Assessment Schedule – 2018 Final

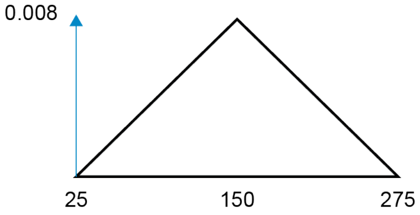
Mathematics and Statistics (Statistics): Apply probability distributions in solving problems (91586)

Evidence Statement

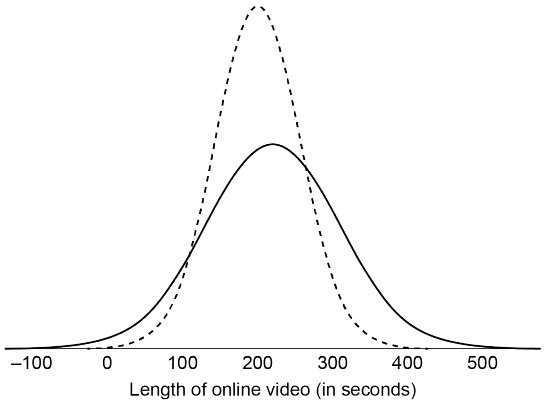
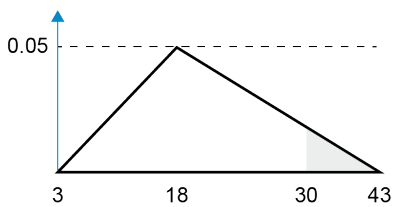
Q	Expected coverage	Achievement (u)	Merit (r)	Excellence (t)
ONE (a)(i)	Poisson distribution $\lambda = 1.3$ per hour $P(X \leq 1)$ $= 0.2725 + 0.3543 = 0.6268$	Correct probability calculated.		
(ii)	Poisson distribution $\lambda = 3.9$ per three hours (8 a.m. to 11 a.m.) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.0992 = 0.9008$	Correct probability calculated for (a)(ii). OR Calculated $\lambda = 5.2$, $P(X \geq 2)$ $= 1 - P(X \leq 1)$ $= 0.9657973$ OR At least one invalid assumption identified in context.	Correct probability calculated for (a)(ii). AND At least one invalid assumption identified in context	Correct probability calculated for (a)(ii). AND At least one invalid assumption identified and discussed in context for (a)(iii).
(iii)	Possible invalid assumptions: • Emails arrive randomly – more likely that emails arrive around set times, or use of automated emails sent by companies, so may not be random throughout the day. • The mean number of emails received per hour (or the rate) is constant – the number of emails received is more likely to vary with the time of day, the day of the week (work days vs weekend), and the time of the year. • Independence – emails received can be replies to emails sent, as part of an ongoing discussion, or as part of a group email. *Accept other reasonable specific discussion around why the Poisson model may not be suitable, linked to the conditions required to apply the distribution.* ***Do not accept that emails could arrive at the same time.***			
(b)(i)	$E(M) = 0 \times 0.09 + 1 \times 0.63 + 2 \times 0.22 + 3 \times 0.06 = 1.25$ Expected cost = \$130 × 1.25 = \$162.50	$E(M)$ correctly calculated.	Expected cost correctly calculated.	
(ii)	$VAR(M) = (0^2 \times 0.09 + 1^2 \times 0.63 + 2^2 \times 0.22 + 3^2 \times 0.06) - 1.25^2$ $= 0.488$ $SD(M) = \sqrt{0.488} = 0.698$ which is smaller than $SD(N) = 1.4$ One reason why $SD(N)$ is larger is that there is likely to be a greater range of outcomes for the random variable, e.g. people with more than three email accounts (3 is the max outcome for M). *Accept other reasons that account for greater variability for N.*	$SD(M)$ correctly calculated.	$SD(M)$ compared to $SD(N)$. AND One reason given for why $SD(N)$ is larger in context	
(iii)	$VAR(M) = 0.488$ *from (b)(ii)* $VAR(N) = 1.4^2 = 1.96$ $VAR(M + N) = 1.893^2 = 3.583$ $VAR(M) + VAR(N) = 0.488 + 1.96 = 2.448$ $2.448 \neq 3.583$ Therefore M and N are not independent.	$VAR(N)$ correctly calculated from $SD(N)$. OR $VAR(M + N)$ correctly calculated from $SD(M + N)$.	$VAR(M) + VAR(N)$ correctly calculated.	Correct explanation given as to why M and N are not independent, supported by correct calculations and statements.

