

Hypothesis Testing Principles

- 1 A coin is suspected of being biased towards heads. The probability of heads is p . A test is to be carried out.
Write down suitable null and alternative hypotheses for this test. [2]
- 2 For each claim, state whether a one-tailed or a two-tailed test is appropriate, giving a reason.
 - (a) A manufacturer claims that a new battery lasts longer than the old type. [1]
 - (b) A technician suspects that the mean thickness of a component has changed from 4.5 mm. [1]
 - (c) A council believes that the proportion of households recycling has fallen below 0.6. [1]
- 3 The random variable $X \sim B(8, p)$. Jo tests $H_0: p = 0.5$ against $H_1: p > 0.5$ at the 5% significance level. She observes $x = 7$.
Find $P(X \geq 7)$ and state Jo's conclusion. [3]
- 4 In a hypothesis test, explain what is meant by
 - (a) the critical region, [1]
 - (b) the significance level. [1]
- 5 A random sample of 16 observations is taken from $N(\mu, 4^2)$ and has mean 52.5. A test of $H_0: \mu = 50$ against $H_1: \mu > 50$ is carried out.
Calculate the test statistic and, at the 5% level, state the conclusion. [3]
- 6 A hypothesis test gives a p-value of 0.032. State the conclusion about H_0 at the
 - (a) 5% significance level, [1]
 - (b) 1% significance level. [1]
- 7 The random variable $X \sim B(20, p)$. A two-tailed test of $H_0: p = 0.3$ against $H_1: p \neq 0.3$ is carried out at the 5% significance level, with 2.5% in each tail.
 - (a) Find the critical region for X . [5]
 - (b) State the actual significance level of the test. [2]
- 8 A seed supplier claims that 20% of its seeds produce red flowers. A gardener suspects the true proportion is higher. She plants 15 seeds and 7 produce red flowers.
Test, at the 5% significance level, the gardener's suspicion. State your hypotheses clearly. [5]
- 9 A coin is tossed 100 times and lands heads 62 times. Using a normal approximation with a continuity correction, test at the 5% significance level whether the coin is biased. State your hypotheses. [5]
- 10 A machine is supposed to fill bags with a mean mass of 500 g. The masses are normally distributed with standard deviation 6 g. A sample of 25 bags has mean mass 497.2 g.
Test, at the 5% significance level, whether the mean mass has changed. [5]
- 11 A pharmaceutical company claims that 40% of patients recover within a week using its drug. A doctor tests $H_0: p = 0.4$ against $H_1: p < 0.4$ using a sample of 25 patients and obtains a p-value of 0.036.
 - (a) Write down the conclusion at the 5% level, in context. [2]
 - (b) Explain why the conclusion differs at the 1% level. [2]
- 12 After carrying out a hypothesis test, a student writes: 'The p-value is 0.03, so there is a 3% probability that the null hypothesis is true.'
 - (a) Explain why this statement is incorrect. [2]
 - (b) State the correct meaning of a p-value of 0.03. [2]

- 13** $X \sim N(\mu, 9^2)$. A random sample of 36 observations is used to test $H_0: \mu = 120$ against $H_1: \mu < 120$ at the 5% significance level.
Find the critical region for the sample mean $\bar{X}$. [5]
- 14** A call centre believes that 15% of its calls are complaints. After a change of policy a manager takes a random sample of 40 calls and carries out a two-tailed test of $H_0: p = 0.15$ at the 10% significance level, with 5% in each tail.
(a) Find the critical region. [5]
(b) State the actual significance level. [2]
(c) In the sample, 11 calls are complaints. State the conclusion in context. [2]
- 15** A school claims that 70% of its students achieve a pass grade. An inspector believes the true proportion is lower. In a random sample of 18 students, 9 achieved a pass grade.
Carry out a test at the 5% significance level, stating your hypotheses and your conclusion in context. [6]
- 16** The masses of eggs from a farm are normally distributed with standard deviation 3.2 g. The farmer claims that the mean mass is 60 g. A sample of 40 eggs has a total mass of 2376 g.
Test the farmer's claim at the 5% significance level, using a two-tailed test. [6]
- 17** A candidate claims that 40% of voters support her. In a random sample of 250 voters, 92 support her. A journalist believes that the true proportion is lower than 40%.
Use a normal approximation, with continuity correction, to test the journalist's belief at the 5% significance level. [6]
- 18** A coin is tossed n times, and a test of $H_0: p = 0.5$ against $H_1: p > 0.5$ is carried out at the 5% significance level, where p is the probability of heads.
Show that if the coin lands heads on every one of the first n tosses then H_0 can be rejected if $n \geq 5$, but not if $n = 4$. [5]
- 19** The random variable $X \sim B(30, 0.4)$. Observing $x = 6$, a student tests $H_0: p = 0.4$ against $H_1: p \neq 0.4$ at the 5% significance level.
Find the p -value for this two-tailed test and state the conclusion. [5]
- 20** A test of $H_0: \mu = 60$ against $H_1: \mu > 60$ is based on the mean $\bar{X}$ of a random sample of size 25 from $N(\mu, 5^2)$. H_0 is rejected when $\bar{X} > c$.
Given that the significance level of the test is 1%, find c . [4]
- 21** Historically 10% of the components made by a machine are defective. After servicing, an engineer suspects the machine now produces more defective components. She tests a random sample of 50 components at the 5% significance level.
(a) Find the critical region for the number of defective components in the sample. [5]
(b) Find the actual significance level and explain why it is less than 5%. [2]
(c) The sample contains 9 defective components. State the conclusion, and explain why a conclusion of 'the machine is now working as before' is not justified. [2]
- 22** A test of $H_0: p = 0.1$ against $H_1: p > 0.1$ is carried out on a random sample of 50 items from a production line, using the critical region $X \geq 10$ where $X \sim B(50, p)$.
Each week a different independent sample is tested, and H_0 is true throughout.
(a) Write down the probability that H_0 is wrongly rejected in a given week. [1]
(b) Find the probability that H_0 is wrongly rejected at least once in 5 weeks. [3]
(c) Find the least number of weeks for which this probability exceeds 0.5. [3]

- 23** A student's attempt at a hypothesis test is shown below.
A company says that 30% of customers prefer brand X. In a sample of 20 customers, 3 prefer brand X. Test at the 5% level whether fewer customers prefer brand X.
 $H_0: p = 0.3$, $H_1: p = 0.3$ is false. $X \sim B(20, 0.3)$. $P(X = 3) = 0.0716$. Since $0.0716 > 0.05$, accept H_0 : the company is correct.
Identify three errors in the student's work and give the correct solution. [7]
-
- 24** The time, T minutes, taken by a train to complete a journey is stated to have a mean of 60. A random sample of 64 journeys gives $\Sigma T = 3984$ and $\Sigma T^2 = 249\,000$.
Use the sample to estimate the variance of T , and hence test at the 5% significance level whether the mean journey time is different from 60 minutes. State why a normal approximation is justified. [8]
-
- 25** The random variable $X \sim B(200, p)$. A two-tailed test of $H_0: p = 0.35$ against $H_1: p \neq 0.35$ is carried out at the 5% significance level.
(a) Use a normal approximation with continuity correction to find the critical region. [7]
(b) Use the binomial distribution to find the actual significance level of this critical region. [3]
-
- 26** A two-stage test of $H_0: p = 0.5$ against $H_1: p > 0.5$ uses $X_1 \sim B(10, p)$ from a first sample. If $X_1 \geq 8$, H_0 is rejected; if $X_1 \leq 4$, H_0 is not rejected; otherwise a second independent sample $X_2 \sim B(10, p)$ is taken, and H_0 is rejected if $X_1 + X_2 \geq 13$.
Find the overall significance level of the test. [8]
-
- 27** A die is thought to be biased so that sixes occur less often than for a fair die. Let p be the probability of a six, and let N be the number of rolls needed to obtain the first six. A test of $H_0: p = \frac{1}{6}$ against $H_1: p < \frac{1}{6}$ is to be carried out at a significance level of at most 5%, rejecting H_0 if $N \geq k$.
(a) Show that $P(N \geq k) = \left(\frac{5}{6}\right)^{k-1}$ under H_0 . [2]
(b) Find the least k for which the significance level does not exceed 5%, and state the actual significance level. [5]
(c) The first six occurs on the 20th roll. State the conclusion. [2]
-
- 28** A random sample of size n is taken from $N(\mu, 15^2)$ and the sample mean is found to be 104. A test is carried out of $H_0: \mu = 100$.
(a) For the alternative hypothesis $H_1: \mu > 100$, show that H_0 is rejected at the 5% significance level if $n \geq 39$ and find the smallest such n . [5]
(b) Find the least n at which H_0 is rejected when $H_1: \mu \neq 100$ is used instead, and explain why it differs. [4]
-
- 29** A coin is tossed n times to test $H_0: p = 0.5$ against $H_1: p \neq 0.5$, where p is the probability of heads, at the 5% significance level with at most 2.5% in each tail.
(a) Show that when $n = 5$ there is no outcome that leads to rejection of H_0 . [3]
(b) Find the smallest n for which a critical region exists, and write it down. [4]
(c) For $n = 6$, find the actual significance level and explain why it is much smaller than 5%. [3]
-
- 30** The lifetime of a type of bulb is normally distributed with standard deviation 12 hours. A company tests $H_0: \mu = 80$ hours against $H_1: \mu < 80$ hours using the mean $\bar{X}$ of a random sample of n bulbs, rejecting H_0 if $\bar{X} < 77$.
(a) Show that the significance level of the test is $P(Z < -\frac{\sqrt{n}}{4})$. [3]
(b) Find the least n such that the significance level is at most 1%. [3]
(c) Using this n , a sample gives $\bar{x} = 76.4$. State the conclusion, and say why the result would be reported differently if n were reduced to 16. [5]

Type I and Type II Errors

- 1 In a hypothesis test, state what is meant by
 - (a) a Type I error, [1]
 - (b) a Type II error. [1]
- 2 A smoke detector is tested. H_0 : the detector is working properly. H_1 : the detector is faulty. Describe, in context,
 - (a) a Type I error, [1]
 - (b) a Type II error. [1]
- 3 $X \sim B(10, p)$. A test of $H_0: p = 0.5$ against $H_1: p > 0.5$ has critical region $X \geq 8$. Find the probability of a Type I error. [2]
- 4 A test is carried out at the 5% significance level, and the test statistic has a continuous distribution.
 - (a) Write down the probability of a Type I error. [1]
 - (b) State what happens to this probability if the significance level is changed to 1%. [1]
- 5 $X \sim B(10, p)$. A test of $H_0: p = 0.5$ against $H_1: p > 0.5$ has critical region $X \geq 9$. Given that in fact $p = 0.8$, find the probability of a Type II error. [2]
- 6 A judge assumes a defendant is innocent (H_0) unless there is sufficient evidence of guilt (H_1). State whether each of the following is a Type I error or a Type II error.
 - (a) An innocent person is found guilty. [1]
 - (b) A guilty person is found not guilty. [1]
- 7 $X \sim B(20, p)$. A test of $H_0: p = 0.3$ against $H_1: p > 0.3$ has critical region $X \geq 10$.
 - (a) Find the probability of a Type I error. [2]
 - (b) Given that actually $p = 0.5$, find the probability of a Type II error. [2]
- 8 The number of faults per day in a system is modelled by $Po(\lambda)$. A test of $H_0: \lambda = 4$ against $H_1: \lambda > 4$ rejects H_0 when the number of faults in a day is at least 9. Find the probability of a Type I error, and the probability of a Type II error if in fact $\lambda = 6$. [4]
- 9 A random sample of 25 observations is taken from $N(\mu, 5^2)$. The null hypothesis $H_0: \mu = 50$ is rejected in favour of $H_1: \mu > 50$ if the sample mean $\bar{X} > 52$. Find the probability of a Type I error, and the probability of a Type II error when $\mu = 53$. [4]
- 10 A hospital screens patients for a disease. H_0 : the patient does not have the disease. H_1 : the patient has the disease. The screening test rejects H_0 if the reading is above a threshold.
 - (a) Describe the consequence of a Type I error and of a Type II error. [2]
 - (b) Explain which error is more serious, and what could be done to the threshold to make it less likely. [2]
- 11 Two tests are proposed for $H_0: p = 0.5$ against $H_1: p > 0.5$. Test A takes 10 observations and rejects H_0 if $X \geq 8$. Test B takes 20 observations and rejects H_0 if $X \geq 15$.
 - (a) Find the probability of a Type I error for each test. [3]
 - (b) Find the probability of a Type II error for each test when $p = 0.7$. [3]
- 12 A sample of 25 is used to test $H_0: p = 0.4$ against $H_1: p < 0.4$ at the 5% significance level, where $X \sim B(25, p)$.
 - (a) Find the critical region. [3]
 - (b) Find the probability of a Type II error when $p = 0.2$. [2]

