

Sampling Distributions

1 (3 marks)

(a) statistic (calculated from a sample)

B1

(b) parameter (describes the whole population)

B1

(c) statistic

B1

Answer: (a) statistic (b) parameter (c) statistic

3 marks

2 (2 marks)

$$E(\bar{X}) = \mu = 50$$

B1

$$SD(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{8}{4} = 2$$

B1

Answer: mean 50, standard deviation 2

2 marks

3 (2 marks)

Mean 120

B1

Variance $\frac{15^2}{9} = 25$, so $\bar{X} \sim N(120, 25)$

B1

Answer: $\bar{X} \sim N(120, 25)$

2 marks

4 (2 marks)

The probability distribution of the statistic (e.g. the sample mean)

B1

over all possible random samples of the given size taken from the population

B1

Answer: distribution of the statistic over all possible samples of a fixed size

2 marks

5 (3 marks)

Samples: (1,1) (1,2) (1,3) (2,1) (2,2) (2,3) (3,1) (3,2) (3,3)

B1

Means: 1, 1.5, 2, 1.5, 2, 2.5, 2, 2.5, 3

B1 B1

Answer: 9 equally likely samples with means 1, 1.5, 2, 1.5, 2, 2.5, 2, 2.5, 3

3 marks

6 (3 marks)

$$\bar{X} \sim N(50, 2^2)$$

B1

$$P(\bar{X} > 52) = P(Z > 1)$$

M1

$$= 0.1587$$

A1

Answer: 0.1587

3 marks

7 (4 marks)

Possible means 1, 1.5, 2, 2.5, 3

B1

$$P(\bar{X} = 1) = \frac{1}{2}^2 = \frac{1}{4}, P(\bar{X} = 1.5) = 2 \times \frac{1}{2} \times \frac{3}{10} = \frac{3}{10}$$

M1 A1

$$P(\bar{X} = 2) = 2 \times \frac{1}{2} \times \frac{1}{5} + \frac{3}{10}^2 = \frac{29}{100}, P(\bar{X} = 2.5) = 2 \times \frac{3}{10} \times \frac{1}{5} = \frac{3}{25}, P(\bar{X} = 3) = \frac{1}{25}$$

A1

$$\text{Answer: } P(\bar{X} = 1) = \frac{1}{4}, P(\bar{X} = 3/2) = \frac{3}{10}, P(\bar{X} = 2) = \frac{29}{100}, P(\bar{X} = 5/2) = \frac{3}{25}, P(\bar{X} = 3) = \frac{1}{25}$$

4 marks

8 (4 marks)

The median is 1 if at least two heads occur, otherwise 0

M1

$$P(\text{median} = 1) = P(2 \text{ heads}) + P(3 \text{ heads}) = 3(0.4)^2(0.6) + (0.4)^3$$

M1

$$= \frac{44}{125}$$

A1

$$P(\text{median} = 0) = \frac{81}{125}$$

A1

$$\text{Answer: } P(M = 0) = \frac{81}{125}, P(M = 1) = \frac{44}{125}$$

4 marks

9 (4 marks)

$n = 50$ is large, so by the central limit theorem $\bar{X}$ is approximately normal E1

$$\bar{X} \approx N(3.5, \frac{35}{12 \times 1/50}) = N(3.5, 0.0583), \text{ standard deviation } 0.242 \quad \text{B1}$$

$$z = \frac{3.9 - 3.5}{0.242} = 1.656 \quad \text{M1}$$

$$P(\bar{X} > 3.9) = 0.0488 \quad \text{A1}$$

Answer: 0.0488 4 marks

10 (3 marks)

By the central limit theorem, for a large sample the sample mean is approximately normal, whatever the shape of the population E1

Mean 3.2 B1

$$\text{Variance } \frac{2.4^2}{80} = 0.0720, \text{ so } \bar{X} \approx N(3.2, 0.0720) \quad \text{B1}$$

Answer: $\bar{X} \approx N(3.2, 0.0720)$ 3 marks

11 (4 marks)

$T = \text{sum of 12 masses: } T \sim N(12 \times 1005, 12 \times 6^2) = N(12060, 432)$ M1 A1

$$z = \frac{12100 - 12060}{\sqrt{432}} = 1.925 \quad \text{M1}$$

$$P(T > 12100) = 0.0271 \quad \text{A1}$$

Answer: 0.0271 4 marks

12 (2 marks)

$$\frac{12}{\sqrt{n}} = 1.5 \quad \text{M1}$$

$$\sqrt{n} = 8 \text{ so } n = 64 \quad \text{A1}$$

Answer: $n = 64$ 2 marks

13 (3 marks)

$\bar{P} \geq 0.6$ means at least 12 heads of 20 M1

$X \sim B(20, 0.5)$ B1

$$P(X \geq 12) = 0.2517 \quad \text{A1}$$

Answer: 0.2517 3 marks

14 (7 marks)

$$P(Z > 1.96) = 0.025, \text{ so } 83.92 = \mu + 1.96 \times \frac{12}{\sqrt{n}} \quad \text{B1 M1}$$

$$P(Z < -0.9) = 0.1841, \text{ so } 78.2 = \mu - 0.9 \times \frac{12}{\sqrt{n}} \quad \text{B1 M1}$$

$$\text{Subtracting: } 5.72 = 2.86 \times \frac{12}{\sqrt{n}}, \text{ so } \sqrt{n} = 6 \text{ and } n = 36 \quad \text{M1 A1}$$

$$\mu = 83.92 - 1.96 \times 2 = 80 \quad \text{A1}$$

Answer: $\mu = 80, n = 36$ 7 marks

15 (9 marks)

$$(a) E(X) = 2 \times \frac{1}{4} + 4 \times \frac{1}{2} + 8 \times \frac{1}{4} = \frac{9}{2} \quad \text{M1 A1}$$

$$\text{Var}(X) = E(X^2) - (E(X))^2 = 25 - \frac{9^2}{2} = \frac{19}{4} \quad \text{A1}$$

$$\text{Answer: } E(X) = \frac{9}{2}, \text{ Var}(X) = \frac{19}{4} \quad \text{3 marks}$$

(b) List the 9 ordered pairs: the possible means are 2, 3, 4, 5, 6 and 8 M1

$$P(\bar{X} = 2) = \frac{1}{16}, P(\bar{X} = 3) = \frac{1}{4}, P(\bar{X} = 4) = \frac{1}{4}, P(\bar{X} = 5) = \frac{1}{8}, P(\bar{X} = 6) = \frac{1}{4}, P(\bar{X} = 8) = \frac{1}{16} \quad \text{A1 A1}$$

$$\text{Answer: } P(\bar{X} = 2) = \frac{1}{16}, P(\bar{X} = 3) = \frac{1}{4}, P(\bar{X} = 4) = \frac{1}{4}, P(\bar{X} = 5) = \frac{1}{8}, P(\bar{X} = 6) = \frac{1}{4}, P(\bar{X} = 8) = \frac{1}{16} \quad \text{3 marks}$$

$$(c) E(\bar{X}) = \frac{9}{2} \quad \text{B1}$$

$$\text{Var}(\bar{X}) = \frac{19}{8} \quad \text{M1}$$

$$\frac{\text{Var}(X)}{2} = \frac{\frac{19}{4}}{2} = \frac{19}{8}, \text{ equal} \quad \text{A1}$$

Answer: both verified 3 marks

16 (8 marks)

(a) $T = X_1 + X_2$ takes the values 4, 5, 6, 7, 8 B1

$$P(T = 4) = 0.2^2 = 0.04; P(T = 5) = 2 \times 0.2 \times 0.5 = 0.20 \quad \text{M1 A1}$$

$$P(T = 6) = 2 \times 0.2 \times 0.3 + 0.5^2 = 0.37 \quad \text{A1}$$

$$P(T = 7) = 2 \times 0.5 \times 0.3 = 0.30; P(T = 8) = 0.3^2 = 0.09 \quad \text{A1}$$

Answer: $P(T = 4, 5, 6, 7, 8) = 0.04, 0.20, 0.37, 0.30, 0.09$ 5 marks

$$(b) E(T) = 4 \times 0.04 + 5 \times 0.20 + 6 \times 0.37 + 7 \times 0.30 + 8 \times 0.09 = 6.2 \quad \text{M1 A1}$$

$$E(X) = 2 \times 0.2 + 3 \times 0.5 + 4 \times 0.3 = 3.1, \text{ and } 2 \times 3.1 = 6.2 = E(T) \quad \text{B1}$$

