OXFORD
INTERNATIONAL AQA EXAMINATIONS

Topic Test: OxfordAQA A level Further Mathematics

Statistics

Name: ______________________

Class: ______________________

Date: ______________________

Time: **92 minutes**

Marks: **76 marks**

Comments:

1 (a) The statistic T is derived from a random sample taken from a population which has an unknown parameter θ. T is an unbiased estimator for θ.

What does the statement "T is an unbiased estimator for θ" imply?

(1)

(b) A random sample of size n is taken from each of two independent populations.

The first population has mean μ and variance σ^2, and $\overline{X}$ denotes the sample mean.

The second population has mean $\frac{\mu}{3}$ and variance $b\sigma^2$, where b is a positive constant, and $\overline{Y}$ denotes the sample mean.

Two unbiased estimators for μ are defined by

$$T_1 = 4\overline{X} - a\overline{Y} \text{ and } T_2 = \frac{1}{9}(8\overline{X} + 3\overline{Y})$$

(i) Determine the value of a.

(3)

(ii) Show that $\text{Var}(T_1) = \frac{\sigma^2}{n}(16 + 81b)$ and find a simplified expression for $\text{Var}(T_2)$.

(5)

(iii) Calculate the relative efficiency of T_2 with respect to T_1 and decide, giving a reason, which of T_1 or T_2 is the more efficient estimator for μ.

(4)

(Total 13 marks)

2 The weight of sand in a bag can be modelled by a normal random variable with unknown mean μ kilograms and known standard deviation 0.4 kilograms.

The sand in each of a random sample of 25 bags from a large batch is weighed. The mean weight of sand in these 25 bags is found to be 19.9 kg.

(a) Construct a 98% confidence interval for the mean weight of sand in bags in the batch.

(4)

(b) Hence comment on the claim that bags in the batch contain an average of 20 kg of sand.

(2)

(Total 6 marks)

3 The weight, X kilograms, of sand in a bag can be modelled by a normal random variable with unknown mean μ and known standard deviation 0.4.

(a) The sand in each of a random sample of 25 bags from a large batch is weighed.
The **total** weight of sand in these 25 bags is found to be 497.5 kg.

(i) Construct a 98% confidence interval for the mean weight of sand in bags in the batch.

(5)

(ii) Hence comment on the claim that bags in the batch contain an average of 20 kg of sand.

(2)

(iii) State why use of the Central Limit Theorem is **not** required in answering part (a)(i).

(1)

(b) The weight, Y kilograms, of cement in a bag can be modelled by a normal random variable with mean 25.25 and standard deviation 0.35.

A firm purchases 10 such bags. These bags may be considered to be a random sample.

(i) Determine the probability that the **mean** weight of cement in the 10 bags is **less** than 25 kg.

(4)

(ii) Calculate the probability that the weight of cement in **each** of the 10 bags is **more** than 25 kg.

(4)

(Total 16 marks)

4 The waiting time at a hospital's A&E department may be modelled by a normal distribution with mean μ and standard deviation $\frac{\mu}{2}$.

The department's manager wishes a 95% confidence interval for m to be constructed such that it has a width of at most 0.2 μ.

Calculate, to the nearest 10, an estimate of the minimum sample size necessary in order to achieve the manager's wish.

(Total 5 marks)

5 A town council wanted residents to apply for grants that were available for home insulation. In a trial, a random sample of 200 residents was encouraged, either in a letter or by a phone call, to apply for the grants. The outcomes are shown in the table.

	Applied for grant	Did not apply for grant	Total
Letter	30	130	160
Phone call	14	26	40
Total	44	156	200

(a) The council believed that a phone call was more effective than a letter in encouraging people to apply for a grant. Use a χ^2-test to investigate this belief at the 5% significance level.

(8)

(b) After the trial, all the residents in the town were encouraged, either in a letter or by a phone call, to apply for the grants. It was found that there was no association between the method of encouragement and the outcome. State, with a reason, whether a Type I error, a Type II error or neither occurred in carrying out the test in part (a).

(2)

(Total 10 marks)

6 A gardener decided to compare the variability in the heights of sunflowers grown on the sunny side of her garden with the variability in the heights of those grown on the shady side of her garden.

She selected a sample of 13 sunflowers on the sunny side and a sample of 10 sunflowers on the shady side of her garden. She measured the height of each sunflower, in metres, correct to two decimal places. Each sample may be regarded as a random sample.

