| **Question** | **Scheme** | **Marks** |
| --- | --- | --- |
| **Section A – Representation and summary of data (26 marks)** | | |
| **1** | mean =  or 60.8 = 1.4*x* 20 (o.e.) | M1 |
| = 57.7142… awrt **57.7** | A1 |
|  |  |
| standard deviation =  or 6.60 = 1.4*x* | M1 |
| = 4.7142… awrt **4.71** | A1 |
|  |  | **(4 marks)** |
| **2(a)** | 2.8 + 5.6 + 2.3 + 9.4 + 0.5 + 1.8 + 84.6 = 107 | M1 |
| Mean = 107 ÷ 28 (= 3.821…) **(awrt 3.8)** | A1 |
|  |  | **(2)** |
| **2(b)** | It will have no effect since | B1 |
| one is 4.5 under what it should be and the other is 4.5 above what it should be | dB1 |
|  |  | **(2)** |
|  |  | **(4 marks)** |
| **3(a)** | 14, 5 | M1 A1 |
|  |  | **(2)** |
| **3(b)** | 21 + 45 + 3 = 69 | M1 A1 |
|  |  | **(2)** |
|  |  | **(4 marks)** |
| **4(a)** | [Range = 48 – 9] = **39** | B1 |
|  |  | **(1)** |
| **4(b)** | [IQR = 25 – 12 ]= **13** | B1 |
|  |  | **(1)** |
| **4(c)** | Median =  == awrt **68.5**° | M1 A1 |
|  |  | **(2)** |
| **4(d)** | Lower Quartile = = **63** **(\*)** | M1 A1cso |
|  |  | **(2)** |
| **4(e)** | 63 – 1.5 × (75 – 63) = 45 75 + 1.5 × (75 – 63) = 93 | M1A1 |
|  | No data above 93 and no data below 45 or 55>45 etc or there are no outliers | A1 |
|  | 40  50  60  70  80  90 | M1  A1ft |
|  |  | **(5)** |
| **4(f)** | Median for the 70°angle is closer (to 70°)[ than the 20° median is to 20°]  The range/IQR for the 70° angle box plot is smaller/shorter  Therefore, students were more accurate at drawing the 70° angle. | B1  B1  dB1 |
|  |  | **(3)** |
|  |  | **(14 marks)** |
| **Section B – Binomial distribution and hypothesis testing (50 marks)** | | |
| **1(a)** |  | M1 |
|  | A1 |
|  | A1 |
|  | A1d |
|  |  | **(4)** |
| **1(b)** | H0:  H1: | B1 |
| or CR *X*  5 | M1 |
|  | A1 |
| Insufficient evidence to reject H0, Accept H0, Not significant. 6 does not lie in the Critical region. | M1d |
| No evidence that increasing the batch size has **reduced** the **percentage** ofbroken **pots (oe)**  **or**  evidence that there is **no change** in the **percentage** ofbroken **pots (oe)** | A1cso |
|  |  | **(5)** |
|  |  | **(9 marks)** |
| **2** |  | B1 |
| or CR of *X* < 3 | M1A1 |
| [0.0285 < 0.05] significant, reject | M1dep |
| There is evidence to support the supplier’s **claim**  or The probability of a **ball** failing the bounce **test** is **less** than **0.2** | A1cso |
|  |  | **(5 marks)** |
| **3(a)** | *X* ~ B(25,0.5) may be implied by calculations in part a or b | M1 |
| P( *X*  7) = 0.0216 |  |
| P(*X*  18) = 0.0216 |  |
| CR *X*  7;  *X*  18 | A1A1 |
|  |  | **(3)** |
| **3(b)** | P(rejecting H0) = 0.0216 + 0.0216 | M1 |
| = 0.0432 awrt 0.0432/0.0433 | A1 |
|  |  | **(2)** |
|  |  | **(5 marks)** |
| **4** |  | B1 |
|  | B1 |
| *X*~B(30,0.5) Using correct Bin | M1 |
| P(*X* 21) = 1 – P(*X* 20) or P(*X* 19) = 0.9506  P(*X* 20) = 0.0494 | M1 |
| = 1 – 0.9786 |  |
| = 0.0214 CR *X* 20 | A1 |
| so significant/reject /in Critical region | M1 dep |
| Evidence to suggest **David’s claim is incorrect**  or The weather **forecast** produced by the local **radio** is better than those achieved by **tossing/flipping a coin** | A1 |
|  |  | **(7 marks)** |
| **5** | H0 : *p* = 0.2 H1 : *p* > 0.2 | B1 |
| Under H0, *X* ~ Bin (10, 0.2) | B1 |
| P (*X* ≥ 4) = 1 – P(*X* ≤ 3) OR P (*X* ≤ 4) = 0.9672 | M1 |
| = 1 – 0.8791 P (*X* ≥ 5) = 0.0328 |  |
| = 0.1209 CR *X* ≥ 5 | A1 |
| 0.1209 > 0.05  Insufficient evidence to reject H0 so teacher’s claim is supported | M1A1ft |
|  |  | **(6 marks)** |
| **6(a)** | *X* ~ B(30, 0.25) | B1 |
| P(*X* ≤ 10) – P(*X* ≤ 4) = 0.8943 – 0.0979  = 0.7964 | M1A1 |
|  |  | **(3)** |
| **6(b)** | H0 : *p* = 0.25 H1 : *p* < 0.25 | B1 |
| B(15, 0.25) | M1 A1 |
| P(*X*  1) = 0.0802 |