Answer: $E(T) = 6.2 = 2E(X)$ 3 marks

17 (4 marks)

$$\bar{X} \sim N(100, \frac{15^2}{40}), \text{ standard deviation } 2.37 \quad \text{B1}$$

$$P(\bar{X} > c) = 0.05, \text{ so } \frac{c - 100}{2.37} = 1.645 \quad \text{M1 B1}$$

$$c = 104 \quad \text{A1}$$

Answer: 104 4 marks

18 (6 marks)

(a) Pairs with sum 8: (1,7), (3,5), (5,3), (7,1), each with probability $\frac{1}{16}$ M1 A1

$$\text{So } P(\bar{X} = 4) = \frac{4}{16} = \frac{1}{4} \quad \text{A1}$$

$$(b) P(\bar{X} \geq 5) = P(5) + P(6) + P(7) = \frac{3}{16} + \frac{2}{16} + \frac{1}{16} = \frac{3}{8} \quad \text{M1 A1}$$

$$\text{Var}(\bar{X}) = \frac{\text{Var}(X)}{2} = \frac{5}{2} = \frac{5}{2} \quad \text{A1}$$

Answer: (b) $\frac{3}{8}, \frac{5}{2}$ 6 marks

19 (5 marks)

$$\bar{A} \sim N(52, \frac{36}{16}) \text{ and } \bar{B} \sim N(50, \frac{25}{25}) \quad \text{B1}$$

$$\bar{A} - \bar{B} \sim N(2, 2.25 + 1) = N(2, 3.25) \quad \text{M1 A1}$$

$$P(\bar{A} - \bar{B} > 0) = P(Z > \frac{0 - 2}{\sqrt{3.25}}) = P(Z > -1.109) \quad \text{M1}$$

$$= 0.8664 \quad \text{A1}$$

Answer: 0.8664 5 marks

20 (4 marks)

$$E(X) = \text{Var}(X) = 2.5; n = 40 \text{ is large, so } \bar{X} \approx N(2.5, \frac{2.5}{40}) \text{ by the central limit theorem} \quad \text{M1 A1}$$

$$z = \frac{3 - 2.5}{\sqrt{0.0625}} = 2.000 \quad \text{M1}$$

$$P(\bar{X} > 3) = 0.0228 \quad \text{A1}$$

Answer: 0.0228 4 marks

21 (6 marks)(a) 27 equally likely samples. $R = 0$ for 3 samples (all equal) M1 A1 $R = 2$ when the sample contains both 0 and 2: $27 - 8$ (no 2s) $- 8$ (no 0s) $+ 1$ (all 1s) = 12 samples M1 A1 $R = 1$ for the remaining 12 samples A1 $P(R = 0) = \frac{1}{9}$, $P(R = 1) = \frac{4}{9}$, $P(R = 2) = \frac{4}{9}$ (check: probabilities sum to 1)(b) $E(R) = 0 \times \frac{1}{9} + 1 \times \frac{4}{9} + 2 \times \frac{4}{9} = \frac{4}{3}$ A1**Answer:** $P(R=0) = \frac{1}{9}$, $P(R=1) = \frac{4}{9}$, $P(R=2) = \frac{4}{9}$; $E(R) = \frac{4}{3}$ 6 marks**22 (10 marks)**(a) $(X_1, X_2) = (2, 2)$: $Y = 2$, probability 0.16; $(2, 7)$: $Y = 5.75$, probability 0.24 M1 A1 $(7, 2)$: $Y = 3.25$, probability 0.24; $(7, 7)$: $Y = 7$, probability 0.36 A1 A1**Answer:** $Y = 2, 3.25, 5.75, 7$ with probabilities **0.16, 0.24, 0.24, 0.36** 4 marks(b) $E(Y) = 2 \times 0.16 + 3.25 \times 0.24 + 5.75 \times 0.24 + 7 \times 0.36 = 5$ M1 A1 $E(Y^2) = 4 \times 0.16 + 3.25^2 \times 0.24 + 5.75^2 \times 0.24 + 49 \times 0.36 = 28.75$, so $\text{Var}(Y) = 28.75 - 25 = 3.75$ A1**Answer:** $E(Y) = 5$, $\text{Var}(Y) = 3.75$ 3 marks(c) $E(X) = 2 \times 0.4 + 7 \times 0.6 = 5 = E(Y)$, so Y is unbiased B1 $\text{Var}(X) = (4 \times 0.4 + 49 \times 0.6) - 25 = 6$, so $\text{Var}(\bar{X}) = 3$ M1 $\text{Var}(\bar{X}) = 3 < 3.75 = \text{Var}(Y)$, so $\bar{X}$ is the better estimator E1**Answer:** Y is unbiased; $\bar{X}$ has the smaller variance and is better 3 marks**23 (6 marks)**(a) $\bar{X} \sim N(500, 0.8^2)$ when working correctly B1 $P(\bar{X} < c) = 0.01$: $c = 500 - 2.326 \times 0.8$ M1 $c = 498$ A1**Answer:** $c = 498$ 3 marks(b) $\bar{X} \sim N(497, 0.8^2)$ B1 $P(\bar{X} < 498) = P(Z < \frac{498 - 497}{0.8})$ M1 $= 0.9227$ A1**Answer:** **0.9227** 3 marks**24 (6 marks)** $n = 64$ is large, so by the central limit theorem $\bar{X} \approx N(12, \frac{6^2}{64}) = N(12, 0.75^2)$ E1 B1 $P(11 < \bar{X} < 13.5) = P(-1.333 < Z < 2)$ M1 A1 $= 0.8860$ A1For $n = 4$ the sample is too small for $\bar{X}$ to be approximately normal when X is skewed E1**Answer:** **0.8860** 6 marks

25 (9 marks)

(a) $M = 1$ if at least two 1s: $P = 3\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)^3 = \frac{1}{2}$ M1 A1

$M = 3$ if at least two 3s: $P = 3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right) + \left(\frac{1}{4}\right)^3 = \frac{5}{32}$ M1 A1

$M = 2$: $P = 1 - \frac{1}{2} - \frac{5}{32} = \frac{11}{32}$ A1

(b) $E(M) = 1 \times \frac{1}{2} + 2 \times \frac{11}{32} + 3 \times \frac{5}{32} = \frac{53}{32}$ M1 A1

Population mean $\mu = \frac{7}{4}$, so $E(M) = \frac{53}{32} \neq \mu$ B1

(c) The sample median is a biased estimator of the mean for this skewed population, while $E(\bar{X}) = \mu$ E1

Answer: $E(M) = \frac{53}{32}$, $\mu = \frac{7}{4}$ 9 marks

26 (9 marks)

(a) 16 equally likely ordered pairs: $P(\bar{X} = 1) = \frac{1}{16}$, $P(\bar{X} = 3/2) = \frac{1}{8}$, $P(\bar{X} = 2) = \frac{3}{16}$, $P(\bar{X} = 5/2) = \frac{1}{4}$, $P(\bar{X} = 3) = \frac{3}{16}$, $P(\bar{X} = 7/2) = \frac{1}{8}$, $P(\bar{X} = 4) = \frac{1}{16}$ M1 A1 A1

(b) 12 equally likely ordered pairs of different values: $P(\bar{X} = 3/2) = \frac{1}{6}$, $P(\bar{X} = 2) = \frac{1}{6}$, $P(\bar{X} = 5/2) = \frac{1}{3}$, $P(\bar{X} = 3) = \frac{1}{6}$, $P(\bar{X} = 7/2) = \frac{1}{6}$ M1 A1 A1

(c) With replacement $\text{Var} = \frac{5}{8}$; without replacement $\text{Var} = \frac{5}{12}$ M1 A1

Without replacement the spread is smaller because the two values cannot be equal and so cannot both be extreme; the formula $\frac{\sigma^2}{n}$ only holds for independent observations E1

Answer: $\text{Var} \frac{5}{8}$ (with), $\frac{5}{12}$ (without) 9 marks

27 (9 marks)

(a) $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$, so $E(\bar{X}) = \frac{1}{n} \times n\mu = \mu$ M1 A1

