Pearson Edexcel

Additional Assessment Materials

Summer 2021

Pearson Edexcel GCE in Mathematics

9MA0 (Applied) (Public release version)

Resource Set 1: Topic 4

Statistical distributions (Test 2)

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

All the material in this publication is copyright

© Pearson Education Ltd 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- Additional Assessment Materials have come from past papers both published (those materials available publicly) and unpublished (those currently under padlock to our centres) presented in a different format to allow teachers to adapt them for use with candidate.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. A local sports centre has showers with two temperature settings, warm and hot.

On the warm setting, the water temperature may be modelled by a Normal distribution with mean 30 °C and standard deviation 2 °C.

(a) Using the model, find the probability that the next time the shower is used on the warm setting ,the water temperature is

(i) exactly 31 °C,

(ii) more than 31 °C.

(2)

i) We have that $X \sim N(30, 2^2)$ and we want to find $P(X=31)$.

$P(X=31) = P\left(Z = \frac{31-30}{2}\right) = P\left(Z = \frac{1}{2}\right) = 0$ from the normal distribution table.

ii) Now we want to find $P(X>31) = P\left(Z > \frac{31-30}{2}\right) = P(Z > 1/2) = 1 - \Phi\left(\frac{1}{2}\right) = 1 - 0.69146$

$= \underline{\underline{0.309}}$

The sports centre manager thinks that a water temperature of more than 33 °C is too high for the warm setting.

She tests the water temperature on the warm setting on 5 randomly selected days.

Given that the probability of the water temperature being more than 33 °C is 0.0668.

(b) find the probability of the water temperature being more than 33 °C

(i) on only the first of these 5 days,

(2)

We have a binomial distibution with $n=5$ and $p = 0.0668$.

We find the probability by doing $p^1(1-p)^4$ Since we have temp $>33°C$ on 1 day and not $>33°C$ on the 4 others. $\Rightarrow$ $0.0668(1-0.0668)^4 = \underline{\underline{0.0507}}$

(ii) on more than 1 of these 5 days. $n=5,\ p=0.00668$

(3)

let $X \sim B(5, 0.0668)$ then $P(X>1) = P(X \leq 1) = 1 - 0.9610$

$= \underline{\underline{0.0390}}$

On the hot setting, the water temperature may be modelled by a Normal distribution with standard deviation 1.5 °C.

The probability that the water temperature is more than 42 °C is 0.0005.

(c) Find the mean water temperature on this setting, giving your answer to 1 decimal place.

(4)

We know that $X \sim N(\mu, 1.5^2)$ and that $P(X>42) = 0.0005$.

$\Rightarrow P\left(Z > \frac{42-\mu}{2}\right) = 0.0005$.

Then the z-value for 0.0005 from the normal distibution table is 3.2905.

Then $3.2905 = \frac{42-\mu}{2} \Rightarrow \mu = \underline{\underline{37.1°C}}$

(Total for Question 1 is 11 marks)

---

2. A machine cuts strips of metal to length L cm, where L is normally distributed with standard deviation 0.5 cm.

Strips with length either less than 49 cm or greater than 50.75 cm **cannot** be used.

Given that 2.5% of the cut lengths exceed 50.98 cm,

(*a*) find the probability that a randomly chosen strip of metal **can** be used.

(5)

We have that $L \sim N(\mu, 0.5^2)$ and $P(L > 50.98) = 0.025$.

$\Rightarrow P\left(Z > \frac{50.98 - \mu}{0.5}\right) = 0.025 \Rightarrow Z = 1.9600$ from normal distribution table.

Then $\frac{50.98 - \mu}{0.5} = 1.9600 \Rightarrow \mu = 50 \Rightarrow L \sim N(50, 0.5^2)$

$\Rightarrow P(49 < L < 50.75) = P(L < 50.75) - P(L < 49) = 0.9332 - 0.0228 = \underline{0.9104}$

$P\left(Z < \frac{50.75-50}{0.5}\right) = \Phi(1.5)$ $P\left(Z < \frac{49-50}{0.5}\right) = 1 - \Phi(2)$

Ten strips of metal are selected at random.

