Mock exam review - statistics

[72 marks]

At the end of a school day, the Headmaster conducted a survey asking students in how many classes they had used the internet.

The data is shown in the following table.

Number of classes in which the students used the internet	0	1	2	3	4	5	6
Number of students	20	24	30	k	10	3	1

1a. State whether the data is discrete or continuous. *[1 mark]*

Markscheme

discrete ***A1***

[1 mark]

The mean number of classes in which a student used the internet is 2.

1b. Find the value of k. *[4 marks]*

Markscheme

$\frac{24+60+3k+40+15+6}{88+k} = 2$ ***M1A1***

Note: Award ***M1*** for substitution into the formula for the mean, award **A1** for a correct equation.

attempt to solve their equation ***(M1)***

$k = 31$ ***A1***

[4 marks]

1c. It was not possible to ask every person in the school, so the Headmaster arranged the student names in alphabetical order and then asked every 10th person on the list. *[1 mark]*

Identify the sampling technique used in the survey.

Markscheme

systematic ***A1***

[1 mark]

As part of a study into healthy lifestyles, Jing visited Surrey Hills University. Jing recorded a person's position in the university and how frequently they ate a salad. Results are shown in the table.

	Salad meals per week			
	0	1–2	3–4	>4
Students	45	26	18	6
Professors	15	8	5	12
Staff and Administration	16	13	10	6

Jing conducted a χ^2 test for independence at a 5 % level of significance.

2a. State the null hypothesis. *[1 mark]*

Markscheme

number of salad meals per week is independent of a person's position in the university ***A1***

Note: Accept "not associated" instead of independent.

[1 mark]

2b. Calculate the p-value for this test. *[2 marks]*

Markscheme

0.0201 (0.0201118...) ***A2***

[2 marks]

2c. State, giving a reason, whether the null hypothesis should be accepted. *[2 marks]*

Markscheme

0.0201 < 0.05 ***R1***

the null hypothesis is rejected ***A1***

Note: Award ***(R1)*** for a correct comparison of their p-value to the test level, award ***(A1)*** for the correct interpretation from that comparison.

Do not award ***(R0)(A1)***.

[2 marks]

Ms Calhoun measures the heights of students in her mathematics class. She is interested to see if the mean height of male students, μ_1, is the same as the mean height of female students, μ_2. The information is recorded in the table.

Male height (cm)	150	148	143	152	151	149	147	
Female height (cm)	148	152	154	147	146	153	152	150

At the 10 % level of significance, a t-test was used to compare the means of the two groups. The data is assumed to be normally distributed and the standard deviations are equal between the two groups.

3a. State the null hypothesis. *[1 mark]*

Markscheme

$\mu_1 - \mu_2 = 0$ ***A1***

Note: Accept equivalent statements in words.

[1 mark]

3b. State the alternative hypothesis. *[1 mark]*

Markscheme

$\mu_1 - \mu_2 \neq 0$ ***A1***

Note: Accept equivalent statements in words.

[1 mark]

3c. Calculate the p-value for this test. *[2 marks]*

Markscheme

0.296 (0.295739...) ***A2***

[2 marks]

3d. State, giving a reason, whether Ms Calhoun should accept the null hypothesis. *[2 marks]*

Markscheme

$0.296 > 0.1$ ***R1***

fail to reject the null hypothesis, there is no difference between the mean height of male and female students ***A1***

Note: Award ***(R1)*** for a correct comparison of their p-value to the test level, award ***(A1)*** for the correct interpretation from that comparison. Do not award ***R0A1***.

[2 marks]

The weights of apples on a tree can be modelled by a normal distribution with a mean of 85 grams and a standard deviation of 7.5 grams.

4a. Find the probability that an apple from the tree has a weight greater than 90 grams. *[2 marks]*

Markscheme

* This sample question was produced by experienced DP mathematics senior examiners to aid teachers in preparing for external assessment in the new MAA course. There may be minor differences in formatting compared to formal exam papers.

Let the weight of an apple be X

$\mathrm{P}(X > 90) = 0.252 (0.252492\ldots)$ **(M1)A1**

[2 marks]

A sample of apples are taken from 2 trees, A and B, in different parts of the orchard.

The data is shown in the table below.