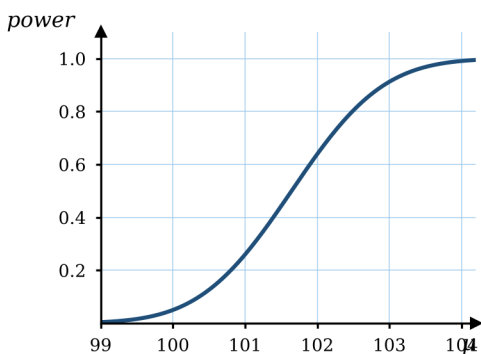
- 13** A hypothesis test has a probability of a Type I error of 0.04 and a probability of a Type II error of 0.30 for a particular alternative value.
- (a) Find the probability that H_0 is not rejected when H_0 is true. [1]
- (b) Find the power of the test for this alternative value. [2]
-
- 14** A random sample of 16 observations is taken from $N(\mu, 8^2)$ to test $H_0: \mu = 100$ against $H_1: \mu \neq 100$ at the 5% significance level.
- Find the acceptance region for the sample mean $\bar{X}$, and hence the probability of a Type II error when $\mu = 106$. [6]
-
- 15** A test of $H_0: p = 0.3$ against $H_1: p \neq 0.3$ for $X \sim B(20, p)$ has critical region $X \leq 1$ or $X \geq 11$.
- Find the probability of a Type II error when $p = 0.5$. [5]
-
- 16** For $X \sim B(20, p)$, a researcher tests $H_0: p = 0.5$ against $H_1: p > 0.5$. She is considering the critical regions $X \geq 15$ and $X \geq 16$.
- (a) Find the probability of a Type I error for each critical region. [3]
- (b) Given that $p = 0.7$, find the probability of a Type II error for each. [3]
- (c) A student says that the second critical region is better because it has the smaller significance level. Comment. [2]
-
- 17** $X \sim B(30, p)$ and $H_0: p = 0.2$ is tested against $H_1: p > 0.2$. The critical region $X \geq 10$ is compared with the critical region $X \geq 11$.
- For $p = 0.4$, find the probability of a Type II error for each critical region, and state which has the smaller probability of a Type I error. [6]
-
- 18** A batch is either good ($p = 0.2$ of items are defective) or bad ($p = 0.4$ defective). A buyer takes a sample of 25 items and rejects the batch if at least 8 are defective. 80% of batches supplied are good.
- Find the probability that a randomly chosen batch is wrongly rejected or wrongly accepted. [6]
-
- 19** A test of $H_0: p = 0.4$ against $H_1: p > 0.4$ uses $X \sim B(100, p)$ and rejects H_0 when $X \geq 50$. Use a normal approximation with a continuity correction to find
- (a) the probability of a Type I error, [4]
- (b) the probability of a Type II error when $p = 0.5$. [4]
-
- 20** Samples are taken from $N(\mu, 5^2)$ to test $H_0: \mu = 50$ against $H_1: \mu > 50$ at the 5% significance level. Test A uses $n = 25$ and test B uses $n = 100$.
- For each test find the probability of a Type II error when $\mu = 52$, and comment on the effect of the sample size. [7]
-
- 21** A company claims that 15% of its packets contain a prize. A consumer group thinks the true proportion is higher and tests a random sample of 50 packets, using a 5% significance level.
- (a) Find the critical region and state the actual probability of a Type I error. [5]
- (b) Find the probability of a Type II error if in fact the proportion is 0.25. [2]
- (c) Hence find the probability that the test makes the correct decision if the proportion really is 0.25, and give its statistical name. [3]
-
- 22** The number of customers arriving per hour at a shop is modelled by $Po(\lambda)$. A manager tests $H_0: \lambda = 6$ against $H_1: \lambda \neq 6$, using a single hour's count X and a 5% significance level with at most 2.5% in each tail.
- (a) Find the critical region and the probability of a Type I error. [6]
- (b) Find the probability of a Type II error when $\lambda = 9$. [2]
- (c) Find the probability of a Type II error when $\lambda = 3$ and comment on the two answers. [3]

- 23** A student's work on a test of $H_0: p = 0.4$ against $H_1: p < 0.4$ for $X \sim B(20, p)$ is shown. Critical region $X \leq 4$, because $P(X \leq 4) = 0.0510$. $P(\text{Type I error}) = P(X \leq 4 \mid p = 0.2) = 0.6296$. $P(\text{Type II error}) = P(X \geq 5 \mid p = 0.4) = 0.9490$. Find the three mistakes in the student's work and give the correct values, taking $p = 0.2$ as the alternative value in the Type II calculation. [7]
- 24** The lifetime T hours of a type of component is modelled by an exponential distribution with parameter λ , so that $P(T > t) = e^{-\lambda t}$. A test of $H_0: \lambda = 0.1$ against $H_1: \lambda < 0.1$ is carried out using the lifetime of one component, rejecting H_0 if $T > 30$.
 (a) Find the probability of a Type I error. [2]
 (b) Find the probability of a Type II error when $\lambda = 0.05$. [2]
 (c) A second component is tested independently and H_0 is rejected if both lifetimes exceed 30. Find the new probabilities of a Type I error and of a Type II error when $\lambda = 0.05$, and comment. [5]
- 25** A supplier delivers batches in which the proportion of defective items is either 0.05 (acceptable) or 0.25 (unacceptable). A buyer takes 20 items from each batch and rejects the batch if more than 2 are defective.
 (a) Find the probability that an acceptable batch is rejected (the producer's risk). [2]
 (b) Find the probability that an unacceptable batch is accepted (the consumer's risk). [2]
 (c) Find the smallest number c such that rejecting a batch when $X > c$ gives a producer's risk of at most 0.05, and find the consumer's risk for this rule. [4]
- 26** A coin is tossed until the first head appears, and N is the number of tosses needed. A test of $H_0: p = 0.5$ against $H_1: p < 0.5$, where p is the probability of a head, rejects H_0 if $N \geq 6$.
 (a) Find the probability of a Type I error. [2]
 (b) Find the probability of a Type II error when $p = 0.3$. [2]
 (c) Find the smallest value of k so that the rule 'reject H_0 if $N \geq k$ ' has $P(\text{Type I}) \leq 0.1$, and find the corresponding $P(\text{Type II})$ when $p = 0.3$. [4]
- 27** A sample of size n is taken from $N(\mu, 10^2)$ to test $H_0: \mu = 100$ against $H_1: \mu > 100$, rejecting H_0 when $\bar{X} > c$. The significance level is to be 5%, and the probability of a Type II error when $\mu = 105$ is to be at most 10%.
 (a) Show that $c = 100 + \frac{16.45}{\sqrt{n}}$. [3]
 (b) Show that the Type II condition requires $\frac{c - 105}{\frac{10}{\sqrt{n}}} \leq -1.2816$. [3]
 (c) Hence find the smallest sample size n . [5]
- 28** The number of defects per metre of cable is modelled by $Po(\lambda)$. Two metres are inspected, giving a total $X \sim Po(2\lambda)$. A test of $H_0: \lambda = 3$ against $H_1: \lambda = 5$ rejects H_0 if $X \geq c$.
 (a) Write down the distribution of X under each hypothesis. [2]
 (b) Express $P(\text{Type I}) + P(\text{Type II})$ in terms of c . [3]
 (c) Find the value of c which minimises this sum, giving the minimum value. [5]
- 29** A random sample X_1, X_2 is taken from a continuous uniform distribution on $[0, \theta]$. A test of $H_0: \theta = 8$ against $H_1: \theta > 8$ rejects H_0 if $\max(X_1, X_2) > k$.
 (a) Show that $P(\max(X_1, X_2) \leq x) = \frac{x^2}{\theta^2}$ for $0 \leq x \leq \theta$. [2]
 (b) Find k so that the significance level is 5%. [2]
 (c) Find the probability of a Type II error when $\theta = 10$. [2]
 (d) Find the least number of observations n for which the test, with the same rule $\max > k_n$ chosen to keep the 5% level, has Type II error probability below 0.1 when $\theta = 10$. [6]

- 30** A laboratory screens many candidate treatments, each tested on 20 patients with $X \sim B(20, p)$. $H_0: p = 0.3$ (no improvement) is tested against $H_1: p > 0.3$ at the 5% significance level, and H_0 is rejected for $X \geq c$. In fact, 90% of the candidate treatments have $p = 0.3$, and the other 10% have $p = 0.5$.
- Find c and the actual probability of a Type I error. [4]
 - Find the probability that a treatment with $p = 0.5$ is correctly detected. [2]
 - A treatment has been flagged as improved (H_0 rejected). Find the probability that it actually has $p = 0.3$. [4]
 - Explain why this probability is much larger than the significance level of the test. [2]

Power of a Test

- In a hypothesis test, state
 - what is meant by the power of the test, [1]
 - how the power is related to the probability of a Type II error. [1]
- For a particular alternative value, a test has a probability of a Type II error of 0.35. Write down the power of the test. [1]
- $X \sim B(10, p)$. A test of $H_0: p = 0.5$ against $H_1: p > 0.5$ has critical region $X \geq 8$. Find the power of the test when $p = 0.8$. [2]
- A test is carried out using a random sample.
 - State, with a reason, the effect on the power of the test of increasing the sample size, keeping the significance level the same. [1]
 - State the power of a test at a value of the parameter in H_0 , for a continuous test statistic with significance level 5%. [1]
- $X \sim \text{Po}(\lambda)$. A test of $H_0: \lambda = 3$ against $H_1: \lambda > 3$ has critical region $X \geq 6$. Find the power of the test when $\lambda = 5$. [2]
- A test has a power of 0.85 when $p = 0.7$ and a significance level of 0.04.
 - Write down the probability of a Type II error when $p = 0.7$. [1]
 - Write down the probability of a Type I error. [1]
- $X \sim B(20, p)$. A test of $H_0: p = 0.5$ against $H_1: p < 0.5$ has critical region $X \leq 5$.
 - Find the significance level of the test. [2]
 - Find the power of the test when $p = 0.3$. [2]
- A test of $H_0: p = 0.5$ against $H_1: p > 0.5$ uses $X \sim B(15, p)$ and has critical region $X \geq 12$.
 - Find the power of the test when $p = 0.6$, $p = 0.7$ and $p = 0.8$. [4]
 - Describe how the power changes as p moves further from 0.5, and explain why. [2]
- A random sample of 25 observations is taken from $N(\mu, 5^2)$. The null hypothesis $H_0: \mu = 50$ is tested against $H_1: \mu > 50$ at the 5% significance level. Find the power of the test when $\mu = 52$. [4]
- A random sample of 36 observations is taken from $N(\mu, 6^2)$ to test $H_0: \mu = 80$ against $H_1: \mu \neq 80$ at the 5% significance level. Find the power of the test when $\mu = 83$. [4]
- The lifetime T of a component has probability density function $\lambda e^{-\lambda t}$ for $t \geq 0$. A test of $H_0: \lambda = 0.1$ against $H_1: \lambda < 0.1$ rejects H_0 when one observed lifetime exceeds 30. Find the power of the test when $\lambda = 0.05$. [2]