(*b*) Find the probability fewer than 4 of these strips **cannot** be used.

(2)

The probability that a strip cannot be used will be $1 - 0.9104 = 0.0896$.

Then $X \sim B(10, 0.0896)$ and we want $P(X \leq 3) = \underline{0.991}$ from binomial distribution.

A second machine cuts strips of metal of length X cm, where X is normally distributed with standard deviation 0.6 cm.

A random sample of 15 strips cut by this second machine was found to have a mean length of 50.4 cm.

(*c*) Stating your hypotheses clearly and using a 1% level of significance, test whether or not the mean length of all the strips, cut by the second machine, is greater than 50.1 cm.

(5)

$H_0: \mu = 50.1\text{cm}$ v.s. $H_1: \mu > 50.1\text{cm}$

Then $\bar{X} \sim N\left(50.1, \frac{0.6^2}{15}\right) \Rightarrow P(\bar{X} > 50.4) = P\left(Z > \frac{50.4 - 50.1}{\sqrt{0.6^2/15}}\right) = P(Z > 1.94) = 1 - \Phi(1.94)$ $= \underline{0.02619}$

Then $0.02619 > 0.01$ $\alpha = 1\%$ sig. level

$\Rightarrow$ we do not reject H_0, and we conclude that there is insufficient evidence than $\underline{\mu > 50.1 \text{ cm}}$.

(Total for Question 2 is 12 marks)

3. A large number of cyclists take part in a cycling time trial.

A random sample of these cyclists are selected and their times, in minutes, are summarised in the following statistics

$$\sum x = 1680, \qquad \sum x^2 = 47654.4, \qquad n = 60$$

(a) Calculate, for this sample, the value of

(i) the mean time,

let the mean be $\bar{x}$, then $\bar{x} = \frac{\sum x}{n} = \frac{1680}{60} = 28$ minutes. **(1)**

(ii) the standard deviation of the times.

$\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{47654.4}{60} - 28^2} = \underline{3.2}$ minutes **(2)**

Historically, the mean time for cyclists on this time trial has been 27 minutes and 30 seconds.

Lucy is watching the time trial and believes that the mean time of cyclists in this time trial is greater than the mean time of cyclists in previous time trials.

The times of cyclists on this time trial are modelled by a Normal distribution with standard deviation 3 minutes.

(b) Test, at the 5% level of significance, whether or not this sample provides evidence to support Lucy's belief. You should state your hypotheses and show your working clearly.

let $H_0 : \mu = 27.5$ v.s. $H_1 : \mu > 27.5$ and $X \sim N\left(27.5, \frac{3^2}{60}\right)$ **(5)**

Then $P(X > 28) = P\left(Z > \frac{28-27.5}{\sqrt{3^2/60}}\right) = P(Z > 1.29) = \Phi(1.29) = 1 - 0.90147 = 0.09853 = \underline{0.10}$

Then at $\alpha = 0.05$, $0.05 < 0.10 \Rightarrow$ do not reject H_0 and we conclude that there is insufficient evidence to support Lucy's claim.

Speedy Wheels cycling club entered its 5 fastest riders and 5 beginners to take part in the time trial.

The fastest 20% of the cyclists in the time trial are invited to compete in a race the following week.

(i) Explain, with specific reference to the parameter p, why the distribution B(10,0.2) might not be reasonable to model the number of these *Speedy Wheels* cycling club members who are invited to compete in the race.

(2)

The assumption of Constant probability is unreasonable as the 5 fastest riders and 5 beginners will have differing chances, i.e. the fastest riders will be more likely to have a faster time.

(ii) Suggest how to improve the model for the number of these *Speedy Wheels* cycling club members invited to compete in the race.

We should use two independent binomial models with different values of p for each type of rider. **(1)**

(Total for Question 3 is 11 marks)

---

4. A company has a customer services call centre. The company believes that the time taken to complete a call to the call centre may be modelled by a normal distribution with mean 16 minutes and standard deviation σ minutes.

Given that 10% of the calls take longer than 22 minutes,

(a) show that, to 3 significant figures, the value of σ is 4.68.

