2.4 Further Statistics: Hypothesis Testing, Type I and Type II Errors

Total Marks

AQA 7367 · Calculator allowed · about 220 minutes · Write-in edition

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

Hypothesis Testing for a Poisson or Binomial Parameter

Revision Library

For a hypothesis test on a Poisson mean lambda or a binomial probability p, a null hypothesis H0 states a fixed value of the parameter, and an alternative hypothesis H1 states that the parameter has increased, has decreased (a one-tailed test), or is simply different (a two-tailed test). A test statistic X, assumed to follow the distribution stated in H0, is compared against a critical region: the set of values of X for which H0 is rejected in favour of H1.

Because X is discrete, the probability of X falling in a proposed critical region can rarely be made to equal a chosen significance level exactly. The usual convention is to choose the largest critical region whose probability does not exceed the significance level requested. This probability is called the actual significance level of the test, and it is also equal to P(Type I error), the probability of wrongly rejecting a true H0.

A Type II error occurs when H0 is not rejected even though H1 is actually true. Its probability depends on the true value of the parameter, and is found by calculating the probability that X falls outside the critical region, using the true (alternative) value of the parameter in place of the value stated in H0. The power of a test is defined as 1 - P(Type II error): the probability of correctly rejecting a false H0. Power increases as the true parameter value moves further from the value stated in H0, and is also affected by the significance level chosen and the sample size used.

1 A market-research company uses a 12-sided spinner which should show each of the numbers 1 to 12 with equal probability. To check whether the spinner is biased towards even numbers, an assistant spins it 12 times and records X, the number of times an even number occurs. If the spinner is fair, X ~ B(12, 0.5). In one trial, an even number occurs on 9 of the 12 spins.

(a) Calculate the two-tailed p-value associated with this result. **(4)**

(b) State the conclusion of the test at the 5% significance level, with a reason. **(1)**

(Total for Question 1 is 5 marks)

2 A garden centre in Exeter finds that, historically, 30% of tomato seedlings supplied by a particular grower are affected by blight in damp weather. A random sample of 20 seedlings is taken, and X is the number affected. A new fungicide is applied to the grower's stock, and the centre wants to test, at the 5% significance level, whether the fungicide has reduced the proportion affected below 0.3.

(a) State suitable hypotheses, then find the critical region for the test and state the actual significance level. **(4)**

(b) After treatment, 1 of the 20 sampled seedlings is found to be affected by blight. State, with a reason, the conclusion of the test. **(2)**

(c) Suppose that the fungicide has in fact reduced the true proportion of seedlings affected to 0.15. Find the probability that the test results in a Type II error. **(3)**

(d) Hence state the power of the test in this case. **(1)**

(e) State, without further calculation, what would happen to the probability of a Type II error found in part (c) if the significance level of the test were increased from 5% to 10%. Justify your answer. **(2)**

(Total for Question 2 is 12 marks)

3 The number of typing errors per page of a large legal document is modelled, under normal conditions, by $X \sim Po(10)$. A proofreading test uses the critical region $X \leq 3$ or $X \geq 18$ to test whether the mean number of errors per page differs from 10.

(a) Find the probability of a Type I error for this test. **(5)**

(Total for Question 3 is 5 marks)

4 Referring again to the garden centre's fungicide trial in Question 2, where X ~ B(20, 0.3) under H0, suppose instead that the garden centre wants the actual significance level of the one-tailed (decrease) test to not exceed 1%.

(a) Find the resulting critical region, and state the actual significance level. **(3)**

(b) Explain briefly why this stricter test is less likely to detect a genuine reduction in the proportion of affected seedlings than the 5% test used in Question 2(a). **(2)**

(Total for Question 4 is 5 marks)

5 A textile mill in Bradford inspects rolls of fabric for flaws. Under normal production, the number of flaws found on a 20m roll is modelled by $X \sim Po(8)$. Following a change of yarn supplier, the mill wants to test, at the 5% significance level, whether the mean number of flaws per roll has decreased.