- 12** $X \sim \text{Po}(\lambda)$. A test of $H_0: \lambda = 6$ against $H_1: \lambda \neq 6$ has critical region $X \leq 1$ or $X \geq 12$. Find the power of the test when $\lambda = 12$. [3]
- 13** A single observation X is taken from a continuous uniform distribution on $[0, \theta]$. A test of $H_0: \theta = 10$ against $H_1: \theta < 10$ rejects H_0 if $X < 0.5$.
 (a) Find the significance level of the test. [2]
 (b) Find the power of the test when $\theta = 4$. [2]
 (c) Find the value of θ for which the power of the test is 0.5. [2]
- 14** A test of $H_0: p = 0.3$ against $H_1: p > 0.3$ uses $X \sim B(30, p)$ at the 5% significance level. Find the critical region and the power of the test when $p = 0.5$. [6]
- 15** A test of $H_0: \mu = \mu_0$ against $H_1: \mu > \mu_0$ is based on the mean of a sample of size n from $N(\mu, \sigma^2)$ at the 5% significance level.
 Show that the power of the test when $\mu = \mu_0 + \delta$ is $P(Z > 1.645 - \frac{\delta\sqrt{n}}{\sigma})$. Hence find the power when $n = 64$, $\sigma = 8$ and $\delta = 3$. [6]
- 16** Two tests of $H_0: p = 0.5$ against $H_1: p > 0.5$ are proposed. Test A uses $X \sim B(10, p)$ with critical region $X \geq 8$. Test B uses $X \sim B(20, p)$ with critical region $X \geq 15$.
 (a) Find the significance level of each test. [3]
 (b) Find the power of each test when $p = 0.75$. [3]
 (c) Which test would you recommend? Give a reason. [2]
- 17** $X \sim \text{Po}(\lambda)$. A test of $H_0: \lambda = 10$ against $H_1: \lambda \neq 10$ is carried out at the 5% significance level with at most 2.5% in each tail. Find the critical region and the power of the test when $\lambda = 14$. [7]
- 18** $X \sim B(10, p)$. A test of $H_0: p = 0.5$ against $H_1: p > 0.5$ has critical region $X \geq 8$. Find the power of the test for $p = 0.6, 0.7, 0.8$ and 0.9 , and sketch the power function for $0.5 \leq p \leq 1$, labelling the value at $p = 0.5$. [7]
- 19** The graph shows the power of a test of $H_0: \mu = 100$ against $H_1: \mu > 100$, based on the mean of 25 observations from $N(\mu, 5^2)$, as a function of the true mean μ .

 (a) Write down the value of the power at $\mu = 100$ and explain its meaning. [2]
 (b) Calculate the power of the test when $\mu = 102$. [3]
 (c) Calculate the smallest μ for which the power is at least 0.9. [2]
- 20** A random sample of 25 observations is taken from $N(\mu, 5^2)$ to test $H_0: \mu = 100$ against $H_1: \mu > 100$ at the 5% significance level. Find the value of μ for which the power of the test is 0.8. [5]

- 21** A random sample of 36 observations is taken from $N(\mu, 12^2)$ to test $H_0: \mu = 150$ against $H_1: \mu \neq 150$ at the 5% significance level.
- Find the critical region for the sample mean. [3]
 - Show that the power of the test when $\mu = 155$ is 0.7 to 1 decimal place. [4]
 - Find the value of μ above 150 for which the power is 0.9, ignoring the lower tail. [3]

- 22** A coin is tossed 5 times to test $H_0: p = 0.5$ against $H_1: p > 0.5$, where p is the probability of heads and X is the number of heads. Test 1 rejects H_0 if $X = 5$. Test 2 rejects H_0 if $X \geq 4$.
- Show that the power function of Test 1 is p^5 and that of Test 2 is $5p^4 - 4p^5$. [3]
 - Find the significance level of each test. [3]
 - Find the power of each test when $p = 0.8$ and comment on the comparison. [4]

- 23** The power of a test of $H_0: p = 0.4$ against $H_1: p > 0.4$ using $X \sim B(20, p)$ and critical region $X \geq 12$ is calculated by a student.
- Student's solution: 'Power = $P(X \geq 12 | p = 0.4) = 0.0565$, so the power is 5.65%. This is low, so the test is poor. Also $P(\text{Type II error}) = 1 - 0.0565 = 0.9435$ when $p = 0.6$.'
- Identify the two errors in the student's solution and find the correct power and the correct probability of a Type II error when $p = 0.6$. [6]

- 24** The number of flaws in a length of fabric is modelled by $Po(\lambda)$. Five independent lengths are inspected, and S is the total number of flaws. A test of $H_0: \lambda = 2$ against $H_1: \lambda > 2$ is carried out at the 5% significance level.
- Find the critical region for S , the actual significance level, and the power of the test when $\lambda = 3$. [7]

- 25** A test of $H_0: p = 0.5$ against $H_1: p > 0.5$ uses $X \sim B(100, p)$ at the 5% significance level, with a normal approximation.
- Use a normal approximation with a continuity correction to find the critical region. [4]
 - Use a normal approximation to estimate the power of the test when $p = 0.65$. [4]
 - Use the binomial distribution to find the exact significance level and exact power. [2]

- 26** A single observation X takes the values 1, 2, 3, 4 or 5. Its probability distribution under the null hypothesis H_0 and under the alternative hypothesis H_1 are shown in the table. A test rejects H_0 if X lies in a critical region C , and the significance level must not exceed 10%.

x	1	2	3	4	5
$P(X = x)$ under H_0	0.40	0.30	0.20	0.07	0.03
$P(X = x)$ under H_1	0.05	0.15	0.20	0.30	0.30

- List every possible critical region C of size at most 0.10 that contains no more than two values, and find the power of each. [4]
 - State which region gives the most powerful test and explain, using the ratio of the probabilities under H_1 and H_0 , why it is chosen. [3]
- 27** A statistician plans to test $H_0: p = 0.5$ against $H_1: p > 0.5$ at the 5% significance level (largest critical region $X \geq c$ with significance level at most 0.05) using $X \sim B(n, p)$. She requires a power of at least 0.8 when $p = 0.7$.
- She considers $n = 20$, $n = 30$ and $n = 40$. For each, find c and the power of the test when $p = 0.7$, and state which is the smallest of these three sample sizes that meets her requirement. [5]

- 28** A random sample of 25 observations is taken from $N(\mu, 3^2)$ to test $H_0: \mu = 40$ against $H_1: \mu \neq 40$ at the 5% significance level.
- (a) Show that the power of the test when $\mu = 40 + \delta$ is $P(Z > 1.96 - \frac{\delta}{0.6}) + P(Z < -1.96 - \frac{\delta}{0.6})$. [5]
- (b) Find the power when $\delta = 1.176$, and explain why it is close to 0.5. [3]
- (c) Find the least value of $\delta > 0$ for which the power is at least 0.9, correct to 2 decimal places. [4]
-
- 29** A factory tests a machine setting on each of 3 independent days, using the same test each day. On each day the test has significance level 0.05 and, if the setting is faulty, power 0.7. The machine is stopped if the test rejects H_0 (setting correct) on at least 2 of the 3 days.
- (a) Find the overall probability that the machine is stopped when the setting is correct. [4]
- (b) Find the overall probability that the machine is stopped when the setting is faulty. [3]
- (c) Find the daily power q needed for the overall power to be 0.9, giving q to 3 decimal places. [4]
-
- 30** $X \sim B(4, p)$. A test of $H_0: p = 0.5$ against $H_1: p > 0.5$ rejects H_0 if $X \geq 3$. Let $\pi(p)$ denote the power function of the test.
- (a) Show that $\pi(p) = 4p^3 - 3p^4$. [3]
- (b) Prove that π is an increasing function of p for $0 < p < 1$. [4]
- (c) Find the significance level of the test and the value of p for which the power is 0.5, giving p to 3 decimal places. [4]

Poisson Hypothesis Tests

- 1** The number of typing errors per page in a report is modelled by a Poisson distribution with mean λ . A proofreader believes that the mean number of errors per page is less than 3. Write down suitable hypotheses for a test of her belief. [2]
-
- 2** $X \sim \text{Po}(\lambda)$. A test of $H_0: \lambda = 3.5$ against $H_1: \lambda < 3.5$ is carried out at the 5% significance level, and $x = 1$ is observed. Find the p -value and state the conclusion. [3]
-
- 3** Telephone calls arrive at a switchboard at a mean rate of 2.4 per hour. Assuming a Poisson model, write down the distribution of the number of calls in
- (a) 5 hours, [1]
- (b) 30 minutes. [1]
-
- 4** $X \sim \text{Po}(\lambda)$. A test of $H_0: \lambda = 4$ against $H_1: \lambda > 4$ is carried out at the 5% significance level, and $x = 9$ is observed. Find the p -value and state the conclusion. [3]
-
- 5** State two conditions that must hold for the number of events in a fixed interval to be modelled by a Poisson distribution. [2]
-
- 6** $X \sim \text{Po}(\lambda)$. A test of $H_0: \lambda = 5$ against $H_1: \lambda < 5$ is carried out at the 5% significance level. Find the critical region. [2]
-
- 7** Emails arrive at a company at a mean rate of 4 per hour. After an advertising campaign the manager believes the rate has increased. In a 3-hour period 18 emails arrive. Test the manager's belief at the 5% significance level, stating your hypotheses. [5]

- 8** $X \sim \text{Po}(\lambda)$. A two-tailed test of $H_0: \lambda = 7$ against $H_1: \lambda \neq 7$ is carried out at the 5% significance level, with at most 2.5% in each tail.
 (a) Find the critical region. [5]
 (b) Find the actual significance level. [2]
-
- 9** A road has an average of 0.8 potholes per 100 m. After resurfacing a 300 m stretch, an engineer finds no potholes. She claims that resurfacing has reduced the number of potholes. Test her claim at the 5% significance level. [5]
-
- 10** Customers enter a shop at a mean rate of 1.5 per 10 minutes. During a promotion, 11 customers enter in a 40-minute period, and the manager believes that the rate has changed. Test the manager's belief at the 5% significance level. [5]
-
- 11** The number of accidents per month at a junction is modelled by $\text{Po}(\lambda)$. A council claims $\lambda = 30$. A resident believes the true value is higher, and observes 40 accidents in a month. Use a normal approximation with a continuity correction to test the resident's belief at the 5% significance level. [5]
-
- 12** $X \sim \text{Po}(\lambda)$. A test of $H_0: \lambda = 8$ against $H_1: \lambda \neq 8$ has critical region $X \leq 2$ or $X \geq 14$. Find the probability of a Type I error. [3]
-
- 13** $X \sim \text{Po}(\lambda)$. A two-tailed test of $H_0: \lambda = 9$ against $H_1: \lambda \neq 9$ is carried out at the 5% significance level, and $x = 3$ is observed. Find the p-value for this two-tailed test and state the conclusion. [4]
-
- 14** The table shows the number of accidents at a level crossing on each of 8 consecutive days. Previously the mean was 1.5 accidents per day. Following the installation of a barrier, the operator claims that the mean has fallen. Test the operator's claim at the 5% significance level, stating your hypotheses clearly.
- | Day | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----|---|---|---|---|---|---|---|---|
| | 0 | 1 | 0 | 2 | 1 | 0 | 0 | 0 |
- [6]
-
- 15** $X \sim \text{Po}(\lambda)$. A test of $H_0: \lambda = 12$ against $H_1: \lambda > 12$ is carried out at the 5% significance level. Find the critical region and the power of the test when $\lambda = 16$. [6]
-
- 16** A call centre states that its mean rate of calls is 0.9 per minute. A supervisor believes the rate is higher.
 (a) In a 10 minute period there are 14 calls. Test the supervisor's belief at the 5% significance level. [5]
 (b) In a different 20 minute period there are 28 calls. Carry out the same test and comment on the difference in conclusion. [4]
-
- 17** The number of cars passing a checkpoint in 10 minutes is modelled by $\text{Po}(\lambda)$. Historically $\lambda = 50$. On one day 36 cars pass in 10 minutes and a trader claims that the rate has changed. Use a normal approximation with a continuity correction to test this claim at the 5% significance level. [5]
-
- 18** Vehicles passing a point on a rural road during the day are counted over several days. A researcher proposes that the number of vehicles per hour follows a Poisson distribution.
 (a) Give one reason why the Poisson model might be unsuitable at a point near a school. [2]
 (b) Hourly counts over 10 hours are 4, 8, 3, 7, 6, 9, 2, 6, 5, 5. Calculate the mean and variance, and comment on the suitability of a Poisson model. [4]
-

