

Sampling Distributions

- 1** State whether each of the following is a statistic or a parameter:
 (a) the mean mass of 30 apples picked from a crate
 (b) the mean mass of all apples grown on a farm
 (c) the largest value in a random sample. [3]

- 2** A population has mean 50 and standard deviation 8. Random samples of size 16 are taken. Write down the mean and the standard deviation of the sample mean $\bar{X}$. [2]

- 3** $X \sim N(120, 15^2)$. A random sample of 9 observations is taken. State the distribution of the sample mean $\bar{X}$. [2]

- 4** Explain what is meant by the sampling distribution of a statistic. [2]

- 5** A population consists of the numbers 1, 2 and 3, each occurring with probability $\frac{1}{3}$. A random sample of size 2 is taken with replacement. List all possible samples and find the mean of each. [3]

- 6** $X \sim N(50, 10^2)$. A random sample of 25 observations is taken. Find $P(\bar{X} > 52)$. [3]

- 7** The random variable X has the distribution shown. Two independent observations X_1 and X_2 are made. Find the sampling distribution of the sample mean $\bar{X} = \frac{X_1 + X_2}{2}$.

x	1	2	3
P	1/2	3/10	1/5

[4]

- 8** A biased coin lands heads with probability 0.4. It is tossed three times and the number of heads is recorded, scoring 1 for heads and 0 for tails. Find the sampling distribution of the median of the three scores. [4]

- 9** A fair six-sided die is rolled 50 times. The score X on a single roll has mean 3.5 and variance $35/12$. Use the central limit theorem to find the probability that the mean score of the 50 rolls exceeds 3.9. [4]

- 10** The times taken by customers to be served in a shop are skewed, with mean 3.2 minutes and standard deviation 2.4 minutes. Explain why the mean of a sample of 80 customers can be modelled by a normal distribution and state the distribution of the sample mean. [3]

- 11** The mass of a bag of sugar is normally distributed with mean 1005 g and standard deviation 6 g. A box holds 12 randomly chosen bags. Find the probability that the total mass of the bags in a box exceeds 12 100 g. [4]

- 12** A population has standard deviation 12. Find the sample size n needed so that the sample mean has a standard deviation of 1.5. [2]

- 13** A fair coin is tossed 20 times. Let $\bar{P}$ be the proportion of heads. Find $P(\bar{P} \geq 0.6)$. [3]

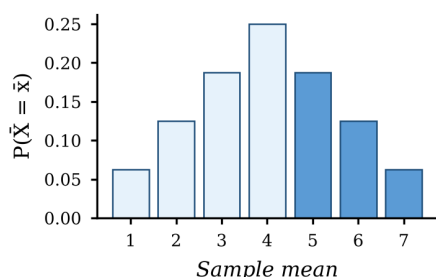
- 14** The random variable X is normally distributed with mean μ and standard deviation 12. The mean of a random sample of n observations of X is denoted by $\bar{X}$. Given that $P(\bar{X} > 83.92) = 0.025$ and $P(\bar{X} < 78.2) = 0.1841$, find the values of μ and n . [7]

- 15** X takes the values 2, 4 and 8 with probabilities $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$ respectively. Two independent observations are taken and their mean $\bar{X}$ is calculated.
- (a) Find $E(X)$ and $\text{Var}(X)$. [3]
- (b) Find the sampling distribution of $\bar{X}$. [3]
- (c) Verify that $E(\bar{X}) = E(X)$ and $\text{Var}(\bar{X}) = \frac{\text{Var}(X)}{2}$. [3]

- 16** A café sells small, medium and large drinks costing £2, £3 and £4. The probabilities that a customer buys a small, medium or large drink are 0.2, 0.5 and 0.3 respectively. Two customers are chosen at random and T is the total amount, in pounds, that they spend.
- (a) Find the sampling distribution of T. [5]
- (b) Find $E(T)$ and verify that $E(T) = 2E(X)$. [3]

- 17** The scores on a test are normally distributed with mean 100 and standard deviation 15. Find the value c such that the mean of a random sample of 40 scores exceeds c with probability 0.05. [4]

- 18** A population consists of the four numbers 1, 3, 5 and 7, each equally likely. A sample of size 2 is chosen with replacement. The sample mean is $\bar{X}$. The diagram shows the sampling distribution of $\bar{X}$, with the bars for $\bar{X} \geq 5$ shaded.
- (a) Show that $P(\bar{X} = 4) = \frac{1}{4}$.
- (b) Find $P(\bar{X} \geq 5)$ and $\text{Var}(\bar{X})$.