| Reject H0 or Significant or 1 1ies in the critical region  There is evidence that the radio **company’s** claim is true.  **Or**  The new transmitter will reduce the proportion of houses unable to receive **radio** | dM1  A1 cso |
|  |  | **(5)** |
|  |  | **(8 marks)** |
| **7(a)** | *X* ~ B (20, 0.25) | M1 |
| P(*X* ≥ 10) = 1 – 0.9861 = 0.0139 | A1 |
| P(*X* ≤ 1) = 0.0243 | A1 |
| (0 ≤) *X* ≤ 1  10 ≤ *X* (≤20) | A1A1 |
|  |  | **(5)** |
| **7(b)** | H0: *p* = 0.25 |  |
| H1: *p* < 0.25 | B1 |
| *X* ~ B (20, 0.25) |  |
| P(*X* ≤ 3) = 0.2252 or CR *X* ≤ 1 | M1A1 |
| Insufficient evidence to reject H0, Accept H0, Not significant. 3 does not lie in the Critical region. | M1d |
| No evidence that the **changes** to the process have **reduced** the **percentage** of **defective articles (oe)** | A1cso |
|  |  | **(5)** |
|  |  | **(10 marks)** |
| **Section C – Sampling (36 marks)** | | |
| **1** | Label females 1 – 100 (or 0 – 99) and males 1 – 300 ( or 0 – 299) | B1 |
| Using *random numbers* for each group | B1 |
| in range 1 – 100 (0 – 99) select 15 females and using 1 – 300 (or 0 – 299) select 45 males | B1 |
|  |  | **(3 marks)** |
| **2(a)** | -(accurate) estimates for each strata / job  **-more** representativeof the population  -reflects population structure | B1 |
|  |  | **(1)** |
| **2(b)** | Total staff = 720 | B1 |
| Managers | M1 |
| Drivers |  |
| Administrators |  |
| Warehouse | A1 |
|  |  | **(3)** |
| **2(c)** | Label all managers 1 – 72 o.e. | B1 |
| Using random numbers in range 1-72 or 0-71 select 4 (managers). | B1 |
|  |  | **(2)** |
|  |  | **(6 marks)** |
| **3(a)** | Analyse / find estimates for a particular subgroup of the population. Stratified guarantees representation of all groups, srs does not. Observe relationships between subgroups – srs does not guarantee equal or proportionate representation. Rare or extreme cases as part of a small subgroups can be represented proportionately in stratified i.e. stratified represents the structure of the population– srs does not allow this. Stratified typically require large sample size compared to srs due to lower variability within subgroups compared to entire population. |  |
| Any 2 distinct reasons | B1B1 |
|  |  | **(2)** |
| **3(b)** | It (a stratified sample) is not biased as the members are chosen randomly. You can estimate the sampling errors (for a stratified sample) It (a stratrified sample) gives more accurate estimates as it is a random process. A quota sample may be (interviewer / process) biased. It’s not possible to estimate/find the sampling errors for a quota sample (whereas you can for a stratified sample) |  |
|  | Any 2 distinct reasons | B1B1 |
|  |  | **(2)** |
|  |  | **(4 marks)** |
|  | Notes:  Award B1B1 two correct, B1B0 one correct. Allow ‘it’ for ‘stratified’. Do not award marks for vague responses such as ‘cheap’, ‘easy’ ‘quick’ ‘random’ etc. Mentioning ‘sampling frame’ alone is not sufficient for a mark. Mentioning ‘non-response are not recorded’ alone is not sufficient for a mark. |  |
| **4(a)** | (This is a sample where) **every** (possible) **sample** (of size *n*) has an **equal chance** of being chosen. | B1 |
|  |  | **(1)** |
| **4(b)** | ‘When it is impossible to provide a **sampling frame**’ or a correct example with an indication | B1 |
| of sampling frame being impossible. |  |
|  |  | **(1)** |
| **4(c)** | A **list/register** of **all** the students. | B1 |
| Number the students (from 0 to 74, 1 to 75 etc.) | B1 |
| Using the **random no. table** read off the nos. and **identify or select the students allocated those nos**. | B1 |
|  |  | **(3)** |
|  |  | **(5 marks)** |
| **5(a)** | Sampling frame within each species of fish in the lake impossible to obtain. | B1 |
|  |  | **(1)** |
| **5(b)** | Quota sampling | B1 |
|  |  | **(1)** |
| **5(c)** | Advantages:  Sample can be obtained quickly  Costs are kept to a minimum  Administration of survey is easy | B1 |