- 19** On weekdays the number of enquiries to a helpdesk in a morning is modelled by $Po(2.5)$. On Saturdays the number in a morning is modelled by $Po(1.2)$. In one week (5 weekday mornings and 1 Saturday morning), assuming independence, 20 enquiries are received. Test, at the 5% significance level, whether the helpdesk is busier than the model suggests. [6]
-
- 20** A factory's quality log shows an average of 40 minor defects per week. After new training, 29 defects are recorded in one week. The manager believes the mean has fallen. Let $X \sim Po(\lambda)$ be the number of defects in a week.
- (a) Use a normal approximation with a continuity correction to test the manager's belief at the 5% significance level. [5]
- (b) Use the Poisson distribution to find the exact p-value and comment on the approximation. [2]
-
- 21** $X \sim Po(\lambda)$. A test of $H_0: \lambda = 2$ against $H_1: \lambda > 2$ is carried out using a single observation of X at the 5% significance level.
- (a) Find the critical region and the actual significance level. [4]
- (b) Find the power of the test when $\lambda = 4$. [2]
- (c) Show that the power is 0.5 when λ is between 5 and 6 and find this value of λ to 2 decimal places. [5]
-
- 22** The number of faults per roll in a fabric is modelled by $Po(\lambda)$. A manager's work on a test of $H_0: \lambda = 2.5$ against $H_1: \lambda > 2.5$ using the faults in 4 rolls is shown. 'Faults in 4 rolls $X \sim Po(2.5)$. Observed $x = 16$. $P(X \geq 16) = 1 - P(X \leq 16) = 0.0000$. This is below 5% so reject H_0 .' Identify two errors in the manager's work and give the correct conclusion. [6]
-
- 23** Two similar machines run for the same length of time. Machine A produces 3 faulty items and machine B produces 12 faulty items. The numbers of faults are independent Poisson variables with means λ_A and λ_B . It is known that, given a total of 15 faulty items, the number from machine A is $B(15, \frac{\lambda_A}{\lambda_A + \lambda_B})$. Use this result to test, at the 5% significance level, whether the machines have different fault rates. [6]
-
- 24** Breakdowns on a production line occur as a Poisson process at a mean rate of 2 per day. A manager believes that the rate has decreased after maintenance. The time T days until the third breakdown is recorded.
- (a) Explain why the event $T > t$ is equivalent to there being at most 2 breakdowns in the first t days. [2]
- (b) The third breakdown occurs after 3.5 days. Test the manager's belief at the 5% significance level. [5]
-
- 25** The number of customer complaints per day is modelled by $Po(\lambda)$. The manager records the total number of complaints, S , over 4 days. A test of $H_0: \lambda = 3$ against $H_1: \lambda \neq 3$ is carried out at the 5% significance level with at most 2.5% in each tail. Find the critical region for S and the probability of a Type II error when $\lambda = 5$. [8]
-
- 26** The number of defects per day on a line is $Po(\lambda)$. A daily test of $H_0: \lambda = 3$ against $H_1: \lambda > 3$ rejects H_0 when the day's count is at least 8.
- (a) Find the probability of a Type I error on a given day. [2]
- (b) Over a working week of 5 independent days with $\lambda = 3$, find the probability of at least one false rejection. [2]
- (c) Find the least number of days after which the probability of at least one false rejection exceeds 0.5. [3]

- 27** A mean of 0.4 defects per unit is claimed for a process. A quality engineer inspects n units and records the total number of defects, which she models as $Po(0.4n)$ under the claim. She tests $H_0: \lambda = 0.4$ against $H_1: \lambda < 0.4$, rejecting H_0 only if no defects at all are found, at a significance level of at most 5%.
- (a) Show that she needs $n \geq 8$ for the test to have a significance level of at most 5%. [4]
- (b) With $n = 8$, find the actual significance level and the power when the true mean is 0.1 defects per unit. [4]
- (c) Comment on why the test has such low power, and suggest how a different critical region could improve it. [2]
-
- 28** $X \sim Po(\lambda)$. A two-tailed test of $H_0: \lambda = 25$ against $H_1: \lambda \neq 25$ is carried out at the 5% significance level.
- (a) Use a normal approximation with continuity correction to find the critical region. [5]
- (b) Use the Poisson distribution to find the actual significance level of this region. [2]
- (c) Use exact Poisson probabilities to find the critical region with at most 2.5% in each tail, and comment on the difference. [5]
-
- 29** Faults in a long cable occur as a Poisson process at a mean rate of λ per kilometre. A technician is told $\lambda = 1.5$ and believes the true rate is lower. She inspects the cable until the 4th fault is found, which is at a distance of 4.2 km from the start. Let D be the distance to the 4th fault.
- (a) Show that $P(D > d) = P(Y \leq 3)$, where $Y \sim Po(1.5d)$ under H_0 . [2]
- (b) Carry out a test of $H_0: \lambda = 1.5$ against $H_1: \lambda < 1.5$ at the 5% significance level. [4]
- (c) Find the least distance D , to 1 decimal place, beyond which the 4th fault would lead to rejection of H_0 . [4]
-
- 30** Two machines produce faults as independent Poisson processes with rates λ_A and λ_B per hour. Machine A runs for 10 hours and produces 14 faults. Machine B runs for 8 hours and produces 5 faults. Given the total number of faults T , the number from machine A is $B(T, \frac{10\lambda_A}{10\lambda_A + 8\lambda_B})$.
- Test, at the 5% significance level, whether machine A has a higher fault rate than machine B. State your hypotheses and the distribution used. [8]

Chi-Squared Goodness-of-Fit Tests

- 1** A six-sided die is rolled many times and the results are used to test whether the die is fair.
- (a) Write down suitable hypotheses. [2]
- (b) State the degrees of freedom and the critical value of the test statistic at the 5% significance level. [2]
-
- 2** A die is rolled 120 times. The scores 1 to 6 occur 18, 22, 15, 25, 21 and 19 times respectively. Find the expected frequencies if the die is fair and calculate the value of $\sum \frac{(O - E)^2}{E}$. [3]
-
- 3** A genetic model predicts that offspring show a dominant and a recessive characteristic in the ratio 3 : 1. Of 200 offspring, 165 show the dominant characteristic and 35 show the recessive characteristic.
- Calculate the expected frequencies and the test statistic $\sum \frac{(O - E)^2}{E}$. [3]
-
- 4** State the number of degrees of freedom in each of the following goodness-of-fit tests, after any pooling needed.
- (a) Testing a binomial $B(5, 0.4)$ model with 6 classes, none needing to be pooled. [1]
- (b) Testing a Poisson model with the mean estimated from the data, 7 classes, the last two being pooled. [1]

- 5 In a goodness-of-fit test, state the rule that is usually applied to the expected frequencies before the test statistic is calculated, and what is done if a class does not satisfy it. [2]

- 6 Kelly rolls a tetrahedral die and records the number on which it lands. She calculates that if the die is fair, the expected frequency of each number is 45. The table shows three of the four observed frequencies; the fourth is missing.

Number	1	2	3	4
Frequency	52	38	41	x

- (a) Find the value of x. [2]
 (b) Explain why there are 3 degrees of freedom for a goodness-of-fit test of a fair die. [1]
 (c) Calculate the test statistic and state whether the die appears biased at the 5% significance level. [3]

- 7 A die is rolled 120 times, giving the frequencies shown. Test, at the 5% significance level, whether the die is fair.

Score	1	2	3	4	5	6
Frequency	14	22	17	28	24	15

[6]

- 8 A genetic theory predicts that four types of pea, A, B, C and D, occur in the ratio 9 : 3 : 3 : 1. In an experiment, 550 peas were grown with the results shown. Test the theory at the 5% significance level.

Type	A	B	C	D
Frequency	312	108	96	34

[6]

- 9 A hospital records the number of births on each day of the week over a period, giving the results shown. Test, at the 5% significance level, whether births are equally likely on each day.

Day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Births	38	45	52	60	57	44	34

[6]

- 10 A fair coin is thought to have been tossed in sets of four tosses. In 160 sets the number of heads was recorded as shown. Test, at the 5% significance level, whether the number of heads in a set follows a binomial distribution $B(4, 0.5)$.

Heads	0	1	2	3	4
Frequency	8	38	62	41	11

[6]

- 11** The number of flaws in each of 120 sheets of glass is recorded. The frequencies are shown. A manager suggests that the number of flaws per sheet follows $Po(1.5)$. Test this suggestion at the 5% significance level.

Count	0	1	2	3	4	5 or more
Frequency	30	38	28	14	7	3

[7]

- 12** A Poisson distribution $Po(2.4)$ is used to model the number of faults per item in a sample of 150 items. The expected frequencies are shown, with 2 d.p. for some values. Find the values of s, t and u, giving your answers to 2 decimal places.

Count	0	1	2	3	4 or more
Expected	s	32.66	39.19	t	u

[6]

- 13** A goodness-of-fit test with 7 classes (no parameters estimated) gives a test statistic of 11.3.
- (a) State the number of degrees of freedom and the critical values of χ^2 at the 10%, 5% and 1% significance levels. [3]
- (b) State the conclusion at each of these significance levels. [2]

- 14** From each of 80 batches a sample of 5 items is inspected and the number of defective items recorded, as shown. A manager suggests that the number of defectives per sample follows a binomial distribution. Test the manager's suggestion at the 5% significance level, estimating p from the data.

Defectives	0	1	2	3	4	5
Samples	28	31	15	4	2	0

[9]

- 15** The number of customers arriving at a counter in each of 200 one-minute intervals is shown. Test, at the 5% significance level, whether a Poisson distribution is a suitable model.

Arrivals	0	1	2	3	4	5	6+
Minutes	38	62	49	30	13	5	3

[9]

- 16** A coin is spun 5 times and the number of heads is recorded. This is repeated 150 times, giving the frequencies shown. Rosa believes that the number of heads follows $B(5, 0.5)$, and calculates the expected frequencies shown, where r and s are missing.

Heads	0	1	2	3	4	5
Frequency	5	20	38	47	30	10
Expected	r	23.44	s	s	23.44	r

- (a) Find the values of r and s, giving your answers to 2 decimal places. [2]
- (b) Test, at the 5% significance level, whether $B(5, 0.5)$ is a suitable model. [5]
- (c) Estimate p from the data, and test at the 5% significance level whether $B(5, p)$ with this value of p is a suitable model. State which model fits better. [5]

- 17** A discrete random variable X has probability distribution $P(X = x) = kx$ for $x = 1, 2, 3, 4$. In a sample of 100 observations the values 1, 2, 3 and 4 occurred 12, 20, 31 and 37 times respectively.
Find k , and test at the 5% significance level whether the distribution is a good model. [7]

- 18** Show that $\sum \frac{(O - E)^2}{E} = \sum \frac{O^2}{E} - N$, where $N = \sum O = \sum E$.
Hence calculate the test statistic for observed frequencies 20, 31, 29 against expected frequencies 25, 25, 30. [5]

- 19** Two fair dice are rolled together 360 times and the total score is recorded each time, as shown.
Test, at the 5% significance level, whether the dice are behaving as fair dice.

Total	2	3	4	5	6	7	8	9	10	11	12
Frequency	8	21	29	36	52	61	49	41	33	20	10

[7]

- 20** A die is rolled repeatedly until a six appears, and the number of rolls is recorded. This is repeated 120 times, with the results shown. Test, at the 5% significance level, whether the die is fair, using the geometric distribution with $p = \frac{1}{6}$.

Throws	1	2	3	4	5	6	7 or more
Frequency	38	28	17	11	9	7	10

[7]

- 21** The number of boys in each of 100 families with four children is recorded as shown. A sociologist proposes that the number of boys follows a binomial distribution $B(4, p)$.

Boys	0	1	2	3	4
Families	4	18	40	27	11

- (a) Estimate p from the data. [2]
 (b) Test the sociologist's proposal at the 5% significance level. [7]
 (c) Explain why the number of degrees of freedom is reduced by 1 because p was estimated. [1]

- 22** A student's solution to a goodness-of-fit test is shown.
 Test whether a die is fair using 90 rolls with frequencies 10, 12, 20, 9, 18, 21. ' H_0 : the die is not fair. $E = 90 \div 6 = 15$ each. $\sum (O - E)^2/E = 4.8$ (using the values $O - E$ rather than squares). $\nu = 6$. Critical value 12.592. $4.8 < 12.592$ so the die is fair.'
 Identify three errors in the student's work and give the correct conclusion. [6]

- 23** The heights of 120 adult men in a town are grouped as shown. The sample mean is 170 cm and the sample standard deviation is 8 cm. A researcher suggests that heights are normally distributed.

