

Sampling Distributions

- 1 A school has 1200 pupils. The head teacher records the time taken to travel to school by a random sample of 50 pupils. State, with a reason, whether each of the following is a parameter or a statistic.
 - (a) The mean travel time of all 1200 pupils. [1]
 - (b) The mean travel time of the 50 pupils in the sample. [1]
- 2 A population has mean 50 and standard deviation 6. The mean $\bar{X}$ of a random sample of size 9 is calculated. Write down $E(\bar{X})$ and the standard error of $\bar{X}$. [3]
- 3 Explain what is meant by the sampling distribution of a statistic. [2]
- 4 $X \sim N(20, 4^2)$. The mean $\bar{X}$ is found for a random sample of size 4. State the distribution of $\bar{X}$. [2]
- 5 A population has $E(X) = 12$ and $\text{Var}(X) = 9$. For a random sample of size 25, find $E(\bar{X})$ and the standard error of $\bar{X}$. [2]
- 6 Give two reasons why a sample may be used instead of a census. [2]
- 7 A population consists of the values 1, 3 and 5, each equally likely. A random sample of size 2 is taken with replacement and the sample mean $\bar{X}$ is calculated.
 - (a) List all 9 possible samples and their means, and hence write down the sampling distribution of $\bar{X}$. [4]
 - (b) Show that $E(\bar{X})$ equals the population mean. [2]
- 8 A large number of tiles are either 10 cm wide or 20 cm wide, and 40% of the tiles are 10 cm wide. A random sample of three tiles is taken and M is the mode of the three widths. Find the sampling distribution of M . [4]
- 9 In a large population, 30% of people own a particular brand of phone. A random sample of 50 people is taken and $\hat{P}$ is the proportion in the sample who own the phone. Find $E(\hat{P})$, $\text{Var}(\hat{P})$ and the standard error of $\hat{P}$. [4]
- 10 $X \sim N(100, 12^2)$. A random sample of 9 observations is taken. Find $P(\bar{X} > 104)$. [3]
- 11 $X \sim N(5, 2^2)$. T is the sum of a random sample of 16 observations. Find $P(T < 70)$. [4]
- 12 The standard deviation of a population is σ . Explain what happens to the standard error of the sample mean when the sample size is multiplied by 4, and why a larger sample gives a better estimate of the population mean. [3]
- 13 A random sample of size 2 is taken with replacement from the numbers 1, 2, 3, 4 (each equally likely). R is the range of the sample. Find the sampling distribution of R and $E(R)$. [5]
- 14 A random sample of size 2 is taken without replacement from the population 1, 2, 3, 4, 5. Find the possible values of the sample mean $\bar{X}$ with their probabilities, and show that $\text{Var}(\bar{X}) = \frac{3}{4}$. Compare this with the variance if the sampling were with replacement. [7]
- 15 $X \sim U[0, 1]$. A random sample of 3 observations is taken and M is the smallest value. Find $P(M > 0.5)$ and the cumulative distribution function of M . [5]
- 16 In a large population 35% of voters support a policy. A random sample of 40 voters is taken and $\hat{P}$ is the proportion in the sample who support it. Find $P(\hat{P} \geq 0.45)$ exactly, and by using a normal approximation. [6]

- 17** The mean $\bar{X}_A$ of a sample of 10 observations from $N(50, 5^2)$ and the mean $\bar{X}_B$ of an independent sample of 15 observations from $N(48, 6^2)$ are calculated. Find $P(\bar{X}_A > \bar{X}_B)$. [6]
- 18** $X_1, X_2, \dots, X_n$ are independent random variables, each with mean μ and variance σ^2 . Show that the sample mean $\bar{X}$ has $E(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$. [5]
- 19** A population has standard deviation 10. A student states: "For a sample of size 25 the sample mean has variance $\frac{10^2}{25}$... so the standard error is 4." Identify the error, state the correct standard error, and find the sample size required to halve it. [5]
- 20** A large bag contains $2p$ coins and $10p$ coins, with three times as many $2p$ coins as $10p$ coins. A random sample of two coins is taken and their values, in pence, are X_1 and X_2 . The statistic Y is defined by $Y = \frac{X_1 + 3X_2}{4}$. [4]
 (a) Find the sampling distribution of Y . [4]
 (b) Show that Y is an unbiased estimator of the mean value of a coin in the bag, and find $\text{Var}(Y)$. [3]
- 21** A population consists of the numbers 1, 2, 3 and 4, each equally likely. A random sample of size 3 is taken with replacement and M is the sample median. Find the sampling distribution of M and $E(M)$. [7]
- 22** The mass of a type of apple has standard deviation 8 g. The mean mass $\bar{X}$ of a random sample of n apples is used to estimate the population mean μ . Assuming $\bar{X}$ is normally distributed, find the smallest n such that $P(|\bar{X} - \mu| < 1) \geq 0.95$. [6]
- 23** A population consists of the numbers 1, 2, 3, 4 and 5, each equally likely. A random sample of size 2 is taken with replacement and M is the sample maximum. [3]
 (a) Show that $P(M = m) = \frac{2m - 1}{25}$ for $m = 1, 2, 3, 4, 5$. [3]
 (b) Find $E(M)$ and explain why M is a biased estimator of the greatest value in the population. [4]
- 24** A population consists of the values 1, 3 and 5, each equally likely, and has variance $\frac{8}{3}$. A random sample of size 2 is taken with replacement and $S^2 = \frac{1}{n-1} \sum (X - \bar{X})^2$ is the sample variance. List the possible values of S^2 with their probabilities and show that $E(S^2)$ equals the population variance. [6]
- 25** Independent random samples of sizes 10 and 40 are taken from a population with mean μ and variance σ^2 , with sample means $\bar{X}_1$ and $\bar{X}_2$. The combined mean of all 50 observations is C . Show that $C = \frac{\bar{X}_1 + 4\bar{X}_2}{5}$, and find $E(C)$ and $\text{Var}(C)$ in terms of μ and σ^2 . Compare $\text{Var}(C)$ with $\text{Var}(\bar{X}_2)$. [7]
- 26** The random variable X takes the values 0, 2 and 4 with probabilities 0.5, 0.3 and 0.2 respectively. A random sample of size 2 is taken and $\bar{X}$ is the sample mean. [5]
 (a) Find the sampling distribution of $\bar{X}$. [5]
 (b) Show that $\text{Var}(\bar{X}) = \frac{1}{2} \text{Var}(X)$. [3]

