**Paper 3B/4B: Further Statistics 1 Mark Schemes**

|  |  |  |  |
| --- | --- | --- | --- |
| **Question** | **Scheme** | **Marks** | **AOs** |
| **1** | Ho : = 5 (= 2.5) H1 : > 5 (> 2.5) | B1 | 2.5 |
| *X* ~ Po (2.5) | B1 | 3.3 |
| **Method 1:** | **Method 2:** |  |  |
|  = 1 − 0.9858 | P(*X* 5) = 0.1088 P(*X* 6) = 0.042 | M1 | 1.1b |
|  = 0.0142 | CR *X*  6 | A1 | 1.1b |
| 0.0142 < 0.05 7  6 or 7 is in critical region or 7 is significantReject H0. There is evidence at the 5% significance level that the level of pollution has increased.**or**There is evidence to support the scientists claim is justified | A1cso | 2.2b |
| **(5 marks)** |
| **Notes:** |
| **B1:** Both hypotheses correctusing **or and** 5 **or** 2.5**B1:** Realising that the model Po(2.5) is to be used. This may be stated or used**M1:** Using or writing **or** a correct CR **or** P(*X* 5) = awrt 0.109 **and** P(*X* 6) = awrt 0.042 **A1:** awrt 0.0142 **or** CR*X*  6 **or** *X* > 5**M1:** A fully correct solution and drawing a correct inference in context |

|  |  |  |  |
| --- | --- | --- | --- |
| **Question** | **Scheme** | **Marks** | **AOs** |
| **2(a)** | P(*X* 1) = 1 – P(*X* = 0) 1 – P(*X* = 0) = 0.049 | B1 | 3.1b |
| P(*X* = 0) = 0.951 | B1 | 1.1b |
| *x*5 = 0.951 *x* = 0.99 | M1 | 3.1b |
|  *p* = 0.01 | A1 | 1.1b |
| *X* ~B(1000, 0.01) | M1 | 3.3 |
|  Mean = *np* = 10 | A1ft | 1.1b |
|  Variance = *np*(1 – *p*) = 9.9 | A1ft | 1.1b |
|  | **(7)** |  |
| **(b)** | *X* ~ Po(“10”) then require: P(*X* > 6) = 1 – P (*X* 6) | M1 | 3.4 |
| = 1 – 0.1301 |  |  |
|   = 0.870 | A1 | 1.1b |
|  | **(2)** |  |
| **(c)** | The approximation is valid as : the number of calls is large | B1 | 2.4 |
| The probability of connecting to the wrong agent is small | B1 | 2.4 |
|  | **(2)** |  |
| **(d)** | The answer is accurate to 2 decimal place | **B1** | 3.2b |
|  | **(1)** |  |
| **(12 marks)** |
| **Notes:** |
| **(a)****B1:** Realising that the P(at least 1 call ) = 1 – P(*X* = 0)**B1:** Calculating P(*X* = 0) = 0.951**M1:** Forming the equation  may be implied by *p* = 0.01**A1:** 0.01 only**M1:** Realising the need to use the modelB(1000, 0.01) This may be stated or used **A1:** Mean =10 or ft their *p* but onlyif 0 < *p* < 1**A1:** Var = 9.9 or ft their *p* but only if 0 < *p* < 1 |
| **(b)****M1:** Using the model Po(“their 10”) (this may be written or used) and 1 – P (*X* 6) **A1:** awrt 0.870 Award M1 A1 for awrt 0.870 with no incorrect working |
| **(c)****B1:** Explaining why approximation is valid -need the context of number and calls**B1:** Need the context connecting, wrong agent |
| **(d)****B1:** Evaluating the accuracy of their answer in (b). Allow 2 significant figures |
| **Question** | **Scheme** | **Marks** | **AOs** |
| **3(a)** | Expected value for 2 = | M1 | 3.4 |
|  = 28.3015… | A1 | 1.1b |
| Expected value for 4 or more = 150 – (53.8 + 56.6 + 28.3 + 8.9) = 2.4 | A1ft | 1.1b |
| H0: Bin(20, 0.05) is a suitable modelH1: Bin(20, 0.05) is not a suitable model | B1 | 2.5 |
| Combining last two groups

|  |  |
| --- | --- |
|  |  **3** |
| **Observed frequency** | 19 |
| **Expected frequency** | 11.3 |

