GCE A LEVEL

1305U50-1

Z22-1305U50-1

THURSDAY, 23 JUNE 2022 – AFTERNOON

FURTHER MATHEMATICS – A2 unit 5
FURTHER STATISTICS B

1 hour 45 minutes

1305U501
01

ADDITIONAL MATERIALS

In addition to this examination paper, you will need:

- a WJEC pink 16-page answer booklet;
- a Formula Booklet;
- a calculator;
- statistical tables (RND/WJEC Publications).

INSTRUCTIONS TO CANDIDATES

Use black ink or black ball-point pen.

Answer **all** questions.

Sufficient working must be shown to demonstrate the **mathematical** method employed.

Answers without working may not gain full credit.

Unless the degree of accuracy is stated in the question, answers should be rounded appropriately.

INFORMATION FOR CANDIDATES

The maximum mark for this paper is 80.

The number of marks is given in brackets at the end of each question or part-question.

You are reminded of the necessity for good English and orderly presentation in your answers.

© WJEC CBAC Ltd.

stemup.co.uk

Reminder: Sufficient working must be shown to demonstrate the **mathematical** method employed.

1. Rachel records the times taken, in minutes, to cycle into town from her house on a random sample of 10 days. Her results are shown below.

15·5 14·9 16·2 17·3 14·8 14·2 16·0 14·2 15·5 15·1

Assuming that these data come from a normal distribution with mean μ and variance 0·9, calculate a 90% confidence interval for μ. [5]

2. Geraint is a beekeeper. The amounts of honey, X kg, that he collects annually, from each hive are modelled by the normal distribution $\mathrm{N}\left(15, 5^2\right)$. At location *A*, Geraint has three hives and at location *B* he has five hives. You may assume that the amounts of honey collected from the eight hives are independent of each other.

(a) (i) Find the probability that Geraint collects more than 14 kg of honey from the first hive at location *A*.

(ii) Find the probability that he collects more than 14 kg of honey from exactly two out of the three hives at location *A*. [4]

(b) Find the probability that the total amount of honey that Geraint collects from all eight hives is more than 160 kg. [4]

(c) Find the probability that Geraint collects at least twice as much honey from location B as from location A. [7]

© WJEC CBAC Ltd.

stemup.co.uk

3. A statistics teacher wants to investigate whether students from the north of a county and students from the south of the same county feel similarly stressed about examinations. The teacher carries out a psychometric test on 10 randomly selected students to give a score between 0 (low stress) and 100 (high stress) to measure their stress levels before a set of examinations. The results are shown in the table below.

Student	Area	Stress Level
Heledd	North	67
Mair	North	55
Hywel	South	26
Gwyn	South	70
Liam	South	36
Marcin	South	57
Gosia	South	32
Kestutas	North	64
Erica	North	60
Tomos	North	22

(a) State one reason why a Mann-Whitney test is appropriate. [1]

(b) Conduct a Mann-Whitney test at a significance level as close to 5% as possible. State your conclusion clearly. [6]

(c) How could this investigation be improved? [1]

TURN OVER

© WJEC CBAC Ltd.

stemup.co.uk

4. The Department of Health recommends that adults aged 18 to 65 should take part in at least 150 minutes of aerobic exercise per week. The results of a survey show that 940 out of 2000 randomly selected adults aged 18 to 65 in Wales take part in at least 150 minutes of aerobic exercise per week.

(a) Calculate an approximate 95% confidence interval for the proportion of adults aged 18 to 65 in Wales who take part in at least 150 minutes of aerobic exercise per week. [6]

(b) Give two reasons why the interval is approximate. [2]

(c) Suppose that a 99% confidence interval is required, and that the width of the interval is to be no greater than 0·04. Estimate the minimum additional number of adults to be surveyed to satisfy this requirement. [4]

5. A laboratory carrying out screening for a certain blood disorder claims that the average time taken for test results to be returned is 38 hours. A reporter for a national newspaper suspects that the results take longer, on average, to be returned than claimed by the laboratory. The reporter finds the time, x hours, for 50 randomly selected results, in order to conduct a hypothesis test. The following summary statistics were obtained.

$$\sum x = 2163 \qquad \sum x^2 = 98508$$

(a) Calculate the *p*-value for the reporter's hypothesis test, and complete the test using a 5% level of significance. Hence write a headline for the reporter to use. [9]

(b) Explain the relevance or otherwise of the Central Limit Theorem to your answer in part (a). [1]

(c) Briefly explain why a random sample is preferable to taking a batch of 50 consecutive results. [1]

(d) On another occasion, the reporter took a different random sample of 10 results.