The independent random variables X and Y denote the height, in metres, of sunflowers grown on the sunny side and on the shady side of her garden respectively. You may assume that X and Y are normally distributed with variances ${\sigma_X}^2$ and ${\sigma_Y}^2$ respectively.

The following results were obtained.

$$\sum(x-\bar{x})^2 = 4.68 \quad \text{and} \quad \sum(y-\bar{y})^2 = 8.10$$

(a) Calculate unbiased estimates of ${\sigma_X}^2$ and ${\sigma_Y}^2$.

(1)

(b) The gardener believed that the height of sunflowers grown on the shady side was more variable than the height of sunflowers grown on the sunny side.

Carry out an F-test, using a significance level of 5%, to test this belief.

(6)

(Total 7 marks)

7 It is claimed that a new drug is effective in the prevention of sickness in holiday-makers. A sample of 100 holiday-makers was surveyed, with the following results.

	Sickness	**No sickness**	**Total**
Drug taken	24	56	80
No drug taken	11	9	20
Total	35	65	100

Assuming that the 100 holiday-makers are a random sample, use a χ^2 test, at the 5% level of significance, to investigate the claim.

(Total 8 marks)

8 An office worker believed that, on average, his journey time from home to the office in the morning was longer than his journey time back home in the evening. In order to test this belief, he recorded his journey times, to the nearest minute, for a random sample of 8 morning journeys and a random sample of 10 evening journeys. The results are summarised below.

Mornings $n = 8$ $\sum m = 204$ $\sum(m - \bar{m})^2 = 470$

Evenings $n = 10$ $\sum e = 180$ $\sum(e - \bar{e})^2 = 300$

(a) State **two** assumptions that need to be made so that an independent-samples t-test is valid.

(2)

(b) Making these assumptions, investigate, at the 2.5% level of significance, the office worker's belief.

(9)

(Total 11 marks)

Mark schemes

1 (a) E(T) = θ

B1

1

(b) (i) $\mathrm{E}(T_1) = \mu$

$\Rightarrow \mathrm{E}(4\bar{X} - a\bar{Y}) = 4\mathrm{E}(\bar{X}) - a\mathrm{E}(\bar{Y}) = \mu$

M1

$\Rightarrow 4\mu - a\frac{\mu}{3} = \mu$

A1

$\Rightarrow 3 = \frac{a}{3} \Rightarrow a = 9$

A1

3

(ii) $\mathtt{Var}(T_1) = 16\mathtt{Var}(\bar{X}) + a^2\mathtt{Var}(\bar{Y})$

M1

$= 16\frac{\sigma^2}{n} + 81b\frac{\sigma^2}{n}$

M1

$= \frac{\sigma^2}{n}(16 + 81b)$

AG

A1

$\mathtt{Var}(T_2) = \frac{64}{81} \times \frac{\sigma^2}{n} + \frac{9}{81} \times \frac{b\sigma^2}{n}$

M1

$= \frac{b\sigma^2}{81n}(64 + 9b)$

A1

5

(iii) $\mathrm{RE}(T_2 \mathtt{wrt} T_1) = \dfrac{\{\mathtt{Var}(T_2)\}^{-1}}{\{\mathtt{Var}(T_1)\}^{-1}} = \dfrac{\mathtt{Var}(T_1)}{\mathtt{Var}(T_2)}$

M1

$= 81 \times \dfrac{(16 + 81b)}{(64 + 9b)}$

ft on Var(T_2)

A1ft

$81 \times 16 + 81^2 b > 64 + 9b$

($\because$ {9,16,64,81} $\subset Z^+$ and $b > 0$ [given])

E1ft

$\Rightarrow$ RE(T_2wrt T_1) > 1

$\Rightarrow$ RE T_2 more efficient than T_1.