$\text{Var}(\bar{X}) = \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) = \frac{1}{n^2} \times n\sigma^2$ (independence) M1 A1
 $= \frac{\sigma^2}{n}$ A1

(b) $\text{SD} = \frac{\sigma}{\sqrt{n}}$; with 4n: $\frac{\sigma}{\sqrt{4n}} = \frac{\sigma}{2\sqrt{n}} = \text{half}$ M1 A1

(c) $\frac{\sigma}{\sqrt{25}} = 3$ so $\sigma = 15$; $\frac{15}{\sqrt{n}} = 0.75$ gives $\sqrt{n} = 20$, $n = 400$ M1 A1

Answer: (c) $n = 400$ 9 marks

28 (9 marks)

(a) $X = \text{number defective} \sim B(n, 0.3)$; $\bar{P} = \frac{X}{n}$: $E(\bar{P}) = 0.3$, $\text{Var}(\bar{P}) = \frac{0.3 \times 0.7}{n} = \frac{0.21}{n}$ M1 A1 A1

(b) $n = 100$: $\bar{P} \approx N(0.3, 0.0021)$, standard deviation 0.0458 M1
 $P(0.24 < \bar{P} < 0.36) = P(|Z| < 1.309) = 0.8096$ M1 A1

(c) Need $\frac{0.05}{\sqrt{0.21/n}} \geq 1.96$ M1 B1

$n \geq 0.21 \times \frac{(1.96)^2}{(0.05)^2} = 322.7$, so $n = 323$ A1

Answer: (a) 0.3, $\frac{0.21}{n}$ (b) 0.8096 (c) $n = 323$ 9 marks

29 (8 marks)

(a) $M \leq m$ requires all three scores $\leq m$: $P = \left(\frac{m}{6}\right)^3$ M1 A1

$P(M = m) = \left(\frac{m}{6}\right)^3 - \left(\frac{m-1}{6}\right)^3$: $P(M = 1) = \frac{1}{216}$, $P(M = 2) = \frac{7}{216}$, $P(M = 3) = \frac{19}{216}$, $P(M = 4) = \frac{37}{216}$, $P(M = 5) = \frac{61}{216}$, $P(M = 6) = \frac{91}{216}$ M1 A1

(b) $E(M) = \sum m P(M = m) = \frac{119}{24} \approx 4.96$ M1 A1

(c) The greatest face is 6 but $E(M) = \frac{119}{24} < 6$, so M is biased E1

The bias is negative (M underestimates the largest value on average) E1

Answer: $E(M) = \frac{119}{24}$ 8 marks

30 (8 marks)

(a) $T = A_1 + \dots + A_4 + C_1 + C_2 + C_3$: $E(T) = 4 \times 78 + 3 \times 36 = 420$ M1 A1

$\text{Var}(T) = 4 \times 11^2 + 3 \times 8^2 = 676$ M1 A1

$P(T < 450) = P\left(Z < \frac{450 - 420}{\sqrt{676}}\right) = 0.8757$ A1

(b) $P(T > 480) = P(Z > 2.308) = 0.0105$ M1 A1

(c) Masses are independent, and the people are a random sample (normally distributed) E1

Answer: (a) 0.8757 (b) 0.0105 8 marks

Estimators and Bias**1 (2 marks)**

A statistic T calculated from a random sample used to estimate θ B1

T is unbiased if $E(T) = \theta$ B1

Answer: $E(T) = \theta$ 2 marks

2 (3 marks)

Mean = $\frac{65}{5} = 13$ B1

$\sum (x - \bar{x})^2 = 1 + 4 + 4 + 1 + 0 = 10$ M1

$s^2 = \frac{10}{4} = 2.5$ A1

Answer: mean 13, variance 2.5 3 marks

3 (3 marks)

Mean = $\frac{420}{10} = 42$ B1

$S_{xx} = 17850 - \frac{420^2}{10} = 210$ M1

$s^2 = \frac{210}{9} = 23.3$ A1

Answer: mean 42, $s^2 = 23.3$ 3 marks

4 (3 marks)

$E(T) = \frac{1}{4}(E(X_1) + 3E(X_2))$ M1

$= \frac{1}{4}(\mu + 3\mu)$ A1

$= \mu$, so T is unbiased A1

Answer: $E(T) = \mu$ 3 marks

5 (2 marks)

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

B2

Answer: $s^2 = (\sum x^2 - (\sum x)^2/n)/(n - 1)$

2 marks**6 (2 marks)**

The sample proportion $\frac{X}{n}$ is unbiased for p

M1

$$= \frac{6}{40} = 0.15$$

A1

Answer: 0.15

2 marks**7 (4 marks)**

$$\sum fx = 59, \sum fx^2 = 135$$

B1 B1

$$\bar{x} = \frac{59}{40} = 1.475$$

A1

$$s^2 = \frac{1}{39}(135 - \frac{59^2}{40}) = 1.23$$

A1

Answer: mean 1.475, $s^2 = 1.23$

4 marks**8 (4 marks)**

$$\bar{y} = \frac{36}{12} = 3, \text{ so } \bar{x} = 53$$

M1 A1

$$s_y^2 = \frac{1}{11}(215 - \frac{36^2}{12}) = 9.73$$

M1

Subtracting a constant does not change the variance, so $s_x^2 = 9.73$

A1

Answer: mean 53, $s^2 = 9.73$

4 marks**9 (4 marks)**

$$E(T) = \frac{1}{6}(\mu + 2\mu + 3\mu) = \mu$$

M1 A1

$$\text{Var}(T) = \frac{1}{36}(1^2 + 2^2 + 3^2)\sigma^2$$

M1

$$= \frac{14}{36}\sigma^2 = \frac{7}{18}\sigma^2$$

A1

Answer: unbiased; $\text{Var}(T) = 7\sigma^2/18$

4 marks**10 (4 marks)**

$$E(X) = np$$

B1

$$E(\bar{P}) = \frac{1}{n} \times np = p, \text{ so it is unbiased}$$

M1 A1

$$\text{Var}(\bar{P}) = \frac{1}{n^2} \times np(1-p) = \frac{p(1-p)}{n}$$

A1

Answer: $E = p$, $\text{Var} = p(1-p)/n$

4 marks**11 (4 marks)**

$$\bar{x} = \frac{15}{6} = 2.5, \text{ an unbiased estimate of } \lambda$$

M1 A1

$$\text{Var}(\bar{X}) = \frac{\lambda}{n}, \text{ estimated by } \frac{2.5}{6}$$

M1

$$\text{Standard error} = \sqrt{2.5/6} = 0.645$$

A1

Answer: 2.5; 0.645

4 marks**12 (3 marks)**

$$\text{Standard error} = \frac{6}{\sqrt{25}} = 1.2$$

M1 A1

Halving needs $\sqrt{n}$ doubled, so $n = 100$

B1

Answer: 1.2; $n = 100$

3 marks

13 (5 marks)

$$E(T_1) = \frac{\mu + \mu}{2} = \mu \text{ and } E(T_2) = \frac{\mu + 3\mu}{4} = \mu, \text{ so both unbiased}$$

M1 A1

$$\text{Var}(T_1) = \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{\sigma^2}{2}$$

M1

$$\text{Var}(T_2) = \frac{1}{16}(\sigma^2 + 9\sigma^2) = \frac{5\sigma^2}{8}$$

M1

$$\frac{1}{2} < \frac{5}{8}, \text{ so } T_1 \text{ has the smaller variance and is the better estimator}$$

A1

Answer: T_1

5 marks

14 (6 marks)

$$(a) E(T) = (a + b)\mu$$

M1

$$T \text{ unbiased iff } (a + b)\mu = \mu, \text{ i.e. } a + b = 1$$

A1

$$(b) \text{Var}(T) = (a^2 + b^2)\sigma^2 \text{ with } b = 1 - a$$

M1

$$= (a^2 + 1 - 2a + a^2)\sigma^2 = \sigma^2(2a^2 - 2a + 1)$$

A1

$$\frac{d}{da}: 4a - 2 = 0$$

M1

$$a = \frac{1}{2}, \text{ the sample mean is the best choice}$$

A1

Answer: $a = 1/2$

6 marks

15 (6 marks)

$$E(X_i^2) = \text{Var}(X_i) + \mu^2 = \sigma^2 + \mu^2$$

B1

$$E(\bar{X}^2) = \text{Var}(\bar{X}) + \mu^2 = \frac{\sigma^2}{n} + \mu^2$$

B1

$$E(\sum X_i^2 - n\bar{X}^2) = n(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)$$