- 19** The reaction times (ms) of group A are $N(52, 6^2)$ and of group B are $N(50, 5^2)$. A random sample of 16 from A and an independent random sample of 25 from B are taken, with sample means $\bar{A}$ and $\bar{B}$. Find $P(\bar{A} > \bar{B})$. [5]

- 20** The number of faults X on a sheet of metal has a Poisson distribution with mean 2.5. The faults on 40 randomly chosen sheets are counted. Use the central limit theorem to find the probability that the mean number of faults per sheet exceeds 3. [4]

- 21** A population takes the values 0, 1 and 2, each with probability $\frac{1}{3}$. A random sample of size 3 is taken with replacement and R is the range of the sample.
- (a) Find the sampling distribution of R.
- (b) Find $E(R)$. [6]

- 22** The random variable X takes the values 2 and 7 with probabilities 0.4 and 0.6 respectively. Two independent observations X_1 and X_2 are made and the statistic $Y = \frac{X_1 + 3X_2}{4}$ is calculated.
- (a) Find the sampling distribution of Y. [4]
- (b) Find $E(Y)$ and $\text{Var}(Y)$. [3]
- (c) Show that Y is an unbiased estimator of $E(X)$, and say, with a reason, whether Y or $\bar{X}$ is the better estimator. [3]

- 23** A machine fills bags with a mass that is normally distributed with standard deviation 4 g. When working correctly the mean is 500 g. A quality inspector takes a random sample of 25 bags each hour and stops the machine if the sample mean is below c grams.
- (a) Find c so that the probability of stopping the machine unnecessarily is 0.01. [3]
- (b) The machine develops a fault and the true mean falls to 497 g. Find the probability that the inspector's check stops the machine. [3]
-
- 24** The daily number of orders X received by a shop has mean 12 and standard deviation 6 and is skewed. Over a random sample of 64 days, find the probability that the sample mean number of orders is between 11 and 13.5. Explain why the same method would be unreliable for a sample of 4 days. [6]
-
- 25** A population takes the values 1, 2 and 3 with probabilities $\frac{1}{2}$, $\frac{1}{4}$ and $\frac{1}{4}$. A random sample of size 3 is taken and M is the sample median.
- (a) Find the sampling distribution of M .
- (b) Find $E(M)$ and compare it with the population mean.
- (c) State what your answer suggests about M as an estimate of the population mean. [9]
-
- 26** A population consists of the numbers 1, 2, 3 and 4. Samples of size 2 are taken at random, and the sample mean is $\bar{X}$.
- (a) Find the sampling distribution of $\bar{X}$ when sampling is with replacement.
- (b) Find it when sampling is without replacement.
- (c) Find $\text{Var}(\bar{X})$ in each case and comment. [9]
-
- 27** $X_1, X_2, \dots, X_n$ are independent observations from a population with mean μ and variance σ^2 , and $\bar{X}$ is their mean.
- (a) Prove that $E(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$.
- (b) Hence show that the standard deviation of the sample mean is halved when the sample size is quadrupled.
- (c) A sample of 25 has $\bar{X}$ standard deviation 3. How large must the sample be for this standard deviation to be 0.75? [9]
-
- 28** A manufacturer says that 30% of its components are defective. In a random sample of n components, $\bar{P}$ is the proportion that are defective.
- (a) Write down $E(\bar{P})$ and $\text{Var}(\bar{P})$ in terms of n .
- (b) For $n = 100$, use a normal approximation to find $P(0.24 < \bar{P} < 0.36)$.
- (c) Find the smallest n such that $P(|\bar{P} - 0.3| < 0.05) \geq 0.95$ using a normal approximation (ignore continuity corrections). [9]
-
- 29** A fair six-sided die is rolled 3 times and M is the largest of the three scores.
- (a) Show that $P(M \leq m) = \left(\frac{m}{6}\right)^3$ for $m = 1, 2, \dots, 6$, and hence find the sampling distribution of M .
- (b) Find $E(M)$.
- (c) Explain why M is a biased estimator of the greatest value on the die, and state the sign of the bias. [8]
-
- 30** The masses of adults are normally distributed with mean 78 kg and standard deviation 11 kg. The masses of children are normally distributed with mean 36 kg and standard deviation 8 kg. A lift carries 4 randomly chosen adults and 3 randomly chosen children.
- (a) Find the probability that their total mass is less than 450 kg.
- (b) The lift is overloaded if the total exceeds 480 kg. Find the probability that it is overloaded.
- (c) State two assumptions you have made. [8]

