left(\sum_{i=1}^{8} X_i^2 - 8\bar{X}^2\right)\right]$$

$$= \frac{1}{8}E\left(X_1^2 + X_2^2 + \dots + X_8^2 - 8\bar{X}^2\right)$$

$$= \frac{1}{8}\left[E(X_1^2) + \dots + E(X_8^2) - E(8\bar{X}^2)\right]$$

$$= \frac{1}{8}\left[8E(X^2) - 8E(\bar{X}^2)\right]$$

$E(Y) = E(X^2) - E(\bar{X}^2)$ *

$\Rightarrow E(Y) = Var(X) + E(X)^2 - Var(\bar{X}) - E(\bar{X})^2$

$= Var(X) - Var(\bar{X})$

$Var(X) = E(X^2) - E(X)^2$

$\Rightarrow Var(X) + E(X)^2 = E(X^2)$

$\Rightarrow Var(\bar{X}) + E(\bar{X})^2 = E(\bar{X}^2)$

*

$$E(Y) = \sigma^2 - \frac{\sigma^2}{n}$$

$$= \sigma^2 - \frac{\sigma^2}{8}$$

$$\therefore E(Y) = \frac{7\sigma^2}{8} \quad [2]$$

(d) $\text{bias} = E(Y) - \sigma^2$

$$= \frac{7\sigma^2}{8} - \sigma^2$$

$$= -\frac{\sigma^2}{8}$$

$$|\text{bias}| = \frac{\sigma^2}{8} \quad [2]$$

8 (a) let X = no. of sixes obtained by each student in 30 rolls

$$X \sim B\left(30, \tfrac{1}{6}\right)$$

$$E(X) = np = 30\left(\tfrac{1}{6}\right) = 5$$

$$Var(X) = np(1-p) = 5\left(\tfrac{5}{6}\right) = \frac{25}{6}$$

By the Central Limit Theorem, $\overline{X} \approx\sim N\left(5, \frac{25/6}{50}\right)$ [3]

$$\therefore \overline{X} \sim N\left(5, \tfrac{1}{12}\right)$$

(b) critical values $= \pm 1.9600$

Rejecting H_0 at upper tail,

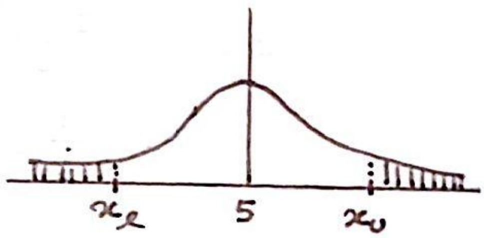
$$\frac{x_U - 5}{\sqrt{(1/12)}} \geqslant 1.96$$

$$x_U \geqslant 1.96\sqrt{\tfrac{1}{12}} + 5$$

Rejecting H_0 at lower tail,

$$\frac{x_L - 5}{\sqrt{1/12}} \leqslant -1.96$$

$$x_L \leqslant 5 - 1.96\sqrt{\tfrac{1}{12}}$$



$\therefore$ The critical region is $\overline{X} \leqslant 4.43$ and $\overline{X} \geqslant 5.57$. (3 SF) [4]