2.4 Further Statistics: Hypothesis Testing, Type I and Type II Errors: 30-minute Pack

AQA 7367 · Calculator allowed · about 30 minutes

Total Marks

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)

© Revision Library | revisionlibrary.co.uk

2 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 2 is 5 marks)

---

3 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 $X \sim Geo(0.25)$. 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 3 is 9 marks)

---

4 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 4 is 9 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.4 Further Statistics: Hypothesis Testing, Type I and Type II Errors: 30-minute Pack

Question 1

(a) attempts $P(X \geq 9)$ using $X \sim B(12, 0.5)$ **M1**

(a) $P(X \geq 9) = 299/4096 =$ 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) attempts $P(X \leq 3)$ with $\lambda = 10$ **M1**

(a) $P(X \leq 3) =$ awrt 0.0103 **A1**

(a) attempts $P(X \geq 18) = 1 - P(X \leq 17)$ with $\lambda = 10$ **M1**

(a) $P(X \geq 18) =$ 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 3

(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 ≥ 12; Type II error 0.600

Question 4

(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.

© Revision Library | revisionlibrary.co.uk