- 27** $X_1, \dots, X_n$ are independent observations from $U[0, \theta]$. M is the largest of the n observations.
- (a) Show that the probability density function of M is $g(m) = \frac{nm^{n-1}}{\theta^n}$ for $0 \leq m \leq \theta$. [3]
- (b) Show that $E(M) = \frac{n\theta}{n+1}$. [3]
- (c) Find the least n such that $P(M > 0.95\theta) > 0.9$. [3]
-
- 28** X_1 and X_2 are independent observations from an exponential distribution with parameter λ , and $\bar{X} = \frac{X_1 + X_2}{2}$.
- (a) Show that the sum $S = X_1 + X_2$ has probability density function $f_S(s) = \lambda^2 s e^{-\lambda s}$ for $s > 0$. [4]
- (b) Hence show that $\bar{X}$ has cumulative distribution function $G(x) = 1 - (1 + 2\lambda x)e^{-2\lambda x}$ for $x \geq 0$. [4]
- (c) For $\lambda = 1$, find $P(\bar{X} < 1)$ and compare it with $P(X < 1)$ for a single observation. [3]
-
- 29** A population consists of the six values 2, 4, 4, 6, 8 and 10. A random sample of size 3 is taken without replacement and $\bar{X}$ is the sample mean.
- (a) Find the population mean μ and variance σ^2 . [3]
- (b) There are 20 equally likely samples. Given that the sum of the squares of the 20 possible sample means is $\frac{6040}{9}$, use $\text{Var}(\bar{X}) = E(\bar{X}^2) - \mu^2$ to find $\text{Var}(\bar{X})$, and verify that $\text{Var}(\bar{X}) = \frac{\sigma^2}{n} \times \frac{N-n}{N-1}$. [5]
-
- 30** X_1 and X_2 are independent random variables, each uniformly distributed on $[0, 1]$, and $R = |X_1 - X_2|$ is the sample range.
- (a) By considering the unit square, show that the cumulative distribution function of R is $G(r) = 2r - r^2$ for $0 \leq r \leq 1$. [4]
- (b) Hence find the probability density function of R , and $E(R)$. [4]
- (c) Find the probability that the range of the sample exceeds 0.8, and the median of R . [2]

Central Limit Theorem

- 1** State the central limit theorem. [2]
-
- 2** A population has mean 10 and variance 16. The mean $\bar{X}$ of a random sample of size 64 is calculated. Write down the approximate distribution of $\bar{X}$. [2]
-
- 3** Each observation X has mean 5 and standard deviation 2. T is the sum of 100 independent observations. State the approximate distribution of T . [2]
-
- 4** The distribution of the times to complete a task is skewed. A random sample of 50 times is taken. Explain why the sample mean can still be modelled using a normal distribution. [2]
-
- 5** A population has mean 20 and standard deviation 6. Find the probability that the mean of a random sample of 36 observations exceeds 21. [3]
-
- 6** $X \sim U[0, 6]$. For a random sample of 48 observations, state the approximate distribution of $\bar{X}$. [3]
-
- 7** The lifetime of a component is exponentially distributed with mean 2 years. Use the central limit theorem to find the approximate probability that the mean lifetime of a random sample of 50 components exceeds 2.4 years. [4]

- 8** The mass of a box of tiles has mean 12.5 kg and standard deviation 1.5 kg. Find the approximate probability that 40 randomly chosen boxes have a total mass greater than 510 kg. [4]
- 9** $X \sim \text{Po}(4)$. Use the central limit theorem to estimate $P(\bar{X} < 3.8)$ for a random sample of 60 observations of X . [4]
- 10** A fair six-sided die is rolled 30 times. Use a normal approximation, with a continuity correction, to estimate the probability that the total score is more than 115. [5]
- 11** Each person in a large population independently has probability 0.2 of owning a bicycle. Use the central limit theorem to estimate the probability that the proportion owning a bicycle in a sample of 100 people exceeds 0.27. [4]
- 12** The number of faulty items in a batch of 200 is modelled by $X \sim B(200, 0.35)$. The mean $\bar{X}$ of the numbers of faulty items in 50 independent batches is calculated. Use the central limit theorem to estimate $P(\bar{X} > 71)$. [5]
- 13** A population has standard deviation 3. Find the smallest sample size n for which the probability that the sample mean is within 0.5 of the population mean is at least 0.95. [5]
- 14** A biased spinner lands on the numbers 1, 2, 3, 4 and 5 with probabilities 0.15, 0.25, 0.2, 0.3 and 0.1 respectively. The spinner is spun 60 times and the mean $\bar{X}$ of the numbers is found. Find an estimate of the probability that $\bar{X}$ is greater than 3.2. [6]
- 15** The lifetime of a battery is exponentially distributed with mean 30 hours. A torch uses a new battery as soon as the previous one fails. Use the central limit theorem to estimate the probability that 50 batteries last for more than 1600 hours in total. [5]
- 16** Independent random samples of size 40 and 50 are taken from two populations with means 20 and 19 and standard deviations 5 and 4 respectively, giving sample means $\bar{X}_A$ and $\bar{X}_B$. Estimate $P(\bar{X}_A > \bar{X}_B)$. [6]
- 17** The number of emails received per day is modelled by $X \sim \text{Po}(3)$. Use the central limit theorem, with a continuity correction, to estimate the probability that the total number of emails received over 40 days is at least 130. [5]
- 18** A random sample of 120 observations is taken from a geometric distribution with parameter 0.4. Estimate the probability that the sample mean is greater than 2.8. [5]
- 19** A fair ten-sided die, with faces numbered 1 to 10, is rolled 50 times and $\bar{X}$ is the mean score.
 (a) Use the central limit theorem to find, to 3 significant figures, the value of k such that $P(\bar{X} < k) = 0.05$. [5]
 (b) Explain why the central limit theorem is relevant in part (a). [1]
- 20** The lengths of rods produced by a machine have mean 40 mm and standard deviation 3 mm. A student says: "By the central limit theorem, since there are 100 rods in a sample, the length of each rod is approximately normally distributed." Explain the error, and find the probability that the mean length of the 100 rods exceeds 40.5 mm. [5]
- 21** A fair die is rolled repeatedly until a six has been obtained three times, and X is the total number of rolls. This is done by each of 40 people and the mean $\bar{X}$ of the 40 values of X is calculated. Estimate the probability that $\bar{X}$ is less than 17.5. [6]
- 22** The continuous random variable X has probability density function $f(x) = \frac{3x^2}{8}$ for $0 \leq x \leq 2$, and $f(x) = 0$ otherwise. The mean $\bar{X}$ of a random sample of 50 observations of X is calculated. Estimate $P(\bar{X} > 1.6)$. [7]