(i) State, with a reason, what type of hypothesis test the reporter should use on this occasion.

(ii) State one assumption required to carry out this test. [2]

© WJEC CBAC Ltd.

stemup.co.uk

6. A zoologist knows that the median body length of adults in a species of fire-bellied toads is 4·2 cm. The zoologist thinks he has discovered a new subspecies of fire-bellied toads. If there is sufficient evidence to suggest the median body length differs from 4·2 cm, he will continue his studies to confirm whether he has discovered a new subspecies. Otherwise, he will abandon his studies on fire-bellied toads.

The lengths of 10 randomly selected adult toads from the group being investigated are given below.

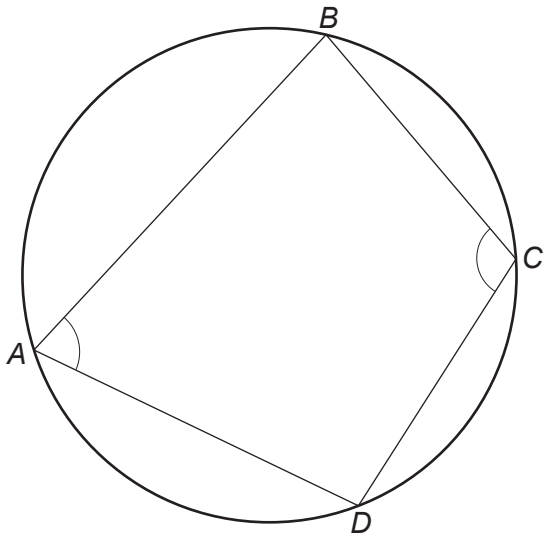
5·0 3·2 4·9 4·0 3·3 4·2 6·1 4·3 4·8 5·9

Carry out a suitable Wilcoxon signed rank test at a significance level as close to 1% as possible and give your conclusion in context. [8]

© WJEC CBAC Ltd.

stemup.co.uk

7.



The diagram above shows a cyclic quadrilateral *ABCD,* where $\hat{BAD} = \alpha$, $\hat{BCD} = \beta$ and $\alpha + \beta = 180°$. These angles are measured.

The random variables X and Y denote the measured values, in degrees, of $\hat{BAD}$ and $\hat{BCD}$ respectively. You are given that X and Y are independently normally distributed with standard deviation σ and means α and β respectively.

(a) Calculate, correct to two decimal places, the probability that $X + Y$ will differ from 180° by less than σ. [4]

(b) Show that $T_1 = 45° + \frac{1}{4}(3X - Y)$ is an unbiased estimator for α and verify that it is a better estimator than X for α. [6]

(c) Now consider $T_2 = \lambda X + (1 - \lambda)(180° - Y)$.

(i) Show that T_2 is an unbiased estimator for α for all values of λ.

(ii) Find Var(T_2) in terms of λ and σ.

(iii) Hence determine the value of λ which gives the best unbiased estimator for α. [9]

END OF PAPER

© WJEC CBAC Ltd.

stemup.co.uk

BLANK PAGE

© WJEC CBAC Ltd.

stemup.co.uk

Surname	StemUp Model Solution
Answer booklet	Write the question number in the left-hand margin at the start of each answer.

