Additional Assessment Materials

Summer 2021

Pearson Edexcel GCE in As Mathematics

8MA0_21 (Public release version)

Resource Set 1: Topic 1

Statistical sampling

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

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- Additional Assessment Materials have come from past papers both published (those materials available publicly) and unpublished (those currently under padlock to our centres) presented in a different format to allow teachers to adapt them for use with candidate.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. Sara is investigating the variation in daily maximum gust, t kn, for Camborne in June and July 1987.

   She used the large data set to select a sample of size 20 from the June and July data for 1987. Sara selected the first value using a random number from 1 to 4 and then selected every third value after that.

   (*a*) State the sampling technique Sara used.

   (1)

   (*b*) From your knowledge of the large data set, explain why this process may not generate a sample of size 20.

   (1)

   The data Sara collected are summarised as follows

   $$n = 20 \qquad \Sigma\, t = 374 \qquad \Sigma\, t^2 = 7600$$

   (*c*) Calculate the standard deviation.

   (2)

   (Total for Question 1 is 4 marks)

2. A sixth form college has 84 students in Year 12 and 56 students in Year 13

   The head teacher selects a stratified sample of 40 students, stratified by year group.

   (*a*) Describe how this sample could be taken.

   (3)

   The head teacher is investigating the relationship between the amount of sleep, s hours, that each student had the night before they took an aptitude test and their performance in the test, p marks.

   For the sample of 40 students, he finds the equation of the regression line of p on s to be

   $$p = 26.1 + 5.60s$$

   (*b*) With reference to this equation, describe the effect that an extra 0.5 hours of sleep may have, on average, on a student's performance in the aptitude test.

   (1)

   (*c*) Describe one limitation of this regression model.

   (1)

   (Total for Question 2 is 5 marks)

3. A lake contains three different types of carp.

There are an estimated 450 mirror carp, 300 leather carp and 850 common carp.

Tim wishes to investigate the health of the fish in the lake.

He decides to take a sample of 160 fish.

(*a*)　Give a reason why stratified random sampling cannot be used.

(1)

(*b*)　Explain how a sample of size 160 could be taken to ensure that the estimated

　populations of each type of carp are fairly represented.

　You should state the name of the sampling method used.

(2)

As part of the health check, Tim weighed the fish.

His results are given in the table below.

Weight (*w* kg)	Frequency (f)	Midpoint (*m* kg)
$2 \leq w < 3.5$	8	2.75
$3.5 \leq w < 4$	32	3.75
$4 \leq w < 4.5$	64	4.25
$4.5 \leq w < 5$	40	4.75
$5 \leq w < 6$	16	5.5

(You may use $\sum fm = 692$　and　$\sum fm^2 = 3053$)

(*c*)　Calculate an estimate for the standard deviation of the weight of the carp.

(2)

Tim realised that he had transposed the figures for 2 of the weights of the fish.

He had recorded in the table 2.3 instead of 3.2 and 4.6 instead of 6.4

(d)   Without calculating a new estimate for the standard deviation, state what effect

        (i) using the correct figure of 3.2 instead of 2.3

        (ii) using the correct figure of 6.4 instead of 4.6

    would have on your estimated standard deviation.

    Give a reason for each of your answers.

(2)

(Total for Question 3 is 7 marks)

4. Sara was studying the relationship between rainfall, r mm, and humidity, h %, in the UK. She takes a random sample of 11 days from May 1987 for Leuchars from the large data set.

She obtained the following results.

h	93	86	95	97	86	94	97	97	87	97	86
r	1.1	0.3	3.7	20.6	0	0	2.4	1.1	0.1	0.9	0.1

Sara examined the rainfall figures and found

$$Q_1 = 0.1 \qquad Q_2 = 0.9 \qquad Q_3 = 2.4$$

A value that is more than 1.5 times the interquartile range (IQR) above Q_3 is called an outlier.

(*a*) Show that $r = 20.6$ is an outlier.

(1)

(*b*) Give a reason why Sara might   (i) include

(ii) exclude

this day's reading.

(2)

Sara decided to exclude this day's reading and drew the following scatter diagram for the remaining 10 days' values of r and h.
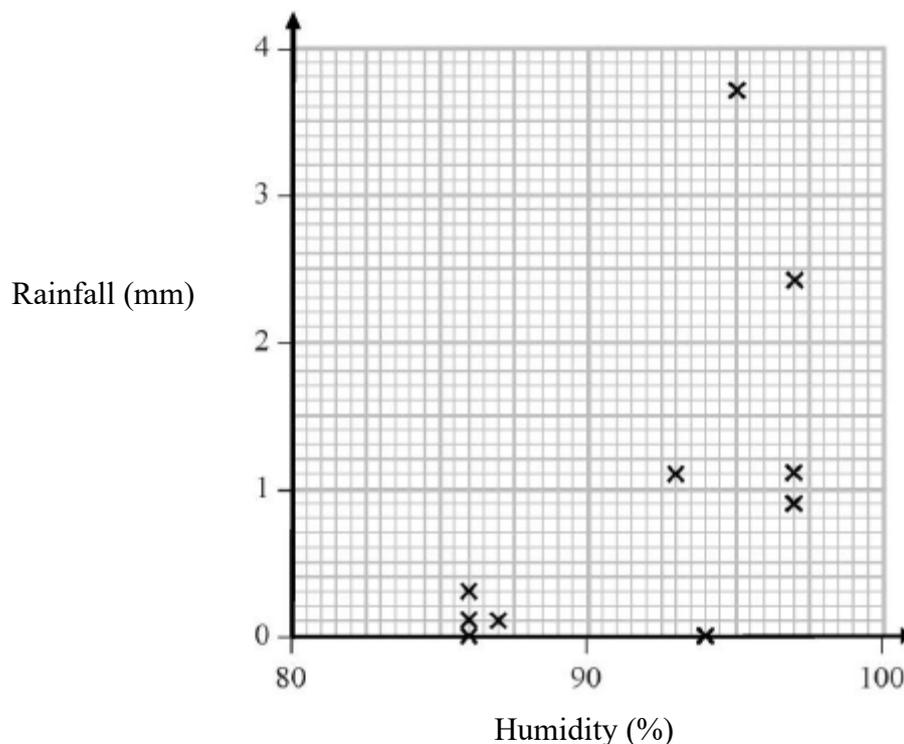
(*c*) Give an interpretation of the correlation between rainfall and humidity.

(1)

The equation of the regression line of r on h for these 10 days is $r = -12.8 + 0.15h$.

(*d*) Give an interpretation of the gradient of this regression line.

(1)

(*e*) (i) Comment on the suitability of Sara's sampling method for this study.

(ii) Suggest how Sara could make better use of the large data set for her study.

(2)

(Total for Question 4 is 7 marks)