Weight of apples from tree A (g)	90	85	84	79	87	88	91	88
Weight of apples from tree B (g)	77	84	86	83	80	81	85	

The owner of the orchard wants to know whether the mean weight of the apples from tree $\mathrm{A}(\mu_A)$ is greater than the mean weight of the apples from tree $\mathrm{B}(\mu_B)$ so sets up the following test:

$\mathrm{H}_0 : \mu_A = \mu_B$ and $\mathrm{H}_1 : \mu_A > \mu_B$

4b. Find the p-value for the owner's test. *[2 marks]*

Markscheme

p-value $= 0.0189 (0.018947\ldots)$ **(M1)A1**

[2 marks]

4c. The test is performed at the 5% significance level. *[2 marks]*

State the conclusion of the test, giving a reason for your answer.

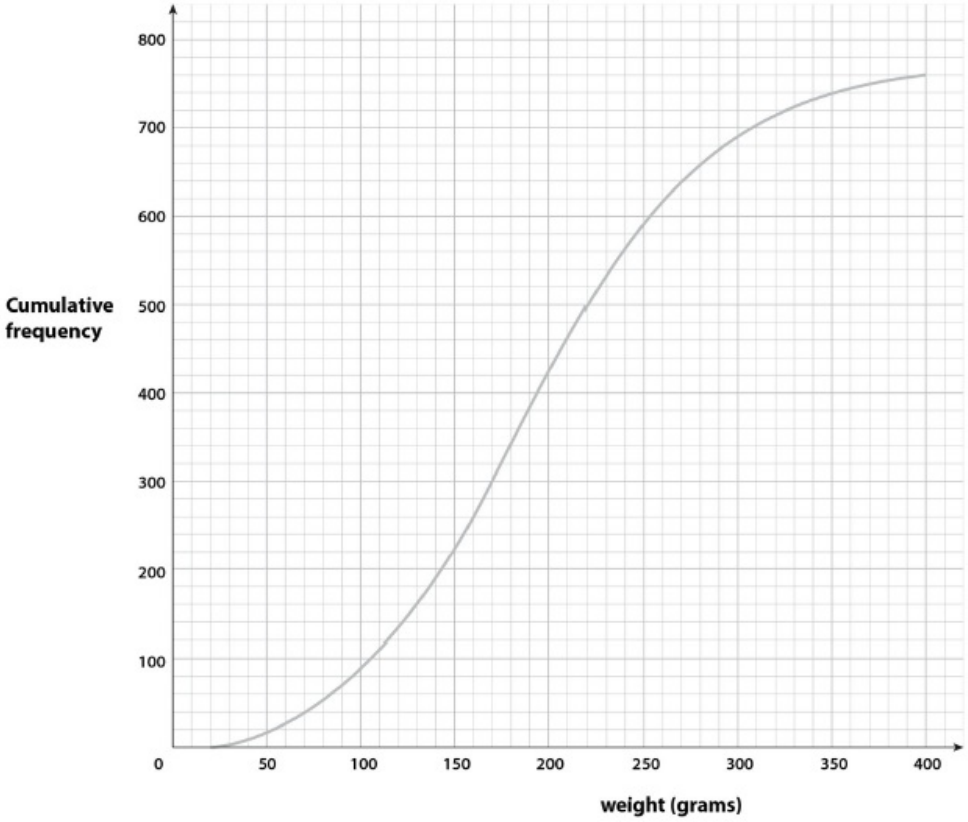
Markscheme

$0.0189 < 0.05$ **R1**

Sufficient evidence to reject the null hypothesis (that the weights of apples from the two trees are equal) **A1**

[2 marks]

A food scientist measures the weights of 760 potatoes taken from a single field and the distribution of the weights is shown by the cumulative frequency curve below.



5a. Find the number of potatoes in the sample with a weight of more than 200 grams. *[2 marks]*

Markscheme

* This sample question was produced by experienced DP mathematics senior examiners to aid teachers in preparing for external assessment in the new MAA course. There may be minor differences in formatting compared to formal exam papers.

$760 - 420 = 340\ (\text{g})$ **(M1)A1**

[2 marks]

5b. Find the median weight. *[1 mark]*

Markscheme

Median $= 190$ (g) **A1**

[1 mark]

5c. Find the lower quartile. *[1 mark]*

Markscheme

Lower quartile $= 135 - 140$ (g) **A1**

[1 mark]

5d. Find the upper quartile. *[1 mark]*

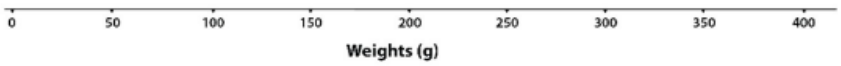
Markscheme

Upper quartile $= 242 - 247$ (g) **A1**

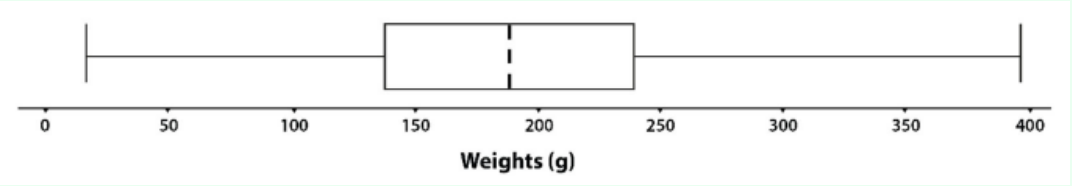
[1 mark]