- 23** A parcel contains 30 items of type A, each with mass of mean 5 kg and standard deviation 1 kg, and 20 items of type B, each with mass of mean 8 kg and standard deviation 2 kg. All masses are independent. Use the central limit theorem to estimate the probability that the total mass exceeds 330 kg. [6]
- 24** The lifetime of a component is exponentially distributed with mean 2 years. Packs of 50 components are tested and the mean lifetime of each pack is recorded. Estimate the probability that, of 5 independent packs, at least 2 have a mean lifetime exceeding 2.1 years. [7]
- 25** In a large population 30% of people have a certain blood type. A random sample of 150 people is taken. Estimate the probability that the proportion with this blood type exceeds 0.38. Find also the smallest sample size for which the standard error of the sample proportion is less than 0.03. [6]
- 26** The random variable X has an exponential distribution with mean 1. The mean $\bar{X}$ of a random sample of 30 observations is used to estimate probabilities. It is known that $30\bar{X}$ has a gamma distribution with shape parameter 30, and that the exact value of $P(\bar{X} > 1.3)$ is 0.0591.
 (a) Use the central limit theorem to estimate $P(\bar{X} > 1.3)$. [4]
 (b) Find the percentage error of the approximation and suggest why it is not very accurate. [3]
- 27** The mass of an adult passenger has mean 72 kg and standard deviation 14 kg. A lift has maximum load 1000 kg. Using the central limit theorem, find the probability that 12 randomly chosen passengers exceed the maximum load, and find the largest number of passengers for which this probability is at most 0.05. [8]
- 28** X is a random variable with mean 50 and variance 100. A random sample of n observations is taken, and the central limit theorem is used. Show that the probability that the sample mean differs from 50 by more than 2 is approximately $2[1 - \Phi(0.2\sqrt{n})]$, and find the least n for which this probability is less than 0.01. [8]
- 29** A machine is kept running by replacing each component as soon as it fails. The lifetime of each component is exponentially distributed with mean 100 hours, independently. Using the central limit theorem, find the smallest number n of components that must be bought so that the probability that the total lifetime is at least 5000 hours is at least 0.95. [10]
- 30** $X_1, \dots, X_n$ are independent observations from $U[0, 1]$.
 (a) Find the least n such that the probability that $\bar{X}$ lies between 0.45 and 0.55 is at least 0.9, using the central limit theorem. [5]
 (b) For $n = 12$, show that $T = X_1 + \dots + X_{12} - 6$ has mean 0 and variance 1, and explain how this could be used to generate approximately normal random numbers. [4]

Point Estimation

- 1** Explain the difference between an estimator and an estimate. [2]
- 2** A random sample of 5 observations is 5, 7, 9, 4, 10. Find unbiased estimates of the population mean and variance. [3]
- 3** A random sample of 30 observations has $\sum x = 240$. Find an unbiased estimate of the population mean. [2]
- 4** In a random sample of 60 voters, 18 support a new policy. Find an unbiased estimate of the proportion of all voters who support it. [2]

- 5** The numbers of faults found in 5 randomly chosen rolls of cloth are 3, 1, 4, 2 and 5. Assuming the number of faults per roll is Poisson distributed, find an unbiased estimate of the Poisson parameter. [2]
- 6** The mean of a random sample of lifetimes of a component is 4 years. Assuming an exponential distribution, estimate its parameter λ . [2]
- 7** A random sample of 40 observations has $\Sigma x = 1640$ and $\Sigma x^2 = 70000$. Calculate unbiased estimates of the population mean and variance. [4]
- 8** The masses, x grams, of 25 randomly chosen eggs are coded using $y = x - 100$. It is found that $\Sigma y = 75$ and $\Sigma y^2 = 1000$. Find unbiased estimates of the mean and variance of the masses. [5]
- 9** The number of calls received by an office in each of 50 randomly chosen hours is summarised: 0 calls in 12 hours, 1 call in 20 hours, 2 calls in 11 hours, 3 calls in 5 hours and 4 calls in 2 hours. Assuming a Poisson model, estimate the parameter and hence estimate the probability of no calls in an hour. [5]
- 10** A random sample from $U[0, \theta]$ is 3.2, 5.1, 4.4, 2.8 and 4.9. Use the sample mean to estimate θ , and state whether the estimate is consistent with the largest observation. [4]
- 11** A random sample of 36 observations has unbiased variance estimate $s^2 = 12.96$. Find an estimate of the standard error of the sample mean. [3]
- 12** Two random samples from the same population have sizes 20 and 30 and means 15 and 16 respectively. Find an unbiased estimate of the population mean using all 50 observations. [3]
- 13** The number of trials needed to get the first success in a repeated experiment is modelled by a geometric distribution with parameter p . The numbers of trials needed on 5 independent occasions were 3, 5, 2, 4 and 6. Estimate p . [3]
- 14** The random variable X has probability density function $f(x) = (\theta + 1)x^\theta$ for $0 \leq x \leq 1$, where $\theta > 0$. A random sample gives a mean of 0.7. Use the method of moments to estimate θ . [5]
- 15** A random sample from the continuous uniform distribution on $[-\theta, \theta]$ is $-1.8, 2.5, -0.7, 3.1, -2.9, 1.2, 0.6, -2.2$. The population mean is 0 and the variance is $\frac{\theta^2}{3}$. Use the sample variance to estimate θ . [5]
- 16** The numbers of accidents at a junction in 8 randomly chosen weeks were 2, 0, 3, 1, 1, 4, 0, 1. Assume a Poisson model. Write down the likelihood of the data in terms of λ , and hence show that the maximum likelihood estimate of λ is the sample mean. [6]
- 17** The number of flaws per metre of a cable has a Poisson distribution. In a random sample of 40 one-metre lengths the total number of flaws was 62. Estimate the probability that a one-metre length has at least 3 flaws. [5]
- 18** In a survey, 45 of 120 people in town A and 30 of 100 people in town B own a pet. Find unbiased estimates of the proportions in each town, an unbiased estimate of the difference between them, and the estimated standard error of this difference. [6]
- 19** A random sample of 10 observations has mean 12 and unbiased variance estimate $s^2 = 4$. An eleventh observation of 15 is added to the sample. Find the new sample mean and new unbiased variance estimate. [6]