Dependent on previous E1

E1ft

4

[13]

2 (a) 98% (0.98) $\Rightarrow$ z = **2.32 to 2.33**

AWFW *(2.3263)*

B1

CI for μ is $\bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}$

Used with z (2.05 to 2.58),
$\bar{x}$ (19.9) and σ (0.4)
and ÷ $\sqrt{n}$ with n > 1

M1

Thus $19.9 \pm 2.3263 \times \frac{0.4}{\sqrt{25}}$

z (2.05 to 2.06 or 2.32 to 2.33 or 2.57 to 2.58),
$\bar{x}$ (19.9) and σ (0.4) and ÷ $\sqrt{25 \text{ or } 24}$

A1

Hence **19.9 ± 0.2**

CAO / AWRT *(0.186104)*

or

(19.7, 20.1)

AWRT

A1

4

(b) **Clear correct comparison of 20 with CI**

F on CI providing it contains 20

eg 20 is within CI **or** LCL < 20 < UCL

*Quoting values for CI is **not** required*

BF1

so

Agree with claim **or** no reason to doubt claim

OE; dependent on previous BF1

Bdep1

2

[6]

3 (a) (i) $\overline{x} = \frac{497.5}{25} =$ **<u>19.9</u>**

CAO

B1

98% (0.98) ⇒ z = **<u>2.32 to 2.33</u>**

AWFW *(2.3263)*

B1

CI for μ is $\overline{x} \pm z \times \frac{\sigma}{\sqrt{n}}$

Used with z (2.05 to 2.58),

$\overline{x}$ (497.5 or 19 to 21) and σ (0.4)

and ÷ $\sqrt{n}$ with n > 1

M1

Thus $19.9 \pm 2.3263 \times \frac{0.4}{\sqrt{25}}$

z (2.05 to 2.06 or 2.32 to 2.33 or 2.57 to 2.58),

$\overline{x}$ (19.9) and σ (0.4) and ÷ $\sqrt{25 \text{ or } 24}$

A1

Hence **<u>19.9 ± 0.2</u>**

CAO / AWRT *(0.186104)*

or

<u>(19.7, 20.1)</u>

AWRT

A1

5

(ii) **Clear correct comparison** of **20** with **CI**

F on CI providing it contains 20

eg 20 is within CI **or** LCL < 20 < UCL

*Quoting values for CI is **not** required*

BF1

Agree with claim **or** no reason to doubt claim

OE; dependent on previous BF1

Bdep1

2

(iii) **Weight** of sand in a bag or X / x or **original distribution** or **parent population**

It / mean / data / sample / information / sand is normal ⇒ B0

is **normal**

*Reference **only to** sample size **or** standard deviation ⇒ B0*

B1

1

(b) (i) $Y \sim N(25.25,\ 0\ 35^2)$

Accept percentage equivalent probabilities

V(mean) = **0.35^2 / 10 or 0.0122 to 0.0123**

CAO / AWFW (0.01225)

or

SD (mean) = **<u>0.35 / √10 or 0.11 to 0.111</u>**

CAO / AWFW (0.11068)

B1

$$P\left(\bar{Y} < 25\right) = P\left(Z < \frac{25 - 25.25}{0.35/\sqrt{10}}\right)$$

*Standardising 25 using 25.25 and **0.35 / √10 OE** but allow (25.25 − 25)*

M1

= $P(Z < -2.25877) = 1 - P(Z < 2.25877)$

***Correct** area change*

*May be implied by a **correct** answer **or** an answer < **0.5***

m1

= 1 − (0.98809 to 0.98778)

= **<u>0.011 to 0.013</u>**

AWFW (0.01195)

(0.987 to 0.989) ⇒ B1 M1 m0 A0

A1

4

(ii) $P(Y > 25) = P\left(Z > \frac{25 - 25.25}{0.35}\right)$

Standardising 25 using 25.25 and 0.35 but allow (25.25 − 25)

M1

$= P(Z > -0.71429) = P(Z < 0.71429)$

= **<u>0.761 to 0.764</u>**

AWFW (0.76247)

(0.236 to 0.239) ⇒ M1 A0

A1

$P(Y > 25 \text{ in each of } 10) = \underline{\boldsymbol{p^{10}}}$

Any p^{10} providing $0 < p < 1$

*May be implied by a **correct** answer*

M1

= **<u>0.065 to 0.068</u>**

AWFW (0.06641)

A1

4

[16]

4

95% $\Rightarrow$ z = **1.96**

CAO (AWRT from calculator)

B1

Require $2 \times \frac{1.96\sigma}{\sqrt{n}} \leq 0.2\mu$

Used; may be implied

Allow 'no 2 ×'

Allow '= sign' throughout

M1

Thus $2 \times \frac{1.96}{\sqrt{n}} \times \frac{\mu}{2} \leq 0.2\mu$

Use of $\sigma = \frac{\mu}{2}$; *may be implied*

Allow 'no 2 ×'

