Assessment Schedule – 2018

Mathematics and Statistics (Statistics): Apply probability concepts in solving problems (91585)

Evidence Statement

<table>
<tr><th>Q</th><th>Expected Coverage</th><th>Achievement (u)</th><th>Merit (r)</th><th>Excellence (t)</th></tr>
<tr><td>ONE
(a)(i)</td><td>Table created from information given
<table>
<tr><td></td><td></td><td colspan="2">Positive health claim</td><td></td></tr>
<tr><td></td><td></td><td>Yes</td><td>No</td><td></td></tr>
<tr><td rowspan="2">Website targeted at</td><td>General population</td><td>13</td><td>33</td><td>46</td></tr>
<tr><td>Teenagers</td><td>8</td><td>16</td><td>24</td></tr>
<tr><td></td><td></td><td>21</td><td>49</td><td>70</td></tr>
</table>
P(positive health claim and targeted at general population)
$= \frac{13}{70}$
Therefore, the events “a website makes a positive health claim” and “a website is targeted at the general population” are not mutually exclusive, as the joint probability of these events is not zero.</td><td>Correct probability calculated.</td><td>Correct probability calculated.
AND
Correct reasoning about events not being mutually exclusive.</td><td></td></tr>
<tr><td>(ii)</td><td>P(health claim | teenager) $= \frac{8}{24} = 0.3333$
P(health claim | general population) $= \frac{13}{46} = 0.2826$
$\frac{0.3333}{0.2826} = 1.18$
Estimate that website is 1.18 times as likely to make a positive health claim if the website is targeted at teenagers, compared to if the website is targeted at the general population.
This is not more than twice as likely, so claim is not supported.
Note: This is NOT relative risk. Do not penalise a candidate for using RR in their answers, but this is not a risk question! Accept statement made that 0.333 is not double 0.2826 without calculation of ratio.</td><td>At least one correct conditional probability calculated.</td><td>Both conditional probabilities calculated and compared to reach conclusion that claim is not supported.</td><td></td></tr>
<tr><td>(b)(i)</td><td>P(correct) $= \frac{18+38+8}{100} = \frac{64}{100} = 64\%$</td><td>Correct percentage calculated.</td><td></td><td></td></tr>
</table>

(ii)	In addition to having a low rate of correct predictions (64%), only 8 / 22 of the people predicted to purchase Brand C actually purchased Brand C, and 18 / 35 of the people predicted to purchase Brand A actually purchased Brand A. So, the model has a low rate of correct predictions, both overall, and for two of the three brands. *Accept other valid reasoning.* *Note: Over half of the customers use Brand B (52 / 100), which the model has good prediction rate for (38 out of 43 predictions correct).*	At least one correct additional proportion calculated as part of reasoning.	At least one correct additional proportion calculated and used appropriately to support reasoning about a potential issue with the model.	At least two correct additional proportions calculated and used appropriately to support reasoning about a potential issue with the model.
(iii)	$3\times\frac{64}{100}\times\frac{63}{99}\times\frac{36}{98}=0.4488$ Need to assume the people were randomly selected, or assume that the events that each person was incorrectly predicted the brand of toothpaste, are independent, to calculate joint probability (use multiplicative principle). *Accept for Achievement:* $3\times 0.64\times 0.64\times 0.36$ (probability calculated using sampling with replacement).	Probability correctly calculated for only one arrangement OR probability calculated with replacement.	Correct probability calculated.	Correct probability calculated and explanation of need for independence provided.

NØ	N1	N2	A3	A4	M5	M6	E7	E8
No response; no relevant evidence.	Reasonable start / attempt at one part of the question.	1 of u	2 of u	3 of u	1 of r	2 of r	1 of t	2 of t

<table>
<tr><th>Q</th><th>Expected Coverage</th><th>Achievement (u)</th><th>Merit (r)</th><th>Excellence (t)</th></tr>
<tr><td>TWO
(a)</td><td>Using probability tree
P(North Island) = $0.129 \times 0.711 + 0.871 \times 0.77 =$ 0.7624
Using two-way table
P(North Island) $= \frac{266}{349} = 0.762$</td><td>Correct probability calculated.</td><td></td><td></td></tr>
<tr><td>(b)</td><td>P $(N \mid B) = 0.711$
P $(N) = 0.762$
Different answers suggest non-independence of events N and B.
Note we are dealing with a model / theoretical situation here, so do not accept discussion of the similarity of 0.711 and 0.762.</td><td>P $(N \mid B)$ correctly calculated.</td><td>P $(N \mid B)$ correctly calculated and explanation of non-independence of events.</td><td></td></tr>
<tr><td>(c)</td><td><table>
<tr><th></th><th>Boys</th><th>Girls</th><th>Co-ed</th><th>Total</th></tr>
<tr><td>Private</td><td>1</td><td>1</td><td>16</td><td>18</td></tr>
<tr><td>State</td><td>44</td><td>52</td><td>235</td><td>331</td></tr>
<tr><td>Total</td><td>45</td><td>53</td><td>251</td><td>349</td></tr>
</table>
P(state girls’ school) $= \frac{52}{349}$
The candidate may also approach the problem by logically working through the information provided (no tables or diagrams), and presenting this reasoning to support their answer.</td><td>Two correct counts / proportions related to events either shown in a table or diagram or in working.</td><td>Table or diagram completed or reasoning provided that correctly determines the number of state girls’ schools found.</td><td>Table or diagram completed or reasoning provided that correctly determines the number of state girls’ schools found
AND
correct probability calculated.</td></tr>
<tr><td>(d)(i)</td><td>Expected counts for a random sample of 50 schools
Year 9 to 13 – 33.8 schools
Other – 16.2 schools
The observed counts for the student are different from the expected counts.
Accept reasoning that 67% of schools are Year 9 to 13, but 50% of sample have schools Year 9 to 13.</td><td>Expected counts calculated.</td><td>Expected counts calculated and compared to observed counts, with explanation that the differences appear to be large.</td><td></td></tr>
</table>