 | M1 | 2.1 |
| ** | B1 | 1.1b |
| Critical value, χ2 (0.05) = 7.815 | B1 | 1.1a |
| Test statistic =  | M1 | 1.1b |
|  = 8.117 | A1 | 1.1b |
| In critical region, sufficient evidence to reject H0, accept H1Significant evidence at 5% level to reject the manager’s model | A1 | 3.5a |
|  | **(10)** |  |
| **(b)** |  = 4 – 2 = 2 |  |  |
| 4 classes due to pooling | B1 | 2.4 |
| 2 restrictions (equal total and mean/proportion) | B1 | 2.4 |
|  | **(2)** |  |
| **(c)** | H0: Binomial distribution is a good modelH1: Binomial distribution is not a good model | B1 | 3.4 |
| Critical value, χ2 (0.05) = 5.991Test statistic is not in critical region, insufficient evidence to reject H0There is evidence that the Binomial distribution is a good model | B1 | 3.5a |
|  | **(2)** |  |
| **(14 marks)** |

|  |
| --- |
| **Question 3 notes:** |
| **(a)****M1:** Using the binomial model may be implied by 28.3**A1:** awrt 28.3**A1:** awrt 2.4 or ft their “28.3”**B1:** Both hypotheses correct using the correct notation or written out in full **M1:** For recognising the need to combine groups**B1:** Number of degrees of freedom = 3 may be implied by a correct CV**B1:** awrt 7.82**M1:** Attempting to find or  may be implied by awrt 8.12**A1:** awrt 8.12**A1:** Evaluating the outcome of a model by drawing a correct inference in context |
| **(b)****B1:** Explaining why there are 4 classes**B1:** Explanation of why 2 is subtracted |
| **(c)****B1:** Correct hypotheses for the refined model**B1:** The CV awrt 5.99 anddrawing the correct inference for the refined model  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Question** | **Scheme** | **Marks** | **AOs** |
| **4** | Po(2.3)*n* = 100** **=** 2.3 |  |  |
| CLT  | M1A1 | 3.1a1.1b |
|  | M1 | 3.4 |
|  = P(Z > 1.318..) |  |  |
|  = 0.09632… | A1 | 1.1b |
|  | **(4)** |  |
| **(4 marks)** |
| **Notes:** |
| **M1:** For realising the need to use the CLT to set  normal with correct mean  May be implied by using the correct normal distribution**A1:** For fully correct normal stated or used**M1:** Use of the normal model to find . Can be awarded for   **or** awrt 1.32**A1:** awrt 0.0963 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Question** | **Scheme** | **Marks** | **AOs** |
| **5(a)** |  | M1 | 3.3 |
|  = 0.05940… = awrt **0.0594** | A1 | 1.1b |
|  | **(2)** |  |
| **(b)** | The model is only valid if: |  |  |
| the games (trials) are **independent** | B1 | 3.5b |
| the probability of winning a prize, 0.15, is **constant** for each game | B1 | 3.5b |
|  | **(2)** |  |
| **(c)** |  | M1A1 | 3.1b1.1b |
| Solving: 2*p* = 1 – *p* | M1 | 1.1b |
| *p* =  (> 0.15) so Mary has the greater chance of winning a prize | A1 | 3.2a |
|  | **(4)** |  |
| **(8 marks)** |
| **Notes:** |
| **5(a)****M1:** For selecting an appropriate model negative binomial or B(7, 0.15) with an extra success in 8th trial e.g.  Allow  may be implied by awrt 0.0594 **A1:** awrt 0.0594 |
| **(b)****B1:** Stating the first assumption that games are independent**B1:** Stating the second assumption that the probability remains constant |
| **(c)****M1:** Forming an equation for the mean or for the standard deviation**A1:** Both equations correct**M1:** Solving the 2 equations leading to 2*p* = 1 – *p***A1:** For *p* = followed by a correct deduction |