Estimators and Bias

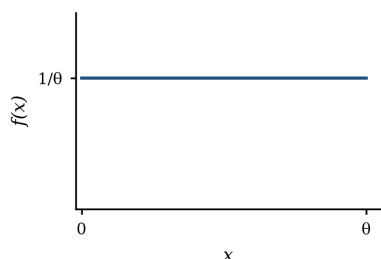
- 1** Explain what is meant by an unbiased estimator of a population parameter θ . [2]
- 2** A random sample of 5 observations is 12, 15, 11, 14, 13. Find unbiased estimates of the population mean and variance. [3]
- 3** A random sample of 10 observations has $\Sigma x = 420$ and $\Sigma x^2 = 17\,850$. Find unbiased estimates of the mean and variance of the population. [3]
- 4** X_1 and X_2 are independent observations from a population with mean μ . Show that $T = \frac{X_1 + 3X_2}{4}$ is an unbiased estimator of μ . [3]
- 5** State the formula, in terms of n , Σx and Σx^2 , for the unbiased estimate s^2 of a population variance. [2]
- 6** A sample of 40 items contains 6 defectives. Write down an unbiased estimate of the proportion p of defective items in the population. [2]
- 7** The table shows the number of cars x owned by each of 40 randomly chosen households. Find unbiased estimates of the mean and variance of the number of cars per household. [4]
- | x | 0 | 1 | 2 | 3 | 4 |
|-----|---|----|----|---|---|
| f | 8 | 14 | 11 | 5 | 2 |
- 8** A random sample of 12 values of x is coded using $y = x - 50$, giving $\Sigma y = 36$ and $\Sigma y^2 = 215$. Find unbiased estimates of the population mean and variance of x . [4]
- 9** X_1, X_2, X_3 are independent observations from a population with mean μ and variance σ^2 . $T = \frac{X_1 + 2X_2 + 3X_3}{6}$. Show that T is unbiased for μ and find $\text{Var}(T)$ in terms of σ^2 . [4]
- 10** $X \sim B(n, p)$. Show that the sample proportion $\bar{P} = \frac{X}{n}$ is an unbiased estimator of p and find its variance. [4]
- 11** A random sample of 6 observations from a Poisson distribution with mean λ is 3, 1, 4, 2, 2, 3. Find an unbiased estimate of λ and the standard error of this estimate (estimated using the sample mean). [4]
- 12** Samples of size n are taken from a population with variance σ^2 . The standard error of the sample mean is defined as $\frac{s}{\sqrt{n}}$. A sample of 25 values has an unbiased variance estimate $s^2 = 36$. Find the standard error and the sample size needed to halve it. [3]
- 13** X_1 and X_2 are independent observations from a population with mean μ and variance σ^2 . Two estimators of μ are $T_1 = \frac{X_1 + X_2}{2}$ and $T_2 = \frac{X_1 + 3X_2}{4}$. Show that both are unbiased, and find which has the smaller variance. [5]

- 14** X_1 and X_2 are independent observations from a population with mean μ and variance σ^2 . $T = aX_1 + bX_2$.
 (a) Show that T is unbiased for μ if and only if $a + b = 1$.
 (b) With $a + b = 1$, show that $\text{Var}(T) = \sigma^2(2a^2 - 2a + 1)$ and find the value of a that minimises it. [6]

- 15** For a random sample $X_1, \dots, X_n$ from a population with mean μ and variance σ^2 , let $T = \frac{1}{n} \sum (X_i - \bar{X})^2$.
 Given that $\sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$, show that $E(T) = \frac{n-1}{n} \sigma^2$, and hence state the bias of T as an estimator of σ^2 . [6]

- 16** Two independent random samples are taken from the same population. Sample 1 has $n = 12$, $\sum x = 540$ and $\sum x^2 = 24\,580$. Sample 2 has $n = 18$, $\sum x = 810$ and $\sum x^2 = 36\,950$. Combine the samples to find unbiased estimates of the population mean and variance. [5]

- 17** A random variable X is uniformly distributed on the interval $[0, \theta]$, with probability density function $f(x) = \frac{1}{\theta}$ for $0 \leq x \leq \theta$, as shown in the diagram. A random sample $X_1, \dots, X_n$ is taken.
 Show that $2\bar{X}$ is an unbiased estimator of θ and find its variance in terms of θ and n .



[5]

- 18** The masses (kg) of 8 randomly chosen parcels are shown. Find unbiased estimates of the population mean and variance, and an estimate of the standard error of the sample mean.

4.1	5.3	3.8	4.7	5.0	4.4	5.6	4.9
-----	-----	-----	-----	-----	-----	-----	-----

[5]

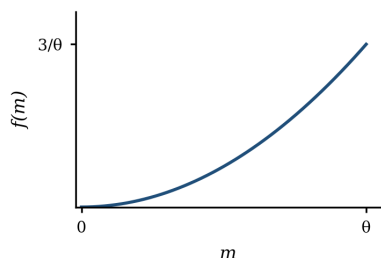
- 19** A sample of 12 observations has a sample variance, calculated with divisor n , of 7.5. Find an unbiased estimate of the population variance and its standard deviation estimate. Show how the estimates change if the sample size is 21 with the same variance (divisor n). [4]

- 20** X_1, X_2 and X_3 are independent Poisson random variables with mean λ . Two estimators of λ are $T_1 = \frac{X_1 + X_2 + X_3}{3}$ and $T_2 = \frac{X_1 + 2X_2}{3}$. Show that both are unbiased and determine which is more efficient. [5]

- 21** A population has mean μ and variance σ^2 . Three independent observations X_1, X_2, X_3 are made. Consider $T_1 = \bar{X}$, $T_2 = \frac{X_1 + 2X_2 - X_3}{2}$ and $T_3 = \frac{2X_1 + X_3}{3}$.
 (a) Show that all three are unbiased for μ .
 (b) Find the variance of each.
 (c) State, with a reason, which should be used. [9]