5e. The weight of the smallest potato in the sample is 20 grams and the weight of the largest is 400 grams. *[2 marks]*

Use the scale shown below to draw a box and whisker diagram showing the distribution of the weights of the potatoes. You may assume there are no outliers.



Markscheme



M1A1

Note: The **M1** is for a box and whisker plot and the **A1** for all 5 statistics in the right places.

[2 marks]

A dice manufacturer claims that for a novelty die he produces the probability of scoring the numbers 1 to 5 are all equal, and the probability of a 6 is two times the probability of scoring any of the other numbers.

6a. Find the probability of scoring a six when rolling the novelty die. *[3 marks]*

Markscheme

* This sample question was produced by experienced DP mathematics senior examiners to aid teachers in preparing for external assessment in the new MAA course. There may be minor differences in formatting compared to formal exam papers.

Let the probability of scoring $1, \ldots, 5$ be p,

$5p + 2p = 1 \Rightarrow p = \frac{1}{7}$ **(M1)(A1)**

Probability of $6 = \frac{2}{7}$ **A1**

[3 marks]

6b. Find the probability of scoring more than 2 sixes when this die is rolled 5 times. *[4 marks]*

Markscheme

Let the number of sixes be X

$X \sim \mathrm{B}\left(5, \frac{2}{7}\right)$ **(M1)**

$\mathrm{P}(X > 2) = \mathrm{P}(X \geq 3)$ or $\mathrm{P}(X > 2) = 1 - \mathrm{P}(X \leq 2)$ **(M1)**

$= 0.145\ (0.144701\ldots)$ **(M1)A1**

[4 marks]

To test the manufacture's claim one of the novelty dice is rolled 350 times and the numbers scored on the die are shown in the table below.

Number scored	Frequency
1	32
2	57
3	47
4	58
5	54
6	102

6c. Find the expected frequency for each of the numbers if the manufacturer's claim is true. *[2 marks]*

Markscheme

Expected frequency is $350 \times p$ or $350 \times 2p$ **(M1)**

Number scored	Frequency	Expected frequency
1	32	50
2	57	50
3	47	50
4	58	50
5	54	50
6	102	100

A1

[2 marks]

A χ^2 goodness of fit test is to be used with a 5% significance level.

6d. Write down the null and alternative hypotheses. *[2 marks]*

Markscheme

H_0 : The manufacture's claim is correct **A1**
H_1 : The manufacturer's claim is not correct **A1**

[2 marks]

6e. State the degrees of freedom for the test. *[1 mark]*

Markscheme

Degrees of freedom $= 5$ **A1**

[1 mark]

6f. Determine the conclusion of the test, clearly justifying your answer. *[4 marks]*

Markscheme

p-value $= 0.0984\ (0.0984037\ldots)$ **(M1)A1**

$0.0984 > 0.05$ **R1**

Hence insufficient evidence to reject the manufacture's claim. **A1**

[4 marks]

A calculator generates a random sequence of digits. A sample of 200 digits is randomly selected from the first 100 000 digits of the sequence. The following table gives the number of times each digit occurs in this sample.

digit	0	1	2	3	4	5	6	7	8	9
frequency	17	21	15	19	25	27	19	23	18	16

It is claimed that all digits have the same probability of appearing in the sequence.

7a. Test this claim at the 5% level of significance. *[7 marks]*

Markscheme

H_0: The sequence contains equal numbers of each digit. ***(A1)***

H_1: The sequence does not contain equal numbers of each digit. ***(A1)***

$\chi^2_{\text{calc}} = \frac{(9+1+25+1+25+49+1+9+4+16)}{20} = 7$ ***(M1)(A1)***

The number of degrees of freedom is 9. ***(A1)***

$\chi^2_{0.95;\,5} = 16.919$ ***(A1)***

$\chi^2_{\text{calc}} < 16.919$. Hence H_0 is accepted. ***(A1)***

[7 marks]

7b. Explain what is meant by the 5% level of significance. *[2 marks]*

Markscheme

The probability of rejecting H_0 when it is true ***(A1)***

is 0.05. ***(A1)***

Note: Award ***(A1)(A1)*** for "the probability of a type I error is 0.05."

