Cambridge Assessment International Education

# Example Candidate Responses – Paper 4

## Cambridge International AS & A Level Further Mathematics 9231

For examination from 2022

Cambridge Pathway

# Contents

# Introduction

The main aim of this booklet is to exemplify standards for those teaching Cambridge International AS & A Level Further Mathematics 9231 and to show how different levels of candidates' performance relate to the subject's curriculum and assessment objectives.

In this booklet, candidate responses have been chosen from the June 2022 series to exemplify a range of answers for all the questions on the question paper.

For each question, the response is annotated with a clear explanation of where and why marks were awarded or omitted. This is followed by examiner comments on how the answer could have been improved. In this way, it is possible for you to understand what candidates have done to gain their marks and what they could do to improve their answers. There is also a list of common mistakes candidates made in their answers for each question.

This document provides illustrative examples of candidate work with examiner commentary. These help teachers to assess the standard required to achieve marks beyond the guidance of the mark scheme. Please also refer to the June 2022 Examiner Report for further detail and guidance.

The questions and mark schemes used here are available to download from the School Support Hub. These files are:

**9231 June 2022 Question Paper 43**

**9231 June 2022 Mark Scheme 43**

Past exam resources and other teaching and learning resources are available on the School Support Hub:

www.cambridgeinternational.org/support

## How to use this booklet

This booklet goes through the paper one question at a time. The candidate answers are set in a table. In the left-hand column are the candidate answers, and in the right-hand column are the examiner comments.

| Example Candidate Response – 1 | Examiner comments |
|---|---|
| **1** The times taken by members of a large quiz club to complete a challenge have a normal distribution with mean $\mu$ minutes. The times, $x$ minutes, are recorded for a random sample of 8 members of the club. The results are summarised as follows, where $\bar{x}$ is the sample mean.<br>$\bar{x} = 33.8 \qquad \Sigma(x-\bar{x})^2 = 94.5$<br>Find a 95% confidence interval for $\mu$. [4]<br>$n=8<30$ $\qquad S_x^2 = \frac{1}{n-1}(\Sigma(x-\bar{x})^2) = \frac{1}{7}\times 94.5$<br>$P=0.975 \rightarrow z=1.96$ **1** $\qquad = 13.5$<br>$33.8 \pm 1.96\cdot\sqrt{\frac{13.5}{8}}$<br>$33.8 \pm 2.546$<br>$31.3 \le \mu \le 36.3$<br>$\therefore [31.3, 36.3]$ is the confidence interval for $\mu$. | **1** The expression for the confidence interval is of the correct form, but this candidate makes the common and critical error of using the $z$-value 1.96 instead of the $t$-value 2.365. The information in the question concerns a small sample with unknown population variance and … is |

**Answers** are by real candidates in exam conditions. These show you the types of answers for each level. Discuss and analyse the answers with your learners in the classroom to improve their skills.

**Examiner comments** are alongside the answers. These explain where and why marks were awarded. This helps you to interpret the standard of Cambridge exams so you can help your learners to refine their exam technique.

# Question 1

| Example Candidate Response – 1 | Examiner comments |
|---|---|
| **1** The times taken by members of a large quiz club to complete a challenge have a normal distribution with mean $\mu$ minutes. The times, $x$ minutes, are recorded for a random sample of 8 members of the club. The results are summarised as follows, where $\overline{x}$ is the sample mean.<br><br>$\overline{x} = 33.8 \qquad \sum(x-\overline{x})^2 = 94.5$<br><br>Find a 95% confidence interval for $\mu$. [4]<br><br>$n=8<30$ $\qquad S_x^2 = \frac{1}{8-1}\left(\sum(x-\overline{x})^2\right) = \frac{1}{7}\times 94.5$<br>$P=0.975 \rightarrow z=1.96$ **1** $\qquad = 13.5$<br>$33.8 \pm 1.96\cdot\sqrt{\frac{13.5}{8}}$<br>$33.8 \pm 2.546$<br>$31.3 \leq \mu \leq 36.3$<br>$\therefore$ $[31.3, 36.3]$ is the confidence interval for $\mu$. | **1** The expression for the confidence interval is of the correct form, but this candidate makes the common and critical error of using the $z$-value 1.96 instead of the $t$-value 2.365. The information in the question concerns a small sample with unknown population variance and this means that a $z$-distribution is not valid.<br><br>**Total mark awarded = 1 out of 4** |

# Question 2

## Example Candidate Response – 1

2 A scientist is investigating the size of shells at various beach locations. She selects four beach locations and takes a random sample of shells from each of these beaches. She classifies each shell as large or small. Her results are summarised in the following table.

| | | Beach location | | | | |
|---|---|---|---|---|---|---|
| | | *A* | *B* | *C* | *D* | Total |
| Size of shell | Large | 68 | 69 | 96 | 81 | 314 |
| | Small | 28 | 55 | 64 | 39 | 186 |
| | Total | 96 | 124 | 160 | 120 | 500 |

Test, at the 10% significance level, whether the size of shell is independent of the beach location. [7]