Height (cm)	<155	155-162	162-169	169-176	176-183	>183
Frequency	4	14	32	38	21	11

- (a) Calculate the expected frequencies for a normal distribution with these parameters. [4]
 (b) Test the researcher's suggestion at the 5% significance level. [6]

- 24** A researcher claims that a die was rolled 600 times and that the results were completely consistent with a fair die. The goodness-of-fit statistic for a 10-sided die is calculated as 1.2.
- (a) State the number of degrees of freedom, and use tables or a calculator to find $P(\chi^2_9 \leq 1.2)$, giving your answer to 3 decimal places. [2]
- (b) Use the lower-tail critical value to explain why the data might be regarded as suspiciously good, and suggest what this might indicate. [4]

- 25** In a population of 200 individuals the genotypes AA, Aa and aa occur as shown. A genetic model states that the probabilities of the genotypes are θ^2 , $2\theta(1 - \theta)$ and $(1 - \theta)^2$ respectively.

Genotype	AA	Aa	aa
Frequency	58	102	40

- (a) Show that the proportion of A genes in the sample is 0.5450, and use it as an estimate of θ . [2]
- (b) Test the model at the 5% significance level. [4]

- 26** In 100 trials, the number of successes in a group of 6 is recorded as shown. The mean number of successes per group is 3.020.

Number	0	1	2	3	4	5	6 or more
Frequency	3	9	24	31	18	10	5

- (a) Test, at the 5% level, whether $B(6, p)$ with p estimated from the mean is a suitable model. [6]
- (b) Test, at the 5% level, whether a Poisson model with the same mean is suitable, and state which model fits better. [6]

- 27** In a binomial experiment with n trials, X successes are observed. A chi-squared test with two classes (success and failure) is used to test $H_0: p = p_0$.

- (a) Show that the test statistic is $\frac{(X - np_0)^2}{np_0(1 - p_0)}$. [5]
- (b) Deduce that the statistic is the square of the normal test statistic for the binomial proportion without continuity correction, and say what this implies for the critical value with $\nu = 1$. [4]
- (c) Verify with $n = 100$, $X = 62$ and $p_0 = 0.5$ that both calculations give the same value of χ^2 . [3]

- 28** A computer program generates 240 numbers, each intended to be uniformly distributed on the interval $[0, 1]$. The interval is split into 8 equal parts and the numbers falling in each part are counted as shown (each part is the interval from the previous upper limit to the given upper limit).

Interval	0-0.125	-0.25	-0.375	-0.5	-0.625	-0.75	-0.875	-1
Count	34	25	33	28	36	24	31	29

- (a) Test, at the 5% significance level, whether the numbers are uniformly distributed. [5]
- (b) Find the greatest number of equal parts into which $[0, 1]$ could have been split while keeping all expected frequencies at least 5, and state the degrees of freedom in that case. [3]
- (c) Explain why using many more classes than needed is not always advisable. [2]

- 29** The waiting times of 140 patients, in minutes, are summarised as shown. A manager proposes an exponential model for the waiting time, with probability density function $\lambda e^{-\lambda t}$ for $t \geq 0$. Using the class mid-points 2.5, 7.5, 15, 30 and 55 to estimate the mean, estimate λ and test the suitability of the model at the 5% significance level.

Time (min)	0-5	5-10	10-20	20-40	over 40
Frequency	52	32	31	20	5

[10]

- 30** A die is rolled 100 times and gives scores 1 to 6 with frequencies 14, 17, 15, 13, 16 and 25. A second experiment gives exactly four times each frequency (400 rolls). A goodness-of-fit test for a fair die is carried out at the 5% significance level in each case.
- Calculate the test statistic and state the conclusion for each experiment. [7]
 - Show that multiplying all frequencies by a constant k multiplies the statistic by k . [2]
 - Explain what this indicates about the effect of sample size on the test. [2]

Chi-Squared Tests for Association

- 1** A teacher wishes to find out whether choice of A level subject is associated with gender.
- Write down suitable null and alternative hypotheses for a χ^2 test. [2]
 - The data are put in a contingency table with 2 rows (gender) and 4 columns (subject). State the number of degrees of freedom. [1]

- 2** Write down the number of degrees of freedom for a χ^2 test of association on a contingency table with
- 2 rows and 2 columns, [1]
 - 3 rows and 5 columns, [1]
 - 4 rows and 3 columns, after two of the columns have been combined. [1]

- 3** In a contingency table the row total for 'Female' is 80, the column total for 'Walks to school' is 45 and the grand total is 200. Calculate the expected frequency for the cell 'Female, walks to school' under the hypothesis of no association. [2]

- 4** The table shows how 80 children travel to school. Calculate the row totals, the column totals and the expected frequency in each cell if there is no association between sex and the method of travel.

	Cycle	Walk
Boys	18	12
Girls	22	28

[3]

- 5** In a 2×2 contingency table the observed frequencies are 30, 20, 20, 30 and the expected frequencies are all 25. Calculate the value of the test statistic $\sum \frac{(O - E)^2}{E}$. [2]
- 6** In a χ^2 test of association, one cell has an expected frequency of 3.2.
- State what should be done before the test statistic is calculated. [1]
 - State the effect of this on the degrees of freedom. [1]

- 7 Students are taught using one of two methods and then pass or fail an exam, as shown. Test, at the 5% significance level, whether the result is associated with the teaching method.

	Pass	Fail
Method A	48	22
Method B	36	34

[7]

- 8 A survey records the usual method of commuting of 200 workers, as shown. Test, at the 5% significance level, whether commuting method is independent of age group.

	Car	Bus	Train
Under 30	34	28	38
30 and over	52	20	28

[7]

- 9 A company asks 230 customers in three regions which of three colours they prefer for a new product, with results shown. Test, at the 5% significance level, whether colour preference is associated with region.

	Red	Blue	Green
North	22	31	17
Midlands	35	22	28
South	18	27	30

[7]

- 10 A sample of 100 children were asked whether they prefer chess or football. The table shows the numbers choosing each, where a and b are unknown. The column totals are 40 (chess) and 60 (football). Find a and b, and the expected frequency for 'Girls, chess' if choice is independent of gender.

	Chess	Football
Boys	24	a
Girls	b	24

[5]

- 11 A χ^2 test of association on a 3×3 contingency table gives a test statistic of 10.2.

- (a) State the degrees of freedom and the critical values at the 5% and 1% levels. [3]
 (b) State the conclusion of the test at each level, in context for variables 'age group' and 'type of film preferred'. [2]

- 12 The table shows the results of a survey of 59 people. Show that a χ^2 test cannot be carried out on the table as it stands. Combine the first two columns and test, at the 5% significance level, for association.

	Low	Medium	High
Group 1	3	9	18
Group 2	7	12	10

[8]

- 13** A school council records the favourite colour of 236 students in three year groups, as shown. A test of association between favourite colour and year group is to be carried out.

	Red	Blue	Green	Yellow	Black	Total
Year 7-9	30	18	16	22	6	92
Year 10-11	24	29	13	9	10	85
Sixth form	8	22	16	7	6	59
Total	62	69	45	38	22	236

- (a) Write down suitable hypotheses. [2]
- (b) Only one cell needs to be checked to show that no expected frequency is below 5. Identify this cell and calculate its expected frequency. [3]
- (c) The test statistic is 19.99. Using a 1% significance level, complete the test, stating the degrees of freedom, the critical value and your conclusion. [3]

- 14** Two groups of patients, A and B, rate their treatment as shown. Test at the 5% significance level whether rating is associated with group, and state which category contributes most to the test statistic.

	Poor	Fair	Good	Excellent
Group A	18	24	9	4
Group B	10	20	22	14

[7]

- 15** A radio station asks 300 listeners in three age groups to choose their preferred music type, with the results shown.

	Rock	Pop	Classical
Age 18-30	45	30	25
Age 31-50	30	40	30
Age 51+	20	35	45

- (a) Test, at the 5% significance level, whether music preference is associated with age group. [7]
- (b) Identify the cell that contributes most to χ^2 and describe the pattern it shows. [3]

- 16** In a trial, 80 patients received either a treatment or a control and were assessed as shown.

	Improved	No change
Treatment	22	18
Control	13	27

- (a) Calculate the test statistic $\sum \frac{(O - E)^2}{E}$ for independence of outcome and group. [4]
- (b) For a 2×2 table Yates' continuity correction replaces $(O - E)^2$ by $(|O - E| - 0.5)^2$. Calculate the corrected statistic and test at the 5% level. [4]

- 17** A table of percentages of a sample of 200 people is used by a student to calculate a χ^2 statistic for association between two variables, treating the percentages as the observed frequencies. The result is 8.4 ($\nu = 2$).

Explain why this procedure is incorrect and state what should be done instead. If the student's percentages came from the actual frequencies by multiplying by 2, what is the correct statistic?

[5]

- 18** A survey asks 150 shoppers in three stores whether they would recommend the store (yes or no). Store 1: 34 yes, 16 no. Store 2: 27 yes, 23 no. Store 3: 16 yes, 34 no. Test, at the 5% significance level, whether the proportion recommending is the same in all three stores, and find the p-value of the result.

[7]

- 19** A survey of 190 people gives the table shown.

Find the value of x, and test, at the 5% significance level, whether opinion is independent of sex.

	Agree	Neutral	Disagree
Male	30	20	x
Female	25	35	40

[8]

- 20** The hair and eye colours of 210 people are recorded, as shown. A test is carried out for an association between hair colour and eye colour.

	Brown eyes	Blue eyes	Green eyes
Brown hair	25	12	22
Black hair	28	33	26
Fair hair	15	26	23

- (a) Carry out the test at the 5% level, stating your conclusion.
 (b) Carry out the test at the 10% level, stating your conclusion, and comment on the two conclusions.

[7]

[3]

- 21** Pupils in three year groups report how often they read for pleasure, as shown.

	Never	Rarely	Often	Daily
Year 9	12	20	25	8
Year 10	18	22	15	5
Year 11	9	14	19	22

- (a) Show that no expected frequency is below 5 and test, at the 5% significance level, for an association between year group and reading frequency.
 (b) Find the two cells with the largest contributions to χ^2 and describe what they show.

[7]

[4]

- 22** A student's work on a χ^2 test of association is shown for a 3×4 table with $N = 120$.
 'Expected frequency = row total $\times$ column total, for example $40 \times 30 = 1200$. $\nu = 3 \times 4 = 12$.
 Test statistic = $\Sigma(O - E)/E = 6.1$. Critical value 21.026. $6.1 < 21.026$ so there is a relationship and one variable causes the other.'

Identify four errors in the student's work.

[7]

- 23** The 2×2 table has first row a and $50 - a$, and second row 30 and 20 , where a is an integer with $25 \leq a \leq 49$.
- (a) Show that the row totals are 50 and 50 and that the column totals are $a + 30$ and $70 - a$. [2]
- (b) Find the smallest value of a for which a χ^2 test of independence, at the 5% significance level, finds an association. [5]

- 24** A 2×2 contingency table has observed frequencies $24, 16$ in the first row and $16, 24$ in the second. A second study finds exactly twice every frequency.
- (a) Test for independence in the first study at the 5% significance level. [4]
- (b) Show that the statistic for the second study is double that of the first, and state the conclusion for the second study. [3]
- (c) Explain why the same pattern of data leads to different conclusions. [1]

- 25** Samples of water from three sites are classified by pollution level, as shown.

	Low	Medium	High
Site A	48	22	30
Site B	30	38	32
Site C	22	40	38

- (a) Test, at the 5% significance level, whether pollution level is associated with site. [6]
- (b) The standardised residual of a cell is $(O - E)/\sqrt{E}$. Calculate the residual in each cell and state which cells have a residual of magnitude greater than 1.96 . [4]

- 26** A survey of 212 people records their age group and their weekly exercise level (low, medium or high), as shown.

	Low	Medium	High
Under 20	30	22	18
20-39	26	25	19
40-59	14	20	26
60 and over	3	4	5

- (a) Explain why the last row cannot be used as it stands in a χ^2 test. [2]
- (b) Combine the last two rows and test, at the 5% significance level, whether exercise level is associated with age group. [6]
- (c) Explain why the two rows chosen for combining are sensible. [1]

- 27** A 2×2 contingency table has observed frequencies a, b in the first row and c, d in the second, with total $N = a + b + c + d$.