Question number

1

$\bar{x} = 153.7 \div 10 = 15.37$

$SE = \sqrt{(0.9 \div 10)} = 0.3$

90% CI: $\bar{x} \pm z \times SE$ with $z = 1.6449$

$= 15.37 \pm 1.6449 \times \sqrt{(0.9 \div 10)}$

$= 15.37 \pm 0.49347...$

90% confidence interval for μ is $[14.88, 15.86]$

2(a)(i)

$X \sim N(15, 5^2)$

$P(X > 14) = P(Z > (14 - 15) \div 5) = P(Z > -0.2)$

$= 0.57925... = 0.5793$

2(a)(ii)

P(more than 14 kg from exactly two of three hives)

$= 3 \times 0.5793^2 \times 0.4207$

$= 0.42352... = 0.4235$

2(b)

Let $T = X_1 + X_2 + X_3 + ... + X_8$

$E(T) = 8 \times 15 = 120$

$Var(T) = 8\ Var(X) = 8 \times 25 = 200$

$P(T > 160) = P(Z > (160 - 120) \div \sqrt{200}) = P(Z > 2.8284...)$

$= 0.0023388... = 0.00234$ (3 s.f.)

Question number	
2(c)	$A = X_1 + X_2 + X_3 \sim N(45, 75)$
	$B = X_4 + X_5 + X_6 + X_7 + X_8 \sim N(75, 125)$
	Require $P(B \geq 2A)$. Consider $U = B - 2A$
	$E(U) = 75 - 2 \times 45 = -15$
	$Var(U) = Var(B) + 2^2\, Var(A) = 125 + 4 \times 75 = 425$
	$U \sim N(-15, 425)$
	$P(U \geq 0) = P(Z > (0 - (-15)) \div \sqrt{425}) = P(Z > 0.72760...)$
	$= 0.23342... = 0.2334$
3(a)	There is no knowledge of the underlying distribution of the stress levels (normality cannot be assumed), so a non-parametric test is appropriate.
3(b)	H_0: students from the north and the south of the county are similarly stressed ($\eta_N = \eta_S$)
	H_1: students from the north and the south of the county are not similarly stressed ($\eta_N \neq \eta_S$)
	In order: 22N 26S 32S 36S 55N 57S 60N 64N 67N 70S
	For each North score, count the South scores below it:
	$U = 0 + 3 + 4 + 4 + 4 = 15$
	(check: $U' = 1 + 1 + 1 + 2 + 5 = 10$ and $15 + 10 = 25 = 5 \times 5$)
	$m = n = 5$, two-tailed, level as close to 5% as possible (5.56%):
	lower critical value = 3, upper critical value = $25 - 3 = 22$
	Since $3 < 15 < 22$, there is insufficient evidence to reject H_0.
	There is not enough evidence to say that students from the north and from the south have different stress levels.
3(c)	Use a bigger sample size, with more students tested from each area.

Question number	
4(a)	$\hat{p} = 940 \div 2000 = 0.47$
	$ESE = \sqrt{\hat{p}(1-\hat{p}) \div n} = \sqrt{0.47 \times 0.53 \div 2000} = 0.011160...$
	95% confidence limits are $\hat{p} \pm z \times ESE$
	$= 0.47 \pm 1.96 \times 0.011160...$
	$= 0.47 \pm 0.021874...$
	Approximate 95% confidence interval is $[0.448, 0.492]$
4(b)	1. An estimate $\hat{p}$ has been used in place of p when calculating the standard error.
	2. The binomial distribution has been approximated by the normal distribution.
4(c)	99%: $z = 2.5758...$ Width $= 2 \times 2.5758... \times \sqrt{0.47 \times 0.53 \div n} \leq 0.04$
	$2.5758... \times \sqrt{0.47 \times 0.53 \div n} \leq 0.02$
	$\sqrt{n} \geq 2.5758... \times \sqrt{0.47 \times 0.53} \div 0.02 = 64.279...$
	$n \geq 4131.88...$ so $n = 4132$
	Minimum additional number $= 4132 - 2000 = 2132$ adults
5(a)	$\bar{x} = 2163 \div 50 = 43.26$
	$s^2 = (1 \div 49) \times (98508 - 2163^2 \div 50) = 100.7473...$ $(s = 10.0372...)$
	$H_0: \mu = 38$ $H_1: \mu > 38$
	Under H_0, $\bar{X} \sim N(38, 100.7473... \div 50)$
	$z = (43.26 - 38) \div (10.0372... \div \sqrt{50}) = 3.7055...$
	p-value $= P(\bar{X} > 43.26 \mid H_0 \text{ true}) = P(Z > 3.7055...) = 0.000105$
	Since $0.000105 < 0.05$, there is strong evidence to reject H_0.
	There is strong evidence to reject the laboratory's claim that the average time taken for test results to be returned is 38 hours: on average they take longer.
	Headline: "Lab takes longer than it claims to return blood test results"