M1 A1

$$= (n - 1)\sigma^2, \text{ so } E(T) = \frac{n - 1}{n}\sigma^2$$

A1

$$\text{Bias} = E(T) - \sigma^2 = -\frac{\sigma^2}{n} \text{ (T underestimates } \sigma^2)$$

A1

Answer: bias = $-\sigma^2/n$

6 marks

16 (5 marks)

$$n = 30, \sum x = 1350, \sum x^2 = 61530$$

B1

$$\bar{x} = \frac{1350}{30} = 45$$

M1 A1

$$s^2 = \frac{1}{29}(61530 - \frac{1350^2}{30}) = 26.9$$

M1 A1

Answer: mean 45, $s^2 = 26.9$

5 marks

17 (5 marks)

$$E(X) = \int_0^\theta \frac{x}{\theta} dx = \frac{\theta}{2}$$

M1 A1

$$E(2\bar{X}) = 2E(X) = \theta, \text{ so it is unbiased}$$

A1

$$\text{Var}(X) = \int_0^\theta \frac{x^2}{\theta} dx - (\frac{\theta}{2})^2 = \frac{\theta^2}{3} - \frac{\theta^2}{4} = \frac{\theta^2}{12}$$

M1

$$\text{Var}(2\bar{X}) = 4 \times \frac{\theta^2}{12n} = \frac{\theta^2}{3n}$$

A1

Answer: unbiased; Var = $\theta^2/(3n)$

5 marks

18 (5 marks)

$$\Sigma x = 37.8, \Sigma x^2 = 181.16$$

B1

$$\bar{x} = 4.725$$

A1

$$s^2 = \frac{1}{7}(181.16 - \frac{37.8^2}{8}) = 0.3650$$

M1 A1

$$\text{Standard error} = \sqrt{\frac{s^2}{8}} = 0.214$$

A1

Answer: mean 4.725, $s^2 = 0.3650$, SE = 0.214

5 marks

19 (4 marks)

$$\Sigma(x - \bar{x})^2 = 12 \times 7.5 = 90$$

M1

$$s^2 = \frac{90}{11} = 8.18, s = 2.86$$

A1

$$n = 21: \Sigma(x - \bar{x})^2 = 157.5, s^2 = \frac{157.5}{20} = 7.88$$

M1 A1

Answer: 8.18; 7.88

4 marks

20 (5 marks)

$$E(T_1) = \frac{3\lambda}{3} = \lambda, E(T_2) = \frac{3\lambda}{3} = \lambda, \text{ both unbiased}$$

M1 A1

$$\text{Var}(T_1) = \frac{1}{9} \times 3\lambda = \frac{\lambda}{3}$$

M1

$$\text{Var}(T_2) = \frac{1}{9}(\lambda + 4\lambda) = \frac{5\lambda}{9}$$

M1

$$\frac{1}{3} = \frac{3}{9} < \frac{5}{9}, \text{ so } T_1 \text{ is more efficient}$$

A1

Answer: T_1

5 marks

21 (9 marks)

$$(a) E(T_1) = \mu; E(T_2) = \frac{\mu + 2\mu - \mu}{2} = \mu; E(T_3) = \frac{3\mu}{3} = \mu$$

M1 A1 A1

$$(b) \text{Var}(T_1) = \frac{\sigma^2}{3}$$

B1

$$\text{Var}(T_2) = \frac{1}{4}(1 + 4 + 1)\sigma^2 = \frac{3\sigma^2}{2}$$

M1 A1

$$\text{Var}(T_3) = \frac{1}{9}(4 + 1)\sigma^2 = \frac{5\sigma^2}{9}$$

M1 A1

$$(c) T_1: \text{smallest variance } (\frac{1}{3} < \frac{5}{9} < \frac{3}{2}), \text{ so it is the most efficient}$$

E1

Answer: T_1 (variance $\sigma^2/3$)

9 marks

22 (7 marks)

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

M1

$$= \frac{\sigma^2}{n} + \mu^2, \text{ which is not } \mu^2, \text{ so } \bar{X}^2 \text{ is biased (bias } \frac{\sigma^2}{n})$$

A1 E1

$$(b) \bar{X}^2 - \frac{\sigma^2}{n} \text{ has expectation } \mu^2$$

M1 A1

$$(c) 12^2 - \frac{100}{25}$$

M1

$$= 144 - 4 = 140$$

A1

Answer: 140

7 marks

23 (7 marks)

$$(a) E((X - Y)^2) = \text{Var}(X - Y) + (E(X - Y))^2 = 2\sigma^2 + 0$$

M1 A1

$$\text{So } E\left(\frac{(X - Y)^2}{2}\right) = \sigma^2$$

A1

$$(b) \frac{1}{2k} \sum (X_i - Y_i)^2$$

M1 A1

(c) Differences: 0.7, -1.1, 0.5, -0.9; squares 0.49, 1.21, 0.25, 0.81, total 2.76

M1

$$\text{Estimate} = \frac{2.76}{8} = 0.345$$

A1

Answer: 0.345

7 marks

24 (6 marks)

$$(a) \frac{12}{60} = \frac{80}{N}, \text{ so } N = \frac{80 \times 60}{12} = 400$$

M1 A1

(b) The marked fish mix randomly with the rest, with no births, deaths or migration in between; marked fish are not harmed or less likely to be caught

B1 B1

(c) The estimate $\frac{80 \times 60}{R}$ is undefined (infinite) when no marked fish are recaptured

E1

Mark more fish and/or take a larger second sample so that R is large and the estimate is more reliable

E1

Answer: (a) 400

6 marks

25 (6 marks)

$$\text{Bias} = E(S) - \mu = \mu + 1 - \mu = 1$$

M1 A1

$$\text{Var}(S) = \frac{\sigma^2}{2}$$

B1

$$\text{MSE} = \text{Var} + \text{bias}^2 = \frac{\sigma^2}{2} + 1$$

M1 A1

$$\text{MSE}(\bar{X}) = \frac{\sigma^2}{2} < \frac{\sigma^2}{2} + 1: S \text{ is worse; the bias adds 1 to the error whatever the size of } \sigma^2$$

E1

Answer: bias 1, MSE $\sigma^2/2 + 1$

6 marks

26 (7 marks)

$$(a) E(X^2) = \text{Var}(X) + (E(X))^2 = \lambda + \lambda^2$$

M1 A1

$$E(X(X - 1)) = E(X^2) - E(X) = \lambda + \lambda^2 - \lambda = \lambda^2$$

A1

(b) $X(X - 1)$ is unbiased for λ^2

B1

(c) $E(X^2) = \lambda^2 + \lambda \neq \lambda^2$, bias = λ (positive)

M1 A1

$$(d) 5 \times 4 = 20$$

B1

Answer: (d) 20

7 marks

27 (9 marks)

$$(a) \sum (X_i - \bar{X})^2 = \sum X_i^2 - 2\bar{X}\sum X_i + n\bar{X}^2$$

M1

$$= \sum X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 = \sum X_i^2 - n\bar{X}^2$$

A1

$$(b) E(\sum X_i^2) = n(\sigma^2 + \mu^2), E(n\bar{X}^2) = n\left(\frac{\sigma^2}{n} + \mu^2\right) = \sigma^2 + n\mu^2$$

M1 A1 A1

$$E(\sum (X_i - \bar{X})^2) = n\sigma^2 - \sigma^2 = (n - 1)\sigma^2$$

M1

$$\text{So } E(S^2) = \sigma^2$$

A1

$$(c) \text{Dividing by } n \text{ gives } E = \frac{n-1}{n}\sigma^2 < \sigma^2$$

M1

so it underestimates σ^2 on average

E1

Answer: $E(S^2) = \sigma^2$; dividing by n gives $E = \frac{n-1}{n}\sigma^2 < \sigma^2$