[2 marks]

Sergio is interested in whether an adult's favourite breakfast berry depends on their income level. He obtains the following data for 341 adults and decides to carry out a χ^2 test for independence, at the 10% significance level.

		Income level		
		Low	Medium	High
Favourite berry	Strawberry	21	39	30
	Blueberry	39	67	42
	Other berry	32	45	26

8a. Write down the null hypothesis. *[1 mark]*

Markscheme

The favourite breakfast/berry (of adults) is independent of (their) income (level). ***A1***

[1 mark]

8b. Find the value of the χ^2 statistic. *[2 marks]*

Markscheme

$\chi^2 = 2.27 \ \ (2.26821\ldots)$ ***A2***

[2 marks]

The critical value of this χ^2 test is 7.78.

8c. Write down Sergio's conclusion to the test in context. Justify your answer. *[2 marks]*

Markscheme

EITHER

$2.27 < 7.78$ **OR** $2.27 <$ critical value ***R1***

OR

$0.687 > 0.1$ (using p-value)

THEN

(Do not reject H_0)

Insufficient evidence (at the 10% significance level) that the favourite berry depends on income level. ***A1***

Note: Do not award ***R0A1***. Accept "χ^2" in place of their "2.27", provided an answer was seen in part (b). Their conclusion must be consistent with their χ^2 (or a correct p-value) and their hypothesis.

[2 marks]

Manny and Annabelle, mathematics teachers at Burnham High School, give their students the same examination. A random sample of the examination scores were collected from each of their classes.

Examination scores from Manny's class	76	77	82	84	88	90	91	98
Examination scores from Annabelle's class	68	79	81	89	91	92	92	95

Annabelle uses these scores to conduct a two-tailed t-test to compare the means of the two classes, at the 5% level of significance. It is assumed the examination scores for both classes have the same variance and are normally distributed.

The null hypothesis is $\mu_1 = \mu_2$, where μ_1 is the mean examination score from Manny's class and μ_2 is the mean examination score from Annabelle's class.

9a. Write down the alternative hypothesis. *[1 mark]*

Markscheme

$(\mathrm{H}_1 :)\ \mu_1 \neq \mu_2$ **A1**

Note: Accept an equivalent statement in words referring to μ_1 and μ_2 as defined in the question.

[1 mark]

9b. Find the p-value for this test. Give your answer correct to five decimal places. *[2 marks]*

Markscheme

$0.97652 \quad (0.976516\ldots)$ **A2**

[2 marks]

Annabelle concludes there is insufficient evidence to reject the null hypothesis.

9c. State whether Annabelle's conclusion is correct. Give a reason for your answer. *[2 marks]*

Markscheme

$0.97652 > 0.05 \quad (0.977 > 0.05)$ **R1**

Annabelle's conclusion is correct. **A1**

Note: Do not award ***R0A1***. Answer must reference Annabelle's conclusion; do not accept an answer, without context, of "fail to reject H_0" for the ***A1*** mark.

[2 marks]

Leo is investigating whether a six-sided die is fair. He rolls the die 60 times and records the observed frequencies in the following table:

Number on die	1	2	3	4	5	6
Observed frequency	8	7	6	15	12	12

Leo carries out a χ^2 goodness of fit test at a 5% significance level.

10a. Write down the null and alternative hypotheses. *[1 mark]*

Markscheme

H_0 : The die is fair **OR** $\mathrm{P}(\text{any number}) = \frac{1}{6}$ **OR** probabilities are equal

H_1 : The die is not fair **OR** $\mathrm{P}(\text{any number}) \neq \frac{1}{6}$ **OR** probabilities are not equal ***A1***

[1 mark]

10b. Write down the degrees of freedom. *[1 mark]*

Markscheme

5 ***A1***

[1 mark]

10c. Write down the expected frequency of rolling a 1. *[1 mark]*

Markscheme

10 ***A1***

[1 mark]

10d. Find the p-value for the test. *[2 marks]*

Markscheme

(p-value $=$) 0.287 $(0.28724163\ldots)$ ***A2***

[2 marks]

10e. State the conclusion of the test. Give a reason for your answer. *[2 marks]*

Markscheme

$0.287 > 0.05$ ***R1***

EITHER

Insufficient evidence to reject the null hypothesis ***A1***

OR

Insufficient evidence to reject that the die is fair ***A1***

Note: Do not award ***R0A1***. Condone "*accept* the null hypothesis" or "the die is fair". Their conclusion must be consistent with their p-value and their hypothesis.

[2 marks]