- 20** A student calculates an estimate of the population variance from the sample 6, 8, 11, 7, 13 using $\frac{1}{n}\sum(x - \bar{x})^2$. Explain why this is not the best choice, and find the estimate the student should use. Find also the estimate of the population standard deviation that results. [5]
-
- 21** The ages of 100 randomly chosen customers are grouped: 0 to under 10: 8, 10 to under 20: 22, 20 to under 30: 35, 30 to under 40: 25, 40 to under 50: 10. Using mid-class values, estimate the population mean and variance. [6]
-
- 22** X has an exponential distribution with parameter λ . A random sample $x_1, \dots, x_n$ is taken. Show that the maximum likelihood estimate of λ is $\frac{1}{\bar{x}}$. Five observations are 2.1, 0.8, 3.4, 1.5 and 2.2. Use them to estimate λ and $P(X > 3)$. [8]
-
- 23** The number of trials X until the first success has a geometric distribution with $P(X = x) = p(1 - p)^{x-1}$, $x = 1, 2, \dots$. A random sample of n observations has total $\sum x_i = T$. Show that the maximum likelihood estimate of p is $\frac{n}{T}$. In 40 independent experiments the total number of trials was 130. Estimate p and the probability that an experiment needs more than 4 trials. [8]
-
- 24** A random sample $x_1, \dots, x_n$ is taken from $U[0, \theta]$.
 (a) Write down the likelihood function $L(\theta)$ and explain why it is maximised by taking θ as small as possible, subject to a restriction. [4]
 (b) State the maximum likelihood estimate of θ , and find it for the sample 3.2, 5.1, 4.4, 2.8, 4.9. [3]
 Explain why this estimator tends to underestimate θ .
-
- 25** The continuous random variable X is uniformly distributed on $[a, b]$. A random sample gives $\bar{x} = 5.60$ and $s^2 = 1.282$. Use the method of moments, equating the mean and variance of X to the sample values, to estimate a and b . [6]
-
- 26** Independent random samples of sizes 12 and 18 from the same population have unbiased variance estimates 6.4 and 9.2. Find a pooled unbiased estimate of the population variance, and show that it is better than simply averaging the two estimates by comparing the weights given to each sample. [6]
-
- 27** The random variable X has probability density function $f(x) = \theta x^{\theta-1}$ for $0 < x < 1$, where $\theta > 0$. A random sample $x_1, \dots, x_n$ is taken.
 (a) Show that the maximum likelihood estimator of θ is $\hat{\theta} = -\frac{n}{\sum \ln x_i}$. [5]
 (b) For the sample 0.4, 0.7, 0.9, 0.55, 0.8, 0.65 find the maximum likelihood estimate to 3 significant figures, and compare it with the method of moments estimate, given $E(X) = \frac{\theta}{\theta + 1}$. [5]
-
- 28** A random sample $x_1, \dots, x_n$ is taken from $N(\mu, \sigma^2)$.
 (a) Show that the maximum likelihood estimates are $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$. [7]
 (b) For the sample 4.1, 5.6, 6.2, 3.8, 5.3, 4.9, 6.5, 5.0, find the maximum likelihood estimates and the unbiased estimate of the variance, and explain why they differ. [4]
-

- 29** The random variable X has probability density function $f(x) = \frac{2(\theta - x)}{\theta^2}$ for $0 \leq x \leq \theta$, and $f(x) = 0$ otherwise.
- (a) Show that $E(X) = \frac{\theta}{3}$, and hence write down the method of moments estimator of θ . [4]
- (b) A random sample is 0.5, 0.9, 1.4, 0.3, 2.6. Find the estimate of θ , check it against the data, and estimate $P(X > 1)$. [5]
-
- 30** Events occur in a Poisson process of constant rate λ per hour. The numbers of events observed in three separate periods of length 2, 3 and 1 hours were 6, 11 and 3 respectively.
- (a) Write down the likelihood of these observations and show that the maximum likelihood estimate of λ is the total number of events divided by the total time. [6]
- (b) Hence estimate the probability that no events occur in a 30-minute period, and the expected number of events in a working day of 8 hours. [3]

Unbiased Estimators

- 1** Define what is meant by an unbiased estimator of a population parameter θ . [2]
-
- 2** A population has mean μ . $X_1, X_2, \dots, X_n$ is a random sample. Show that $\bar{X}$ is an unbiased estimator of μ . [2]
-
- 3** X_1 and X_2 are independent observations from a population with mean μ . Show that $T = 3X_1 - 2X_2$ is an unbiased estimator of μ . [2]
-
- 4** X_1 and X_2 are independent observations from a population with mean μ . $U = \frac{X_1 + 2X_2}{2}$. Find $E(U)$, the bias of U as an estimator of μ , and the constant k such that kU is unbiased. [3]
-
- 5** $X \sim B(n, p)$ with n known. Show that $\hat{p} = \frac{X}{n}$ is an unbiased estimator of p . [2]
-
- 6** X_1, X_2 and X_3 are independent observations from a population with mean μ . $T = \frac{X_1 + X_2 + X_3}{4}$. Find the bias of T as an estimator of μ . [2]
-
- 7** X_1 and X_2 are independent random variables, each with mean μ and variance σ^2 . $T_1 = \frac{X_1 + X_2}{2}$ and $T_2 = \frac{X_1 + 2X_2}{3}$. Show that both are unbiased estimators of μ and state, with working, which is the better estimator. [6]
-
- 8** X_1 and X_2 are independent observations from a population with mean μ and variance σ^2 . $T = aX_1 + (1 - a)X_2$. Show that T is an unbiased estimator of μ for every value of a , and find the value of a for which $\text{Var}(T)$ is smallest. [5]
-
- 9** A random sample $X_1, \dots, X_n$ is taken from $U[0, \theta]$. Show that $2\bar{X}$ is an unbiased estimator of θ and find its variance. [4]
-
- 10** A population has mean μ and variance $\sigma^2 = 10$. A sample of size 5 is taken and $T = \frac{1}{5} \sum (X_i - \bar{X})^2$. Given that $E(\sum (X_i - \bar{X})^2) = (n - 1)\sigma^2$, find $E(T)$ and the bias of T as an estimator of σ^2 . [3]

- 11** $X \sim B(10, p)$. An estimator of p is $T = \frac{X+1}{12}$. Find $E(T)$ and the bias of T , and state for which value of p the estimator is unbiased. [5]
- 12** The statistic T is used to estimate a parameter θ and $E(T) = 2\theta + 3$. Write down an unbiased estimator of θ based on T and find its variance in terms of $\text{Var}(T)$. [3]
- 13** X_1, X_2, X_3 is a random sample from $\text{Po}(\lambda)$. Two estimators of λ are $A = \frac{X_1 + X_2 + X_3}{3}$ and $B = \frac{X_1 + 2X_2 + 3X_3}{6}$. Show that both are unbiased and determine which has the smaller variance. [5]
- 14** A random sample $X_1, \dots, X_n$ is taken from a population with mean μ and variance σ^2 . Show that $\bar{X}^2$ is a biased estimator of μ^2 , find its bias, and hence find an unbiased estimator of μ^2 that uses the unbiased sample variance s^2 . [6]
- 15** Independent samples of sizes 10 and 15 are taken from a population with mean μ and variance σ^2 , giving sample means $\bar{X}_1$ and $\bar{X}_2$. $T = a\bar{X}_1 + (1-a)\bar{X}_2$. Show that T is unbiased for any a , and find the value of a that minimises $\text{Var}(T)$. [7]
- 16** $X_1, \dots, X_n$ is a random sample from $U[0, \theta]$ and M is the largest observation, with $E(M) = \frac{n\theta}{n+1}$. Find the bias of M as an estimator of θ , and construct an unbiased estimator. For $n = 5$ and a sample maximum of 8.4, find the corresponding estimate of θ . [5]
- 17** A population has known mean μ and unknown variance σ^2 . Show that $T = \frac{1}{n} \sum (X_i - \mu)^2$ is an unbiased estimator of σ^2 . [4]
- 18** X_1, X_2 and X_3 are independent observations from a population with mean μ and variance σ^2 . $T = aX_1 + bX_2 + cX_3$. Given that T is unbiased and that $a = 0.2$ and $b = 0.3$, find c , and find $\text{Var}(T)$ in terms of σ^2 . [5]
- 19** $X_1, \dots, X_n$ is a random sample from a Poisson distribution with parameter λ . Show that $\bar{X}^2$ is a biased estimator of λ^2 , find its bias, and find an unbiased estimator of λ^2 in terms of $\bar{X}$ and n . [5]
- 20** T_1 and T_2 are independent unbiased estimators of the same parameter θ , with $\text{Var}(T_1) = 4$ and $\text{Var}(T_2) = 9$. The combined estimator is $C = aT_1 + (1-a)T_2$. Show that C is unbiased, find the value of a that minimises $\text{Var}(C)$, and find the minimum variance. [7]
- 21** $X_1, \dots, X_n$ is a random sample from a population with mean μ and variance σ^2 . Show that $E(\sum (X_i - \bar{X})^2) = (n-1)\sigma^2$ and hence that $s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$ is an unbiased estimator of σ^2 . [7]
- 22** $X \sim \text{Po}(\lambda)$. Show that $E(X^2) = \lambda + \lambda^2$, and hence that $T = X^2 - X$ is an unbiased estimator of λ^2 . Find the value of this estimate when the single observation is $x = 5$, and compare it with x^2 . [6]
- 23** $X \sim B(n, p)$ with $E(X) = np$ and $\text{Var}(X) = np(1-p)$. Show that $T = \frac{X(X-1)}{n(n-1)}$ is an unbiased estimator of p^2 . For $n = 20$ and $x = 8$, compare this estimate with $(x/n)^2$. [7]