$H_0$: size of shell is independent of beach location
$H_1$: size of shell is dependent of beach location

Expected values :

| | A | B | C | D |
|---|---|---|---|---|
| Large | 60.288 | 77.872 | 100.48 | 75.36 |
| Small | 35.712 | 46.128 | 59.52 | 44.64 |

test statistic $= \sum \frac{(O_i - E_i)^2}{E_i} = 7.041$
critical value $= \chi_3^2 (0.90) = 6.251$
$7.041 > 6.251$, reject $H_0$
There is sufficient evidence to prove that the size of shell is dependent of the beach location

## Examiner comments

1 The hypotheses and the calculations are all correct here. However, the conclusion illustrates a common error. The conclusion needs to be written both in context and with a level of uncertainty in the words used. A hypothesis test cannot prove anything; it only provides sufficient evidence to suggest. A statement such as ‘there is insufficient evidence to suggest that the size of shell is independent of the beach location’ is a more appropriate wording. In general, the words ‘prove’, ‘show that’ etc. should be avoided, as well as the even more definite statement ‘the size of the shell is dependent on the beach location’.

**Total mark awarded = 6 out of 7**

## Example Candidate Response – 2

**2** A scientist is investigating the size of shells at various beach locations. She selects four beach locations and takes a random sample of shells from each of these beaches. She classifies each shell as large or small. Her results are summarised in the following table.

| | | Beach location | | | | |
|---|---|---|---|---|---|---|
| | | *A* | *B* | *C* | *D* | Total |
| Size of shell | Large | 68 | 69 | 96 | 81 | 314 |
| | Small | 28 | 55 | 64 | 39 | 186 |
| | Total | 96 | 124 | 160 | 120 | 500 |

Test, at the 10% significance level, whether the size of shell is independent of the beach location. [7]

1 $H_0$: the size of shell is independent of the beach location.

$E=$

~~$E = 60.288 \to E$~~

$68 \Rightarrow E = 60.288$ | $69 \Rightarrow \frac{124\times314}{500} = 77.872$ | $96 \Rightarrow \frac{160\times314}{500} = 100.48$

$81 \Rightarrow \frac{120\times314}{500} = 75.36$ | $28 \Rightarrow \frac{96\times186}{500} = 35.712$ | $55 \Rightarrow \frac{124\times186}{500} = 46.128$

$64 \Rightarrow \frac{160\times186}{500} = 59.52$, $39 \Rightarrow \frac{120\times186}{500} = 44.64$

$X^2 = \sum \frac{(O-E)^2}{E}$

$X^2 \Rightarrow = \cancel{60.288}\ \frac{(68-60.288)^2}{60.288} + \frac{(69-77.872)^2}{77.872} + \frac{(96-100.48)^2}{100.48}$

$+ \frac{(81-75.36)^2}{75.36} + \frac{(28-35.712)^2}{35.712} + \frac{(55-46.128)^2}{46.128} + \frac{(64-59.52)^2}{59.52}$

$+ \frac{(39-44.64)^2}{44.64} = 7.04$ 2

3 $1\times3=3$ ~~3-1=2~~ degree of freedom $= 3-1 = 2$

$X^2_{2,\,0.9} = 4.605$

$7.04 > 4.605$

$\therefore$ reject $H_0$, size of shell is not independent of the beach location

## Examiner comments

1 This is a correct expression of the null hypothesis, but there is no reference to the alternative hypothesis. Both the null and alternative hypotheses must be stated when any hypothesis test is being used.

2 The test statistic is correctly calculated as 7.04.

3 The number of degrees of freedom for a 2 by 4 contingency table is 1 x 3. The candidate calculates this, but then thinks that 1 has to be subtracted from it. This is a muddling of the methods for finding the number of degrees of freedom in different situations. It is common to see incorrect values of the tabular value used in a chi-squared test. Candidates are advised to pay attention to how to find the correct tabular value, because the value is crucial to the test.

**Total mark awarded = 4 out of 7**

# Question 3

## Example Candidate Response – 1

**3** George throws two coins, $A$ and $B$, at the same time. Coin $A$ is biased so that the probability of obtaining a head is $a$. Coin $B$ is biased so that the probability of obtaining a head is $b$, where $b < a$. The probability generating function of $X$, the number of heads obtained by George, is $G_X(t)$. The coefficients of $t$ and $t^2$ in $G_X(t)$ are $\frac{5}{12}$ and $\frac{1}{12}$ respectively.

**(a)** Find the value of $a$. [2]

| $X$ | 0 | 1 | 2 |
|---|---|---|---|
| $P$ | | $\frac{5}{12}$ | $\frac{1}{12}$ |

1

$X=2 \quad P = ab \quad ab = \frac{1}{12}$ 2

$X=1 \quad P = a - ab = \frac{5}{12}$ 3

$a - \frac{1}{12} = \frac{5}{12} \quad a = \frac{1}{2}$ 4

## Examiner comments

1 The candidate leaves a gap in the probability distribution table corresponding to $x = 0$. This does not matter in this part of the question, but it becomes a problem in part (b).