(a) State suitable hypotheses, then find the critical region for the test and state the actual significance level. **(4)**

(b) Suppose that, in fact, the new yarn has reduced the true mean number of flaws per roll to 5. Find the probability of a Type II error. **(3)**

(c) Hence state the power of the test in this case. **(1)**

(d) The mill's quality manager wants the power of the test to be at least 0.5 when the true mean is 5, without inspecting more than one roll. Suggest one change to the test that could achieve this, and state one disadvantage of your suggestion. **(2)**

(Total for Question 5 is 10 marks)

6 The number of faults in a length of fabric is modelled by a Poisson distribution with mean λ. A test of H0: $\lambda = 5$ against H1: $\lambda > 5$ is carried out at the 5% significance level, using a single observation X of the number of events.

(a) Find the critical region for the test. **(3)**

(b) State the probability of a Type I error for this test, to 3 significant figures. **(2)**

(c) Given that the true value of λ is 6, find the probability of a Type II error, to 3 significant figures. **(2)**

(Total for Question 6 is 7 marks)

7 The random variable X ~ B(25, p). A test of H0: p = 0.25 against H1: p < 0.25 is carried out at the 10% significance level.

(a) Find the critical region for the test. **(2)**

(b) State the probability of a Type I error, to 3 significant figures. **(2)**

(c) Find the probability of a Type II error when the true value of p is 0.1, to 3 significant figures. **(2)**

(Total for Question 7 is 6 marks)

8 A fair coin-like device is claimed to land on 'success' with probability 0.15. A tester believes the probability of success is actually lower. Let X be the number of trials needed to get the first success, so under the claim X ~ Geo(0.15). The first success occurs on trial 16.

(a) State suitable hypotheses for the test, in terms of p. **(2)**

(b) Find P(X ≥ 16) assuming H0 is true, and test the tester's belief at the 10% level. **(3)**

(c) Find the critical region for the test at the 10% level, and the probability of a Type II error if the true value of p is 0.08, to 3 significant figures. **(4)**

(Total for Question 8 is 9 marks)

9 The number of customers arriving per minute is modelled by a Poisson distribution with mean λ. A test of H0: $\lambda = 8$ against H1: $\lambda < 8$ is carried out at the 10% significance level, using a single observation X of the number of events.

(a) Find the critical region for the test. **(3)**

(b) State the probability of a Type I error for this test, to 3 significant figures. **(2)**

(c) Given that the true value of λ is 3, find the probability of a Type II error, to 3 significant figures. **(2)**

(Total for Question 9 is 7 marks)

10 The random variable X ~ B(12, p). A test of H0: p = 0.2 against H1: p > 0.2 is carried out at the 10% significance level.

(a) Find the critical region for the test. **(2)**

(b) State the probability of a Type I error, to 3 significant figures. **(2)**

(c) Find the probability of a Type II error when the true value of p is 0.7, to 3 significant figures. **(2)**

(Total for Question 10 is 6 marks)

11 The number of customers arriving per minute is modelled by a Poisson distribution with mean λ. A test of H0: $\lambda = 4$ against H1: $\lambda > 4$ is carried out at the 10% significance level, using a single observation X of the number of events.

(a) Find the critical region for the test. **(3)**

(b) State the probability of a Type I error for this test, to 3 significant figures. **(2)**

(c) Given that the true value of λ is 10, find the probability of a Type II error, to 3 significant figures. **(2)**

(Total for Question 11 is 7 marks)

12 A fair coin-like device is claimed to land on 'success' with probability 0.25. A tester believes the probability of success is actually lower. Let X be the number of trials needed to get the first success, so under the claim X ~ Geo(0.25). The first success occurs on trial 11.

(a) State suitable hypotheses for the test, in terms of p. **(2)**

(b) Find $P(X \geq 11)$ assuming H0 is true, and test the tester's belief at the 5% level. **(3)**

(c) Find the critical region for the test at the 5% level, and the probability of a Type II error if the true value of p is 0.08, to 3 significant figures. **(4)**

(Total for Question 12 is 9 marks)

13 The number of accidents per month on a stretch of road is modelled by a Poisson distribution with mean λ. A test of H0: $\lambda = 2$ against H1: $\lambda > 2$ is carried out at the 5% significance level, using a single observation X of the number of events.