- 24** $X_1, \dots, X_n$ is a random sample from an exponential distribution with mean θ , so that the probability density function is $f(x) = \frac{1}{\theta}e^{-x/\theta}$, $x \geq 0$. Let m be the smallest of the n observations. Show that nm is an unbiased estimator of θ and compare its variance with that of $\bar{X}$. [8]
-
- 25** $X_1, \dots, X_n$ is a random sample from $U[0, \theta]$ and M is the largest observation, with probability density function $\frac{nm^{n-1}}{\theta^n}$ on $[0, \theta]$. Show that $\text{Var}(M) = \frac{n\theta^2}{(n+1)^2(n+2)}$. Hence find the variance of the unbiased estimator $T = \frac{n+1}{n}M$ and compare it with the variance of $2\bar{X}$ when $n = 10$. [9]
-
- 26** A student says: "Since s^2 is an unbiased estimator of σ^2 , s is an unbiased estimator of σ ." By considering random samples of size 2, taken with replacement, from the population 1, 3, 5 (each value equally likely, $\sigma^2 = \frac{8}{3}$), show that the student is wrong. [7]
-
- 27** $X_1, \dots, X_n$ ($n \geq 2$) is a random sample from an exponential distribution with parameter λ . The sum $S = \sum X_i$ has probability density function $g(s) = \frac{\lambda^n s^{n-1} e^{-\lambda s}}{(n-1)!}$ for $s > 0$.
- (a) Show that $E(\frac{1}{S}) = \frac{\lambda}{n-1}$. [4]
- (b) Hence show that $\frac{1}{\bar{X}}$ is a biased estimator of λ , find its bias, and construct an unbiased estimator of λ . [4]
- (c) A sample of 8 observations has mean 2.5. Find the biased and the unbiased estimates of λ . [2]
-
- 28** $X_1, \dots, X_n$ are independent observations from a population with mean μ and variance σ^2 . $T = \sum a_i X_i$ where the a_i are constants.
- (a) Show that T is an unbiased estimator of μ if and only if $\sum a_i = 1$, and find $\text{Var}(T)$. [3]
- (b) By considering $\sum (a_i - \frac{1}{n})^2$, prove that among all unbiased estimators of this form, $\bar{X}$ has the smallest variance. [5]
-
- 29** $X_1, \dots, X_n$ is a random sample from $U[0, \theta]$, and M is the largest observation. For an estimator T of θ the mean squared error is $\text{MSE}(T) = \text{Var}(T) + [\text{bias}(T)]^2$. Use $\text{Var}(M) = \frac{n\theta^2}{(n+1)^2(n+2)}$ and $E(M) = \frac{n\theta}{n+1}$.
- (a) Show that $\text{MSE}(M) = \frac{2\theta^2}{(n+1)(n+2)}$. [4]
- (b) For $n = 10$ compare, in terms of θ^2 , the mean squared errors of M , of the unbiased estimator $\frac{n+1}{n}M$ (variance $\frac{\theta^2}{n(n+2)}$), and of $2\bar{X}$ (variance $\frac{\theta^2}{3n}$). Comment. [5]
-
- 30** $X_1, \dots, X_n$ is a random sample from $\text{Po}(\lambda)$ and $S = \sum X_i$, so that $S \sim \text{Po}(n\lambda)$. The parameter of interest is $\theta = e^{-\lambda}$, the probability that a single observation equals 0.
- (a) Show that $E(z^S) = e^{n\lambda(z-1)}$ for a constant z , and hence show that $U = (1 - \frac{1}{n})^S$ is an unbiased estimator of θ . [6]
- (b) For $n = 10$ and $S = 13$, find the estimate given by U and the estimate $e^{-\bar{X}}$, and explain which is preferred. [3]

Confidence Intervals

- 1 A 95% confidence interval for a population mean μ is (12.3, 14.7). Find the sample mean and the width of the interval. [2]
- 2 Write down the critical values of the standard normal distribution used to construct 90%, 95% and 99% confidence intervals. [3]
- 3 A random sample of 16 observations from a normal population with standard deviation 4 has mean 50. Find a 95% confidence interval for the population mean. [3]
- 4 State what happens to the width of a confidence interval for a mean (with σ known) when:
 - (a) the confidence level is increased, [1]
 - (b) the sample size is increased. [1]
- 5 A random sample of 36 observations from a population with standard deviation 3 has mean 20. Find a 99% confidence interval for the population mean. [3]
- 6 A confidence interval for a mean is (8.4, 9.6). Find the sample mean and the margin of error. [2]
- 7 Explain what is meant by a 95% confidence interval for a population mean. [3]
- 8 The masses, in grams, of a random sample of 8 apples from a large orchard are 142, 155, 148, 160, 151, 139, 146 and 157. The standard deviation of the masses of all the apples is known to be 8 g. Find a 98% confidence interval for the mean mass of the apples. [4]
- 9 A 95% confidence interval for the mean of a normal population, based on a random sample of 25 observations, is (34.2, 37.8). Find the population standard deviation, assuming it is known. [4]
- 10 The standard deviation of a population is 4. Find the smallest sample size needed for a 95% confidence interval for the mean to have a margin of error of at most 0.5. [3]
- 11 A random sample of 12 observations from $N(\mu, 3^2)$ has mean 48.5. Find a 98% confidence interval for μ . [3]
- 12 A random sample of 100 observations from a population with standard deviation 10 has mean 50. A confidence interval for the population mean is (48.04, 51.96). Find the confidence level of this interval. [4]
- 13 A random sample of 100 observations has mean 65 and unbiased variance estimate $s^2 = 144$. Since the sample is large, use s as an estimate of the population standard deviation to find a 95% confidence interval for the population mean. [3]
- 14 A student states that a 95% confidence interval for a mean is (4.2, 5.8) and concludes: "There is a 95% probability that the population mean lies between 4.2 and 5.8." Explain why this statement is not correct, and give a correct interpretation. [3]
- 15 A 95% confidence interval for μ , based on a random sample from a normal population with known variance, is (10.4, 13.6). Without knowing the sample size or the standard deviation, find the 99% confidence interval for μ from the same sample. [5]
- 16 A manufacturer claims that the mean lifetime of a bulb is 1200 hours. A random sample of 36 bulbs, from a population with standard deviation 90 hours, has a mean lifetime of 1165 hours. Construct a 95% confidence interval for the mean and comment on the claim. [5]