- (a) Show that the expected frequency in the top-left cell is $E_{11} = \frac{(a+b)(a+c)}{N}$ and that $a - E_{11} = \frac{ad - bc}{N}$. [3]
- (b) Hence show that the test statistic is $\chi^2 = \frac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$. [4]
- (c) Use the formula to calculate χ^2 for $a = 18, b = 32, c = 36, d = 24$. [2]

- 28** Eight patients received drug A and eight received drug B. Of those given A, 7 recovered; of those given B, 3 recovered.
Let X be the number of the 10 patients who recovered who were given drug A. Assuming the drug has no effect on recovery, X has a hypergeometric distribution: $P(X = x) = \frac{C(8, x) C(8, 10 - x)}{C(16, 10)}$.
- State the null hypothesis, and find the expected frequencies for a χ^2 test. Explain why the usual χ^2 test is unreliable here. [4]
 - Use the hypergeometric distribution to find $P(X \geq 7)$ and test, at the 5% level, whether drug A is more effective. [5]
-
- 29** Two treatments, A and B, are compared in two hospitals. The numbers of patients who survived and died are shown.
- | | Treatment | Survived | Died |
|------------|-----------|----------|------|
| Hospital 1 | A | 81 | 6 |
| Hospital 1 | B | 234 | 36 |
| Hospital 2 | A | 192 | 71 |
| Hospital 2 | B | 55 | 25 |
- For each hospital, calculate the proportion of patients surviving on each treatment, and state which treatment appears better in each hospital. [4]
 - Combine the two hospitals into a single table and calculate the proportion surviving on each treatment. [3]
 - Carry out a χ^2 test of association between treatment and survival on the combined table, at the 5% significance level, and explain how the results in parts (a) and (b) can disagree. [5]
-
- 30** A 2×2 table of 100 observations has frequencies $25 + d$, $25 - d$ in the first row and $25 - d$, $25 + d$ in the second row, where d is a positive integer less than 25.
- Show that the χ^2 statistic for independence is $\frac{4d^2}{25}$. [3]
 - Find the least value of d for which the test is significant at the 5% level. [3]
 - With Yates' continuity correction, the statistic becomes $\frac{4(d - 0.5)^2}{25}$. Find the least value of d for which the corrected test is significant at 5%, and state the effect of the correction. [4]
-
- ## Tests for Population Means
-
- 1** The lengths of bolts produced by a machine are normally distributed with standard deviation 0.4 mm. The machine is set to produce bolts of mean length 25 mm. An engineer thinks that the mean length is now greater than 25 mm.
Write down suitable hypotheses for a test of the engineer's belief. [2]
-
- 2** A t-test is to be carried out on a small sample to test a hypothesis about a population mean. State the assumption about the population that is required for the test to be valid. [2]
-
- 3** A random sample of 16 observations from a normal population has mean 54.2 and unbiased variance estimate $s^2 = 12.96$. A test of $H_0: \mu = 52$ against $H_1: \mu > 52$ is carried out. Calculate the value of the test statistic, and state the degrees of freedom. [3]
-

- 4** State the degrees of freedom for each t-test.
- (a) A one-sample test using 12 observations. [1]
- (b) A paired test using 10 pairs of observations. [1]
- (c) A two-sample test with pooled variance using samples of size 8 and 11. [1]
-
- 5** Five students sit two tests, and their scores on test 1 are 12, 15, 11, 18, 14 and on test 2 are 15, 14, 15, 20, 19.
Find the difference (test 2 – test 1) for each student and calculate the mean difference. [3]
-
- 6** State, with a reason, whether a z-test or a t-test is appropriate in each case, assuming the population is normal.
- (a) $n = 9$, σ^2 known to be 25. [1]
- (b) $n = 12$, σ unknown, estimated by s from the sample. [1]
- (c) $n = 150$, σ unknown, estimated by s . [1]
-
- 7** A machine is set to fill packets with 12.0 g of powder. A sample of 8 packets has masses (in g) 12.10, 11.80, 12.60, 12.30, 11.90, 12.50, 12.20, 12.40. Assuming that the masses are normally distributed, test at the 5% significance level whether the mean mass differs from 12.0 g. [7]
-
- 8** A sample of 12 batteries has lifetimes x hours with $\sum x = 745.2$ and $\sum x^2 = 46\,321.0$. The manufacturer claims a mean lifetime of at least 60 hours. Assuming normality, test at the 5% level whether the mean lifetime is less than 60 hours. [7]
-
- 9** The IQ scores of a population are normally distributed with standard deviation 15. A sample of 50 students from a school has mean score 103.2.
- (a) Test, at the 5% significance level, whether the mean IQ score of students at the school differs from 100. [4]
- (b) Find a 95% confidence interval for the mean and explain how it agrees with your conclusion. [3]
-
- 10** A drug is thought to reduce resting heart rate. The heart rates (beats per minute) of 8 patients before and after taking the drug are shown.
Assuming the differences are normally distributed, test at the 5% significance level whether the drug reduces heart rate.
- | Patient | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---------|----|----|----|----|----|----|----|----|
| Before | 74 | 80 | 68 | 85 | 77 | 90 | 72 | 81 |
| After | 71 | 79 | 65 | 84 | 70 | 88 | 72 | 76 |
- [8]
-
- 11** A random sample of 100 observations of a variable has mean 49.1 and standard deviation $s = 6.2$ (calculated with divisor $n - 1$). Use a z-test to test at the 5% significance level whether the population mean is less than 50, and explain why this is justified without assuming normality. [5]
-
- 12** Two machines fill bags with sugar. Samples of 40 and 50 bags give mean masses of 82.4 g and 79.8 g respectively. The masses have known variances of 36 g^2 and 49 g^2 respectively. Test at the 5% significance level whether the mean masses differ. [6]
-
- 13** Samples of 10 and 12 plants are given fertiliser A and fertiliser B respectively. The heights have means 54.6 cm and 51.2 cm and unbiased standard deviations 4.1 cm and 3.7 cm. Assuming normal populations with equal variances, test at the 5% significance level whether fertiliser A gives a greater mean height. [7]
-

- 14** The thickness of a metal sheet is specified as 100 μm . Ten sheets have thicknesses (in μm) 101.8, 103.5, 99.20, 104.1, 102.7, 100.9, 103.8, 102.2, 104.6, 101.4. Assume the thickness is normally distributed.
- (a) Test, at the 1% significance level, whether the mean thickness exceeds 100 μm . [7]
- (b) Find a 95% confidence interval for the mean thickness. [3]
-
- 15** Twelve volunteers have their reaction times (milliseconds) measured before and after a training programme. For the differences $d = \text{before} - \text{after}$, $\Sigma d = 138$ and $\Sigma d^2 = 2120$. Stating your assumption, test at the 1% significance level whether the training reduces the mean reaction time, and find the p-value. [9]
-
- 16** The times (seconds) taken by two groups of athletes to complete a course are: Group A: 23.10, 25.40, 22.80, 24.90, 26.10, 23.70, 25.00; Group B: 21.90, 22.80, 23.50, 21.20, 22.40, 23.00, 21.70, 22.60. Assuming normal populations with equal variances, test at the 5% significance level whether the mean times differ. [9]
-
- 17** Samples of size 15 and 18 from two normal populations with equal variances have means 68.4 and 63.9 and unbiased standard deviations 7.2 and 6.5 respectively.
- (a) Find a 95% confidence interval for the difference between the population means $\mu_1 - \mu_2$. [6]
- (b) Use the interval to state the conclusion of a two-tailed test at the 5% level of $H_0: \mu_1 = \mu_2$, justifying your answer. [2]
-
- 18** A one-tailed t-test of $H_0: \mu = 40$ against $H_1: \mu < 40$ is to be carried out at the 5% significance level using a sample of 10 observations. For the sample, $s = 2.5$. Find the range of values of the sample mean $\bar{x}$ for which H_0 would be rejected. [4]
-
- 19** A random sample from a normal population has mean 23.0 and unbiased standard deviation 6.0. A test of $H_0: \mu = 21$ against $H_1: \mu > 21$ is carried out at the 5% significance level.
- (a) Carry out the test when the sample size is 9. [4]
- (b) Carry out the test when the sample size is 36, with the same mean and standard deviation, and comment. [5]
-
- 20** The waiting time, in minutes, at an accident and emergency department has an unknown distribution with mean μ . A sample of 60 patients has mean waiting time 148.5 minutes and standard deviation $s = 21.0$ minutes. The hospital target is a mean of 140 minutes.
- (a) State, with a reason, why a test of $H_0: \mu = 140$ against $H_1: \mu > 140$ can be carried out using a normal distribution. [1]
- (b) Carry out the test at the 1% significance level. [4]
-
- 21** Eight athletes run a timed test before (set 1) and after (set 2) a training programme. The times (seconds) are before: 52.00, 48.00, 55.00, 60.00, 47.00, 53.00, 58.00, 50.00; after: 50.00, 47.00, 52.00, 56.00, 46.00, 49.00, 55.00, 48.00.
- (a) Carry out a paired t-test at the 5% significance level, two-tailed, for a change in mean time. [6]
- (b) A colleague analyses the same data with a two-sample pooled t-test. Carry this out and compare the conclusions. [4]
- (c) Explain why the paired test is the more appropriate and why it gives a larger test statistic. [2]

- 22** A student's test of whether the mean of a normal population is greater than 50 uses a sample of 9 observations: 53, 49, 57, 52, 55, 48, 54, 56, 51.
Student: 'Mean = 52.78. Sum of squares of deviations = 75.56, so variance = $75.56 \div 9 = 8.395$, $sd = 2.897$. $z = (52.78 - 50)/2.897 = 0.959 < 1.645$ so do not reject H_0 .'
Identify three errors in the student's method and carry out the test correctly at the 5% significance level. [9]
-
- 23** Two teaching programmes are compared. Group 1 (14 students) has mean score 142.6 and unbiased standard deviation 11.3. Group 2 (20 students) has mean score 135.1 and unbiased standard deviation 12.4. Assume normal populations with a common variance.
(a) Test, at the 5% significance level, for a difference in mean score. [6]
(b) Find a 99% confidence interval for the difference in means. [4]
(c) Explain what the interval shows about a test of $H_0: \mu_1 = \mu_2$ at the 1% significance level. [2]
-
- 24** A population is normally distributed with standard deviation 12. A researcher takes a random sample of size n and tests $H_0: \mu = 100$ against $H_1: \mu \neq 100$ at the 5% significance level. The sample mean turns out to be 104.
Find the smallest value of n for which this result leads to rejection of H_0 . [6]
-
- 25** A sample of 8 heights (in cm) from a normal population is 172.4, 168.9, 175.1, 170.3, 173.8, 169.5, 171.9, 174.2. A t-test of $H_0: \mu = 170$ against $H_1: \mu > 170$ is carried out.
(a) Calculate the test statistic. [4]
(b) The heights are converted to millimetres. Calculate the test statistic for $H_0: \mu = 1700$ and show that it is unchanged. [3]
(c) Prove that the t statistic is unchanged when every observation is multiplied by a positive constant k and the hypothesised mean is also multiplied by k . [3]
-
- 26** Two types of tyre are compared. The lifetimes of tyres of type A and type B are normally distributed with known standard deviations 3 000 km and 4 000 km respectively. A sample of 25 tyres of each type is tested. $H_0: \mu_A = \mu_B$ is tested against $H_1: \mu_A > \mu_B$ at the 5% significance level.
(a) Find the critical value for $\bar{X}_A - \bar{X}_B$. [4]
(b) Find the probability of a Type II error when $\mu_A - \mu_B = 2\,500$ km. [3]
-
- 27** Independent random samples of sizes n_1 and n_2 are taken from populations with the same variance σ^2 . Their unbiased variance estimates are s_1^2 and s_2^2 , and the pooled estimate is $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$.
(a) Prove that s_p^2 is an unbiased estimator of σ^2 . [4]
(b) Show that the variance of $\bar{X}_1 - \bar{X}_2$ is $\sigma^2(\frac{1}{n_1} + \frac{1}{n_2})$. [3]
(c) Samples of size 6 and 9 have $s_1^2 = 4.8$ and $s_2^2 = 7.2$. Calculate s_p^2 , and the standard error of the difference in sample means. [4]