|  | Disadvantages:  Not possible to estimate sampling errors  Process not random  Surveyor may not be able to identify species of fish easily | B1 |
|  |  | **(2)** |
| **5(d)** | |  |  | | --- | --- | | Species | Quota | | Trout |  | | Bass |  | | Pike |  |   Fish are caught from the lake until the quota of 17 trout, 7 bass and 6 pike are reached.  If a fish is caught and the species quota is full, then this is ignored. | B1B1B1  B1 |
|  |  | **(4)** |
|  |  | **(8 marks)** |
| **6(a)** | **Quota Sampling:** |  |
| Advantages: Fieldwork can be done **quick**ly, or administering the test is **easy**, or costs are kept to a minimum (**cheap**), or gives estimates for each course.  or OK for large populations or sampling frame not required (o.e.) |  |
| Disadvantages: **Non-random** process or not possible to estimate the sampling errors, or non response not recorded, or interviewer can introduce **bias** in sample choice. (o.e.) | B1 |
| **Stratified Sampling:** |  |
| Advantages: Can give accurate estimates as it is a **random** process, or gives estimates for each course or **representative** of [BUT not “proportional” to] the whole population. (o.e.) |  |
| Disadvantages: Sampling frame required, or strata may not be clear as some students overlap courses or not suitable for large populations. (o.e.) | B1 |
|  |  | **(2)** |
| **6(b)** | Total enrolments = 1000 (may be implied by calculations) | B1 |
| Leisure and Sport= | M1 |
| Information Technology==34 |  |
| Health and Social Care= |  |
| Media Studies==4 | A1 |
|  |  | **(3)** |
| **6(c)** | The college’s information system would be used to identify each student |  |
| and which course they are enrolled on.  i.e. idea of **sampling frame** or **list** for **each course**.  Use of **random numbers** to select required number of students **from each course** | ` |
|  |  | **(2)** |
|  |  | **(7 marks)** |
| **7(a)** | The list of ID numbers | B1 |
|  |  | **(1)** |
| **7(b)** | *F* ~ B(50,0.02) | B1 B1 |
|  |  | **(2)** |
|  |  | **(3 marks)** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Source paper** | **Question number** | **New spec references** | **Question description** | **New AOs** |
| **Section A – Representation and summary of data** | | | | | |
| 1 | S1 2014 | 2 | A 2.3 | Representation and summary of data | 1.1b |
| 2 | S1 Jan 2011 | 2 | A 2.3 | Representation and summary of data | 1.1b, 2.4, 3.1b |
| 3 | S1 Jan 2012 | 1 | A 2.1, 2.3 | Representation and summary of data | 1.1b, 1.2, 3.4 |
| 4 | S1 2015 | 1 | A 2.1, 2.3, 2.4 | Representation and summary of data | 1.1b, 1.2, 2.1, 2.4 |
| **Section B – Binomial distribution and hypothesis testing** | | | | | |
| 1 | S2 2017 | 1 | A 4.1, 5.1 | Binomial distribution, Hypothesis tests | 1.1b, 1.2, 2.2b, 2.4, 2.5 |
| 2 | S2 2014R | 1 | A 4.1, 5.1, 5.2 | Binomial hypothesis test | 1.1b, 2.2b, 2.5, 3.3 |
| 3 | S2 2012 | 2 | A 4.1, 5.1, 5.2 | Hypothesis testing | 1.1b, 1.2, 3.4 |
| 4 | S2 Jan 2012 | 2 | A 4.1, 5.1, 5.2 | Hypothesis testing, Tests on binomial | 1.1b, 2.5, 3.3, 3.5a |
| 5 | S2 Jan 2011 | 2 | A 4.1, 5.1, 5.2 | Hypothesis testing | 1.1b, 2.2b, 2.5, 3.3 |
| 6 | S2 2015 | 2 | A 4.1, 5.1 | Binomial distribution, Hypothesis tests | 1.1b, 2.2b, 2.5, 3.3, 3.4 |
| 7 | S2 2013 | 6 | A 4.1, 5.1, 5.2 | Binomial distribution | 1.1b, 1.2, 2.2b, 2.5, 3.1b, 3.4 |
| **Section C – Sampling** | | | | | |
| 1 | S3 2013R | 1 | A 1.1 | Sampling | 1.1b, 1.2, 2.4 |
| 2 | S3 2017 | 1 | A 1.1 | Sampling | 1.1b, 1.2 |
| 3 | S3 2016 | 1 | A 1.1 | Sampling | 1.2, 2.4 |
| 4 | S3 2014 | 1 | A 1.1 | Sampling | 1.1b, 1.2, 3.1b |
| 5 | S3 2012 | 2 | A 1.1 | Sampling | 1.1b, 1.2, 2.4, 3.1b |
| 6 | S3 2013 | 3 | A 1.1 | Sampling | 1.1b, 1.2, 3.3 |
| 7 | S2 2011 | 1 | A 1.1, 4.1 | Sampling | 1.1b, 1.2 |