- 17** The standard deviation of a population is 3. What is the smallest sample size for which the width of a 99% confidence interval for the mean is at most 1? [4]
- 18** Independent random samples of sizes 40 and 50 are taken from two populations with standard deviations 5 and 4 and give means 20.4 and 18.9 respectively. Find a 95% confidence interval for the difference between the population means, and state whether it suggests the means differ. [6]
- 19** A random sample of 60 observations has $\Sigma x = 372$ and $\Sigma x^2 = 2500$. Use the sample to calculate a 95% confidence interval for the population mean, using the unbiased variance estimate in place of the population variance. [6]
- 20** The random variable $\bar{X}$ is the mean of a random sample of size n from $N(\mu, \sigma^2)$. Starting from $P(-1.96 < Z < 1.96) = 0.95$, where $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, derive the 95% confidence interval for μ . [5]
- 21** The time between arrivals at a clinic is exponentially distributed. In a random sample of 100 inter-arrival times the mean was 2.9 minutes. Use the central limit theorem, and the fact that for an exponential distribution the standard deviation equals the mean, to construct an approximate 95% confidence interval for the mean time between arrivals. [5]
- 22** Two random samples are taken from a population with standard deviation 6. Both have mean 52, and have sizes 20 and 80. Find a 95% confidence interval for the population mean from each, and compare their widths. [6]
- 23** A 95% confidence interval for a population mean, based on a random sample of size 50 from a population with known standard deviation, has width 1.2.
 (a) Find the sample size needed to give a 95% confidence interval of width 0.6. [3]
 (b) With the original sample size of 50, find the confidence level that would give an interval of width 0.6. [4]
- 24** The standard deviation of the masses of packets is known to be 8 g. A 90% confidence interval for the mean mass, based on a random sample of size n , has width 3.29 g. Find n , and hence find the 99% confidence interval if the sample mean is 40 g. [6]
- 25** Two independent random samples of sizes 30 and 50 from the same population (standard deviation 5) have means 12.4 and 13.1. Find a 95% confidence interval for the population mean using both samples combined, and compare its width with the interval from the sample of size 50 alone. [6]
- 26** A random sample of 25 observations from $N(\mu, 9)$ is used to construct a 95% confidence interval for μ . The mean of the sample is 31.2. A second analyst uses the same data but incorrectly uses 9 as the standard deviation (instead of the variance). Find both intervals and explain the discrepancy. [6]
- 27** Ten independent random samples are taken from a normal population and a 95% confidence interval for the population mean is calculated from each. Find the probability that at least 9 of the 10 intervals contain the population mean, giving your answer to 3 significant figures. Find also the expected number of intervals that do not contain the mean, and the probability that all 10 intervals contain the mean. [8]

- 28** The true standard deviation of a population is 8, but a researcher uses $\sigma = 6$ to calculate 95% confidence intervals for the mean from samples of size n . Show that the probability that such an interval contains the population mean is $2\Phi(1.47) - 1$, whatever the value of n , and evaluate it. Find also the probability that the interval contains the mean if the researcher had used $\sigma = 6$ when the true σ is 10. Comment. [8]
-
- 29** A one-sided 95% confidence bound for a mean μ is $U = \bar{x} + 1.645 \frac{\sigma}{\sqrt{n}}$, so that $P(\mu < U) = 0.95$. A random sample of 36 observations from a population with standard deviation 4 has mean 47.
 (a) Find the upper confidence bound and explain why 1.645 is used instead of 1.96. [3]
 (b) Find the smallest sample size for which the upper bound is below 50 when the sample mean is 47. [4]
 (c) Find the smallest sample size for which the upper bound is below 49. [3]
-
- 30** Random samples of size n are taken from $N(\mu, \sigma^2)$ with σ known, and 95% confidence intervals for μ are constructed.
 (a) Show that the width of the interval does not depend on the sample values, and that the probability that the interval contains μ is 0.95. [4]
 (b) Two independent samples give two such intervals. Find the probability that both contain μ , and the probability that exactly one does. [3]
 (c) Find the smallest number of independent intervals for which the probability that at least one fails to contain μ exceeds 0.5. [3]
-

Estimation of Means and Proportions

- 1** Write down the critical t-value used for a two-sided 95% confidence interval when the sample size is:
 (a) 10, [1]
 (b) 16, [1]
 (c) 25. [1]
-
- 2** In a random sample of 200 customers, 60 prefer brand A. Find an approximate 95% confidence interval for the proportion of all customers who prefer brand A. [3]
-
- 3** State the conditions under which a confidence interval for a population mean should be calculated using the t-distribution rather than the normal distribution. [2]
-
- 4** A random sample of 10 observations from a normal population has mean 25 and unbiased standard deviation estimate $s = 4$. Find a 95% confidence interval for the population mean. [3]
-
- 5** In a random sample of 100 people, 40 own a smartphone. Estimate the proportion who own one and the standard error of your estimate. [3]
-
- 6** A random sample of 80 observations gives a sample proportion of 0.35. Find an approximate 90% confidence interval for the population proportion. [3]
-
- 7** A random sample of 8 measurements, assumed to come from a normal population, is 12.1, 13.4, 11.8, 12.9, 13.1, 12.5, 11.9, 12.8. Find a 95% confidence interval for the population mean. [5]
-
- 8** The masses, x grams, of 12 randomly chosen items from a normal population are coded using $y = x - 50$. It is found that $\sum y = 30$ and $\sum y^2 = 210$. Find a 90% confidence interval for the population mean mass. [5]
-