- 22** $X_1, \dots, X_n$ is a random sample from a population with mean μ and variance σ^2 .
 (a) Show that $E(\bar{X}^2) = \mu^2 + \frac{\sigma^2}{n}$, and hence that $\bar{X}^2$ is a biased estimator of μ^2 .
 (b) Given that σ^2 is known, find an unbiased estimator of μ^2 .
 (c) For $n = 25$ and $\sigma^2 = 100$ a sample has mean 12. Calculate the estimate of μ^2 . [7]
-
- 23** Pairs of independent readings (X_i, Y_i) , $i = 1, \dots, k$, are taken from the same population with mean μ and variance σ^2 .
 (a) Show that $\frac{(X_i - Y_i)^2}{2}$ is an unbiased estimator of σ^2 .
 (b) Hence write down an unbiased estimator of σ^2 based on all k pairs.
 (c) Four pairs of readings are (12.1, 11.4), (9.8, 10.9), (15.2, 14.7), (11.0, 11.9). Calculate the estimate. [7]
-
- 24** To estimate the number N of fish in a lake, 80 fish are caught, marked and returned. Later a random sample of 60 fish is caught and 12 of them are marked.
 (a) Estimate N .
 (b) State two assumptions made.
 (c) Explain why the method fails if $R = 0$ and say how the experiment could be improved. [6]
-
- 25** A random sample X_1, X_2 is taken from a population with mean μ and variance σ^2 .
 Find the bias and mean squared error of the estimator $S = \frac{X_1 + X_2}{2} + 1$, and say, with a reason, whether S is better than $\bar{X}$ when σ^2 is small compared to 1. [4]
-
- 26** X has a Poisson distribution with mean λ , and a single observation X is made.
 (a) Show that $E(X(X - 1)) = \lambda^2$.
 (b) Hence write down an unbiased estimator of λ^2 .
 (c) Explain why X^2 is not unbiased, and find its bias.
 (d) Calculate the unbiased estimate of λ^2 when $X = 5$. [7]
-
- 27** $X_1, \dots, X_n$ is a random sample from a distribution with mean μ and variance σ^2 . Let $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$.
 (a) Show that $\sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$.
 (b) Hence prove that $E(S^2) = \sigma^2$.
 (c) Explain why dividing by n instead of $n - 1$ would give an estimator with a negative bias. [9]
-
- 28** Two independent random samples of sizes n_1 and n_2 are taken from a population with mean μ and variance σ^2 . Their means are $\bar{X}_1$ and $\bar{X}_2$. An estimator of μ is $T = w\bar{X}_1 + (1 - w)\bar{X}_2$.
 (a) Show that T is unbiased for every w .
 (b) Find $\text{Var}(T)$ in terms of w , n_1 , n_2 and σ^2 .
 (c) Find the value of w that minimises $\text{Var}(T)$ and show that the minimum variance estimator is the mean of the combined sample.
 (d) For $n_1 = 10$ and $n_2 = 30$, state the optimal w and the minimum variance. [11]

- 29** X_1, X_2, X_3 is a random sample from the uniform distribution on $[0, \theta]$ and M is the largest of the three observations. The diagram sketches the probability density function of M .
- (a) Show that the probability density function of M is $f(m) = \frac{3m^2}{\theta^3}$ for $0 \leq m \leq \theta$.
- (b) Show that $E(M) = \frac{3\theta}{4}$, so that M is biased, and state the bias.
- (c) Find the constant k such that kM is an unbiased estimator of θ .
- (d) Show that $\text{Var}(M) = \frac{3\theta^2}{80}$ and find $\text{Var}(kM)$. Compare with $\text{Var}(2\bar{X}) = \frac{\theta^2}{9}$.



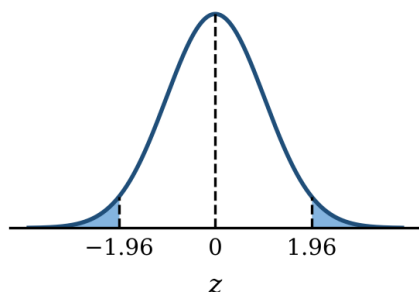
[12]

- 30** $X \sim B(n, p)$ and $\bar{P} = \frac{X}{n}$. A statistician uses $T = \bar{P}(1 - \bar{P})$ to estimate $p(1 - p)$.
- (a) Show that $E(\bar{P}^2) = p^2 + \frac{p(1-p)}{n}$.
- (b) Hence show that $E(T) = \frac{n-1}{n} p(1-p)$.
- (c) Write down an unbiased estimator of $p(1-p)$ in terms of T and n , and evaluate it for $n = 20$ and $X = 6$.