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)
TWO (a)(i)	Binomial distribution $n = 12$, $p = 0.46$ $P(X \leq 6)$ $= 0.7157$	Correct probability calculated for (a)(i).	Correct probability calculated for (a)(i). AND Model identified as binomial and justified with two conditions linked to the context.	
(a)(ii)	Binomial because: • fixed number of trials (12 people) • fixed probability of success (46% of people do not revisit website) • only two outcomes (do not revisit or do revisit) • independent events (whether one person revisits the site does not affect another person visiting the site.).			
(a)(iii)	Let X be the number of visitors that revisit the website. $P(X \geq 1) = 1 - P(X = 0)$ Using two binomial distribution models Website A, $n = 10$, $p = 0.2$ Website B, $n = 10$, $p = 0.45$ For Website A: $P(X_A = 0) = 0.1074$ For Website B: $P(X_B = 0) = 0.0025$ P(zero visitors revisit both A & B) = $0.1074 \times 0.0025 = 0.0003$ $P(X \geq 1) = 1 - 0.0003 = 0.9997$ Independence can be assumed as visitors were randomly redirected to one of the versions of the website.	One correct probability calculated using either of the binomial distribution models.	Correct probability calculated for $P(X = 0)$.	Correct probability calculated for $P(X \geq 1)$. AND Assumption of independence discussed in reference to the use of random redirection.
(b)(i)	Using the counts for between 75 and 225: $\frac{80+125+85}{365} = 79.5\%$. So the owner is wrong, as this is lower than 95%. For the lower limit of 35, there are 45 days in the 25 to 75 interval, so the proportion that is between 25 to 35 is approximately $\frac{10}{50} \times \frac{45}{365} = 2.5\%$. Similarly, for the upper limit of 260, there are 30 days in the 225 to 275 interval, so the proportion that is between 260 to 275 is approximately $\frac{15}{50} \times \frac{30}{365} = 2.5\%$. *Note: These calculations are based on linear interpolation using the uniform distribution, which of course is just an assumption! Accept calculations annotated on the graph.*	Proportion correctly calculated showing that the owner is wrong. OR a reasonable attempt to confirm lower and upper values for the central 95% of page views	Proportion correctly calculated showing that the owner is wrong AND a reasonable attempt to confirm lower and upper values for the central 95% of page views. OR Lower and upper values for the central 95% of page views are confirmed, supported by correct reasoning.	Proportion correctly calculated showing that the owner is wrong AND Lower and upper values for the central 95% of page views are confirmed, supported by correct reasoning.

(ii)	Triangular distribution with parameters a = 25, b = 275, c = 150  Accept a = 24.5, and b = 275.5 using continuity correction.	Triangular distribution parameters estimated from data.		
(iii)	As the number of page views is a discrete variable but the triangular distribution is a continuous variable, a continuity correction could be used. Without a continuity correction, $P(X > 150) = 0.5$. With a continuity correction, $P(X > 150.5) = 0.5 \times 124.5 \times \frac{124.5}{125} \times \frac{1}{125} = 0.496$. These two probabilities are similar, and given that the model itself is based on data from a previous year (365 days) and can only be used to calculate an **estimate** of the probability that the website receives more than 150 page views tomorrow, a continuity correction could not need to be used.	At least one probability correctly calculated or stated. OR Identify discrete variable but the triangular distribution is a continuous variable, a continuity correction could be used and calculates $P(X > 149.5)$ correctly.	Both probabilities correctly calculated or stated. OR $P(X > 1\ 50)$ correct, attempts to use continuity correction in second probability and correct explanation of whether a continuity correction needs to be used or comparing probabilities. OR Clear discussion around concept of estimate and use of the continuity correction.	

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)(i)	Sketch shows approx. symmetric bell shape, centred on 200 with “bell” between 100 and 300. Must be obviously narrower and higher than Company A distribution.  	Sketch demonstrates correct understanding of key features of normal distribution.		
(a)(ii)	Company A: $\mu = 220$, $\sigma = 90$, $P(X > 330) = 0.11081$ Company B: $\mu = 200$, $\sigma = 50$, $P(X > 330) = 0.00466$ Therefore, Company A’s model is better, as the probability is closer to 10%.	One correct probability calculated. OR One correct x value using inverse normal.	Both probabilities correctly calculated. AND Correct conclusion.	
(a)(iii)	Company A: $\mu = 220$, $\sigma = 90$ $P(100 < X < 150) = 0.1271$ $P(100 < X < 300) = 0.7218$ $\frac{P(100 < X < 150)}{P(100 < X < 300)} = \frac{0.1271}{0.7218} = 0.1761$ *Note: P(X < 150) = 0.2184 accept for Achievement and as part of calculation for conditional probability Merit.*	One correct probability calculated.	Two correct probabilities calculated. AND Attempt at conditional probability . Uses 0.2184 and then calculates conditional probability correctly	Correct conditional probability calculated.
(b)	 Using a triangular distribution model (shown above), the height at $x = 30$ is $\frac{13}{25} \times 0.05$ $P(X > 30) = 0.5 \times \frac{13}{25} \times 0.05 \times 13 = 0.169$ P(two advertisements both longer than 30 seconds) = $0.169^2 = 0.029$ *Note: No assumptions needed for independence due to random selection.*	Correct probability calculated for one advertisement longer than 30 seconds.	Correct probability calculated for both advertisements being longer than 30 seconds.	

(c)	Focus of reasons should be on whether they are realistic Possible reasons: • Boundedness – there is a physical limit to the number of raisins that can fit in the packet, while the normal distribution has an infinite upper limit of the normal- the packet can only hold so many raisins as it must be able to close and not break • Skew –the normal distribution assumes that the distribution of raisins per box is symmetrical however it is more realistic that there are more raisins in box than mean. The advertising standards/ the consumer guarantee act would charge company with false advertising if there were fewer than advertised i.e. have more raisins than they state in the packet rather than fewer • Under this model, 95% of boxes could be expected to contain between approximately 140 and 260 raisins each (calculated using two standard deviations either side of the mean). This is a large variation in number of raisins between boxes and not realistic for manufacturing processes. • Considering realistic quality control, packaging usually monitored by weight of the contents. For raisins, being small and tightly packed inside boxes, counting process seems unrealistic. *Accept other reasons that focus on the unrealistic nature of the model. Do not accept reasons based on the number of raisins being discrete and the normal distribution being continuous as this is not an issue of realism!*	At least one calculation or one description or sketch of the "real" distribution of raisins per box given as part of the discussion. OR ONE reason not linked to features of normal or vice versa	ONE reason discussed and supported by a calculation or description or sketch of the "real" distribution of raisins per box.	TWO distinct reasons discussed with each reason supported by a calculation or description or sketch of the "real" distribution of raisins per box.

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 – 7	8 – 14	15 – 18	19 – 24