(a) Find the critical region for the test. **(3)**

(b) State the probability of a Type I error for this test, to 3 significant figures. **(2)**

(c) Given that the true value of λ is 6, find the probability of a Type II error, to 3 significant figures. **(2)**

(Total for Question 13 is 7 marks)

14 The random variable X ~ B(10, p). A test of H0: p = 0.4 against H1: p < 0.4 is carried out at the 5% significance level.

(a) Find the critical region for the test. **(2)**

(b) State the probability of a Type I error, to 3 significant figures. **(2)**

(c) Find the probability of a Type II error when the true value of p is 0.15, to 3 significant figures. **(2)**

(Total for Question 14 is 6 marks)

15 A fair coin-like device is claimed to land on 'success' with probability 0.1. A tester believes the probability of success is actually lower. Let X be the number of trials needed to get the first success, so under the claim X ~ Geo(0.1). The first success occurs on trial 12.

(a) State suitable hypotheses for the test, in terms of p. **(2)**

(b) Find $P(X \geq 12)$ assuming H0 is true, and test the tester's belief at the 10% level. **(3)**

(c) Find the critical region for the test at the 10% level, and the probability of a Type II error if the true value of p is 0.05, to 3 significant figures. **(4)**

(Total for Question 15 is 9 marks)

16 The random variable X ~ B(12, p). A test of H0: p = 0.2 against H1: p > 0.2 is carried out at the 5% significance level.

(a) Find the critical region for the test. **(2)**

(b) State the probability of a Type I error, to 3 significant figures. **(2)**

(c) Find the probability of a Type II error when the true value of p is 0.35, to 3 significant figures. **(2)**

(Total for Question 16 is 6 marks)

17 A fair coin-like device is claimed to land on 'success' with probability 0.15. A tester believes the probability of success is actually lower. Let X be the number of trials needed to get the first success, so under the claim X ~ Geo(0.15). The first success occurs on trial 23.

(a) State suitable hypotheses for the test, in terms of p. **(2)**

(b) Find $P(X \geq 23)$ assuming H0 is true, and test the tester's belief at the 5% level. **(3)**

(c) Find the critical region for the test at the 5% level, and the probability of a Type II error if the true value of p is 0.05, to 3 significant figures. **(4)**

(Total for Question 17 is 9 marks)

18 A fair coin-like device is claimed to land on 'success' with probability 0.15. A tester believes the probability of success is actually lower. Let X be the number of trials needed to get the first success, so under the claim X ~ Geo(0.15). The first success occurs on trial 19.

(a) State suitable hypotheses for the test, in terms of p. **(2)**

(b) Find $P(X \geq 19)$ assuming H0 is true, and test the tester's belief at the 5% level. **(3)**

(c) Find the critical region for the test at the 5% level, and the probability of a Type II error if the true value of p is 0.05, to 3 significant figures. **(4)**

(Total for Question 18 is 9 marks)

19 A fair coin-like device is claimed to land on 'success' with probability 0.3. A tester believes the probability of success is actually lower. Let X be the number of trials needed to get the first success, so under the claim X ~ Geo(0.3). The first success occurs on trial 19.

(a) State suitable hypotheses for the test, in terms of p. **(2)**

(b) Find $P(X \geq 19)$ assuming H0 is true, and test the tester's belief at the 10% level. **(3)**

(c) Find the critical region for the test at the 10% level, and the probability of a Type II error if the true value of p is 0.1, to 3 significant figures. **(4)**

(Total for Question 19 is 9 marks)

20 A fairground game at a summer fete in Durham is designed so that a player wins on any go with probability 0.4, independently of other goes. The organiser wants to test, at the 10% significance level, whether the true win probability differs from 0.4, based on X, the number of wins in 15 goes.

(a) State suitable hypotheses, then find the critical region for this two-tailed test and state the actual significance level. **(6)**

(b) In one trial of 15 goes, there are 3 wins. State, with a reason, the conclusion of the test. **(2)**

(c) State one advantage of using a 10% significance level rather than a 5% significance level for this test, together with a corresponding disadvantage. **(1)**

(Total for Question 20 is 9 marks)

21 Referring again to the textile mill's test in Question 5, where $X \sim Po(8)$ under H0 and the critical region for the 5% one-tailed (decrease) test was found to be $X \leq 3$, suppose instead that the new yarn has reduced the true mean number of flaws per roll to 3 (rather than 5, as in Question 5).