[9]

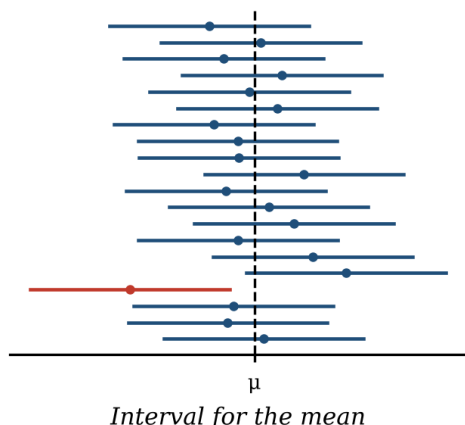
Confidence Intervals

- 1** A population is normally distributed with standard deviation 10. A random sample of 25 observations has mean 40.
Find a 95% confidence interval for the population mean.
- 2** Write down the critical z-values, to 3 decimal places, used for
- (a) a 90% confidence interval
- (b) a 95% confidence interval
- (c) a 99% confidence interval. The diagram shows the central 95% of the standard normal distribution, with 2.5% in each tail.



[3]

- 3 The diagram shows 95% confidence intervals for a population mean μ calculated from 20 independent random samples (the dashed line shows the true value of μ). A 95% confidence interval from one particular sample is (12.3, 14.9) and a student says: "There is a 95% probability that the population mean lies between 12.3 and 14.9."



- (a) Explain what is wrong with the student's statement. [2]
- (b) How many of the 20 intervals in the diagram do not contain μ ? Is this consistent with a 95% confidence level? [1]
-
- 4 A 95% confidence interval for a mean is (23.4, 28.6). Find the sample mean and the margin of error. [2]
-
- 5 State, with a reason, the effect on the width of a confidence interval for a mean of (a) increasing the sample size, (b) increasing the confidence level from 90% to 99%. [2]
-
- 6 A random sample of 36 bags of flour has a mean mass of 1.02 kg. The masses are normally distributed with standard deviation 0.06 kg. Find a 99% confidence interval for the population mean. [4]
-
- 7 A 95% confidence interval for the mean of a normal population with known standard deviation, based on a random sample, is (112.4, 119.8). Find a 90% confidence interval for the population mean, using the same sample. [4]
-
- 8 In a random sample of 200 voters, 56 say that they will vote for Party A. Use a normal approximation to find a 90% confidence interval for the population proportion. [5]
-
- 9 The resistances (ohms) of 8 randomly chosen resistors are shown. The resistance of a resistor is normally distributed with standard deviation 3.2 ohms. Find a 95% confidence interval for the mean resistance.
- | | | | | | | | |
|-----|----|-----|----|-----|----|-----|-----|
| 102 | 98 | 105 | 99 | 101 | 97 | 104 | 100 |
|-----|----|-----|----|-----|----|-----|-----|
- [5]
-
- 10 The standard deviation of a population is 12. Find the smallest sample size needed so that a 95% confidence interval for the mean has a width of at most 6. [4]
-
- 11 A 95% confidence interval for the mean of a normal population based on a random sample of 36 observations is (48.6, 53.4). Find the sample mean and the population standard deviation. [4]
-
- 12 The number of customers entering a shop in each minute is modelled by a Poisson distribution with mean λ . In 6 randomly chosen minutes a total of 54 customers entered. Find an approximate 95% confidence interval for λ using a normal approximation for the sample mean. [5]

- 13** A machine is supposed to fill bottles with a mean of 50 ml. A random sample of 64 bottles has a mean of 49.6 ml. The standard deviation of the volume is known to be 8 ml. Find a 95% confidence interval for the mean and state whether it supports the claim that the mean is 50 ml. [5]
- 14** A random sample of 100 adults has a mean resting pulse of 64.2 beats per minute and an unbiased estimate of the population variance of 18.5. Find a 95% confidence interval for the population mean pulse, stating why it is reasonable to use the normal distribution here. [5]
- 15** The masses of apples are normally distributed with standard deviation 9 g. A random sample of 36 apples is used to construct a confidence interval of the form $(\bar{x} - 2.5, \bar{x} + 2.5)$ for the mean mass. Find the confidence level of this interval to the nearest 0.1%. [4]
- 16** In a survey of n randomly chosen students, a proportion 0.65 say that they would use a new app. A 95% confidence interval for the proportion of all students who would use the app does not contain the value 0.5. Find the least possible value of n . [5]
- 17** A survey company wants a 95% confidence interval for a population proportion with a margin of error of at most 0.03. Show that, whatever the value of the proportion, a sample of size $n = \frac{1.96^2}{4 \times 0.03^2}$ suffices, and find this sample size. [4]
- 18** For a random sample of 50 observations of a variable X , $\sum x = 1830$ and $\sum x^2 = 68\,500$. Find a 99% confidence interval for the population mean. [5]
- 19** A random sample of 10 observations from a normal population has mean 23.4 and unbiased standard deviation estimate $s = 3.1$. Because the population variance is unknown and the sample is small, a t -distribution with 9 degrees of freedom is used. The critical value for a 95% interval is $t = 2.262$. Find a 95% confidence interval for the population mean. [4]
- 20** The number of emails received by an office in a 10-minute interval is modelled by a Poisson distribution with mean λ . Over 5 randomly chosen intervals, a total of 62 emails arrived. Find an approximate 99% confidence interval for λ , and for the mean number of emails per hour. [6]
- 21** The continuous random variable X is uniformly distributed on the interval $[k - 3, k + 5]$, where k is a constant. A random sample of 80 observations of X has mean 41.2. Find a 98% confidence interval for k . [7]
- 22** The daily number of minutes of exercise taken by students is normally distributed with standard deviation 6.4 minutes. A random sample of 64 students has a mean of 47.5 minutes per day.
(a) Find a 95% confidence interval for the mean daily exercise time.
(b) Hence find a 95% confidence interval for the mean weekly exercise time.
(c) A claim says the mean weekly exercise time is 350 minutes. Comment. [7]
- 23** A statistician constructs 95% confidence intervals for a population mean from 100 independent random samples.
(a) State the distribution of the number X of intervals that contain the true mean.
(b) Find the expected number of intervals that contain the true mean.
(c) Find the probability that at least 90 of the intervals contain the true mean.
(d) Find the probability that all 100 intervals contain the true mean, and comment. [7]