- 9** In an opinion poll of 500 randomly chosen voters, 215 said they would vote for candidate A. Find an approximate 95% confidence interval for the proportion of all voters supporting A, and comment on whether A is likely to have a majority. [5]
- 10** A pollster wants a 95% confidence interval for a proportion with a margin of error of at most 0.03, whatever the true proportion. Find the smallest sample size required. [4]
- 11** In a survey 52% of 150 men and 45% of 200 women said they exercise weekly. Find an approximate 95% confidence interval for the difference between the proportions of men and women who exercise weekly. [5]
- 12** A random sample of 45 waiting times has mean 12.2 minutes and unbiased standard deviation estimate 3.1 minutes. A 95% confidence interval for the mean waiting time is required.
 (a) Find the interval using the t-distribution. [2]
 (b) Find the interval using the normal distribution. [2]
 (c) Comment on the difference between the two intervals. [1]
- 13** A 95% confidence interval for the mean of a normal population, based on a random sample of 25 observations and the t-distribution, is (18.4, 21.6). Find the sample mean and the unbiased standard deviation estimate s . [4]
- 14** A machine is set to fill packets with a mean of 100 g. A random sample of 20 packets, from a normal population, has mean 104.5 g and unbiased standard deviation estimate 12 g. Construct a 95% confidence interval for the mean fill, and state whether there is evidence that the machine is incorrectly set. [5]
- 15** Independent random samples of sizes 60 and 75 have means 24.6 and 22.9 and unbiased variance estimates 36 and 49 respectively. Find an approximate 95% confidence interval for the difference between the two population means. [5]
- 16** A secretary asks n members of a club whether they approve of a proposal, and 55% of those asked approve. Assume that the proportion stays at 0.55 however large n is. A 95% confidence interval for the proportion of all members who approve is found not to contain 0.70. Find the least possible value of n . [5]
- 17** A random number device is claimed to produce the digit 1 with probability 0.5. In 400 trials it produced it 232 times. Find a 99% confidence interval for the true probability and comment on the claim. [5]
- 18** A random sample of n trials gives a sample proportion $\hat{p}$ of successes, where the number of successes is $X \sim B(n, p)$. Explain why $\hat{p}$ is approximately $N(p, \frac{p(1-p)}{n})$ for large n , and hence derive the approximate 95% confidence interval $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. [5]
- 19** A 95% confidence interval for a population proportion, calculated using a normal approximation, is (0.36, 0.44). Find the sample proportion, and find the size of the sample that was used. [5]
- 20** A random sample of 10 observations from a normal population has mean 50 and unbiased standard deviation estimate 6. A confidence interval for the population mean, calculated using the t-distribution, is (46, 54). Show that the critical value used was $t = 2.108$ (to 3 decimal places), and use the fact that $t_9 = 1.833$ for 90% and $t_9 = 2.262$ for 95% to say between which two confidence levels this interval lies. The calculator value of the confidence level is 93.6%. [5]

- 21** Five students sat a test before and after a revision course. Their scores (before, after) were (82, 86), (75, 79), (90, 91), (68, 74) and (77, 80). Assuming the differences are normally distributed, find a 95% confidence interval for the mean improvement. [6]
-
- 22** A random sample of 12 observations from a normal population has $\Sigma x = 150.6$ and $\Sigma x^2 = 1905.4$. Find a 95% confidence interval for the population mean. [6]
-
- 23** A random sample of 6 observations from a normal population is 8.1, 9.4, 7.6, 10.2, 8.8, 9.0. Find a 95% confidence interval for the mean using the unbiased standard deviation estimate. Find also the interval that would result from wrongly using the normal critical value 1.96 and comment. [6]
-
- 24** A 95% confidence interval for p , the proportion of defective items, is based on a sample of 40 items of which 2 are defective. The interval is calculated as $0.05 \pm 1.96\sqrt{\frac{0.05 \times 0.95}{40}}$.
- (a) Evaluate the interval and explain why it is unsatisfactory. [4]
- (b) State what condition on the sample would make the normal-approximation interval reliable, and find the smallest sample size for which $n\hat{p} \geq 10$ if $\hat{p} = 0.05$. [2]
-
- 25** Independent random samples from two normal populations with a common variance have sizes 8 and 10, means 34.2 and 31.5, and unbiased standard deviation estimates 2.1 and 2.6. Find a 95% confidence interval for the difference between the population means, using a pooled estimate of the common variance. [7]
-
- 26** In a trial, 46% of 250 patients given a new treatment recovered, compared with 38% of 300 patients given the standard treatment. Find a 95% confidence interval for the difference in recovery proportions, and state what the interval suggests about the new treatment. [6]
-
- 27** A random sample of n observations from a normal population is used to construct a 95% confidence interval for the mean using the t -distribution. Find the smallest n for which the margin of error is at most half of the sample standard deviation s . Use $t_{16} = 2.12$, $t_{17} = 2.11$ and $t_{18} = 2.101$. [6]
-
- 28** A pilot random sample of 15 observations from a normal population has unbiased standard deviation estimate $s = 6.2$ and mean 40.8.
- (a) Find a 95% confidence interval for the population mean from the pilot sample. [4]
- (b) Using $s = 6.2$ as an estimate of σ , find the sample size needed for a 95% confidence interval for the mean with a margin of error of at most 1. [3]
- (c) Explain why a larger sample than the answer to (b) might be advisable. [2]
-
- 29** Three independent opinion polls of the same population, of sizes 400, 600 and 1000, found that 46%, 52% and 49% respectively support a proposal.
- (a) Find a 95% confidence interval for the population proportion using the first poll alone. [3]
- (b) Combine the polls into a single sample and find a 95% confidence interval for the population proportion. [5]
- (c) Compare the widths of the two intervals and state what the combined interval suggests about majority support. [2]
-

- 30** Six plants were measured before and after treatment with a fertiliser. The heights in cm were: before 52.1, 48.3, 55.0, 50.2, 46.8, 53.5 and after 53.4, 49.9, 56.1, 51.0, 48.5, 54.2, respectively. Assume that the differences are normally distributed.
- (a) Find a 95% confidence interval for the mean increase in height using the paired differences. [5]
- (b) Ignoring the pairing, treat the before and after heights as independent samples with a common variance. Their unbiased standard deviation estimates are 3.131 and 2.866. Find the 95% confidence interval for the difference in means using a pooled variance. [5]
- (c) Compare the two intervals and explain why the paired analysis is more appropriate here. [2]