2 The coefficient of $t^2$ in the probability generating function (pgf) is the probability of obtaining 2 heads when the coins are thrown. This is correctly identified as *ab*.

3 The coefficient of *t* in the pgf is the probability of obtaining one head when the two coins are thrown. There are two ways of achieving this; the candidate has only coin *A* showing a head and coin *B* showing a tail. There should be an additional term (1 - *a*) *b* in this expression. Errors such as this in writing down the two equations in (a) and (b) were fairly common.

4 The step of solving two equations in (a) and (b) has been simplified by this candidate's error. Even candidates who had the correct equations made errors in solving them. Algebra such as this should be second nature to the candidate.

Mark for (a) = 0 out of 2

## Example Candidate Response – 1, continued

The random variable $Y$ is the sum of two independent observations of $X$.

**(b)** Find the probability generating function of $Y$, giving your answer as a polynomial in $t$.

$G_X(t) = \frac{5}{12}t + \frac{1}{12}t^2.$ 5

$G_Y = (\frac{5}{12}t + \frac{1}{12}t^2)(\frac{5}{12}t + \frac{1}{12}t^2)$ 6

$= \frac{25}{144}t^2 + \frac{10}{144}t^3 + \frac{1}{144}t^4.$

**(c)** Find Var($Y$). [3]

$\mathrm{Var}(Y) = G''_Y(1) + G'_Y(1) - \{G'_Y(1)\}^2$ 7

$\cancel{= \frac{25}{144}\times 2\times 1 + \frac{10}{144}\times 3\times 2 + \frac{1}{144}\times 4\times 3}$

$= \frac{25}{144}\times 2\times 1 + \frac{10}{144}\times 3\times 2 + \frac{1}{144}\times 4\times 3 + \frac{25}{144}\times 2 + \frac{10}{144}\times 3 + \frac{1}{144}\times 4.$

$- (\frac{7}{12})^2$

$= \frac{157}{144}$

## Examiner comments

5 The omission in the probability distribution table in part (a) is now significant. The pgf written here by the candidate was seen on many scripts. It should be noted that the coefficients of the terms in the pgf always sum to 1. Therefore, there is a missing constant term of 1/2. This is the probability that no heads are obtained when the coins are thrown and so could have been found in that way.

6 This is the correct method, but the result is incorrect because of the missing term in the pgf.
Mark for (b) = 1 out of 3

7 This is the correct formula for finding Var($Y$). The calculation is incorrect because of earlier errors. However, this solution illustrates how errors could easily occur when work is not presented clearly. Candidates are advised to write down the first and second derivatives of the pgf of $Y$. They then substitute $t$ = 1 in each and finally substitute in the formula which is given in List of formula (MF19). This candidate has carried out all the steps in one line, and done so accurately, but it is an unnecessary risk to take. This solution is only just acceptable, because the examiner can see what is happening.
Mark for (c) = 2 out of 3

**Total mark awarded = 3 out of 8**

# Question 4

| Example Candidate Response – 1 | Examiner comments |
|---|---|
| The continuous random variable $X$ has probability density function f given by $$\mathrm{f}(x) = \begin{cases} \frac{3}{8}\left(1+\frac{1}{x^2}\right) & 1 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$ **(a)** Find $\mathrm{E}(\sqrt{X})$. [3] | |
| $E(X^{1/2}) = \int x^{1/2} f(x)\,dx$ <br> $= \int_1^3 x^{1/2}\left[\frac{3}{8}(1+\frac{1}{x^2})\right]dx$ <br> $= \int_1^3 x^{1/2}\left[\frac{3}{8}+\frac{3}{8}(\frac{1}{x^2})\right]dx$ <br> $= \int_1^3 \frac{3}{8}x^{1/2} + \cancel{\frac{3}{8x^{3/2}}}\ \frac{3}{8}x^{-3/2}$ <br> $= \left[\frac{1}{4}x^{3/2} \cancel{+} - \frac{3}{4}x^{-1/2}\right]_1^3$ <br> $= \left[\frac{1}{4}(3)^{3/2} - \frac{3}{4}(3)^{-1/2}\right] - \left[\frac{1}{4}(1)^{3/2} - \frac{3}{4}(1)^{-1/2}\right]$ <br> $= \cancel{0.866+\frac{1}{2}}\ \cancel{1.1547}\ 0.866+\frac{1}{2}$ <br> $\cancel{=0.366}\ = 1.366$ <br> $\cancel{= 1.1547 + 366}$ | 1 The candidate has the correct expression as the integrand and completes the integration successfully. <br> Mark for (a) = 3 out of 3 |
| The random variable $Y$ is given by $Y = X^2$. <br> **(b)** Find the probability density function of $Y$. [4] | |
| Let G(y) be the cdf of Y <br> $P(Y \le y) = \cancel{P(X^2 \le y)}\ P(y \le x^2)$ <br> $= P(y^{1/2} \le x)$ <br> $f(x) \to F(x) \Rightarrow$ ~~diff~~ integrate <br> $F(X) = \int_1^x \frac{3}{8}(1+\frac{1}{x^2})\,dx$ for $1 \le x \le 3$ <br> $= \left[\frac{3}{8}x\ \cdots\right]$ $\left[\frac{3}{8}x + (-\frac{3}{8}x^{-1})\right]_1^x$ <br> $\cancel{= \left[\frac{1}{4}x^{3/2} - \frac{3}{4}x^{-1/2}\right]_1^x}$ <br> $\cancel{= \frac{1}{4}x^{3/2} - \frac{3}{4}x^{-1/2} + \frac{1}{2}}$ <br> $= \frac{3}{8}x - \frac{3}{8}x^{-1} - 0$ <br> $F(x) = 1$, for $x > 3$ | 2 The candidate integrates the pdf of $X$ to find the cdf of $X$. |