9 marks

28 (11 marks)

$$(a) E(T) = w\mu + (1 - w)\mu = \mu \quad \text{M1 A1}$$

$$(b) \text{Var}(T) = w^2 \frac{\sigma^2}{n_1} + (1 - w)^2 \frac{\sigma^2}{n_2} \quad \text{M1 A1}$$

$$(c) \frac{d}{dw}: 2w/n_1 - 2(1 - w)/n_2 = 0 \quad \text{M1 A1}$$

$$w(n_2 + n_1) = n_1, \text{ so } w = \frac{n_1}{n_1 + n_2} \quad \text{M1 A1}$$

$$\text{Then } T = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2} = \text{mean of all } n_1 + n_2 \text{ observations} \quad \text{A1}$$

$$(d) w = \frac{10}{40} = 0.25 \quad \text{B1}$$

$$\text{Minimum variance} = \frac{\sigma^2}{n_1 + n_2} = \frac{\sigma^2}{40} \quad \text{B1}$$

Answer: $w = n_1/(n_1 + n_2)$; (d) 0.25, $\sigma^2/40$ 11 marks

29 (12 marks)

$$(a) P(M \leq m) = \left(\frac{m}{\theta}\right)^3; \text{ differentiate: } f(m) = \frac{3m^2}{\theta^3} \quad \text{M1 A1}$$

$$(b) E(M) = \int_0^\theta \frac{3m^3}{\theta^3} dm = \frac{3\theta}{4} \quad \text{M1 A1}$$

$$\text{Bias} = \frac{3\theta}{4} - \theta = -\frac{\theta}{4} \quad \text{A1}$$

$$(c) k = \frac{4}{3} \quad \text{B1}$$

$$(d) E(M^2) = \int_0^\theta \frac{3m^4}{\theta^3} dm = \frac{3\theta^2}{5} \quad \text{M1 A1}$$

$$\text{Var}(M) = \frac{3\theta^2}{5} - \frac{9\theta^2}{16} = \frac{3\theta^2}{80} \quad \text{A1}$$

$$\text{Var}\left(\frac{4}{3}M\right) = \frac{16}{9} \times \frac{3\theta^2}{80} = \frac{\theta^2}{15} \quad \text{M1 A1}$$

$$\frac{1}{15} < \frac{1}{9}, \text{ so } \frac{4}{3}M \text{ is more efficient than } 2\bar{X} \quad \text{E1}$$

Answer: $k = 4/3$; $\text{Var} = \theta^2/15$ 12 marks

30 (9 marks)

$$(a) E(\bar{P}^2) = \text{Var}(\bar{P}) + (E(\bar{P}))^2 = \frac{p(1-p)}{n} + p^2 \quad \text{M1 A1}$$

$$(b) E(T) = E(\bar{P}) - E(\bar{P}^2) = p - p^2 - \frac{p(1-p)}{n} \quad \text{M1 A1}$$

$$= p(1-p)\left(1 - \frac{1}{n}\right) = \frac{n-1}{n}p(1-p) \quad \text{M1 A1}$$

$$(c) \frac{n}{n-1}T \text{ is unbiased} \quad \text{B1}$$

$$n = 20, X = 6: T = 0.3 \times 0.7 = 0.21 \quad \text{M1}$$

$$\frac{20}{19} \times 0.21 = 0.221 \quad \text{A1}$$

Answer: $20T/19 = 0.221$ 9 marks

Confidence Intervals

1 (4 marks)

$\bar{X} \sim N(\mu, \frac{10^2}{25})$ with standard error $\frac{10}{\sqrt{25}} = 2.00$

B1

z-value for a 95% interval: 1.960

B1

Interval: $40 \pm 1.960 \times 2.00$

M1

(36.1, 43.9)

A1

Answer: (36.1, 43.9)

4 marks

2 (3 marks)

(a) $P(Z < z) = 0.95$: $z = 1.645$

B1

(b) $P(Z < z) = 0.975$: $z = 1.960$

B1

(c) $P(Z < z) = 0.995$: $z = 2.576$

B1

Answer: 1.645, 1.960, 2.576

3 marks

3 (3 marks)

(a) The population mean μ is a fixed constant, not a random variable

E1

95% of intervals constructed in this way (from repeated samples) would contain μ

E1

Answer: 95% of such intervals contain μ

2 marks

(b) 1 interval does not contain μ ; the expected number is $20 \times 0.05 = 1$, so this is consistent

B1

Answer: 1, consistent

1 mark

4 (2 marks)

Sample mean = midpoint = $\frac{23.4 + 28.6}{2} = 26$

B1

Margin of error = $\frac{28.6 - 23.4}{2} = 2.6$

B1

Answer: mean 26, margin 2.6

2 marks

5 (2 marks)

(a) The width decreases: the standard error $\frac{\sigma}{\sqrt{n}}$ is smaller

B1

(b) The width increases: a larger z-value is needed

B1

Answer: (a) narrower (b) wider

2 marks

6 (4 marks)

$\bar{X} \sim N(\mu, \frac{0.06^2}{36})$ with standard error $\frac{0.06}{\sqrt{36}} = 0.0100$

B1

z-value for a 99% interval: 2.576

B1

Interval: $1.02 \pm 2.576 \times 0.0100$

M1

(0.994, 1.05)

A1

Answer: (0.994, 1.05)

4 marks

7 (4 marks)

Sample mean = $\frac{112.4 + 119.8}{2} = 116.1$

B1

$1.96 \times \text{standard error} = \frac{119.8 - 112.4}{2} = 3.7$, so standard error = 1.888

M1

90% interval: $116.1 \pm 1.645 \times 1.888$

M1

(113.0, 119.2)

A1

Answer: (113.0, 119.2)

4 marks

8 (5 marks)

$$\text{Sample proportion } \bar{p} = \frac{56}{200} = 0.280$$

B1

$$\text{Standard error} = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} = \sqrt{\frac{0.280 \times 0.720}{200}} = 0.0317$$

M1 A1

$$90\%: z = 1.645, \text{ so interval } 0.280 \pm 1.645 \times 0.0317$$

M1

$$(0.228, 0.332)$$

A1

Answer: (0.228, 0.332)

5 marks

9 (5 marks)

$$\bar{x} = \frac{806}{8} = 100.75$$

M1 A1

$$\text{Standard error} = \frac{3.2}{\sqrt{8}} = 1.13$$

B1

$$100.75 \pm 1.96 \times 1.13$$

M1

$$(98.5, 103)$$

A1

Answer: (98.5, 103)

5 marks

10 (4 marks)

$$\text{Width} = 2 \times 1.96 \times \frac{12}{\sqrt{n}} \leq 6$$

M1 B1

$$\sqrt{n} \geq \frac{2 \times 1.96 \times 12}{6} = 7.84$$

M1

$$n \geq 61.5, \text{ so } n = 62$$

A1

Answer: n = 62

4 marks

11 (4 marks)

$$\text{Sample mean} = \frac{48.6 + 53.4}{2} = 51$$

B1

$$\text{Margin} = 2.4 = 1.96 \times \frac{\sigma}{\sqrt{36}}$$

M1 B1

$$\sigma = \frac{2.4 \times 6}{1.96} = 7.35$$

A1

Answer: mean 51, $\sigma = 7.35$

4 marks

12 (5 marks)

$$\bar{x} = \frac{54}{6} = 9$$

B1

$$\text{For large samples } \bar{X} \approx N\left(\lambda, \frac{\lambda}{n}\right); \text{ estimate the standard error by } \sqrt{\frac{9}{6}} = 1.22$$

M1 A1

$$9 \pm 1.96 \times 1.22$$

M1

$$(6.60, 11.4)$$

A1

Answer: (6.60, 11.4)

5 marks

13 (5 marks)

$$\text{Standard error } \frac{8}{\sqrt{64}} = 1$$

B1

$$49.6 \pm 1.96 \times 1$$

M1

$$(47.6, 51.6)$$

A1

50 lies within the interval

M1

so the data are consistent with a mean of 50 ml at this level (no evidence against the claim)

A1

Answer: (47.6, 51.6); consistent with 50

5 marks

14 (5 marks)

$n = 100$ is large, so the sample mean is approximately normal and s^2 is a good estimate of σ^2 (central limit theorem) E1

$$\text{Standard error} = \frac{\sqrt{18.5}}{\sqrt{100}} = 0.430 \quad \text{M1 A1}$$

$$64.2 \pm 1.96 \times 0.430 \quad \text{M1}$$

(63.4, 65.0) A1

Answer: (63.4, 65.0) 5 marks

15 (4 marks)

$$\text{Standard error} = \frac{9}{\sqrt{36}} = 1.5 \quad \text{B1}$$

$$z = \frac{2.5}{1.5} = 1.67 \quad \text{M1}$$

$$\text{Confidence level} = P(-z < Z < z) = 2\Phi(1.67) - 1 = 90.4\% \quad \text{M1}$$