(a) Using the same critical region as in Question 5, find the probability of a Type II error, and hence the power of the test, in this case. **(5)**

(b) Compare the power found in part (a) with the power found in Question 5(c), and explain, with reference to the two Poisson distributions involved, why the powers differ. **(3)**

(Total for Question 21 is 8 marks)

22 The number of minor faults found on newly manufactured circuit boards from a particular production line is modelled, under normal conditions, by $X \sim Po(6)$. Quality control engineers want to test, at the 10% significance level, whether the mean number of faults per board has changed from 6 (a two-tailed test).

(a) State suitable hypotheses, then find the critical region for this two-tailed test and state the actual significance level. **(6)**

(b) A randomly chosen board is found to have 10 faults. State, with a reason, the conclusion of the test. **(2)**

(c) Suppose that a fault in the manufacturing process has in fact increased the true mean number of faults per board to 9. Find the probability that the test results in a Type II error. **(3)**

(d) Hence state the power of the test in this case. **(1)**

(e) Comment on the suitability of this test for detecting the increase in mean described in part (c). **(1)**

(Total for Question 22 is 13 marks)

Mark scheme · 2.4 Further Statistics: Hypothesis Testing, Type I and Type II Errors

Question 1

(a) attempts P(X≥9) using X ~ B(12, 0.5) **M1**

(a) P(X≥9) = 299/4096 = awrt 0.0730 **A1**

(a) doubles this probability for a two-tailed test **M1**

(a) p-value = awrt 0.146 **A1**

(a) Answer: p-value = awrt 0.146

(b) 0.146 > 0.05, so do not reject H0; there is insufficient evidence, at the 5% level, that the spinner is biased towards even numbers (ft, in context) **B1**

(b) Answer: Since 0.146 > 0.05, H0 is not rejected; there is insufficient evidence that the spinner is biased towards even numbers.

Question 2

(a) states H0: p = 0.3, H1: p < 0.3, with X ~ B(20, 0.3) under H0, and attempts a lower-tail search comparing cumulative probabilities with 0.05 **M1**

(a) compares P(X≤2) and P(X≤3) with 0.05 **M1**

(a) critical region is X ≤ 2 (i.e. {0, 1, 2}), cao **A1**

(a) actual significance level = awrt 0.0355 **A1**

(a) Answer: Critical region: X ≤ 2 (i.e. {0, 1, 2}). Actual significance level = awrt 0.0355.

(b) 1 lies in the critical region {0, 1, 2} found in part (a) (ft) **B1**

(b) reject H0; there is evidence, at the 5% level, that the fungicide has reduced the proportion of seedlings affected by blight (ft, in context) **B1**

(b) Answer: 1 is in the critical region {0, 1, 2}, so H0 is rejected; there is evidence at the 5% level that the fungicide has reduced the proportion affected.

(c) recognises the Type II error region is the complement of the critical region, i.e. X ≥ 3, using the true distribution B(20, 0.15) **M1**

(c) uses P(X≥3 | p=0.15) = 1 - P(X≤2 | p=0.15) **M1**

(c) awrt 0.595 **A1**

(c) Answer: P(Type II error) = awrt 0.595

(d) power = 1 - 0.595 = awrt 0.405 (ft from part (c)) **B1**

(d) Answer: power = awrt 0.405

(e) the probability of a Type II error would decrease (oe) **B1**

(e) a larger significance level allows a larger critical region to be used; more of the alternative distribution's probability then falls inside this larger critical region (i.e. leads to rejecting H0), leaving less probability in the acceptance region, so P(Type II error) falls (oe) **B1**

(e) Answer: The probability of a Type II error would decrease, because a 10% test uses a larger critical region, so a greater proportion of the B(20, 0.15) distribution falls inside it, correctly leading to rejection of H0.

