Additional Assessment Materials

Summer 2021

Pearson Edexcel GCE in Mathematics

9MA0 (Applied) (Public release version)

Resource Set 1: Topic 4

Statistical distributions (Test 1)

General guidance to Additional Assessment Materials for use in 2021

Context

- Additional Assessment Materials are being produced for GCSE, AS and A levels (with the exception of Art and Design).
- The Additional Assessment Materials presented in this booklet are an optional part of the range of evidence teachers may use when deciding on a candidate's grade.
- 2021 Additional Assessment Materials have been drawn from previous examination materials, namely past papers.
- Additional Assessment Materials have come from past papers both published (those materials available publicly) and unpublished (those currently under padlock to our centres) presented in a different format to allow teachers to adapt them for use with candidate.

Purpose

- The purpose of this resource to provide qualification-specific sets/groups of questions covering the knowledge, skills and understanding relevant to this Pearson qualification.
- This document should be used in conjunction with the mapping guidance which will map content and/or skills covered within each set of questions.
- These materials are only intended to support the summer 2021 series.

1. Helen believes that the random variable C, representing cloud cover from the large data set, can be modelled by a discrete uniform distribution.

(a) Write down the probability distribution for C. **(2)**

The discrete uniform distribution implies that we will have n equally likely outcomes. We know that cloud cover data ranges from 0 to 8

c	0	1	2	3	4	5	6	7	8
$P(C=c)$	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9

(b) Using this model, find the probability that cloud cover is less than 50%. **(1)**

Cloud cover of less than 50% would be anything less than 4 Oktas, which is found by $P(C<4) = 4 \times \frac{1}{9} = \frac{4}{9} = 0.\dot{4}\dot{4}$

Helen used all the data from the large data set for Hurn in 2015 and found that the proportion of days with cloud cover of less than 50% was 0.315.

(c) Comment on the suitability of Helen's model in the light of this information. **(1)**

This is not a suitable model because the value she has found is less than the expected value from part (c) $(0.315 < 0.44)$.

(d) Suggest an appropriate refinement to Helen's model. **(1)**

Helen should alter her model to a non-uniform distribution based model which can change probabilities based on location and time

For example, one may expect a higher cloud coverage in summer compared to spring.

(Total for Question 1 is 5 marks)

2. In an experiment a group of children each repeatedly throw a dart at a target. For each child, the random variable H represents the number of times the dart hits the target in the first 10 throws.

Peta models H as B(10, 0.1).

(a) State two assumptions Peta needs to make to use her model. **(2)**

We are using a binomial distribution so we must assume that
- Each throw must be independent.
- The probability, p, must remain constant throughout ($P = 0.1$ in our case)
- there are a fixed number of trials ($n = 10$ in our case)

(b) Using Peta's model, find P($H \geq 4$). **(1)**

$P(H \geq 4) = 1 - P(H \leq 3) = 1 - 0.9872 = \underline{\underline{0.0128}}$

where 0.9872 is from binomial distribution table for $n = 10$, $x = 3$, and $p = 0.10$.

For each child the random variable F represents the number of the throw on which the dart first hits the target.

Using Peta's assumptions about this experiment,

(c) find P(F = 5). **(2)**

We have $(0.9)^4 \times (0.1) = \underline{\underline{0.066}}$, noting that we have 0.9^4 for the four misses and 0.1^1 for the one hit.

Tom assumes that in this experiment no child will need more than 10 throws for the dart to hit the target for the first time. He models P($F = n$) as

$$P(F = n) = 0.01 + (n - 1) \times \alpha,$$

where α is a constant.

(d) Find the value of α **(4)**

n	1	2	3	...	10
$P(F=n)$	0.01	$0.01+\alpha$	$0.01+2\alpha$	...	$0.01+9\alpha$

We have an arithmetic sequence with $a = 0.01$ and $d = \alpha$.
We know that the sum of our arithmetic sequence / sum of probabilities will equal 1.

$\Rightarrow S_n = 1 = \frac{n}{2}(2a + (n-1)d \Rightarrow 5(0.02 + 9\alpha) = 1$

$\Rightarrow 9\alpha = \frac{1}{5} - 0.02$

$\Rightarrow \alpha = \underline{\underline{0.02}}$

(e) Using Tom's model, find P(F = 5)

(1)

$P(F=5) = 0.01 + (5-1)\,0.02 = \underline{0.09}$

(f) Explain how Peta's and Tom's models differ in describing the probability that a dart hits the target in this experiment.