| Example Candidate Response – 1, continued | Examiner comments |
|---|---|
| $G(y) = \begin{cases} \frac{3}{8}y^{1/2} - \frac{3}{8}y^{-1/2} & 1 \le y^{1/2} \le \sqrt{3} \\ 1 & y > \sqrt{3} \end{cases}$ (3)<br>$g(y) =$ For $1 \le y^{1/2} \le \sqrt{3}$, $\frac{3}{16}y^{-1/2} + \frac{3}{16}y^{-3/2}$ (4)<br>$g(y) = \begin{cases} \frac{3}{16}y^{-1/2} + \frac{3}{16}y^{-3/2} & 1 \le y^{1/2} \le \sqrt{3} \\ 0 & \text{otherwise} \end{cases}$ | (3) The change of variable from $X$ to $Y$ has been applied to find the CDF of $Y$.<br><br>(4) Differentiation of the CDF of $Y$ leads to the correct pdf of $Y$. The error here is that the domain is not correctly defined. We need to see the domain of $y$.<br>Mark for (b) = 3 out of 4 |
| **(c)** Find the 40th percentile of $Y$.<br>40th percentile $= \frac{2}{5}$<br>$G(m) = \frac{2}{5}$<br>$\frac{3}{8}y^{1/2} - \frac{3}{8}y^{-1/2} = \frac{2}{5}$<br>$\frac{3}{8}y^{1/2} - \frac{3}{8y^{1/2}} = \frac{2}{5}$<br>$\frac{3}{8}y - \frac{3}{8} = \frac{2}{5}y^{1/2}$<br>$\frac{3}{8}y - \frac{2}{5}y^{1/2} = \frac{3}{8}$<br>$\frac{3}{8}y - \frac{2}{5}y^{1/2} - \frac{3}{8} = 0$ (5)<br>$(3y-5)(5y+3) = 0$<br>$\cancel{3y-5=0}$ $3y-5=0$ $5y+3=0$<br>$y = \frac{5}{3}$ $y = -\frac{3}{5}$<br>(N/A)<br>the 40th percentile of $Y$ is $\frac{5}{3}$ | (5) This equation is correct, but now the candidate implicitly uses a substitution of $y$ for the square root of $y$, so the final answer of $\frac{5}{3}$ is in fact the value of the square root of $y$ and not $y$ itself. This error occurred quite often and could have been avoided by using an explicit substitution such as $u$ is equal to the square root of $y$.<br>Mark for (c) = 2 out of 3<br><br>**Total mark awarded = 8 out of 10** |

## Example Candidate Response – 2

**4** The continuous random variable $X$ has probability density function f given by

$$f(x) = \begin{cases} \frac{3}{8}\left(1+\frac{1}{x^2}\right) & 1 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

**(a)** Find $\mathrm{E}(\sqrt{X})$.

$E(\sqrt{x}) = \int \sqrt{x}\, f(\sqrt{x})\, dx$ 1

$= \int_1^3 \sqrt{x} \cdot \frac{3}{8}\left(1+\frac{1}{x}\right) dx$ $\quad 4.763$

$= \int_1^3 \frac{3}{8}\sqrt{x} + \frac{\sqrt{x}}{x}\, dx$

$= \int_1^3 \frac{3}{8} x^{\frac{1}{2}} + x^{-\frac{1}{2}}\, dx$

$= \left[\frac{2}{3}\cdot\frac{3}{8} x^{\frac{3}{2}} + 2 \cdot x^{\frac{1}{2}}\right]_1^3$

$= \left[\frac{1}{4} x^{\frac{3}{2}} + 2x^{\frac{1}{2}}\right]_1^3 \quad \approx 2.513$

The random variable $Y$ is given by $Y = X^2$. $\qquad \frac{3}{8} + \frac{3}{8}x^{-2}$

**(b)** Find the probability density function of $Y$.

$$F(x) \begin{cases} \cancel{1 \quad \text{otherwise}} \\ \frac{3}{8}x - \frac{3}{8}x^{-1} & 1 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$ 2