Estimation of Variance

- 1** A random sample of size n is taken from a normal population with variance σ^2 , and S^2 is the unbiased sample variance. State the distribution of $\frac{(n-1)S^2}{\sigma^2}$. [2]
- 2** The random variable W has a chi-squared distribution with 5 degrees of freedom. Write down $E(W)$ and $\text{Var}(W)$. [2]
- 3** A random sample of 10 observations is taken from a normal population. Write down the critical values of the chi-squared distribution needed to find a 95% confidence interval for σ^2 . [3]
- 4** A random sample of 9 observations has unbiased variance estimate $s^2 = 12.5$. Find the value of $(n-1)s^2$. [2]
- 5** A random sample of 6 observations has $\sum x = 42$ and $\sum x^2 = 312$. Find the unbiased estimate of the population variance. [3]
- 6** State the assumption that must be made about the population when a chi-squared confidence interval is constructed for a variance. [1]
- 7** A random sample of 10 observations from a normal population has unbiased variance estimate $s^2 = 8.4$. Find a 95% confidence interval for σ^2 . [4]
- 8** A 95% confidence interval for the variance of a normal population is (3.97, 28.0). Write down the corresponding 95% confidence interval for the standard deviation, to 3 significant figures. [2]
- 9** A random sample of 16 observations from a normal population has $s^2 = 25$. Find a 90% confidence interval for σ^2 . [4]
- 10** A random sample of 11 observations is taken from $N(\mu, 4)$. Find the probability that the unbiased sample variance S^2 exceeds 6.4. [4]
- 11** A random sample of 8 observations from a normal population is 14.2, 15.8, 13.9, 16.4, 15.1, 14.7, 16.0, 15.3. Find a 95% confidence interval for the population variance. [6]
- 12** A random sample of 21 observations from a normal population has $s^2 = 7.5$. A manufacturer claims that $\sigma^2 = 9$. Find a 95% confidence interval for σ^2 and state whether it supports the claim. [5]
- 13** A random sample of 8 observations from a normal population has standard deviation $s = 2.5$. Find a 95% confidence interval for the population standard deviation. [4]

- 14** A random sample of 12 observations from a normal population has $s^2 = 6$. Find a one-sided 95% confidence interval of the form $(0, U)$ for σ^2 , given that the lower 5% point of χ^2_{11} is 4.575. [4]
- 15** A random sample of 21 observations is taken from a normal population with variance σ^2 . Find the probability that the sample variance s^2 lies between $0.6\sigma^2$ and $1.5\sigma^2$. [5]
- 16** A machine is required to produce rods with $\sigma \leq 0.5$ mm. A random sample of 20 rods from a normal population has $s = 0.42$ mm. Find a 95% confidence interval for σ and comment on whether the requirement is met. [5]
- 17** A random sample of 18 observations from a normal population has standard deviation $s = 5.2$. Find a 90% confidence interval for σ , and state whether the value $\sigma = 4$ is plausible. [6]
- 18** A random sample of size n is taken from a normal population with variance σ^2 . Given that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, show how a 95% confidence interval for σ^2 is derived, explaining why the critical values appear in the order shown in the interval. [5]
- 19** A 95% confidence interval for the variance of a normal population, based on a random sample of 15 observations, has lower limit 5.360. Given that the upper 2.5% point of χ^2_{14} is 26.119 and the lower 2.5% point is 5.629, find s^2 and the upper limit of the interval. [5]
- 20** A student finds a 95% confidence interval for σ^2 from a sample of 10 observations with $s^2 = 8.4$, and writes: "Interval = $(\frac{9 \times 8.4}{2.700}, \frac{9 \times 8.4}{19.023}) = (28.0, 3.97)$." Explain the error, and find the correct interval. Explain also why the sample variance 8.4 does not lie at the centre of the correct interval. [5]
- 21** The thicknesses, in mm, of a random sample of 6 washers from a normal population are 5.1, 5.8, 4.9, 5.5, 6.2 and 5.3. Find a 90% confidence interval for the standard deviation of the thicknesses of all the washers. [6]
- 22** A random sample of 15 observations from a normal population has $\Sigma x = 182.4$ and $\Sigma x^2 = 2262.7$. Find a 95% confidence interval for σ^2 . [6]
- 23** Two random samples from normal populations both have $s^2 = 16$, with sizes 10 and 40. Find a 95% confidence interval for σ^2 in each case and comment on the widths. [6]
- 24** A random sample of 26 observations is taken from a normal population with variance $\sigma^2 = 9$. Using the fact that $\text{Var}(\chi^2_\nu) = 2\nu$, show that $\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$, and find the standard deviation of S^2 . Use a normal approximation to estimate $P(S^2 > 12)$. [7]
- 25** Independent random samples of sizes 10 and 15 from normal populations with the same variance σ^2 have unbiased variance estimates 4.2 and 5.1. Find the pooled estimate of σ^2 and a 95% confidence interval for σ^2 based on 23 degrees of freedom. [7]
- 26** A random sample of 10 observations is taken from a normal population with known mean 50. The observations are 48.2, 51.5, 50.3, 47.9, 52.1, 49.4, 50.8, 49.1, 51.2, 49.9. Given that $\frac{\Sigma(x - \mu)^2}{\sigma^2} \sim \chi^2_{10}$ when μ is known, find a 95% confidence interval for σ^2 . Explain why 10 degrees of freedom are used rather than 9. [6]

27 A 95% confidence interval for σ^2 is constructed from a random sample of 10 observations from a normal population, as $(\frac{9s^2}{19.023}, \frac{9s^2}{2.700})$. Show that the interval contains σ^2 with probability 0.95, and find the probability that the upper limit of the interval is less than $2\sigma^2$. [7]

28 $X_1, \dots, X_n$ is a random sample from a normal population with variance σ^2 . Consider estimators of the form $T_k = \frac{1}{k} \sum (X_i - \bar{X})^2$. It is known that $\sum (X_i - \bar{X})^2 / \sigma^2 \sim \chi^2_{n-1}$. The mean squared error is $MSE = \text{Var} + \text{bias}^2$.

(a) Show that $MSE(T_k) = \frac{\sigma^4[2(n-1) + (n-1-k)^2]}{k^2}$. [5]

(b) For $n = 10$, evaluate $MSE(T_k)/\sigma^4$ for $k = 9, 10$ and 11 , and comment. [4]

29 For large ν , $\sqrt{(2\chi^2_\nu)}$ is approximately normal with mean $\sqrt{(2\nu - 1)}$ and standard deviation 1. A random sample of 101 observations from a normal population has $s^2 = 12$.

(a) Use this approximation to estimate the 2.5% upper and lower points of χ^2_{100} . [4]

(b) Hence find an approximate 95% confidence interval for σ^2 , and for σ . [4]

(c) The exact points are 129.561 and 74.222. Comment on the accuracy of the approximation. [2]

30 A process is acceptable if $\sigma \leq 0.5$. A random sample of n observations from a normal population gives $s = 0.42$. The one-sided 95% upper confidence limit for σ^2 is $\frac{(n-1)s^2}{c}$, where c is the lower 5% point of χ^2_{n-1} . The values of c are 10.117 for $n = 20$, 25.695 for $n = 40$ and 42.339 for $n = 60$. For which of these sample sizes is the upper confidence limit for σ below 0.5? [8]