(1)

- Peta assumes that the probability of hitt g the target is constant throughout
- Tom's model has the probability increa ng as n increases.

(Total for Question 2 is 11 marks)

---

3. The discrete random variable D has the following probability distribution

d	10	20	30	40	50
$P(D = d)$	$\frac{k}{10}$	$\frac{k}{20}$	$\frac{k}{30}$	$\frac{k}{40}$	$\frac{k}{50}$

where k is a constant.

(a) Show that the value of k is 600/137.

(2)

The probabilities must sum to 1. $\Rightarrow k\left(\frac{1}{10} + \frac{1}{0} + \frac{1}{30} + \frac{1}{40} + \frac{1}{50}\right) = 1$

$\Rightarrow k(137/600) = 1 \Rightarrow k = \frac{600}{137}$ as require

The random variables $D1$ and $D2$ are independent and each have the same distribution as D.

(b) Find P ($D_1 + D_2 = 80$)

Give your answer to 3 significant figures.

(3)

d = 80, and there are 3 options of d which add to 80.

1. 50 and 30, 2. 40 and 40 or 3. 30 and 50

Therefore we have $2 \times \frac{k}{50} \times \frac{k}{30} + \frac{k}{40} \times \frac{k}{40} = \underline{0.0376}$ using $k = \frac{600}{137}$

$\Rightarrow P(D_1 + D_2 = 80) = \underline{0.0376}$

A single observation of *D* is made.

The value obtained, *d*, is the common difference of an arithmetic sequence.

The first 4 terms of this arithmetic sequence are the angles, measured in degrees, of quadrilateral *Q*

(*c*) Find the exact probability that the smallest angle of *Q* is more than 50° **(5)**

Our angles are a, $a+d$, $a+2d$ and $a+3d$.

The sum $S_4 = 4a + 6d = 360$ $\left(S_4 = \frac{4}{2}(2a + 3d) = 4a + 6d\right)$ and o since 60°. in Q.

$= 2a + 3d = 180$

Then the smallest angle $a > 50$ is given by $d = 10$ and $a = 75$ or

$d = 20$ and $a = 60$

Then $P(D = 10 \text{ or } D = 20) = \frac{3k}{20} = \frac{90}{137}$.

$d = 30$ and $a = 45$ but this doesn't work since a must be greater than 50.

We leave our answer as a fraction since the question requires an exact probability.

(Total for Question 3 is 10 marks)

4. A machine puts liquid into bottles of perfume. The amount of liquid put into each bottle, *D* ml, follows a normal distribution with mean 25 ml

Given that 15% of bottles contain less than 24.63 ml

(a) find, to 2 decimal places, the value of *k* such that P(24.63 < *D* < *k*) = 0.45 **(5)**

A random sample of 200 bottles is taken.

We have $X \sim N(25, \sigma^2)$, so we must find σ^2 first.

Then we know that, for 15%, $Z = -1.0364$ from the 0.1500 (negative) on the formula sheet.

$\Rightarrow \frac{24.63 - 25}{\sigma} = -1.0364 \Rightarrow \sigma = 0.357.$

$\Rightarrow X \sim N(25, 0.357^2)$ then $P(24.63 < D < k) = 0.45 \Leftrightarrow P(D < k) - P(D < 24.63) = 0.45$

$\Rightarrow P(D < k) - 0.15 = 0.45 \Rightarrow P(D < k) = 0.60$ and finding 0.60 in the table

is $0.25. \Rightarrow \frac{k - 25}{0.357} = 0.25 \Rightarrow k = 25.089\ldots = \underline{25.09}$ ml

(b) Using a normal approximation, find the probability that fewer than half of these bottles contain between 24.63 ml and k ml

Recall that we have $n = 200$, and $X \sim B(200, 0.45)$ with $p = 0.45$ from (a). **(3)**

Then $\mu = np = 200 \times 0.45 = 90$ and $\sigma = npq = 200 \times 0.45 \times 0.55 = 49.5 \Rightarrow$ the normal approximation will be $Y \sim N(90, 49.5)$.

Then $P(Y < 100) = P(Y < 99.5)$ with continuity correction.