$G(Y) = P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = F(\sqrt{y})$

$$F(\sqrt{y}) = \begin{cases} 1 & \text{otherwise} \\ \frac{3}{8}\sqrt{y} - \frac{3}{8}\frac{1}{\sqrt{y}} & 1 \leq \sqrt{y} \leq 3 \quad (1 \leq y \leq 9) \\ 0 & \text{otherwise} \end{cases}$$ 3

## Examiner comments

1 This is a common error. The integrand should have f($x$) and not f($x^2$). Candidates mistakenly think that the function itself has to change when finding an expectation.
Mark for (a) = 0 out of 3

2 The candidate integrates the PDF of $X$ to find the CDF of $X$, correctly.

3 The candidate applies the change of variable correctly, giving the CDF of $Y$. However, it is the PDF of $Y$ that is required, so differentiation of the CDF needs to follow. This omission was commonly seen.
Mark for (b) = 2 out of 4

| Example Candidate Response – 2, continued | Examiner comments |
|---|---|
| **(c)** Find the 40th percentile of $Y$.<br><br>$\frac{3}{8}\sqrt{y} - \frac{3}{8}\frac{1}{\sqrt{y}} = 0.4$ **4**<br><br>$\left(\frac{3}{8}\right)^2 y - 2 \times \left(\frac{3}{8}\right)^2 + \left(\frac{3}{8}\right)^2 \frac{1}{y} = 0.16$ **5**<br><br>$\frac{9}{64}y - 2\times\frac{9}{64} + \frac{9}{64}\frac{1}{y} = 0.16$<br><br>$\frac{9}{64}y^2 - \frac{18}{64}y + \frac{9}{64} = 0.16y$<br><br>$9y^2 - 18y + 9 = 10.24y$<br><br>$9y^2 - 28.24y + 9 = 0$<br><br>**6** $y \approx 2.79$<br><br>[3] | **4** This is the correct starting point.<br><br>**5** An alternative method to using a substitution is to square both sides of the equation.<br><br>**6** There is a slight loss of accuracy in the final answer.<br>Mark for (c) = 2 out of 3<br><br>**Total mark awarded =<br>4 out of 10** |

# Question 5

## Example Candidate Response – 1

5 A manager claims that the lengths of the rubber tubes that his company produces have a median of 5.50 cm. The lengths, in cm, of a random sample of 11 tubes produced by this company are as follows.

5.56 5.45 5.47 5.58 5.54 5.52 5.60 5.35 5.59 5.51 5.62

It is required to test at the 10% significance level the null hypothesis that the population median length is 5.50 cm against the alternative hypothesis that the population median length is not equal to 5.50 cm.

Show that both a sign test and a Wilcoxon signed-rank test give the same conclusion and state this conclusion. [9]

~~By sign test:~~ $H_0: M = 5.50$ $H_1: M \neq 5.50$

By sign test:

| | | |
|---|---|---|
| | −5.50 | |
| 5.35 | −0.15 | − |
| 5.45 | −0.05 | − |
| 5.47 | −0.03 | − |
| 5.51 | 0.01 | + |
| 5.52 | 0.02 | + |
| 5.54 | 0.04 | + |
| 5.56 | 0.06 | + |
| 5.58 | 0.08 | + |
| 5.59 | 0.09 | + |
| 5.60 | 0.10 | + |
| 5.62 | 0.12 | + |

$P = 8$ $N = 3$

$\Rightarrow T = \min(P, N) = 3$

$p = \frac{1}{2}$, $n = 11$ let X be random variable of number of negative sign.

$\Rightarrow X \sim B(11, \frac{1}{2})$

$\Rightarrow P(X=3) = C_{11}^{3} \times (\frac{1}{2})^3 \times (\frac{1}{2})^8$

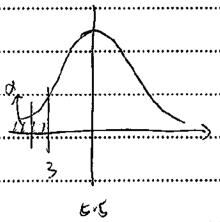
$\approx 0.0806$

$\alpha = \frac{0.10}{2} = 0.05$

$\Rightarrow 0.0806 > 0.05$

$\Rightarrow H_0$ is not rejected

$\Rightarrow$ There is no sufficient evidence to support ~~that~~ the population median length is not equal to 5.50 cm.




## Examiner comments

1 A substantial number of candidates appear unfamiliar with the basic methodology required to carry out a sign test. It is a syllabus item and candidates need to be aware of its existence and how to apply it.

2 The candidate considers the sign differences of each length from 5.50 and correctly finds that there are 8 positives (and 3 negatives).