Question 3

(a) attempts P(X≤3) with λ = 10 **M1**

(a) P(X≤3) = awrt 0.0103 **A1**

(a) attempts P(X≥18) = 1 - P(X≤17) with λ = 10 **M1**

(a) P(X≥18) = awrt 0.0143 **A1**

(a) P(Type I error) = 0.0103 + 0.0143 = awrt 0.0246 **A1**

(a) Answer: P(Type I error) = awrt 0.0246

Question 4

(a) compares P(X≤0), P(X≤1) and P(X≤2) with 0.01 **M1**

(a) critical region is X ≤ 1 (i.e. {0, 1}), cao **A1**

(a) actual significance level = awrt 0.00764 **A1**

(a) Answer: Critical region: X ≤ 1 (i.e. {0, 1}). Actual significance level = awrt 0.00764.

(b) the critical region {0, 1} used here is smaller than the critical region {0, 1, 2} used in Question 2(a), so the acceptance region is correspondingly larger (oe) **B1**

(b) for the same true alternative value of p, a larger proportion of the alternative distribution now falls in the acceptance region, so P(Type II error) is larger (power is lower) than in Question 2 (oe) **B1**

(b) Answer: The critical region {0, 1} is smaller than {0, 1, 2}, so more of the alternative (p = 0.15) distribution falls in the acceptance region here than in Question 2; P(Type II error) is therefore larger, and the power lower, than for the 5% test.

Question 5

(a) states H0: λ = 8, H1: λ < 8, and attempts to compare P(X≤3) and P(X≤4) with 0.05 **M1**

(a) correctly evaluates both cumulative probabilities **M1**

(a) critical region is X ≤ 3 (i.e. {0, 1, 2, 3}), cao **A1**

(a) actual significance level = awrt 0.0424 **A1**

(a) Answer: Critical region: X ≤ 3 (i.e. {0, 1, 2, 3}). Actual significance level = awrt 0.0424.

(b) identifies the acceptance region as X ≥ 4, using the true distribution Po(5) **M1**

(b) uses P(X≥4 | λ=5) = 1 - P(X≤3 | λ=5) **M1**

(b) awrt 0.735 **A1**

(b) Answer: P(Type II error) = awrt 0.735

(c) power = 1 - 0.735 = awrt 0.265 (ft from part (b)) **B1**

(c) Answer: power = awrt 0.265

(d) valid suggestion, e.g. increase the significance level (widen the critical region, for example to a 20% test, giving the critical region X ≤ 5 and power awrt 0.616), so that more of the Po(5) distribution falls inside the (larger) critical region **B1**

(d) valid corresponding disadvantage, e.g. this increases the probability of a Type I error (wrongly concluding the mean has decreased when it has not) (oe) **B1**

(d) Answer: Suggestion: increase the significance level (e.g. to 20%, giving the critical region X ≤ 5), widening the critical region so more of the true Po(5) distribution falls inside it, raising the power above 0.5. Disadvantage: this also increases the probability of a Type I error.

Question 6

(a) uses Po(5) probabilities in the appropriate tail **M1**

(a) $P(X \geq 10) = 0.0318 \leq 0.05$ **A1**

(a) critical region $X \geq 10$ **A1**

(a) Answer: critical region $X \geq 10$

(b) the probability of a Type I error = P(X in the critical region | H0 true) **M1**

(b) 0.0318 **A1**

(b) Answer: 0.0318

(c) P(X not in the critical region | $\lambda = 6$) **M1**

(c) 0.916 **A1**

(c) Answer: 0.916

Question 7

(a) uses B(25, 0.25) cumulative probabilities **M1**

(a) critical region $X \leq 3$ **A1**

(a) Answer: critical region $X \leq 3$

(b) P(X in critical region | $p = 0.25$) **M1**

(b) 0.0962 **A1**

(b) Answer: 0.0962

(c) P(X not in critical region | $p = 0.1$) **M1**

(c) 0.236 **A1**

(c) Answer: 0.236

Question 8

(a) H0: $p = 0.15$ **B1**

(a) H1: $p < 0.15$ **B1**

(a) Answer: H0: $p = 0.15$, H1: $p < 0.15$

(b) $P(X \geq 16) = (1 - 0.15)^{16 - 1}$ **M1**

(b) 0.0874 **A1**

(b) reject H0 **A1**

(b) Answer: 0.0874; reject H0

(c) finds the smallest d with $(1 - 0.15)^{d - 1} \leq 0.1$ **M1**

(c) critical region $X \geq 16$ **A1**

(c) $P(X \leq 15 \mid p = 0.08) = 1 - (1 - 0.08)^{15}$ **M1**

(c) 0.714 **A1**

(c) Answer: $X \geq 16$; Type II error 0.714

Question 9

(a) uses Po(8) probabilities in the appropriate tail **M1**

(a) $P(X \leq 4) = 0.0996 \leq 0.1$ **A1**

(a) critical region $X \leq 4$ **A1**

(a) Answer: critical region $X \leq 4$

(b) the probability of a Type I error = P(X in the critical region | H0 true) **M1**