- 28** Samples from two normal populations give: sample 1, $n_1 = 9$, mean 47.3, $s_1 = 3.1$; sample 2, $n_2 = 15$, mean 41.8, $s_2 = 7.4$. The sample standard deviations suggest the population variances may be unequal. When the variances cannot be assumed equal, the test statistic is $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ with degrees of freedom $\nu = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(\frac{s_1^2}{n_1})^2}{n_1 - 1} + \frac{(\frac{s_2^2}{n_2})^2}{n_2 - 1}}$, rounded down.
- (a) Calculate t and ν . [5]
- (b) Test at the 5% significance level, two-tailed, whether the means differ. [2]
- (c) Calculate the pooled-variance t statistic for the same data, state its degrees of freedom, and comment on why it differs from your answer to part (a). [4]

- 29** A population of measurements is normal with standard deviation $\sigma = 10$. A researcher wants to detect an improvement of $\delta = 4$ using a one-tailed test at the 5% significance level, with a power of 90%. Use the z -based formula $n = \frac{(z_\alpha + z_\beta)^2 \text{Var}}{\delta^2}$, where Var is the variance of the quantity whose mean is tested, $z_\alpha = 1.645$ and $z_\beta = 1.2816$.
- (a) In an independent-samples design with n observations in each group, the variance of $\bar{X}_1 - \bar{X}_2$ is $\frac{2\sigma^2}{n}$. Show that $n = \frac{2\sigma^2(z_\alpha + z_\beta)^2}{\delta^2}$ and find the smallest integer n . [5]
- (b) In a paired design the correlation between the two measurements on each pair is $\rho = 0.6$. Show that $\text{Var}(D) = 2\sigma^2(1 - \rho)$ for each difference D and find the smallest number of pairs needed. [5]
- (c) Compare the total number of observations in the two designs and explain the advantage of pairing. [2]

- 30** A paired experiment on 10 subjects gives the differences d (treatment – control): 1.200, 0.4000, 2.100, –0.3000, 1.800, 0.9000, 1.500, 0.6000, 2.400, 0.2000. Assume the differences are normally distributed with mean μ_d .
- (a) Test $H_0: \mu_d = 0$ against $H_1: \mu_d \neq 0$ at the 5% significance level. [6]
- (b) Find a 95% confidence interval for μ_d . [3]
- (c) Prove that, for a two-tailed test at the 5% level, $H_0: \mu_d = 0$ is rejected if and only if 0 lies outside the 95% confidence interval. [3]

Tests for Population Variances

- 1** The masses of jars of jam are normally distributed. A producer claims that the variance of the mass is 4 g^2 , but an inspector believes it is greater. Write down suitable hypotheses for a test of the inspector's belief. [2]
- 2** A random sample of n observations is taken from a normal population with variance σ^2 , and the sample has unbiased variance estimate s^2 .
- (a) State the distribution of $\frac{(n-1)s^2}{\sigma^2}$. [1]
- (b) State the degrees of freedom when $n = 15$. [1]
- 3** A sample of 20 observations from a normal population has $s^2 = 2.4$. A test of $H_0: \sigma^2 = 2$ against $H_1: \sigma^2 > 2$ is carried out at the 5% significance level. Calculate the test statistic and state the conclusion. [3]

- 4** State the assumption about the population that must be made for a χ^2 test of a population variance to be valid. [2]
-
- 5** Independent samples of sizes 11 and 9 from normal populations have unbiased variance estimates 12.5 and 5.0 respectively. Calculate the variance ratio F, with the larger variance in the numerator, and write down its degrees of freedom. [3]
-
- 6** Use tables or a calculator to write down the upper 5% critical value of the F distribution with (8, 10) degrees of freedom, and of the χ^2 distribution with 12 degrees of freedom. [2]
-
- 7** The lengths of rods made by a machine are normally distributed. The machine is designed to produce a variance of 0.0064 cm^2 . A sample of 25 rods has unbiased variance estimate 0.0098 cm^2 .
Test, at the 5% significance level, whether the variance has increased. [5]
-
- 8** A new method of measuring blood glucose is claimed to have a variance of $2.0 (\text{mmol/l})^2$. A sample of 16 measurements on the same sample gives $s^2 = 1.1$. Test at the 5% level whether the new method is more precise (has smaller variance), assuming normality. [5]
-
- 9** A sample of 8 observations from a normal population is 8.200, 9.100, 7.600, 10.40, 8.800, 9.700, 7.900, 9.300. Test, at the 5% significance level, whether the population variance differs from 1.0. [7]
-
- 10** A random sample of 20 observations from a normal population has $s^2 = 6.5$. Find a 95% confidence interval for the population variance. [5]
-
- 11** Independent random samples of sizes 10 and 12 from two normal populations have unbiased variance estimates 9.6 and 4.2 respectively. Test at the 5% significance level whether the variance of population 1 is greater than that of population 2. [5]
-
- 12** Samples of sizes 9 and 13 from two normal populations have unbiased variance estimates 3.8 and 11.6 respectively. Test at the 5% significance level whether the population variances are equal. [5]
-
- 13** For a sample of 15 observations from a normal population $\Sigma x = 213.0$ and $\Sigma x^2 = 3105.0$. Test at the 5% significance level whether the population variance exceeds 2.5. [7]
-
- 14** A supplier claims that the standard deviation of the thickness of its panels is 2.0 mm. A sample of 12 panels has unbiased variance estimate $s^2 = 4.4 \text{ mm}^2$. Assume the thickness is normally distributed.
(a) Test at the 5% significance level whether the standard deviation is greater than the supplier claims. [5]
(b) Find a 95% confidence interval for σ . [3]
-
- 15** The volumes (ml) of ten drinks dispensed by a machine are 20.40, 19.80, 20.90, 19.10, 21.20, 20.00, 18.90, 20.60, 21.50, 19.50. The machine is said to have a variance of 0.5 ml^2 . Assuming normality, test at the 5% significance level whether the variance differs from 0.5, and find the approximate two-tailed p-value. [8]
-
- 16** Two machines produce components whose lengths (mm) are normally distributed. Samples from machine A and machine B give: A: 15.20, 14.80, 16.10, 15.50, 13.90, 16.80, 15.00, 14.20; B: 14.90, 15.10, 15.40, 14.70, 15.30, 15.00, 15.20, 14.80, 15.10.
Test at the 5% significance level whether the variance of lengths from machine A is greater than that from machine B. [7]
-

- 17** Independent samples of sizes 10 and 12 from normal populations have means 41.5 and 38.2 and unbiased variance estimates 6.8 and 5.2 respectively. A researcher wishes to test whether the population means differ, using a pooled-variance t-test.
- (a) Test at the 10% significance level (two-tailed) whether the population variances can be assumed equal. [4]
- (b) Hence carry out the t-test at the 5% significance level for a difference in means. [6]
-
- 18** A random sample of 18 observations is taken from a normal population with variance σ^2 . A test of $H_0: \sigma^2 = 0.5$ against $H_1: \sigma^2 > 0.5$ is carried out at the 5% significance level. Find the critical region for the sample variance s^2 . [4]
-
- 19** Two independent samples of sizes 8 and 11 are taken from normal populations with the same unknown variance σ^2 . Their unbiased variance estimates are 2.6 and 3.4. Use a pooled estimate to test, at the 5% significance level, $H_0: \sigma^2 = 2.0$ against $H_1: \sigma^2 > 2.0$. [6]
-
- 20** A set of 14 readings from a normal population has $\Sigma x = 392.0$ and $\Sigma x^2 = 11\,401.0$. A manufacturer states that the variance is 4.0. Test this claim at the 5% significance level against the alternative that the variance differs from 4.0. [7]
-
- 21** A test of $H_0: \sigma^2 = 5.0$ against $H_1: \sigma^2 > 5.0$ is carried out on a sample of 20 observations from a normal population, at the 5% significance level.
- (a) Show that H_0 is rejected if $s^2 > 7.933$. [3]
- (b) Find the power of the test when the true variance is 10.0. [4]
- (c) State, with a reason, how the power would change if the sample size were increased. [1]
-
- 22** A student's solution to a test of whether the standard deviation of a normal population exceeds 3 is shown. A sample of 13 observations has $s = 4.1$.
' $H_0: \sigma = 3$. $\chi^2 = (13 \times 4.1)/3 = 17.77$. $\nu = 13$. Critical value = 22.362. $17.77 < 22.362$ so there is no evidence that $\sigma > 3$.'
Identify three errors and carry out the test correctly at the 5% significance level. [6]
-
- 23** A random sample of 25 observations from a normal population has unbiased variance estimate 3.8.
- (a) Find a one-sided 95% confidence interval of the form $\sigma^2 < k$. [3]
- (b) Use your interval to say whether $H_0: \sigma^2 = 6.0$ would be rejected in favour of $H_1: \sigma^2 < 6.0$ at the 5% level, then confirm by carrying out the test. [4]
-
- 24** For large ν , if $X \sim \chi^2_\nu$ then $\sqrt{2X} - \sqrt{2\nu - 1}$ is approximately $N(0, 1)$. A sample of 100 observations from a normal population has $s^2 = 1.45$. A test of $H_0: \sigma^2 = 1$ against $H_1: \sigma^2 > 1$ is carried out.
- (a) Calculate the test statistic χ^2 and the approximate value of $z = \sqrt{2\chi^2} - \sqrt{2\nu - 1}$. [4]
- (b) Test at the 5% significance level using the normal approximation. [2]
- (c) The exact p-value from the χ^2_{99} distribution is 0.0023. Compare it with the p-value from the approximation. [3]
-
- 25** Samples of sizes 8 and 12 from normal populations have unbiased variance estimates $s_1^2 = 2.1$ and $s_2^2 = 6.3$. A test of $H_0: \sigma_1^2 = \sigma_2^2$ against $H_1: \sigma_1^2 < \sigma_2^2$ is carried out at the 5% significance level.
- (a) Explain why the lower-tail critical value of $F_{7,11}$ equals $\frac{1}{F_{11,7}}$ where $F_{11,7}$ is the upper-tail value. [2]
- (b) Find the lower-tail critical value of $F_{7,11}$. [3]
- (c) Carry out the test. [3]

- 26** A sample of size n from a normal population has $s^2 = 1.5\sigma_0^2$, where σ_0^2 is the hypothesised variance. A test of $H_0: \sigma^2 = \sigma_0^2$ against $H_1: \sigma^2 > \sigma_0^2$ is carried out at the 5% significance level.
Find the smallest sample size n for which H_0 would be rejected. [6]
-
- 27** Let s^2 be the unbiased variance estimate from a random sample of size n from $N(\mu, \sigma^2)$, so that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$. You may use the results that a χ^2_ν variable has mean ν and variance 2ν .
- (a) Show that s^2 is an unbiased estimator of σ^2 . [2]
- (b) Show that $\text{Var}(s^2) = \frac{2\sigma^4}{n-1}$. [3]
- (c) Hence find the standard deviation of s^2 as a fraction of σ^2 for $n = 51$, and find how large n must be for it to be below $0.1\sigma^2$. [5]
-
- 28** A 95% confidence interval for σ^2 from a sample of size n is $(\frac{(n-1)s^2}{a}, \frac{(n-1)s^2}{b})$ where a and b are the upper and lower 2.5% points of χ^2_{n-1} .
- (a) Show that the ratio of the upper limit to the lower limit is $\frac{a}{b}$, independent of s^2 . [2]
- (b) For $n = 10$ find this ratio, and comment. [4]
- (c) Find the smallest n such that the upper limit is at most twice the lower limit. [5]
-
- 29** Independent samples of sizes 12 and 10 from normal populations have unbiased variance estimates $s_1^2 = 7.5$ and $s_2^2 = 2.1$. It is claimed that $\sigma_1^2 = 2\sigma_2^2$. You may use the result that $\frac{s_1^2}{\sigma_1^2} \div \frac{s_2^2}{\sigma_2^2} \sim F_{n_1-1, n_2-1}$.
- (a) Show that, if the claim is true, $\frac{s_1^2}{2s_2^2} \sim F_{11,9}$. [3]
- (b) Test, at the 5% significance level, $H_0: \sigma_1^2 = 2\sigma_2^2$ against $H_1: \sigma_1^2 > 2\sigma_2^2$. [4]
- (c) Find a 95% confidence interval for the ratio $\frac{\sigma_1^2}{\sigma_2^2}$. [4]
-
- 30** Independent samples of sizes 10 and 12 are taken from normal populations with variances σ_1^2 and σ_2^2 . A test of $H_0: \sigma_1^2 = \sigma_2^2$ against $H_1: \sigma_1^2 > \sigma_2^2$ is carried out at the 5% significance level using $F = \frac{s_1^2}{s_2^2} \sim F_{9,11}$ under H_0 .
- (a) State the critical value, and show that when $\sigma_1^2 = r\sigma_2^2$, the quantity $\frac{F}{r}$ has an $F_{9,11}$ distribution. [3]
- (b) Find the power of the test when $r = 3$. [3]
- (c) Find, to 1 decimal place, the value of r for which the power is 0.9. [4]