Answer: 90.4% A1

4 marks

16 (5 marks)

$$\text{Standard error of the sample proportion} = \sqrt{\frac{0.65 \times 0.35}{n}} \quad \text{B1}$$

$$\text{The interval does not contain 0.5, so the lower limit exceeds 0.5: } 0.65 - 1.96\sqrt{\frac{0.2275}{n}} > 0.5 \quad \text{M1}$$

$$\sqrt{\frac{0.2275}{n}} < \frac{0.15}{1.96} \quad \text{M1}$$

$$n > 0.2275 \times \left(\frac{1.96}{0.15}\right)^2 = 38.8 \quad \text{A1}$$

The least value is $n = 39$ A1

Answer: $n = 39$ 5 marks

17 (4 marks)

$$\text{Margin of error} = 1.96\sqrt{\frac{p(1-p)}{n}} \leq 0.03 \quad \text{M1}$$

$$p(1-p) \leq \frac{1}{4} \text{ for all } p \text{ (maximum at } p = \frac{1}{2}) \quad \text{M1 A1}$$

$$\text{So } 1.96 \times \frac{1}{2\sqrt{n}} \leq 0.03, \text{ giving } n \geq \left(\frac{1.96}{0.06}\right)^2 = 1067.1, \text{ so } n = 1068 \quad \text{A1}$$

Answer: $n = 1068$ 4 marks

18 (5 marks)

$$\bar{x} = \frac{1830}{50} = 36.6 \quad \text{B1}$$

$$s^2 = \frac{1}{49} \left(68500 - \frac{1830^2}{50} \right) = 31.1 \quad \text{M1 A1}$$

$$\text{Standard error} = \sqrt{\frac{31.1}{50}} = 0.788 \quad \text{M1}$$

$$36.6 \pm 2.576 \times 0.788 = (34.6, 38.6) \quad \text{A1}$$

Answer: (34.6, 38.6) 5 marks

19 (4 marks)

$$\text{Standard error} = \frac{s}{\sqrt{n}} = \frac{3.1}{\sqrt{10}} = 0.980 \quad \text{M1 A1}$$

$$23.4 \pm 2.262 \times 0.980 \quad \text{M1}$$

(21.2, 25.6) A1

Answer: (21.2, 25.6) 4 marks

20 (6 marks)

$$\bar{x} = \frac{62}{5} = 12.4$$

B1

$$\text{Standard error} = \sqrt{\frac{\lambda}{n}} \approx \sqrt{\frac{12.4}{5}} = 1.57$$

M1 A1

$$12.4 \pm 2.576 \times 1.57 = (8.34, 16.5)$$

M1

Per hour (6 intervals): multiply by 6: (50.1, 98.7)

M1 A1

Answer: (8.34, 16.5) per 10 min; (50.1, 98.7) per hour

6 marks

21 (7 marks)

$$E(X) = \frac{(k-3) + (k+5)}{2} = k+1$$

M1 A1

$$\text{Var}(X) = \frac{8^2}{12} = \frac{16}{3}$$

B1

By the central limit theorem $\bar{X} \approx N(k+1, \frac{16}{3/80})$, standard error = 0.2582

M1

98% interval for $k+1$: $41.2 \pm 2.326 \times 0.2582 = 41.2 \pm 0.601$

M1 A1

Subtract 1: (39.6, 40.8)

A1

Answer: (39.6, 40.8)

7 marks

22 (7 marks)

(a) Standard error $\frac{6.4}{8} = 0.8$

B1

$$47.5 \pm 1.96 \times 0.8 = (45.9, 49.1)$$

M1 A1

(b) Multiply the limits by 7: (322, 343)

M1 A1

(c) $350 = 7 \times 50$ is outside the interval

M1

The claim is not supported at the 5% level

A1

Answer: (a) (45.9, 49.1) (b) (322, 343)

7 marks

23 (7 marks)

(a) $X \sim B(100, 0.95)$

B1

(b) $E(X) = 100 \times 0.95 = 95$

B1

(c) $P(X \geq 90) = 1 - P(X \leq 89) = 0.9885$

M1 A1

(d) $0.95^{100} = 0.0059$

M1 A1

Even though each interval has 95% confidence, it is very unlikely that all 100 contain μ

E1

Answer: (c) 0.9885 (d) 0.0059

7 marks

24 (6 marks)

$$\bar{A} - \bar{B} \sim N(\mu_A - \mu_B, \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B})$$

M1

$$\text{Variance} = \frac{36}{40} + \frac{25}{50} = 1.4, \text{ standard error } 1.18$$

M1 A1

Estimate $78.2 - 74.9 = 3.3$

B1

$$3.3 \pm 1.96 \times 1.18 = (0.981, 5.62)$$

A1

0 is not in the interval: evidence that the means differ

E1

Answer: (0.981, 5.62)

6 marks

25 (7 marks)

(a) $\bar{X} \sim N(\mu, \frac{18^2}{30})$; $P(\mu < \bar{x} + 1.645\sigma_{\bar{x}}) = 0.95$

M1

$$b = 112 + 1.645 \times \frac{18}{\sqrt{30}} = 117.4$$

A1

(b) All of the 5% is in one tail, so z satisfies $P(Z < z) = 0.95$ not 0.975

E1

$$(c) z = \frac{118 - 112}{18/\sqrt{30}} = 1.83$$

M1 A1

Confidence level = $\Phi(1.83) = 0.966$

M1 A1

Answer: (a) $\mu < 117.4$ (c) 0.966

7 marks

26 (7 marks)

Mid-points 5, 15, 25, 35, 45: $\sum fx = 2030$, $\sum fx^2 = 60600$

M1 A1

$$\bar{x} = \frac{2030}{80} = 25.375$$

A1

$$s^2 = \frac{1}{79}(60600 - \frac{2030^2}{80}) = 115$$

M1 A1

$$25.375 \pm 1.96 \times \sqrt{\frac{115}{80}} = (23.0, 27.7)$$

M1

Limitation: using mid-points assumes all values in a class equal the mid-point so the mean and variance are only estimates

E1

Answer: (23.0, 27.7)

7 marks

27 (9 marks)

(a) Rearranging: $\bar{X} - z \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z \frac{\sigma}{\sqrt{n}}$

M1 A1

(b) Width = $(\bar{X} + z \frac{\sigma}{\sqrt{n}}) - (\bar{X} - z \frac{\sigma}{\sqrt{n}}) = \frac{2z\sigma}{\sqrt{n}}$

M1 A1

(c) $n = \frac{1200}{6} = 200$

B1

Width = $\frac{2 \times 1.96 \times 25}{\sqrt{200}} = 6.93$

M1 A1

(d) Halving width needs 4 times the sample: $n = 800$ costing £4 800, i.e. £3 600 more

M1 A1

Answer: (c) 6.93 (d) £3 600

9 marks

28 (10 marks)

(a) $\bar{p} = 0.35$; $SE = \sqrt{\frac{0.35 \times 0.65}{400}} = 0.0238$

B1 M1

$0.35 \pm 1.96 \times 0.0238 = (0.303, 0.397)$

M1 A1

(b) $\bar{p}(1 - \bar{p}) = \frac{1}{4} - (\bar{p} - \frac{1}{2})^2 \leq \frac{1}{4}$

M1 A1

So the margin $1.96 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} \leq 1.96 \times \frac{1}{2\sqrt{n}} = \frac{1.96}{2\sqrt{n}}$

A1

(c) $\frac{1.96}{2\sqrt{n}} \leq 0.02$ gives $\sqrt{n} \geq 49$, so $n \geq 2401$

M1 A1

(d) 0.40 lies above the upper limit 0.397, so the claim of being behind 40% is supported

E1

Answer: (a) (0.303, 0.397) (c) 2401

10 marks

29 (9 marks)