3 The sign test depends on using the Binomial distribution, in this case, B(11, 0.5).

4 The probability that needs to be calculated is $P(X \leq 3)$. The error here is that $P(X = 3)$ is found and used.

| Example Candidate Response – 1, continued | Examiner comments |
|---|---|
| By Wilcoxon signed-rank test:<br>5.35   −0.15   −5.5   ⑪    $P = 1+2+4+6+7+8+9+10$<br>5.45   −0.05   ⑤    $= 47$<br>5.47   −0.03   ②    $N = 11+5+3 = 19$<br>5.51   +0.01   ①<br>5.52   0.02   ②    $\Rightarrow T = \min(P, N) = 19$<br>5.54   0.04   ④<br>5.56   0.06   ⑥    $n = 11$, $\alpha = 0.10$, two-tailed<br>5.58   0.08   ⑦    $\Rightarrow C = 13$<br>5.59   0.09   ⑧    $\Rightarrow 19 > 13$<br>5.60   0.10   ⑨    $\Rightarrow H_0$ is not rejected<br>5.62   0.12   ⑩    5<br>$\Rightarrow$ two test given the same conclusion.<br>Overall:<br>There is no sufficient evidence to ~~ref~~ support that the population median length is not equal to 5.50 cm. | 5 The candidate carries out the Wilcoxon signed-rank test correctly with all the steps clearly seen. The conclusion is correct, with an acceptable level of uncertainty.<br><br>**Total mark awarded = 6 out of 9** |

## Example Candidate Response – 2

**5** A manager claims that the lengths of the rubber tubes that his company produces have a median of 5.50 cm. The lengths, in cm, of a random sample of 11 tubes produced by this company are as follows.

5.56 5.45 5.47 5.58 5.54 5.52 5.60 5.35 5.59 5.51 5.62

It is required to test at the 10% significance level the null hypothesis that the population median length is 5.50 cm against the alternative hypothesis that the population median length is not equal to 5.50 cm.

Show that both a sign test and a Wilcoxon signed-rank test give the same conclusion and state this conclusion. [9]

sign test:

$H_0: \mu = 5.50$ cm 1

$H_1: \mu \neq 5.50$ cm

~~sign test~~ 5.56 + 8 positive 2

5.45 − 3 negative

5.47 − ~~11C11 11C8P~~

5.58 + $^{11}C_{11} \times 0.5^{11} + {}^{11}C_{10} \times 0.5^{11} + {}^{11}C_9 \times 0.5^{11}$ 3

5.54 + $= 67 \times 0.5^{11}$

5.52 + $= 0.033$ 4

5.60 + critical value: 1.645   $0.033 < 1.645$.

5.35 − we do not reject Ho

5.59 + there is insufficient evidence to suggest

5.51 + that $\mu \neq 5.50$ cm

5.62 +

~~critical value: 1.645~~

~~critical value:~~

## Examiner comments

1 The hypotheses are stated in the question and involve the population median. In this case, there were no marks for restating these hypotheses, but if there had been, this would have scored zero marks. The symbol used is the accepted symbol for the mean and not the median. The symbol *m* or the words population median would have been acceptable.

2 Application of the sign test proved difficult for most candidates, with many not being aware of the methodology of the test. This example shows two of the typical errors made by candidates who did make some progress.

3 The candidate finds that there are 8 positive differences and attempts to use B(11, 0.5). However, the term for $X = 8$ should also be included.

4 The comparison should be with 0.05. The candidate uses 1.645 which corresponds to a normal distribution.

## Example Candidate Response – 2, continued

Wilcoxon signed-rank test

$H_0: \mu = 5.50$ cm

$H_1: \mu \neq 5.50$ cm

| | −5.50 | P(t) | Q(−) |
|---|---|---|---|
| 5.56 | +0.06 | 6 | |
| 5.45 | −0.05 | | 5 |
| 5.47 | −0.03 | | 3 |
| 5.58 | +0.08 | 7 | |
| 5.54 | +0.04 | 4 | |
| 5.52 | +0.02 | 2 | |
| 5.60 | +0.1 | 9 | |
| 5.35 | −0.15 | | 11 |
| 5.59 | +0.09 | 8 | |
| 5.51 | +0.01 | 1 | |
| 5.62 | +0.12 | 10 | |

$P(t) = 6 + 7 + 4 + 2 + 9 + 8 + 1 + 10 = 47$

$Q(-) = 5 + 3 + 11 = 19$

$T(P, Q) = 19$

$n = 11$ two-tailed 0.1

critical value : 13

since $19 > 13$

we do not reject $H_0$

there is insufficient evidence to suggest $\mu \neq 5.50$

5

## Examiner comments

5 The candidate carries out the Wilcoxon signed-rank test correctly. The conclusion does show an appropriate level of uncertainty, but it is not fully in context, using a symbol which is undefined in the question, and as seen earlier is not acceptable as an alternative to 'population median'.

**Total mark awarded = 6 out of 9**

# Question 6

## Example Candidate Response – 1

**6** A company has two machines, $A$ and $B$, which independently fill small bottles with a liquid. The volumes of liquid per bottle, in suitable units, filled by machines $A$ and $B$ are denoted by $x$ and $y$ respectively. A scientist at the company takes a random sample of 40 bottles filled by machine $A$ and a random sample of 50 bottles filled by machine $B$. The results are summarised as follows.

$$\Sigma x = 1120 \qquad \Sigma x^2 = 31400 \qquad \Sigma y = 1370 \qquad \Sigma y^2 = 37600$$

The population means of the volumes of liquid in the bottles filled by machines $A$ and $B$ are denoted by $\mu_A$ and $\mu_B$.