M1

Thus $\sqrt{n} \geq \frac{1.96}{2}$

Attempt at solution for $\sqrt{n}$ *or n*

M1

Thus $n \geq 96.04$

Thus, to nearest 10; n = **100**

CAO

A1

[5]

5

(a)

O_i	E_i	$(\lvert O_i - E_i \rvert - 0.5)^2 / E_i$
30	35.2	0.6276
14	8.8	2.5102
130	124.8	0.1770
26	31.2	0.7080
	χ^2	**4.0228**

E attempted (at least two correct to 1 d.p.)

M1

Yates' correction attempted; at least one correct value in final column

M1

χ^2 attempted

M1

AWFW 4.02 to 4.03

A1

H_0: No association between method of receiving information and outcome

At least one correct

H_1: Association between method of receiving information and outcome

If "independent" used, it must be the right way round

B1

CV of χ^2 for 1 df = 3.84(1)

B1

4.02 > 3.841 so reject H_0
There is significant evidence of an association between method of receiving information and outcome

Dep on A1 and B1 for CV

A1

Applications higher than expected for telephone calls, so council's belief seems to be true

Dep on previous A1

Context conclusion about council's belief, referring to higher than expected for telephone

Adep1

8

Alternative if Yates' not used

O_i	E_i	$(O_i - E_i)^2 / E_i$
30	35.2	0.7682
14	8.8	3.0727
130	124.8	0.2167
26	31.2	0.8667
	χ^2	**4.9243**

Loses M1 for Yates' and A1 for final χ^2 value but can score all the other 6 marks

Final 2 A1 marks dep on 4.92 to 4.93 and B1 for CV

(b) Type I error was made because

E1

H_0 has been rejected (when it was true)

Dep on previous E1

Edep

SC If 'H_0 accepted' when their χ^2 less than their CV

No error was made because
H_0 has been accepted (when it was true)

(E1)

Dep on previous (E1)

(Edep1)

2

[10]

6

(a) $s_1^2 = \frac{4.68}{12} = 0.39 \quad s_2^2 = \frac{8.10}{9} = 0.9$

Both

B1

1

(b) $H_0 : \sigma_X^2 = \sigma_Y^2$ $H_1 : \sigma_Y^2 > \sigma_X^2$

Both

B1

$F_{calc} = \frac{0.9}{0.39} = 2.31$ AWRT

M1A1

$v_1 = 9$, $v_2 = 12$

Both; (df can be implied by correct CV).

B1

$F_{crit} = 2.796$

B1

Insufficient evidence to reject H_0, so conclude that variability of heights is the same.

E1 ✓ 6

[7]

7

O_i	E_i	$(\lvert O_i - E_i \rvert - 0.5)$ (α)	α^2 / E_i
24	28	3.5	0.4375
56	52	3.5	0.2356
11	7	3.5	1.7500
9	13	3.5	0.9423
			3.3654

E attempted

M1

Yates' correction attempted

M1

χ^2 *attempted*

M1

AWFW 3.36 to 3.37

A1

H_0 : No association between drug and prevention of sickness

H_1 : Association between drug and prevention of sickness

(at least H_0 stated correctly)

B1

$\chi^2_{5\%} = 3.841$

CAO

B1

Accept H_0

A1ft

No evidence at the 5% level of significance to support the claim that the drug is effective against sickness.

E1ft

8

[8]

8 (a) Normal distribution.

E1

Common variance.

E1

2

(b) $H_0 : \mu_m - \mu_e = 0$ $H_1 : \mu_m - \mu_e > 0$

Both

B1

$\bar{m} - \bar{e} = 25.5 - 18 = 7.5$

B1

$s^2 = \dfrac{470 + 300}{10 + 8 - 2} = 48.125$

*Accept [48.12,48.13] M1 requires a decent go at **both** numerator and denominator.*

M1A1

$$\frac{7.5-0}{\sqrt{48.125(8^{-1}+10^{-1})}} = 2.28$$

M1A1

$\nu = 16$ $t_{crit} = 2.121$

(df can be implied by correct CV)

B1B1

Sufficient evidence at 2.5% level to reject H_0 and believe that morning journeys are longer, on average.

A1 ✓ **9**

[11]