(b) 0.0996 **A1**

(b) Answer: 0.0996

(c) P(X not in the critical region | $\lambda = 3$) **M1**

(c) 0.185 **A1**

(c) Answer: 0.185

Question 10

(a) uses B(12, 0.2) cumulative probabilities **M1**

(a) critical region $X \geq 5$ **A1**

(a) Answer: critical region $X \geq 5$

(b) P(X in critical region | $p = 0.2$) **M1**

(b) 0.0726 **A1**

(b) Answer: 0.0726

(c) P(X not in critical region | $p = 0.7$) **M1**

(c) 0.00949 **A1**

(c) Answer: 0.00949

Question 11

(a) uses Po(4) probabilities in the appropriate tail **M1**

(a) $P(X \geq 8) = 0.0511 \leq 0.1$ **A1**

(a) critical region $X \geq 8$ **A1**

(a) Answer: critical region $X \geq 8$

(b) the probability of a Type I error = P(X in the critical region | H0 true) **M1**

(b) 0.0511 **A1**

(b) Answer: 0.0511

(c) P(X not in the critical region | $\lambda = 10$) **M1**

(c) 0.220 **A1**

(c) Answer: 0.220

Question 12

(a) H0: $p = 0.25$ **B1**

(a) H1: $p < 0.25$ **B1**

(a) Answer: H0: $p = 0.25$, H1: $p < 0.25$

(b) $P(X \geq 11) = (1 - 0.25)^{11 - 1}$ **M1**

(b) 0.0563 **A1**

(b) do not reject H0 **A1**

(b) Answer: 0.0563; do not reject H0

(c) finds the smallest d with $(1 - 0.25)^{d - 1} \leq 0.05$ **M1**

(c) critical region $X \geq 12$ **A1**

(c) $P(X \leq 11 \mid p = 0.08) = 1 - (1 - 0.08)^{11}$ **M1**

(c) 0.600 **A1**

(c) Answer: $X \geq 12$; Type II error 0.600

Question 13

(a) uses Po(2) probabilities in the appropriate tail **M1**

(a) $P(X \geq 6) = 0.0166 \leq 0.05$ **A1**

(a) critical region $X \geq 6$ **A1**

(a) Answer: critical region $X \geq 6$

(b) the probability of a Type I error = P(X in the critical region | H0 true) **M1**