**(a)** Test at the 2% significance level whether there is any difference between $\mu_A$ and $\mu_B$. [8]

$H_0: \mu_A - \mu_B = 0 \qquad H_1: \mu_A - \mu_B \neq 0$

$\bar{x} = 28 \qquad S_x^2 = \frac{40}{39}\left(\frac{31400}{40} - 28^2\right) = \cancel{[illegible]}\ \frac{40}{39}$

$\bar{y} = 27.4 \qquad S_y^2 = \frac{50}{49}\left(\frac{37600}{50} - 27.4^2\right) = \frac{62}{49}$

$t = \dfrac{28 - 27.4}{\sqrt{\frac{1}{39} + \frac{62}{49 \cdot 50}}}$ (1)

(2) $= 2.638$

$t_{0.99} = 2.326$

$2.326 < 2.638$

$H_0$ is rejected

It is sufficient to show the different between $\mu_A$ and $\mu_B$. (3)

**(b)** Find the set of values of $\alpha$ for which there would be evidence at the $\alpha\%$ significance level that $\mu_A - \mu_B$ is greater than 0.25. [4]

$t = \dfrac{28 - 27.4 - 0.25}{\sqrt{\frac{1}{39} + \frac{62}{49 \times 50}}}$

$= 1.551$ (4)

$0.9395$

$\alpha = 93.95$

## Examiner comments

(1) This is correct, but 2.638 is not. The $t$ value should be 2.66. It is worth noting here that because the intermediate stage of calculating the denominator is absent, it is not possible to see where the error has occurred, so neither accuracy mark is awarded. If several steps are performed on a calculator, then just an answer given, there is no scope for awarding intermediate accuracy marks. Candidates should be aware that there is a risk involved in such a strategy.

(2) The candidate labels the test statistic as $t$ whereas it should be $z$. The value of 2.326 used in the comparison is a $z$ value, so the mislabelling is condoned in this case.

(3) The conclusion must contain an appropriate level of uncertainty. A statement such as 'there is sufficient evidence to support a difference in population means' would be an appropriate conclusion. Words such as prove and show should be avoided.
Mark for (a) = 5 out of 8

(4) This value is correct, though again it is labelled as $t$. 93.95 is also a correct value, but it needs to be subtracted from 100 to give the limiting least value of alpha as 6.05.
Mark for (b) = 3 out of 4

**Total mark awarded = 8 out of 12**

## Example Candidate Response – 2

A company has two machines, $A$ and $B$, which independently fill small bottles with a liquid. The volumes of liquid per bottle, in suitable units, filled by machines $A$ and $B$ are denoted by $x$ and $y$ respectively. A scientist at the company takes a random sample of 40 bottles filled by machine $A$ and a random sample of 50 bottles filled by machine $B$. The results are summarised as follows.

$$\Sigma x = 1120 \qquad \Sigma x^2 = 31400 \qquad \Sigma y = 1370 \qquad \Sigma y^2 = 37600$$

The population means of the volumes of liquid in the bottles filled by machines $A$ and $B$ are denoted by $\mu_A$ and $\mu_B$.

**(a)** Test at the 2% significance level whether there is any difference between $\mu_A$ and $\mu_B$. [8]

$H_0: \mu_A - \mu_B = 0$

$H_1: \mu_A - \mu_B \neq 0$

$\Sigma x = 1120$ $\quad$ $\Sigma y = 1370$

$\bar{x} = \frac{1120}{40}$ $\quad$ $\bar{y} = 27.4$

$= 28$

$S^2_x = \frac{1}{39}\left(\cancel{W}\ 31400 - \frac{1120^2}{40}\right)$ $\quad$ $S^2_y = \frac{1}{49}\left(37600 - \frac{1370^2}{50}\right)$

$= 1.025641$ $\quad$ $= 1.2653$

$S_x = 1.012739$ $\quad$ $S_y = 1.12486$

$z = \dfrac{(\bar{x} - \bar{y})}{\sqrt{\frac{S^2_x}{n_x} + \frac{S^2_y}{n_y}}}$

$H_1$ $\quad$ $H_0$ $\quad$ $2.326$

1

$= \dfrac{(28 - 27.4)}{\sqrt{\frac{1.025641}{40} + \frac{1.12486}{50}}}$

$= 2.7346781$

$z(0.99) = \cancel{2.236}\ 2.326$

$2.7346781 > 2.326$

2

Accept $H_0$, there is sufficient envidence to show that there is no difference between $\mu_A$ and $\mu_B$

**(b)** Find the set of values of $\alpha$ for which there would be evidence at the $\alpha$% significance level that $\mu_A - \mu_B$ is greater than 0.25. [4]

$H_1$ $\quad$ $H_0$ $\quad$ $1.595228$

$\dfrac{(\bar{x} - \bar{y}) - 0.25}{\sqrt{\frac{1.025641}{40} + \frac{1.12486}{50}}}$

$= 1.595228$

$z(\alpha) \geqslant 1.595228$

$\alpha = 0.9447$

3

$\alpha \geqslant 94.47\%$, $\cancel{\alpha}$ $\alpha \leqslant 5.53\%$

4

## Examiner comments

1 The work up to this point is accurate, but then the number 1.12486 is picked up from earlier, instead of 1.2653. Here, we have an inconsistency with a variance and a standard deviation being used, but this is probably a slip rather than a method error. Slips such as this are not uncommon, and with due care could be avoided.