(a) $50.8 \pm 1.96 \times \frac{4}{5} = (49.2, 52.4)$

M1 A1

(b) H_0 is rejected when $|\frac{\bar{x} - \mu_0}{4/5}| > 1.96$

M1

i.e. $|\bar{x} - \mu_0| > 1.96 \times 0.8$, which is exactly when μ_0 is not in $\bar{x} \pm 1.96 \times 0.8$

M1 A1

(c) $52.6 > 52.4$ so 52.6 is outside the interval: reject H_0

M1 A1

49.5 is inside (49.2, 52.4): do not reject H_0

A1

(d) Greatest μ_0 = upper limit = 52.37

B1

Answer: (a) (49.2, 52.4) (c) reject; do not reject (d) 52.37

9 marks

30 (9 marks)

$$(a) \text{ Width} = 2 \times 2.576 \times \frac{14.8}{\sqrt{n}} \leq 4$$

M1 B1

$$\sqrt{n} \geq \frac{2 \times 2.576 \times 14.8}{4} = 19.1$$

M1

$$n \geq 363.4, \text{ so } n = 364$$

A1

$$(b) \text{ SE} = \frac{16.0}{\sqrt{364}} = 0.839$$

M1

$$62.3 \pm 2.576 \times 0.839 = (60.1, 64.5)$$

A1

Width = 4.32, which is greater than 4

M1 A1

(c) The pilot sample is small, so its estimate of σ is unreliable; a larger true σ gives a wider interval

E1

Answer: (a) $n = 364$

9 marks

Estimation of Population Parameters**1 (3 marks)**

$$\bar{x} = \frac{44}{6} = 7.33$$

B1

$$\Sigma(x - \bar{x})^2 = 23.3$$

M1

$$s^2 = \frac{23.3}{5} = 4.67$$

A1

Answer: mean 7.33, variance 4.67

3 marks

2 (3 marks)

$$\bar{p} = \frac{18}{120} = 0.15$$

B1

$$\text{SE} = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} = \sqrt{\frac{0.15 \times 0.85}{120}}$$

M1

$$= 0.0326$$

A1

Answer: 0.15; SE = 0.0326

3 marks

3 (3 marks)

$$\bar{x} = \frac{340}{20} = 17$$

B1

$$s^2 = \frac{1}{19} \left(6200 - \frac{340^2}{20} \right) = \frac{420}{19}$$

M1

$$= 22.1$$

A1

Answer: mean 17, variance 22.1

3 marks

4 (2 marks)

Estimated mean 2.4 per household

B1

$$\text{Total} \approx 2.4 \times 4200 = 10\,080$$

B1

Answer: 10 080

2 marks

5 (2 marks)

$$s = \sqrt{52.9} = 7.27$$

B1

$$\text{SE} = \frac{s}{\sqrt{30}} = 1.33$$

B1

Answer: $s = 7.27$, $\text{SE} = 1.33$

2 marks

6 (2 marks)

$$\frac{8}{50} = \frac{40}{N}$$

M1

$$N = \frac{40 \times 50}{8} = 250$$

A1

Answer: 250

2 marks

7 (4 marks)

$$\Sigma fx = 887, \Sigma fx^2 = 19721$$

B1 B1

$$\bar{x} = \frac{887}{40} = 22.175$$

A1

$$s^2 = \frac{1}{39} \left(19721 - \frac{887^2}{40} \right) = 1.3276$$

A1

Answer: mean 22.175, variance 1.3276

4 marks

8 (5 marks)

$$\bar{y} = \frac{18}{15} = 1.2$$

M1

$$\bar{x} = 2\bar{y} + 200 = 202.4$$

A1

$$s_y^2 = \frac{1}{14} \left(305 - \frac{18^2}{15} \right) = 20.2$$

M1 A1

$$x = 2y + 200, \text{ so } s_x^2 = 4s_y^2 = 81.0$$

A1

Answer: mean 202.4, variance 81.0

5 marks

9 (6 marks)

$$\Sigma A = 25 \times 8.2 = 205, \Sigma B = 35 \times 9.0 = 315$$

M1

$$\text{Combined mean} = \frac{520}{60} = 8.67$$

A1

$$\Sigma x^2(A) = 24 \times 3.1 + \frac{205^2}{25} = 1755.4, \Sigma x^2(B) = 34 \times 2.7 + \frac{315^2}{35} = 2926.8$$

M1 A1

$$s^2 = \frac{1}{59} \left(4682.2 - \frac{520^2}{60} \right) = 2.98$$

M1 A1

Answer: mean 8.67, variance 2.98

6 marks

10 (4 marks)

$$\text{Both estimates are } \bar{p} = 0.3$$

B1

$$SE(A) = \sqrt{\frac{0.3 \times 0.7}{100}} = 0.0458$$

M1 A1

$$SE(B) = \sqrt{\frac{0.3 \times 0.7}{400}} = 0.0229; B \text{ is more precise because the standard error falls as } n \text{ increases}$$

A1

Answer: SE(A) = 0.0458, SE(B) = 0.0229

4 marks

11 (4 marks)

$$E(X) = \theta + 4\theta + 3(1 - 3\theta)$$

M1

$$= 5\theta + 3 - 9\theta = 3 - 4\theta$$

A1

$$E(\bar{X}) = 3 - 4\theta, \text{ so } E\left(\frac{3 - \bar{X}}{4}\right) = \theta$$

M1

$$\text{Unbiased estimator: } \frac{3 - \bar{X}}{4}$$

A1

Answer: (3 - $\bar{X}$)/4

4 marks

12 (5 marks)

$$\bar{p} = \frac{7}{80} = 0.0875$$

B1

$$\text{Estimated number} = 2000 \times 0.0875 = 175$$

B1

$$SE(\bar{p}) = \sqrt{\frac{0.0875 \times 0.912}{80}} = 0.0316$$

M1

$$\text{Interval for } p: 0.0875 \pm 1.96 \times 0.0316$$

M1

$$\text{Multiply by 2000: (51.2, 299)}$$

A1

Answer: 175; (51.2, 299)

5 marks

13 (5 marks)

$$\Sigma x = 1650, \Sigma x^2 = 340992$$

B1

$$\bar{x} = 206.25$$

A1

$$s^2 = \frac{1}{7}(340992 - \frac{1650^2}{8}) = 97.1$$

M1 A1

$$s = 9.85, SE = \frac{s}{\sqrt{8}} = 3.48$$

A1

Answer: mean 206.25, s = 9.85, SE = 3.48

5 marks

14 (5 marks)(a) Weight each group by its population size: $\frac{600 \times 172.4 + 400 \times 160.8}{1000}$

M1 A1

= 168 cm

A1

(b) The mean of the 50 students is 170.1 cm, which over-represents boys (80% of the sample but 60% of the school)

M1 E1

Answer: 168 cm

5 marks

15 (5 marks)

$$\bar{x} = \frac{37}{10} = 3.70$$

B1

$$s^2 = \frac{1}{9}(153 - \frac{37^2}{10}) = 1.79$$

M1 A1

 $\text{Var}(\bar{X}) = \frac{\lambda}{n}$, which is smaller than the variance of S^2 (for large n)

E1

Use $\bar{x} = 3.70$: it uses a simple unbiased statistic with smaller variance

A1

Answer: $\bar{x} = 3.70, s^2 = 1.79$

5 marks

16 (5 marks)

$$E(X) = \int_0^{\theta} x \times \frac{2x}{\theta^2} dx$$

M1

$$= \frac{2}{\theta^2} \times \frac{\theta^3}{3} = \frac{2\theta}{3}$$

A1

$$E(T) = \frac{3}{2}E(\bar{X}) = \frac{3}{2} \times \frac{2\theta}{3}$$

M1 A1

= θ , so T is unbiased

A1

Answer: $E(X) = 2\theta/3$

5 marks

17 (5 marks) $n = 60$ is large, so s^2 is a good estimate of σ^2 and $\bar{X}$ is approximately normal

E1

$$SE = \sqrt{\frac{225}{60}} = 1.94, z = \frac{124.2 - 120}{1.94} = 2.169$$

M1 A1

$$P(Z > 2.169) = 0.0150 < 0.05$$

M1

Reject H_0

A1

Answer: $z = 2.169$: reject

5 marks

18 (4 marks)

$$(12 - 1) \times 9.6 = 105.6$$

M1

$$(18 - 1) \times 7.8 = 132.6$$

M1

$$\text{Pooled estimate} = \frac{105.6 + 132.6}{28}$$

M1

$$= 8.51$$

A1

Answer: 8.51

4 marks

19 (4 marks)

$$\text{Margin} = 1.96 \sqrt{\frac{0.3 \times 0.7}{n}} \leq 0.05$$

M1 B1

$$n \geq 0.21 \times \left(\frac{1.96}{0.05}\right)^2 = 322.7$$

M1

$$n = 323$$

A1

Answer: n = 323

4 marks

20 (5 marks)