- 24** Two independent random samples are taken from normal populations with known standard deviations. Sample A: $n = 40$, mean 78.2, $\sigma_A = 6.0$. Sample B: $n = 50$, mean 74.9, $\sigma_B = 5.0$. Find a 95% confidence interval for $\mu_A - \mu_B$, and say what the interval suggests about whether the two population means differ. [6]

- 25** Cholesterol readings (mg/dL) are normally distributed with standard deviation 18. A random sample of 30 patients has mean 112.
 (a) Find a 95% one-sided confidence interval of the form $\mu < b$, giving b to 1 decimal place.
 (b) Explain why the critical value for this interval is 1.645 rather than 1.96.
 (c) Find the confidence level of the one-sided interval $\mu < 118$ to 3 significant figures. [7]

- 26** The table shows the ages (in years) of 80 randomly chosen members of a gym, grouped in classes. Use the class mid-points to estimate the mean and the variance of the population ages and hence find a 95% confidence interval for the population mean age. State one limitation of this method.

Age	0 to 10	10 to 20	20 to 30	30 to 40	40 to 50
Frequency	6	18	31	17	8

- 27** A random sample of size n is taken from a normal population with known standard deviation σ and sample mean $\bar{X}$.
 (a) Starting from $P(-z < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z) = c$, derive a confidence interval for μ .
 (b) Show that the width of the interval is $\frac{2z\sigma}{\sqrt{n}}$.
 (c) A company can afford a total sampling cost of £1 200 at £6 per observation. If $\sigma = 25$ find the narrowest 95% confidence interval width achievable.
 (d) State how much more it would cost to halve this width. [9]

- 28** In a poll of $n = 400$ people, 140 support a candidate.
 (a) Find a 95% confidence interval for the proportion supporting the candidate.
 (b) Show that for any sample proportion $\bar{p}$ the margin of error at 95% confidence is at most $\frac{1.96}{2\sqrt{n}}$.
 (c) Use this bound to find the smallest n such that the margin of error is at most 2% whatever the sample proportion.
 (d) A newspaper says: "The candidate is clearly behind 40%." Comment using your interval from (a). [10]

- 29** A random sample of 25 observations from a normal population with $\sigma = 4$ has mean 50.8.
 (a) Find a 95% confidence interval for μ .
 (b) A two-tailed test of $H_0: \mu = \mu_0$ at the 5% level uses the same data. Show that H_0 is rejected if and only if μ_0 lies outside the confidence interval in (a).
 (c) Hence state, without further calculation, the conclusion of the test for $\mu_0 = 52.6$ and for $\mu_0 = 49.5$.
 (d) Find the greatest value of μ_0 (to 2 decimal places) for which H_0 is not rejected. [9]

- 30** A pilot study of 15 observations gives an unbiased variance estimate of 219 (so $s \approx 14.8$). A researcher plans a larger study and wants a 99% confidence interval for the mean of width at most 4 units.
- (a) Use the pilot value of s as an estimate of σ to find the minimum sample size.
- (b) The final study gives mean 62.3 and $s = 16.0$ with the sample size from (a). Find the 99% confidence interval and say whether its width meets the target.
- (c) Explain how the pilot study could have misled the sample-size calculation. [9]

Estimation of Population Parameters

- 1** A random sample of 6 values is 7, 9, 4, 8, 6, 10. Find unbiased estimates of the population mean and variance. [3]
- 2** In a random sample of 120 students, 18 are left-handed. Estimate the proportion of left-handed students in the population and the standard error of your estimate. [3]
- 3** For a random sample of 20 observations, $\Sigma x = 340$ and $\Sigma x^2 = 6\,200$. Find unbiased estimates of the population mean and variance. [3]
- 4** A random sample of 150 households in a town has a mean of 2.4 children. The town has 4200 households. Estimate the total number of children in the town. [2]
- 5** A sample of 30 values has an unbiased estimate of the population variance of 52.9. Find the estimate of the population standard deviation and the standard error of the sample mean. [2]
- 6** To estimate the number of deer in a forest, 40 deer are caught, tagged and released. A later random catch of 50 deer contains 8 tagged deer. Estimate the number of deer in the forest. [2]
- 7** The lengths of 40 randomly chosen leaves, to the nearest cm, are summarised in the table. Find unbiased estimates of the population mean and variance.