2 The comparison with 2.326 is correct, but the conclusion is not correct. Since the calculated test statistic is greater than the tabular value, the null hypothesis should be rejected.
Mark for (a) = 4 out of 8

3 The method up to this point is correct, but the accuracy error from part (a) carries through.

4 The value found is the least possible value of alpha which satisfies the condition given in the question. This correct interpretation was not seen very often.
Mark for (b) = 1 out of 4

**Total mark awarded = 5 out of 12**

## Example Candidate Response – 3

**6** A company has two machines, $A$ and $B$, which independently fill small bottles with a liquid. The volumes of liquid per bottle, in suitable units, filled by machines $A$ and $B$ are denoted by $x$ and $y$ respectively. A scientist at the company takes a random sample of 40 bottles filled by machine $A$ and a random sample of 50 bottles filled by machine $B$. The results are summarised as follows.

$$\Sigma x = 1120 \qquad \Sigma x^2 = 31400 \qquad \Sigma y = 1370 \qquad \Sigma y^2 = 37600$$

The population means of the volumes of liquid in the bottles filled by machines $A$ and $B$ are denoted by $\mu_A$ and $\mu_B$.

**(a)** Test at the 2% significance level whether there is any difference between $\mu_A$ and $\mu_B$. [8]

$H_0: \mu_A - \mu_B = 0 \qquad H_1: \mu_A - \mu_B \neq 0$

$\bar{x} = \frac{\Sigma x}{n_x} = \frac{1120}{40} = 28 \qquad \bar{y} = \frac{\Sigma y}{n_y} = \frac{1370}{50} = 27.4$

$S_x^2 = \frac{1}{n_x - 1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n_x}\right) = \frac{1}{40-1}\left(31400 - \frac{1120^2}{40}\right) = \frac{40}{39}$

$S_y^2 = \frac{1}{50-1}\left(37600 - \frac{1370^2}{50}\right) = \frac{62}{49}$

$S^2 = \frac{\Sigma(x_i - \bar{x})^2 + \Sigma(y - \bar{y})^2}{n_x + n_y - 2} = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$

$= \frac{39 \times \frac{40}{39} + \frac{62}{49} \times 49}{40 + 50 - 2} = \frac{51}{44}$ 1

$z = \frac{\bar{x} - \bar{y} - (\mu_A - \mu_B)}{\sqrt{S^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} = \frac{28 - 27.4}{\sqrt{\frac{51}{44}\left(\frac{1}{40} + \frac{1}{50}\right)}} = 2.627$

2

so $P = 0.9957$ test statistics

critical value ~~1-2~~ ~~1 - 0.02~~ $1 - 0.02 \times \frac{1}{2} = 0.99$

$0.99 < 0.9957$

there is sufficient evidence to show that there is difference between $\mu_A$ and $\mu_B$

**(b)** Find the set of values of $\alpha$ for which there would be evidence at the $\alpha$% significance level that $\mu_A - \mu_B$ is greater than 0.25. [4]

~~$z = \frac{\bar{x} - \bar{y} - (\mu_A - \mu_B)}{\sqrt{S^2 \frac{1}{n_x}}}$~~

$z = \frac{\bar{x} - \bar{y} - (\mu_A - \mu_B)}{\sqrt{S^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} = \frac{28 - 27.4 - 0.25}{\sqrt{\frac{51}{44}\left(\frac{1}{40} + \frac{1}{50}\right)}} = 1.533$ 3

$P = 0.9374$

$1 - 0.9374 = 0.0626 = 6.26\%$ critical value

~~since~~ so $\alpha \in [0, 6.26]$ 4

## Examiner comments

1 The candidate uses a pooled estimate for the variance, assuming that the two populations share a common variance. There is nothing given in the question to suggest that this might be true or a valid assumption. In the absence of a statement in the text that a common population variance is either true or can be assumed, a candidate should use the formula for unequal variances.

2 It is acceptable for candidates to work with a $p$-value instead of a $z$-value as long as the correct comparison is made. This is not seen very often.
Mark for (a) = 3 out of 8

3 The candidate continues to use a pooled variance, so this expression is incorrect.

4 Following from the candidate's 6.26%, the incorrect range is given. It would be [6.26, 100].
Mark for (b) = 0 out of 4

**Total mark awarded = 3 out of 12**

Cambridge Assessment International Education
The Triangle Building, Shaftesbury Road, Cambridge, CB2 8EA, United Kingdom
t: +44 1223 553554
e: info@cambridgeinternational.org    www.cambridgeinternational.org