|  |  |  |  |
| --- | --- | --- | --- |
| **Question** | **Scheme** | **Marks** | **AOs** |
| **6(a)** | gives | M1 | 2.1 |
|  so *k* =  **\*** | A1\*cso | 1.1b |
|  | **(2)** |  |
| **(b)** | P(*X* = 3) = coefficient of *t*3 so  | M1 | 1.1b |
| [ P(*X* = 3) =]  | A1 | 1.1b |
|  | **(2)** |  |
| **(c)** |  | M1 | 2.1 |
| E(*X*) =  | M1 | 1.1b |
|  =  | A1 | 1.1b |
|  | M1A1 | 2.11.1b |
|   | M1 | 1.1b |
| Var(*X*) =  | M1 | 2.1 |
| = \* | A1\*cso | 1.1b |
|  | **(8)** |  |
| **(d)** |  **[** orsub *t*2 for *t*] | M1 | 3.1a |
|  =  | A1 | 1.1b |
|  | **(2)** |  |
| **(14 marks)** |
| **Notes:** |
| **(a)****M1:** Stating **A1\*:** Fully correct proof with no errors cso |
| **(b)****M1:** Attempting to find the coefficient of *t*3. May be implied by obtaining  or awrt 0.11**A1:** , allow awrt 0.111 |

|  |
| --- |
| **Question 6 notes continued:** |
| **(c)****M1:** Attempting to find. Allow Chain rule or multiplying out the brackets and  differentiating**M1:** Substituting *t* = 1 into  **A1:** , allow awrt 1.67 **M1:** Attempting to find **A1:**  **or** o.e.**A1:** o.e.**M1:** Using to find the Variance**A1\*:** cso |
| **(d)****M1:** Realising the need to **or** sub *t*2 for *t***A1:** , or o.e. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Question** | **Scheme** | **Marks** | **AOs** |
| **7(a)** | *X* ~ B(20, 0.2) and seek *c* such that P(*X*  *c*) < 0.10 | M1 | 3.3 |
| [P(*X*  1) = 0.0692] CR is *X*  1 | A1 | 1.1b |
|  | **(2)** |  |
| **(b)** | Size = **0.0692** | B1ft | 1.2 |
|  | **(1)** |  |
| **(c)** | *Y* = no. of spins until red obtained so *Y*~ Geo(0.2) | M1 | 3.3 |
|  so if *p* < 0.2 then mean is larger so seek *d* so that P(*Y*  *d*) < 0.10 | M1 | 2.4 |
|   | M1 | 3.4 |
|  | M1 | 1.1b |
|  *d*  > 11.3.. | A1 | 1.1b |
| CR is ***Y*  12** | A1 | 2.2b |
|  | **(6)** |  |
| **(d)** | Size = [= 0.085899…] = **0.0859** | B1 | 1.1b |
|  | **(1)** |  |
| **(e)(i)** | Power = P(reject H0 when it is false) = P(*X*  1 | *X* ~B(20, *p*)) | M1 | 2.1 |
|  **=**  | M1 | 1.1b |
|  =  \* | A1\*cso | 1.1b |
| **(ii)** |  Power =  | B1 | 1.1b |
|  | **(4)** |  |
| **(f)** | Sam’s test has smaller P(Type I error) (or size) so is better | B1 | 2.2a |
| Power of Sam’s test = 0.1755… | B1 | 1.1b |
| Power of Tessa’s test =  = 0.1673… | B1 | 1.1b |
| So for *p* = 0.15 **Sam’s test** is recommended | B1 | 2.2b |
|  | **(4)** |  |
| **(18 marks)** |

|  |
| --- |
| **Question 7 notes:** |
| **(a)****M1:** Realising the need to use the modelUsing B(20,0.2) with method for finding the CR or  implied by a correct CR**A1:** *X*  1 or *X* < 2 |
| **(b)****B1:** awrt 0.0692 |
| **(c)****M1:** Realising that the model Geo(0.2)is needed. This may be written or used **M1:** Realising the key step that they need to find P(*Y*  *d*) < 0.10**M1:** **U**sing the model **M1:** Using the model  and finding a method to solve leading to a value/range of  values for *d***A1:** For *d*  > 11.3..**A1:** For *Y*  12 or *Y* > 11 (a correct inference) |
| **(d)****B1ft:** awrt 0.0692. ft their answer to part (c) |
| **(e)(i)****M1:** Using B(20, *p*) and realizing they need to find P(*X*  1) o.e. This may be used or written**M1:** Using P(*X* = 0) + P(*X* = 1)**A1\*:** Fully correct proof ( no errors) cso |
| **(ii)****B1:** For   |
| **(f)****B1:** Making a deduction about the tests using the answers to part(b) and (d)**B1:** awrt 0.0176**B1:** awrt 0.167**B1:** A correct inference about which test is recommended  |