(b) 0.0166 **A1**

(b) Answer: 0.0166

(c) P(X not in the critical region | $\lambda = 6$) **M1**

(c) 0.446 **A1**

(c) Answer: 0.446

Question 14

(a) uses B(10, 0.4) cumulative probabilities **M1**

(a) critical region $X \leq 1$ **A1**

(a) Answer: critical region $X \leq 1$

(b) P(X in critical region | p = 0.4) **M1**

(b) 0.0464 **A1**

(b) Answer: 0.0464

(c) P(X not in critical region | p = 0.15) **M1**

(c) 0.456 **A1**

(c) Answer: 0.456

Question 15

(a) H0: p = 0.1 **B1**

(a) H1: p < 0.1 **B1**

(a) Answer: H0: p = 0.1, H1: p < 0.1

(b) $P(X \geq 12) = (1 - 0.1)^{12 - 1}$ **M1**

(b) 0.314 **A1**

(b) do not reject H0 **A1**

(b) Answer: 0.314; do not reject H0

(c) finds the smallest d with $(1 - 0.1)^{d - 1} \leq 0.1$ **M1**

(c) critical region $X \geq 23$ **A1**

(c) $P(X \leq 22 \mid p = 0.05) = 1 - (1 - 0.05)^{22}$ **M1**

(c) 0.676 **A1**

(c) Answer: X ≥ 23; Type II error 0.676

Question 16

(a) uses B(12, 0.2) cumulative probabilities **M1**

(a) critical region $X \geq 6$ **A1**

(a) Answer: critical region $X \geq 6$

(b) P(X in critical region | p = 0.2) **M1**

(b) 0.0194 **A1**

(b) Answer: 0.0194

(c) P(X not in critical region | p = 0.35) **M1**

(c) 0.787 **A1**

(c) Answer: 0.787

Question 17

(a) H0: p = 0.15 **B1**

(a) H1: p < 0.15 **B1**

(a) Answer: H0: p = 0.15, H1: p < 0.15

(b) $P(X \geq 23) = (1 - 0.15)^{23 - 1}$ **M1**

(b) 0.0280 **A1**

(b) reject H0 **A1**

(b) Answer: 0.0280; reject H0

(c) finds the smallest d with $(1 - 0.15)^{d - 1} \leq 0.05$ **M1**

(c) critical region $X \geq 20$ **A1**

(c) $P(X \leq 19 \mid p = 0.05) = 1 - (1 - 0.05)^{19}$ **M1**

(c) 0.623 **A1**

(c) Answer: X ≥ 20; Type II error 0.623

Question 18

(a) H0: p = 0.15 **B1**

(a) H1: p < 0.15 **B1**

(a) Answer: H0: p = 0.15, H1: p < 0.15

(b) $P(X \geq 19) = (1 - 0.15)^{19 - 1}$ **M1**

(b) 0.0536 **A1**

(b) do not reject H0 **A1**

(b) Answer: 0.0536; do not reject H0

(c) finds the smallest d with $(1 - 0.15)^{d - 1} \leq 0.05$ **M1**

(c) critical region $X \geq 20$ **A1**

(c) $P(X \leq 19 \mid p = 0.05) = 1 - (1 - 0.05)^{19}$ **M1**

(c) 0.623 **A1**

(c) Answer: X ≥ 20; Type II error 0.623

Question 19

(a) H0: p = 0.3 **B1**

(a) H1: p < 0.3 **B1**

(a) Answer: H0: p = 0.3, H1: p < 0.3

(b) $P(X \geq 19) = (1 - 0.3)^{19 - 1}$ **M1**

(b) 0.00163 **A1**

(b) reject H0 **A1**

(b) Answer: 0.00163; reject H0

(c) finds the smallest d with $(1 - 0.3)^{d - 1} \leq 0.1$ **M1**

(c) critical region $X \geq 8$ **A1**

(c) $P(X \leq 7 \mid p = 0.1) = 1 - (1 - 0.1)^{7}$ **M1**

(c) 0.522 **A1**

(c) Answer: X ≥ 8; Type II error 0.522

Question 20

(a) states H0: $p = 0.4$, H1: $p \neq 0.4$, with $X \sim B(15, 0.4)$ under H0, and searches the lower tail comparing $P(X \leq 2)$ and $P(X \leq 3)$ with 0.05 **M1**

(a) lower critical region $X \leq 2$ (i.e. {0, 1, 2}), actual lower-tail level = awrt 0.0271 **A1**

(a) searches the upper tail comparing $P(X \geq 9)$ and $P(X \geq 10)$ with 0.05 **M1**

(a) upper critical region $X \geq 10$ (i.e. {10, 11, ..., 15}), actual upper-tail level = awrt 0.0338 **A1**

(a) combined critical region stated as {0, 1, 2} union {10, 11, ..., 15} **A1**

(a) total actual significance level = 0.0271 + 0.0338 = awrt 0.0609 **A1**

(a) Answer: Critical region: {0, 1, 2} union {10, 11, ..., 15}. Total actual significance level = awrt 0.0609.

(b) 3 does not lie in the critical region {0, 1, 2} union {10, ..., 15} found in part (a) (ft) **B1**

(b) do not reject H0; there is insufficient evidence, at the 10% level, that the true win probability differs from 0.4 (ft, in context) **B1**

(b) Answer: 3 is not in the critical region, so H0 is not rejected; there is insufficient evidence at the 10% level that the win probability differs from 0.4.