Length (cm)	20	21	22	23	24
Frequency	3	8	14	9	6

- [4]
- 8** A random sample of 15 observations of x is coded using $y = \frac{x - 200}{2}$. The coded data give $\Sigma y = 18$ and $\Sigma y^2 = 305$. Find unbiased estimates of the population mean and variance of x . [5]
- 9** Two independent random samples are taken from the same population. Sample A has $n = 25$, mean 8.2 and unbiased variance estimate 3.1. Sample B has $n = 35$, mean 9.0 and unbiased variance estimate 2.7. Find the mean of the combined sample and an unbiased estimate of the population variance using the combined data. [6]
- 10** Two surveys estimate the proportion p of people who own a bicycle. Survey A: 30 of 100 own one. Survey B: 120 of 400 own one. Find the standard error of each estimate and say which gives the more precise estimate, with a reason. [4]
- 11** A random variable X takes the values 1, 2 and 3 with probabilities θ , 2θ and $1 - 3\theta$ respectively. Show that $E(X) = 3 - 4\theta$ and hence find an unbiased estimator of θ in terms of the sample mean $\bar{X}$. [4]

- 12** A batch contains 2000 items. A random sample of 80 items is inspected and 7 are defective. Estimate the number of defective items in the batch and give an approximate 95% confidence interval for this number. [5]
-
- 13** The daily sales (in pounds) of a stall on 8 randomly chosen days are shown. Find unbiased estimates of the mean and standard deviation of the daily sales, and the standard error of the mean.
- | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|
| 212 | 198 | 205 | 221 | 190 | 207 | 215 | 202 |
|-----|-----|-----|-----|-----|-----|-----|-----|
- [5]
-
- 14** A school has 600 boys and 400 girls. A random sample of 40 boys has a mean height of 172.4 cm and an independent random sample of 10 girls has a mean height of 160.8 cm.
(a) Find an estimate of the mean height of all students at the school.
(b) Explain why this is preferable to the mean of the 50 sampled students. [5]
-
- 15** A random sample of 10 observations from a Poisson distribution is 3, 5, 2, 4, 6, 3, 2, 5, 4, 3. Both the sample mean and the sample variance are unbiased estimates of λ for a Poisson distribution. Calculate both, and suggest which gives the better estimate, with a reason. [5]
-
- 16** X has probability density function $f(x) = \frac{2x}{\theta^2}$ for $0 \leq x \leq \theta$, and 0 otherwise. A random sample of size n is taken.
Find $E(X)$ and hence show that $T = \frac{3}{2}\bar{X}$ is an unbiased estimator of θ . [5]
-
- 17** A random sample of 60 observations has mean 124.2 and an unbiased variance estimate of 225. Test, at the 5% significance level, the hypothesis $H_0: \mu = 120$ against $H_1: \mu > 120$, using the sample variance as an estimate of σ^2 . State why this is reasonable. [5]
-
- 18** Two independent random samples come from populations with the same variance σ^2 . Sample 1: $n = 12$ and unbiased variance estimate 9.6. Sample 2: $n = 18$ and unbiased variance estimate 7.8. Find the pooled estimate of σ^2 , using $\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$. [4]
-
- 19** A pilot survey suggests that about 30% of people use a certain service. Find the sample size needed so that a 95% confidence interval for the proportion has a margin of error of at most 0.05, using the pilot value. [4]
-
- 20** A random sample of 10 observations is 14, 15, 13, 16, 14, 15, 13, 14, 15, 16. An eleventh observation, 29, is then added.
(a) Find unbiased estimates of the mean and variance before and after adding the 29.
(b) Comment on the effect of the extra observation. [5]
-
- 21** To estimate the size N of a pond's fish population, 100 fish are marked. A first sample of 60 fish contains 8 marked fish. After the fish are mixed and the sample is returned, a second sample of 80 contains 11 marked fish.
(a) Find an estimate of N from each sample.
(b) Combine the two samples to give a single estimate of N.
(c) State two assumptions needed for the estimates to be reliable. [7]
-

- 22** A population of 1000 workers is divided into three strata A, B and C. An independent random sample is taken from each, giving the data in the table (SD means the unbiased estimate of the population standard deviation of the stratum).

(a) Find an estimate of the overall population mean.