Question number	
5(b)	Because $n = 50$ is large, the Central Limit Theorem allows us to assume that the distribution of the sample mean is (approximately) normal, so the normal distribution can be used even though the distribution of times is unknown.
5(c)	Random sampling eliminates bias: 50 consecutive results might all come from the same day, when the process is having a good, or bad, run.
5(d)(i)	A t-test, because the sample size is small ($n = 10$) and the variance is unknown.
5(d)(ii)	The time taken for results to be returned is normally distributed.
6	H_0: median = 4.2 H_1: median ≠ 4.2

x	5.0	3.2	4.9	4.0	3.3	4.2	6.1	4.3	4.8	5.9
$x - 4.2$	0.8	−1.0	0.7	−0.2	−0.9	0	1.9	0.1	0.6	1.7
Rank	5	7	4	2	6	-	9	1	3	8

The zero difference is discarded, so $n = 9$.

$W^+ = 5 + 4 + 9 + 1 + 3 + 8 = 30$ $W^- = 7 + 2 + 6 = 15$

$n = 9$, two-tailed, level as close to 1% as possible:

lower critical value = 2, upper critical value = $45 - 2 = 43$

Since $15 > 2$ (and $30 < 43$), there is insufficient evidence to reject H_0.

There is insufficient evidence that the median body length differs from 4.2 cm, so the zoologist should abandon his studies on fire-bellied toads.

7(a)

$X + Y \sim N(\alpha + \beta, 2\sigma^2) = N(180, 2\sigma^2)$

$P(180 - \sigma < X + Y < 180 + \sigma)$

$= P((180 - \sigma - 180) \div \sqrt{(2\sigma^2)} < Z < (180 + \sigma - 180) \div \sqrt{(2\sigma^2)})$

$= P(-1 \div \sqrt{2} < Z < 1 \div \sqrt{2}) = P(-0.7071 < Z < 0.7071)$

$= 2 \times 0.76025 - 1 = 0.52049... = 0.52$ (2 d.p.)

Question number	
7(b)	$E(T_1) = 45 + (1/4)(3E(X) - E(Y))$, with Y measuring $\beta = 180° - \alpha$
	$E(T_1) = 45 + (3/4)\alpha - (1/4)(180 - \alpha)$
	$E(T_1) = 45 + (3/4)\alpha - 45 + (1/4)\alpha = \alpha$
	$\therefore T_1$ is an unbiased estimator for α
	$Var(T_1) = (9/16)Var(X) + (1/16)Var(Y) = (9/16)\sigma^2 + (1/16)\sigma^2$
	$Var(T_1) = (5/8)\sigma^2$
	$Var(T_1) = (5/8)\sigma^2 < \sigma^2 = Var(X)$, and both are unbiased,
	$\therefore T_1$ is a better estimator than X for α.
7(c)(i)	$E(T_2) = \lambda E(X) + (1 - \lambda)(180° - E(Y)) = \lambda\alpha + (1 - \lambda)(180° - \beta)$
	$E(T_2) = \lambda\alpha + (1 - \lambda)\alpha = \lambda\alpha + \alpha - \lambda\alpha = \alpha$
	$\therefore T_2$ is an unbiased estimator for α for all values of λ.
7(c)(ii)	$Var(T_2) = \lambda^2 Var(X) + (1 - \lambda)^2 Var(180° - Y)$
	$Var(T_2) = \lambda^2\sigma^2 + (1 - \lambda)^2\sigma^2$
7(c)(iii)	$d/d\lambda\ Var(T_2) = 2\lambda\sigma^2 - 2(1 - \lambda)\sigma^2$
	Best estimator has minimum variance: $d/d\lambda\ Var(T_2) = 0$
	$2\lambda\sigma^2 = 2(1 - \lambda)\sigma^2 \Rightarrow 2\lambda = 2 - 2\lambda \Rightarrow \lambda = 1/2$
	$d^2/d\lambda^2\ Var(T_2) = 4\sigma^2 > 0$, therefore a minimum,
	$\therefore \lambda = 1/2$ gives the best unbiased estimator for α.