Non-Parametric Tests

- 1** Non-parametric tests are sometimes used instead of parametric tests.
- (a) State two advantages of non-parametric tests. [2]
- (b) State one disadvantage. [1]
-
- 2** A sign test is to be used to test whether the median time to complete a task is greater than 30 minutes.
Write down suitable hypotheses, where m is the population median. [2]

3 Ten observations are 52, 47, 55, 50, 58, 49, 61, 50, 46, 57. For a sign test of $H_0: m = 50$, find the number of positive signs, the number of negative signs, and the number of observations that must be discarded. [2]

4 Rank the values 14, 9, 21, 9, 17, 5 from smallest (rank 1) to largest, giving tied values the average of the ranks they occupy. [3]

5 In a Wilcoxon signed-rank test with 8 non-zero differences, the sum of the ranks of the positive differences is 26. Find the sum of the ranks of the negative differences. [2]

6 In a Mann-Whitney test, sample A has 6 observations and sample B has 7. The sum of the ranks of sample A is 33. Calculate $U_A = R_A - \frac{m(m+1)}{2}$ and U_B , and check them. [3]

7 A claim is made that the median daily intake of a nutrient is 15.0 g. A nutritionist believes it is higher, and measures 12 people, obtaining 14.2, 15.8, 13.1, 16.4, 15, 17.2, 14.9, 16.1, 15.5, 18, 15, 16.9. Carry out a sign test at the 5% significance level. [6]

8 Ten students sit a test before and after a revision course. Their scores are shown. Use a sign test, at the 5% significance level, to test whether the course affects the scores (two-tailed).

Student	1	2	3	4	5	6	7	8	9	10
Before	62	70	58	66	74	61	69	72	65	68
After	66	74	57	71	79	66	70	77	64	75

[6]

9 A sample of nine observations is 38, 52, 44, 61, 47, 55, 42, 58, 49. Use a Wilcoxon signed-rank test, at the 5% significance level, to test $H_0: m = 45$ against $H_1: m > 45$, where m is the population median. [6]

10 The reaction times (s) of 8 people before and after drinking caffeine are before: 8.2, 7.9, 9.4, 8.8, 7.5, 9, 8.4, 8.9; after: 7.6, 7.7, 8.1, 8.2, 7.6, 8.3, 8, 8.1. Use a Wilcoxon signed-rank test, at the 5% significance level, to test whether caffeine changes the median reaction time (two-tailed). [6]

11 Samples of battery life (hours) from two brands are A: 12.4, 14.1, 11.8, 13, 12.9 and B: 14.8, 13.6, 15.2, 14, 15.9, 13.3. Use a Mann-Whitney test, at the 5% significance level, to test whether brand A tends to have shorter lives than brand B. [6]

12 A sign test is carried out on 40 non-zero differences, of which 28 are positive, to test H_0 : median difference = 0 against H_1 : median difference > 0. Use a normal approximation with a continuity correction to test at the 5% level. [4]

13 A researcher has 9 paired measurements. The differences are strongly skewed with one very large value.
(a) Give a reason why a paired t-test may be unsuitable. [1]
(b) State which non-parametric test could be used instead, and why the Wilcoxon signed-rank test is preferred to the sign test if the differences are symmetric. [2]

14 Eleven observations are 31, 36, 28, 40, 36, 33, 25, 38, 36, 29, 42. Use a Wilcoxon signed-rank test, at the 5% significance level, to test $H_0: m = 33$ against $H_1: m \neq 33$, explaining how you deal with any zero differences and tied ranks. [8]

15 Two groups of workers complete a task in the times (minutes) A: 23, 29, 31, 27, 35, 29, 26 and B: 34, 38, 29, 41, 36, 32, 37, 30. Use a Mann-Whitney test at the 5% significance level to test whether the distributions of times differ (two-tailed). [7]

- 16** In a Wilcoxon signed-rank test with 25 non-zero differences, the sum of the ranks of the negative differences is $T = 82$. For large n , T is approximately normal with mean $\frac{n(n+1)}{4}$ and variance $\frac{n(n+1)(2n+1)}{24}$.
Test at the 5% level whether the population median is greater than the hypothesised value (so that a small T is significant), using a continuity correction. [5]
-
- 17** In a Mann-Whitney test with samples of sizes 12 and 15, $U = 52$ (small U supports H_1 : sample 1 tends to be smaller). For large samples U is approximately normal with mean $\frac{mn}{2}$ and variance $\frac{mn(m+n+1)}{12}$.
Use a normal approximation with a continuity correction to carry out the test at the 5% significance level. [5]
-
- 18** Eight measurements are 4.1, 5.3, 3.8, 6, 4.9, 5.6, 4.4, 5.1. A researcher tests $H_0: m = 4.5$ against $H_1: m > 4.5$ at the 5% significance level.
(a) Carry out a sign test. [4]
(b) Carry out a Wilcoxon signed-rank test and compare the conclusions. [5]
-
- 19** A Wilcoxon signed-rank test is based on $n = 4$ non-zero differences. Under H_0 , each rank 1, 2, 3, 4 is independently positive or negative with probability $\frac{1}{2}$. Let T_+ be the sum of the positive ranks.
(a) List the possible sets of positive ranks that give $T_+ \leq 2$ and find $P(T_+ \leq 2)$. [4]
(b) Find the critical region for T_+ for a one-tailed test at the 10% level and the actual significance level. [4]
-
- 20** A sign test of $H_0: m = m_0$ against $H_1: m \neq m_0$ is carried out on 15 observations, none of which equals m_0 . Let S be the number of observations above m_0 .
Find the critical region at the 5% significance level (at most 2.5% in each tail) and the actual significance level, and state the conclusion if $S = 12$. [6]
-
- 21** A student carries out a Mann-Whitney test of H_1 : sample A tends to be smaller than sample B, using sample A: 12, 15, 9, 20 and sample B: 17, 22, 14, 25, 19, at the 5% significance level.
Student: 'Ranks of A: 1, 2, 4, 7, so $R_A = 14$. $U_A = R_A - \frac{n(n+1)}{2} = 14 - \frac{5 \times 6}{2} = -1$. The critical value is 2 and $-1 \leq 2$, so reject H_0 .'
Explain what is wrong with the student's calculation of U_A , and carry out the test correctly. [7]
-
- 22** Three methods of teaching are compared by test scores: method 1: 18, 22, 25, 19, 21; method 2: 24, 27, 30, 26, 29, 28; method 3: 20, 23, 17, 16, 24.5. The Kruskal-Wallis statistic is $H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N+1)$, which is approximately χ^2_{k-1} under H_0 .
Test, at the 5% significance level, whether the distributions of scores differ. [7]
-
- 23** Under H_0 in a Mann-Whitney test with sample sizes $m = 2$ and $n = 3$, all $C(5, 2) = 10$ ways of placing the two A observations among the 5 ranked values are equally likely. Let U_A be the number of pairs (a, b) with the B observation smaller than the A observation (so $U_A = R_A - \frac{m(m+1)}{2}$).
(a) Complete the list of the 10 arrangements (written in rank order) and the value of U_A for each. [3]
(b) Hence find the distribution of U_A , and show that $E(U_A) = \frac{mn}{2}$. [4]
(c) Find the critical region for a one-tailed test with significance level at most 10%, and state the actual significance level. [3]

- 24** A random sample of 12 observations from a continuous population with median m is arranged in order: $x_{(1)} < x_{(2)} < \dots < x_{(12)}$. The values are 18.9, 19.6, 20.8, 21.4, 22.1, 22.7, 23.1, 24, 25.8, 26.5, 27.3, 28.4.
- Explain why $P(x_{(k)} < m < x_{(13-k)}) = 1 - 2P(S \leq k - 1)$ where $S \sim B(12, 0.5)$. [3]
 - Find the largest k for which the interval $(x_{(k)}, x_{(13-k)})$ is a confidence interval for m with confidence at least 95%, and state the actual confidence level. [4]
 - Write down the interval for the given data. [2]
-
- 25** In a Wilcoxon signed-rank test, $n = 7$ non-zero differences give $T_+ = 5$ (a small value supports $H_1: m < m_0$).
- Explain why there are 128 equally likely sign patterns under H_0 . [2]
 - Show that the number of patterns with $T_+ \leq 5$ is 10 and hence find the exact p-value. [4]
 - Hence state the conclusion at the 5% level and give the critical value that would appear in tables. [2]
-
- 26** In a Mann-Whitney test with sample sizes $m = 4$ and $n = 5$, $U_A = R_A - 10$ has been calculated as 3 (small values support $H_1: A$ tends to be smaller than B).
- Show that there are 126 equally likely arrangements of the ranks under H_0 . [2]
 - Given that the numbers of arrangements giving $U = 0, 1, 2, 3$ are 1, 1, 2, 3 respectively, find the exact p-value. [2]
 - State the conclusion at the 5% level, and explain why a normal approximation would not be suitable here. [3]
-
- 27** In a Wilcoxon signed-rank test with n non-zero differences, $T_+ = \sum_{k=1}^n kI_k$, where $I_k = 1$ if the difference with rank k is positive and 0 otherwise. Under H_0 the I_k are independent with $P(I_k = 1) = \frac{1}{2}$.
- Show that $E(T_+) = \frac{n(n+1)}{4}$. [2]
 - Show that $\text{Var}(T_+) = \frac{n(n+1)(2n+1)}{24}$. [3]
 - For $n = 10$ find the mean and standard deviation of T_+ , and use a normal approximation to estimate the critical value c with $P(T_+ \leq c) \approx 0.05$, comparing it with the exact value from tables. [5]
-
- 28** Under H_0 in a Mann-Whitney test with sample sizes $m = 3$ and $n = 4$, all $C(7, 3) = 35$ arrangements of the ranks are equally likely. Let U be the statistic $R_A - 6$ for the sample of size 3.
- Show that $P(U = 0) = \frac{1}{35}$ and $P(U = 1) = \frac{1}{35}$, and find $P(U = 2)$ and $P(U = 3)$. [4]
 - The complete distribution of U is symmetric about 6. Hence find $P(U \leq 2)$ and $P(U \leq 3)$, and the critical region for a one-tailed test at the 5% significance level. [5]
 - Explain why $E(U) = 6$ and why it is not possible to reach a significance level of 1% with these sample sizes. [3]
-
- 29** Observations are from $N(\mu, \sigma^2)$. To test $H_0: \text{median} = 0$ against $H_1: \text{median} > 0$ at about the 5% level, a sign test is used on a sample of 20 observations.
- Find the critical region for the number S of positive observations, and the actual significance level. [3]
 - Given that in fact $\mu = 0.5\sigma$, show that each observation is positive with probability 0.6915, and find the power of the sign test. [4]
 - If σ were known, a z-test of $H_0: \mu = 0$ at the 5% level could be used. Find its power when $\mu = 0.5\sigma$, and comment on the comparison. [4]

30

Mood's median test compares two samples without assuming normality: the combined sample median is found, and the number of observations in each sample above and below this median is put in a 2×2 table (observations equal to the median are discarded) and a χ^2 test of association is carried out. Sample A: 48, 52, 55, 49, 61, 44, 57, 53, 47, 50. Sample B: 58, 63, 51, 66, 59, 62, 56, 68, 60, 64, 54, 65.

	Sample A	Sample B
Above median	2	9
Below median	8	3

- (a) Show that the combined median is 56.5, and construct the 2×2 table of counts above and below it. [3]
- (b) Find the expected frequencies and the χ^2 statistic, and test at the 5% level whether the two samples come from populations with the same median. [6]
- (c) Apply Yates' correction (replace $(O - E)^2$ by $(|O - E| - 0.5)^2$) and comment on whether the conclusion changes. [3]