(ii)	For this simulation, random sampling produced samples of 50 schools with 25 or fewer Year 9 to 13 schools only 9 times out of 1000 trials. This means the student's results would have been unlikely, given the student did randomly select the 50 secondary schools – only 0.9% of the time would you expect to see a count as low as 25 (or lower) just by random sampling from the 349 schools. The teacher could conclude that something more than just chance produced the observed result of 25 Year 9 to 13 schools in a sample of 50 secondary schools, e.g. maybe the student didn't follow the instructions, maybe the student recorded the results incorrectly, etc. *Note: It should not be concluded that this student definitely made up her sample results, just that it is very unlikely to get the result that she did (getting 25 Year 9 to 13 schools from a random sample of 50). Additionally, given that the whole class did this activity, you could expect to see an "unlikely" result!*	A correct description of the simulation proportion of 9 / 1000.	A correct interpretation of the simulation proportion of 9 / 1000 in terms of the unlikeliness of the observed result IF random sampling was used.	A correct interpretation of the simulation proportion of 9 / 1000 in terms of the unlikeliness of the observed result IF random sampling was used. AND A correct conclusion.

NØ	**N1**	**N2**	**A3**	**A4**	**M5**	**M6**	**E7**	**E8**
No response; no relevant evidence.	Reasonable start / attempt at one part of the question.	1 of u	2 of u	3 of u	1 of r	2 of r	1 of t	2 of t

Q	Expected Coverage	Achievement (u)	Merit (r)	Excellence (t)
THREE (a)	Percentage of world's population living in Asia = 59.7% Percentage of people living in Asia urban areas = 49% Percentage of people in the world that live in an urban area of Asia = $0.597 \times 0.49 = 0.2925$	Correct probability calculated.		
(b)(i)	The student in the question has possibly used the proportion of the South America population that live in urban areas (80%), and the proportion of the North America population that live in urban areas (81%), to claim "likeliness". These are conditional proportions based on what continent you live in, not based on whether you live in an urban or rural area. They have used the wrong conditional probabilities (confusion of the inverse).	Relevant percentages for South America and North America used in response.	Relevant percentages for South America and North America used in response, and confusion of the inverse or similar is explained. OR Correct probability calculated for (b)(ii).	Relevant percentages for South America and North America used in response, and confusion of the inverse or similar is explained. AND Correct probability calculated for (b)(ii).
(b)(ii)	P(North America or South America | Urban) $= \dfrac{0.057 \times 0.8 + 0.078 \times 0.81}{0.5453} = 0.1995$			
(c)(i)	Percentage of land area Europe = 6.8% Percentage of land area North America = 16.4% Combined land area elsewhere = 76.8% However, only 30.1% of Google Street View pictures are from Elsewhere. This shows that Google Street View may not have photos from locations everywhere in the world. *Note: Google Street View photos are not just taken from "streets" anymore!*	Relevant percentages used for either land areas OR game photo locations.	Relevant percentages used for land areas compare to percentages for game photo locations and conclusion made.	
(ii)	P (NA | ST) = P (NA and ST) / P(ST) $= \dfrac{0.297 \times 0.13}{0.402 \times 0.12 + 0.297 \times 0.13 + 0.301 \times 0.36} = 0.198$	One relevant probability correctly calculated as part of response.	Correct probability calculated.	
(iii)	The true probability of a Google Street View photo being taken in Europe is unknown. The *model estimate* for this true probability is 0.402. This *model estimate* is likely not to be the same as the true probability, as the model estimate was based on a large sample of photos. The data collected by the user can be used to calculate an *experimental estimate* of 0.30, which is lower than the *model estimate*.	The experimental estimate is calculated and described as such.	At least two of the three types of probability are described in the response.	All three types of probability are described in the response.

NØ	N1	N2	A3	A4	M5	M6	E7	E8
No response; no relevant evidence.	Reasonable start / attempt at one part of the question.	1 of u	2 of u	3 of u	1 of r	2 of r	1 of t	2 of t

Cut Scores

Not Achieved	Achievement	Achievement with Merit	Achievement with Excellence
0 – 6	7 – 13	14 – 18	19 – 24