(c) any one correctly matched advantage and disadvantage, e.g. advantage: a larger critical region gives higher power, so a genuine change in win probability is more likely to be detected; disadvantage: this also gives a higher probability of a Type I error, wrongly concluding the win probability has changed when it has not (oe) **B1**

(c) Answer: Advantage: a 10% test has a larger critical region and higher power, so it is more likely to detect a genuine change in win probability. Disadvantage: it also has a higher probability of a Type I error (wrongly rejecting a true H0) than a 5% test would.

Question 21

(a) recognises that, since the critical region is $X \leq 3$, the power here equals $P(X \leq 3 \mid \lambda = 3)$, and the Type II error probability is its complement **M1**

(a) attempts to sum $P(X=0)$ to $P(X=3)$ with $\lambda = 3$ **M1**

(a) power = awrt 0.647 **A1**

(a) P(Type II error) = 1 - 0.647 = awrt 0.353 **A1**

(a) clearly labels which value is the power and which is P(Type II error), avoiding confusing the two **B1**

(a) Answer: power = awrt 0.647; P(Type II error) = awrt 0.353

(b) correctly states the power is higher when the true mean is 3 (awrt 0.647) than when the true mean is 5 (awrt 0.265, from Question 5) **B1**

(b) gives a valid reason referencing distance from H0 / overlap of distributions, e.g. a true mean of 3 is further from the hypothesised mean of 8 than a true mean of 5 is, so the null ($\lambda = 8$) and alternative distributions overlap less, and a greater proportion of the alternative distribution falls inside the critical region $X \leq 3$ (oe) **B1**

(b) states the general principle: for a fixed significance level and sample size, the power of a test increases as the true parameter value moves further from the value stated in H0 (oe) **B1**

(b) Answer: The power is higher when the true mean is 3 (0.647) than when it is 5 (0.265, from Question 5), because a true mean of 3 is further from the hypothesised value of 8, so the null and alternative distributions overlap less and more of the alternative distribution falls in the critical region $X \leq 3$; in general, power increases as the true parameter moves further from the value stated in H0.

Question 22

(a) states H0: $\lambda = 6$, H1: $\lambda \neq 6$, and searches the lower tail comparing $P(X \leq 1)$ and $P(X \leq 2)$ with 0.05 **M1**

(a) lower critical region $X \leq 1$ (i.e. {0, 1}), actual lower-tail level = awrt 0.0174 **A1**

(a) searches the upper tail comparing $P(X \geq 10)$ and $P(X \geq 11)$ with 0.05 **M1**

(a) upper critical region $X \geq 11$ (i.e. {11, 12, 13, ...}), actual upper-tail level = awrt 0.0426 **A1**

(a) combined critical region stated as {0, 1} union {11, 12, 13, ...} **A1**

(a) total actual significance level = 0.0174 + 0.0426 = awrt 0.0600 **A1**

(a) Answer: Critical region: {0, 1} union {11, 12, 13, ...}. Total actual significance level = awrt 0.0600.

(b) 10 does not lie in the critical region {0, 1} union {11, 12, ...} found in part (a) (ft) **B1**

(b) do not reject H0; there is insufficient evidence, at the 10% level, that the mean number of faults per board has changed (ft, in context) **B1**

(b) Answer: 10 is not in the critical region, so H0 is not rejected; there is insufficient evidence at the 10% level that the mean has changed.

(c) identifies the acceptance (non-rejection) region as $2 \leq X \leq 10$, using the true distribution Po(9) **M1**

(c) uses $P(2 \leq X \leq 10 \mid \lambda=9) = P(X \leq 10 \mid \lambda=9) - P(X \leq 1 \mid \lambda=9)$ **M1**

(c) awrt 0.705 **A1**

(c) Answer: P(Type II error) = awrt 0.705

(d) power = 1 - 0.705 = awrt 0.295 (ft from part (c)) **B1**

(d) Answer: power = awrt 0.295

(e) valid comment referencing the low power value, e.g. the power is only awrt 0.295, so the test would fail to detect this particular increase in mean on roughly 70% of occasions; the test is not very sensitive to a change of this size (oe) **B1**

(e) Answer: The power is only about 0.295, so the test would fail to detect an increase in mean to 9 on around 70% of occasions; it is not a very sensitive test for a change of this size.