(b) The variance of the estimate is estimated by $\Sigma(\frac{N_h}{N})^2 \frac{s_h^2}{n_h}$. Calculate it and give an approximate 95% confidence interval for the population mean.

Stratum	A	B	C
Population size N_h	500	300	200
Sample size	50	30	20
Sample mean	41.2	38.6	45.1
Sample SD	6.0	5.0	7.0

[8]

- 23** X is uniformly distributed on $[a, b]$. For this distribution $E(X) = \frac{a+b}{2}$ and $\text{Var}(X) = \frac{(b-a)^2}{12}$.

A random sample has mean 14.0 and an unbiased variance estimate 8.0.

(a) Use these to estimate a and b.

(b) Explain why your estimate of a and b might be unreasonable if the sample contains a value outside $[a, b]$.

[6]

- 24** The number of misprints on a page is modelled by a Poisson distribution with mean λ . In a random sample of 100 pages, 37 pages have no misprints and the mean number of misprints per page is 1.1.

(a) Use the proportion of error-free pages to estimate λ , using $P(X = 0) = e^{-\lambda}$.

(b) Compare your estimate with the sample mean.

(c) Explain why you might prefer the estimate from the sample mean.

[6]

- 25** $X \sim B(n, p)$ with n known and p unknown.

(a) Show that $E(X(X-1)) = n(n-1)p^2$.

(b) Hence write down an unbiased estimator of p^2 .

(c) For $n = 12$ and $X = 9$ find the unbiased estimate of p^2 and compare it with $(\bar{p})^2$, where $\bar{p} = \frac{X}{n}$.

[7]

- 26** A researcher wants to estimate the mean time spent online by students. She surveys 100 students leaving the library and finds a mean of 5.4 hours per day and an unbiased variance estimate of 4.2.

(a) Find a 95% confidence interval for the mean time.

(b) Give two reasons why the interval may not be a valid estimate of the mean for all students at the university, and suggest an improvement.

[7]

- 27** $X_1, \dots, X_n$ is a random sample from a population with mean μ and variance σ^2 . The estimator $T = c\bar{X}$ is used to estimate μ , where c is a constant.

(a) Find the bias and variance of T.

(b) Show that the mean squared error is $c^2(\frac{\sigma^2}{n} + \mu^2) - 2c\mu^2 + \mu^2$.

(c) Find the value of c that minimises the mean squared error, and show that it is less than 1.

(d) For $\mu = 10$, $\sigma = 15$ and $n = 9$, evaluate this c and the two mean squared errors (for c and for $c = 1$).

[12]

- 28** Independent random samples of sizes n_1 and n_2 come from populations with the same variance σ^2 . The unbiased variance estimates are s_1^2 and s_2^2 . Let $S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$.

(a) Prove that S_p^2 is an unbiased estimator of σ^2 .

(b) If $\text{Var}(s_1^2) = \frac{2\sigma^4}{n_1 - 1}$ and $\text{Var}(s_2^2) = \frac{2\sigma^4}{n_2 - 1}$, show that $\text{Var}(S_p^2) = \frac{2\sigma^4}{n_1 + n_2 - 2}$.

(c) Hence compare the precision of S_p^2 with that of s_1^2 alone when $n_1 = 8$ and $n_2 = 12$. [9]

- 29** A council wants to estimate the proportion p of households that recycle. A random sample of 400 households gives 252 who recycle.

(a) Find an estimate of p and a 95% confidence interval.

(b) The council claims that 60% recycle. Comment using your interval.

(c) The council wants the margin of error to be at most 0.025 at 95% confidence using the estimate from (a) as a guide. Find the sample size required.

(d) State one source of bias in a postal survey of households and explain how it could affect the estimate. [10]

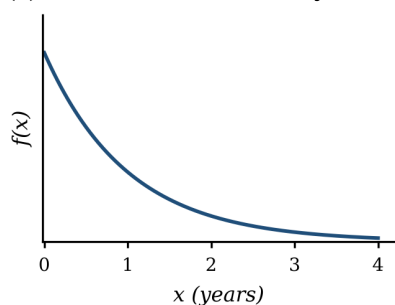
- 30** The lifetime X (years) of a component has probability density function $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$, where $\lambda > 0$ (the diagram shows the shape of f for $\lambda = 1$). You may use $E(X) = \frac{1}{\lambda}$ and $\text{Var}(X) = \frac{1}{\lambda^2}$.

(a) Show that the sample mean $\bar{X}$ is an unbiased estimator of $\frac{1}{\lambda}$ and find its variance.

(b) Eight components last 0.8, 1.9, 0.4, 2.6, 1.1, 0.7, 1.5 and 0.9 years. Estimate $\frac{1}{\lambda}$, and hence estimate λ .

(c) Use a normal approximation for $\bar{X}$ (and your estimate of λ) to find an approximate 95% confidence interval for the mean lifetime $\frac{1}{\lambda}$.

(d) Comment on the validity of using a normal approximation here.



[8]