(a) Before: $\bar{x} = 14.5$, $s^2 = 1.17$

B1 M1

After: $\bar{x} = \frac{174}{11} = 15.8$

M1

$$s^2 = \frac{1}{10} \left(2954 - \frac{174^2}{11} \right) = 20.2$$

A1

(b) The outlier increases the variance estimate greatly (17.3 times) and moves the mean, so the estimates are sensitive to extreme values

E1

Answer: before 14.5, 1.17; after 15.8, 20.2

5 marks

21 (7 marks)

(a) Sample 1: $N \approx \frac{100 \times 60}{8} = 750$

M1 A1

Sample 2: $N \approx \frac{100 \times 80}{11} = 727$

A1

(b) Combined: marked proportion = $\frac{8 + 11}{60 + 80} = \frac{19}{140}$

M1 A1

$$N \approx \frac{100 \times 140}{19} = 737$$

A1

(c) The population is closed (no births, deaths or movements) and marked fish mix randomly and are equally likely to be caught

B1

Answer: (a) 750, 727 (b) 737

7 marks

22 (8 marks)

(a) Weighted mean = $\frac{500 \times 41.2 + 300 \times 38.6 + 200 \times 45.1}{1000}$

M1 A1

$$= 41.2$$

A1

(b) A: $0.5^2 \times \frac{36}{50} = 0.180$, B: $0.3^2 \times \frac{25}{30} = 0.0750$, C: $0.2^2 \times \frac{49}{20} = 0.0980$

M1 A1

Total variance = 0.353, standard error 0.594

M1 A1

$$41.2 \pm 1.96 \times 0.594 = (40.0, 42.4)$$

A1

Answer: mean 41.2; (40.0, 42.4)

8 marks

23 (6 marks)

(a) $a + b = 28$ and $(b - a)^2 = 12 \times 8 = 96$

M1 A1

$$b - a = \sqrt{96} = 9.80$$

M1 A1

$$a = 9.10, b = 18.9$$

A1

(b) The estimates are derived from sample moments only; an observation outside the interval proves the interval estimate is too narrow (the estimates are not constrained by the data)

E1

Answer: a = 9.10, b = 18.9

6 marks

24 (6 marks)

(a) $P(X = 0) = e^{-\lambda}$ estimated by $\frac{37}{100} = 0.37$

M1

$$\lambda = -\ln 0.37 = 0.994$$

M1 A1

(b) The sample mean estimate is 1.1, which differs slightly from 0.994

B1

(c) The sample mean uses all the data, not just whether a page has zero errors, so it has smaller variance

E1

In addition $\bar{x}$ is unbiased for λ whereas $-\ln(\bar{p})$ is biased

E1

Answer: 0.994

6 marks

25 (7 marks)

$$(a) E(X^2) = \text{Var}(X) + (E(X))^2 = np(1-p) + n^2p^2$$

M1 A1

$$E(X(X-1)) = np - np^2 + n^2p^2 - np = n(n-1)p^2$$

M1 A1

$$(b) \frac{X(X-1)}{n(n-1)}$$

B1

$$(c) \frac{9 \times 8}{12 \times 11} = 0.545, \text{ whereas } (\bar{p})^2 = 0.5625$$

M1 A1

Answer: 0.545 vs 0.5625

7 marks

26 (7 marks)

$$(a) SE = \sqrt{\frac{4.2}{100}} = 0.205$$

M1 A1

$$5.4 \pm 1.96 \times 0.205 = (5.00, 5.80)$$

M1 A1

(b) The sample is taken at the library and so over-represents students who study a lot offline (not random, biased)

E1

The sample may not represent all years/faculties (a convenience sample)

E1

Improve by sampling randomly from a list of all students (e.g. a random number generator)

E1

Answer: (5.00, 5.80)

7 marks

27 (12 marks)

$$(a) E(T) = c\mu, \text{ so bias} = (c-1)\mu$$

M1 A1

$$\text{Var}(T) = c^2 \frac{\sigma^2}{n}$$

A1

$$(b) \text{MSE} = \text{Var} + \text{bias}^2 = c^2 \frac{\sigma^2}{n} + (c-1)^2 \mu^2$$

M1

$$= c^2 \left(\frac{\sigma^2}{n} + \mu^2 \right) - 2c\mu^2 + \mu^2$$

A1

$$(c) \frac{d}{dc}: 2c \left(\frac{\sigma^2}{n} + \mu^2 \right) - 2\mu^2 = 0, c = \frac{\mu^2}{\mu^2 + \frac{\sigma^2}{n}}$$

M1 A1

The denominator exceeds the numerator, so $c < 1$

E1

$$(d) \frac{\sigma^2}{n} = 25: c = \frac{100}{125} = 0.8$$

B1

$$\text{MSE}(c=0.8) = 0.64 \times 125 - 2 \times 0.8 \times 100 + 100 = 20$$

M1 A1

$$\text{MSE}(c=1) = 25$$

A1

Answer: $c = \mu^2/(\mu^2 + \sigma^2/n)$; (d) 0.8; 20 vs 25

12 marks

28 (9 marks)

$$(a) E(S_p^2) = \frac{(n_1-1)\sigma^2 + (n_2-1)\sigma^2}{n_1 + n_2 - 2}$$

M1 A1

$$= \sigma^2$$

A1

$$(b) \text{Var}(S_p^2) = \frac{(n_1-1)^2 \text{Var}(s_1^2) + (n_2-1)^2 \text{Var}(s_2^2)}{(n_1 + n_2 - 2)^2}$$

M1 A1

$$= \frac{2\sigma^4[(n_1-1) + (n_2-1)]}{(n_1 + n_2 - 2)^2} = \frac{2\sigma^4}{n_1 + n_2 - 2}$$

M1 A1

$$(c) n_1 = 8, n_2 = 12: \text{Var}(S_p^2) = \frac{2\sigma^4}{18} = 0.111\sigma^4; \text{Var}(s_1^2) = \frac{2\sigma^4}{7} = 0.286\sigma^4$$

M1

Pooling gives less than half the variance, so it is more precise

A1

Answer: proof; $0.111\sigma^4$ vs $0.286\sigma^4$

9 marks

29 (10 marks)

$$(a) \bar{p} = \frac{252}{400} = 0.63$$

B1

$$SE = \sqrt{\frac{0.63 \times 0.37}{400}} = 0.0241$$

M1 A1

$$0.63 \pm 1.96 \times 0.0241 = (0.583, 0.677)$$

M1 A1

(b) 0.60 lies inside the interval, so the data are consistent with the claim

E1

$$(c) 1.96 \sqrt{\frac{0.63 \times 0.37}{n}} \leq 0.025 \text{ gives } n \geq 0.2331 \times \left(\frac{1.96}{0.025}\right)^2 = 1432.8, n = 1433$$

M1 A1

(d) Non-response bias: people who recycle may be more likely to reply, which would overestimate p

E1 E1

Answer: (a) 0.63, (0.583, 0.677) (c) n = 1433

10 marks**30 (8 marks)**

$$(a) E(\bar{X}) = E(X) = \frac{1}{\lambda}, \text{ so it is unbiased}$$

M1 A1

$$\text{Var}(\bar{X}) = \frac{1}{n\lambda^2}$$

A1

$$(b) \bar{x} = \frac{9.9}{8} = 1.24, \text{ so } \hat{\lambda} = \frac{1}{1.24} = 0.808$$

M1 A1

$$(c) SE = \frac{\bar{x}}{\sqrt{n}} = \frac{1.24}{\sqrt{8}} = 0.438$$

M1

$$1.24 \pm 1.96 \times 0.438 = (0.380, 2.10)$$

A1

(d) The exponential distribution is skewed and $n = 8$ is small, so the normal approximation is poor and the interval may be unreliable

E1

Answer: (b) 1.24, 0.808 (c) (0.380, 2.10)

8 marks