(3)

$X \sim N(16, \sigma^2) \Rightarrow P(X > 22) = 0.1 \Rightarrow P\left(Z > \frac{22-16}{\sigma}\right) = 0.1 \Rightarrow P\left(Z > \frac{6}{\sigma}\right) = 0.1$.

The z-value for 0.1 is 1.28 $\Rightarrow 6/\sigma = 1.282 \Rightarrow \sigma = 4.68$ as required.

(b) Calculate the percentage of calls that take less than 13 minutes.

(1)

We want $P(L < 13) = P\left(Z < \frac{16-13}{4.68}\right) = P(Z < 0.64) = 1 - \Phi(0.64) = 1 - 0.73891 = 0.261$

$\Rightarrow$ 26.1%.

A supervisor in the call centre claims that the mean call time is less than 16 minutes. He collects data on his own call times.

- 20% of the supervisor's calls take more than 17 minutes to complete.
- 10% of the supervisor's calls take less than 8 minutes to complete.

Assuming that the time the supervisor takes to complete a call may be modelled by a normal distribution,

(c) estimate the mean and the standard deviation of the time taken by the supervisor to complete a call.

(6)

We know that $P(X > 17) = 0.2$ and $P(X < 8) = 0.1$

$P(X > 17) = P\left(Z > \frac{17-\mu}{\sigma}\right) = 0.2 \Rightarrow Z = 0.8416$ from normal dist. table,

and $P(X < 8) = P\left(Z < \frac{8-\mu}{\sigma}\right) = 0.1 \Rightarrow Z = -1.2816$

$\Rightarrow \frac{17-\mu}{\sigma} = 0.8416$ and $\frac{8-\mu}{\sigma} = -1.2816$ then from solving these

using simillaneous equations we get that $\sigma = 4.24 \Rightarrow \mu = 13.43$.

$\Rightarrow$ Mean is 13.43 minutes and standard deviation is 4.24 minutes.

(d) State, giving a reason, whether or not the calculations in part (c) support the supervisor's claim.

(1)

Yes, it does since $\mu = 13.4$ mins which is less than 16 mins.

(Total for Question 4 is 11 marks)

---

5. A fast food company has a scratchcard competition. It has ordered scratchcards for the competition and requested that 45% of the scratchcards be winning scratchcards.

A random sample of 20 of the scratchcards is collected from each of 8 of the fast food company's stores.

(a) Assuming that 45% of the scratchcards are winning scratchcards, calculate the probability that in at least 2 of the 8 stores, 12 or more of the scratchcards are winning scratchcards.

(5)

$X \sim B(20, 0.45)$ where X is the number of winning scratchcards.

let $Y \sim B(8, p)$ where Y is the number of fast food stores where there is a winning scratchcard. $x = 12, n = 20$ and $p = 0.45$

Then $p = P(X \geq 12) = 1 - P(X \leq 11) = 1 - 0.8692 = 0.1308$.

Then $P(Y \geq 2) = 1 - [P(Y=0) + P(Y=1)] = 1 - [0.3258 + 0.3922] = 0.282$

(b) Write down 2 conditions under which the normal distribution may be used as an approximation to the binomial distribution.

(1)

n must be large and p must be close to 0.5.

A random sample of 300 of the scratchcards is taken. Assuming that 45% of all the scratchcards are winning scratchcards,

(c) use a normal approximation to find the probability that at most 122 of these 300 scratchcards are winning scratchcards.

(4)

$n = 300,\ p = 0.45,\ q = 0.55$ then $X \sim N(np, npq) = N(135, 74.25)$.

and 0.5 for continuity correction

Then $P(X < 122.5) = P\left(Z < \frac{122.5 - 135}{\sqrt{74.25}}\right) = P(Z < -1.45) = \Phi(-1.45) = 0.07353$

Given that 122 of the 300 scratchcards are winning scratchcards,

(d) comment on whether or not there is evidence at the 5% significance level that the proportion of the company's scratchcards that are winning scratchcards is different from 45%. We have a two tailed test and $2 \times 0.07353 = 0.14706 > 0.05$.

=> We have insufficent evidence at $\alpha = 0.05$ to suggest that the proportion of scratchcards is different to 45%.

(Total for Question 5 is 10 marks)