Then $P\left(Z < \frac{99.5 - 90}{\sqrt{49.5}}\right) = P(Z < 1.35) = \Phi(1.35) = 0.91149 = \underline{0.911}$

The machine is adjusted so that the standard deviation of the liquid put in the bottles is now 0.16 ml

Following the adjustments, Hannah believes that the mean amount of liquid put in each bottle is less than 25 ml

She takes a random sample of 20 bottles and finds the mean amount of liquid to be 24.94 ml

(c) Test Hannah's belief at the 5% level of significance.
You should state your hypotheses clearly.

(5)

$H_0 : \mu = 25$ ml v.s. $H_0 : \mu < 25$ (one sided test)

We now have $X \sim N\left(25, \frac{0.16^2}{20}\right)$

Then $P(X < 24.94) = P\left(Z < \frac{24.94 - 25}{\sqrt{\frac{0.16^2}{20}}}\right) = P(Z < -1.68) = 1 - \Phi(1.68) = 1 - 0.95352$
$= \underline{0.04648}$

$\alpha = 0.05 > 0.04648 \Rightarrow$ Reject H_0 and conclude that there is significant evidence to support Hannah's belief.

(Total for Question 4 is 13 marks)

5 A health centre claims that the time a doctor spends with a patient can be modelled by a normal distribution with a mean of 10 minutes and a standard deviation of 4 minutes.

(*a*) Using this model, find the probability that the time spent with a randomly selected patient is more than 15 minutes. **(1)**

We have $X \sim N(10, 4^2)$, then $P(X > 15) = P(Z > \frac{15-10}{4}) = P(Z < 1.25)$

Then $P(Z < 1.25) = 1 - \Phi(1.25) = 1 - 0.89435 = 0.10565 = \underline{\underline{0.106}}$

Some patients complain that the mean time the doctor spends with a patient is more than 10 minutes.

The receptionist takes a random sample of 20 patients and finds that the mean time the doctor spends with a patient is 11.5 minutes.

(*b*) Stating your hypotheses clearly and using a 5% significance level, test whether or not there is evidence to support the patients' complaint. **(4)**

$H_0 : \mu = 10$ mins v.s. $H_1 : \mu > 10$ mins (one sided test)

We have $X \sim N(10, \frac{4^2}{20}) \Rightarrow P(X > 11.5) = P\left(Z > \frac{11.5 - 10}{\sqrt{4^2/20}}\right) = P(Z > 1.67)$

then $P(Z > 1.67) = 1 - \Phi(1.67) = 1 - 0.95254 = 0.04746 < 0.05 = \alpha$

$\Rightarrow$ Evidence to support claim $\Rightarrow$ Reject $H_0 \Rightarrow \mu > \underline{\underline{10}}$ mins.

The health centre also claims that the time a dentist spends with a patient during a routine appointment, T minutes, can be modelled by the normal distribution where $T \sim N(5, 3.5^2)$

(*c*) Using this model,

(i) find the probability that a routine appointment with the dentist takes less than 2 minutes **(1)**

$P(T < 2) = P\left(Z < \frac{2-5}{3.5}\right) = P(Z < -0.86) = \Phi(-0.86) = 0.19489$

$\Rightarrow P(T < 2) = \underline{\underline{0.195}}$

(ii)find P ($T < 2$ | $T > 0$)

$P(T<2) - P(T<0) = 0.195 - 0.1184$

(3)

$$P(T<2 \mid T>0) = \frac{P(0<T<2)}{P(T>0)} = \frac{0.1184}{0.9234} = 0.128$$

(iii) hence explain why this normal distribution may not be a good model for T. **(1)**

We account for negative times, but such time does not exist.

The dentist believes that she cannot complete a routine appointment in less than 2 minutes.

She suggests that the health centre should use a refined model only including values of $T > 2$

(*d*) Find the median time for a routine appointment using this new model, giving your answer correct to one decimal place.

(5)

$P(T>t \mid T>2) = 0.5$

$P(T>t) = 0.5 \times P(T>2)$

$P(T>t) = 0.5 \times 0.8043$

$P(T>t) = 0.40215.$

Then 0.40 corresponds to $z = 0.25$

$$\frac{t-5}{3.5} = 0.25 \Rightarrow t = 5.9 \text{ mins}$$

(Total for Question 5 is 15 marks)