Additional Assessment Materials

Summer 2021

Pearson Edexcel GCE in Mathematics

9MA0 (Applied) (Public release version)

Resource Set 1: Topic 5

Statistical hypothesis testing

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

Additional Assessment Materials, Summer 2021

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- Additional Assessment Materials have come from past papers both published (those materials available publicly) and unpublished (those currently under padlock to our centres) presented in a different format to allow teachers to adapt them for use with candidate.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. Tessa owns a small clothes shop in a seaside town. She records the weekly sales figures, £ w, and the average weekly temperature, t °C, for 8 weeks during the summer.

The product moment correlation coefficient for these data is −0.915.

(a) Stating your hypotheses clearly and using a 5% level of significance, test whether or not the correlation between sales figures and average weekly temperature is negative.

(3)

(b) Suggest a possible reason for this correlation.

(1)

Tessa suggests that a linear regression model could be used to model these data.

(c) State, giving a reason, whether or not the correlation coefficient is consistent with Tessa's suggestion.

(1)

(d) State, giving a reason, which variable would be the explanatory variable.

(1)

Tessa calculated the linear regression equation as $w = 10\,755 - 171t$.

(e) Give an interpretation of the gradient of this regression equation.

(1)

(Total for Question 1 is 7 marks)

---

2. An ornithologist believes that there is a relationship between the tail length, t mm, and the wing length, w mm, of female hook-billed kites. A random sample of size 10 is taken from a database of these kites and the relevant data is given in the table below.

t (mm)	191	197	208	180	188	210	196	191	179	208
w (mm)	284	285	288	273	280	283	288	271	257	289

The ornithologist plans to use a linear regression model based on these data and interpolate or extrapolate as necessary to estimate the wing length of other female hook-billed kites from their tail length.

(a) (i) Explain what is meant by extrapolation.

(1)

(ii) Explain the dangers of extrapolation.

(1)

The ornithologist attempts to calculate the product moment correlation coefficient, r, and obtains a value of 1.3.

(b) Explain how she should be able to identify that this is incorrect without carrying out any further calculations.

(1)

(c) Use your calculator to find the correct value of the product moment correlation coefficient, r.

(1)

(d) Stating your hypotheses clearly test, at the 1% significance level, whether or not there is evidence that the product moment correlation coefficient for the population is positive.

(3)

(e) Explain what your test in part (d) suggests about female hook-billed kites.

(1)

(Total for Question 2 is 8 marks)

3. Sam is investigating the weather throughout the year in Hurn.

He uses the large data set to investigate the daily mean wind direction.

Sam believes that each cardinal wind direction is equally likely in Hurn.

(a) Assuming that Sam is correct,

(i) state the probability that the cardinal wind direction in Hurn on a randomly selected day is NNE,

(1)

(ii) state the distribution that Sam should use to model the probability of each cardinal wind direction in Hurn on a randomly selected day.

(1)

Sam decides to investigate the daily mean wind direction throughout the year.

(b) State a limitation of using the data for Hurn from the large data set as a sampling frame.

(1)

(c) Explain how to use simple random sampling to select 36 days from a year.

(2)

Sam defines the random variable X as the number of days out of 36 on which the daily mean wind direction in Hurn is between the bearings 135° and 225°.

Sam collects data from 36 randomly selected days and finds that $x = 15$.

Sam carries out a hypothesis test at the 10% level of significance.

(d) Given that $H_0 : p = 0.25$ and that the critical region is $\{X \leq 4 \cup X \geq 14\}$,

(i) state the alternative hypothesis, H_1

(ii) giving a reason for your answer, explain what Sam should conclude about the daily mean wind direction in Hurn.

(3)

(Total for Question 3 is 8 marks)

4. A meteorologist believes that there is a relationship between the daily mean windspeed, w kn, and the daily mean temperature, t °C. A random sample of 9 consecutive days is taken from past records from a town in the UK in July and the relevant data is given in the table below.

t	13.3	16.2	15.7	16.6	16.3	16.4	19.3	17.1	13.2
w	7	11	8	11	13	8	15	10	11

The meteorologist calculated the product moment correlation coefficient for the 9 days and obtained $r = 0.609$.

(a) Explain why a linear regression model based on these data is unreliable on a day when the mean temperature is 24 °C.

(1)

(b) State what is measured by the product moment correlation coefficient.

(1)

(c) Stating your hypotheses clearly test, at the 5% significance level, whether or not the product moment correlation coefficient for the population is greater than zero.

(3)

Using the same 9 days, a location from the large data set gave $\bar{t} = 27.2$ and $\bar{w} = 3.5$.

(d) Using your knowledge of the large data set, suggest, giving your reason, the location that gave rise to these statistics.

(1)

(Total for Question 4 is 6 marks)

---

5. Barbara is investigating the relationship between average income (GDP per capita), x US dollars, and average annual carbon dioxide (CO_2) emissions, y tonnes, for different countries.

She takes a random sample of 24 countries and finds the product moment correlation coefficient between average annual CO_2, emissions and average income to be 0.446

(a) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the product moment correlation coefficient for all countries is greater than zero.

(3)

Barbara believes that a non-linear model would be a better fit to the data.
She codes the data using the coding $m = \log_{10} x$ and $c = \log_{10} y$ and obtains the model $c = -1.82 + 0.89m$

The product moment correlation coefficient between c and m is found to be 0.882

(b) Explain how this value supports Barbara's belief.

(1)

(c) Show that the relationship between y and x can be written in the form $y = ax^n$ where a and n are constants to be found.

(5)

(Total for Question 5 is 9 marks)

6. A company sells seeds and claims that 55% of its pea seeds germinate.

(*a*) Write down a reason why the company should not justify their claim by testing all the pea seeds they produce.

(1)

A random selection of the pea seeds is planted in 10 trays with 24 seeds in each tray.

(*b*) Assuming that the company's claim is correct, calculate the probability that in at least half of the trays 15 or more of the seeds germinate.

(3)

(*c*) Write down two conditions under which the normal distribution may be used as an approximation to the binomial distribution.

(1)

A random sample of 240 pea seeds was planted and 150 of these seeds germinated.

(*d*) Assuming that the company's claim is correct, use a normal approximation to find the probability that at least 150 pea seeds germinate.

(3)

(*e*) Using your answer to part (*d*), comment on whether or not the proportion of the company's pea seeds that germinate is different from the company's claim of 55%

(1)

(Total for Question 6 is 9 marks)

---

7. The lifetime, L hours, of a battery has a normal distribution with mean 18 hours and standard deviation 4 hours.

Alice's calculator requires 4 batteries and will stop working when any one battery reaches the end of its lifetime.

(a) Find the probability that a randomly selected battery will last for longer than 16 hours.

(1)

At the start of her exams Alice put 4 new batteries in her calculator. She has used her calculator for 16 hours, but has another 4 hours of exams to sit.

(b) Find the probability that her calculator will not stop working for Alice's remaining exams.

(5)

Alice only has 2 new batteries so, after the first 16 hours of her exams, although her calculator is still working, she randomly selects 2 of the batteries from her calculator and replaces these with the 2 new batteries.

(c) Show that the probability that her calculator will not stop working for the remainder of her exams is 0.199 to 3 significant figures.

(3)

After her exams, Alice believed that the lifetime of the batteries was more than 18 hours. She took a random sample of 20 of these batteries and found that their mean lifetime was 19.2 hours.

(d) Stating your hypotheses clearly and using a 5% level of significance, test Alice's belief.

(5)

(Total for Question 7 is 14 marks)