# 4.11 Hypothesis Testing

## Question Paper

| Course | DP IB Maths |
|---|---|
| Section | 4. Statistics & Probability |
| Topic | 4.11 Hypothesis Testing |
| Difficulty | Very Hard |

**Time allowed:** 140

**Score:** /115

**Percentage:** /100

## Question 1a

An IB student is investigating the study habits of her peers. She prefers to study in the morning but one of her male friends prefers to study in the evening. She decides to conduct a study to find out if the time that a student prefers to study is affected by their gender. She selects 20 boys and 20 girls at random from her year group and asks them to pick their preferred time of study from the options of morning, afternoon or evening.

(a) State the type of sampling used in her investigation.

[1 mark]

## Question 1b
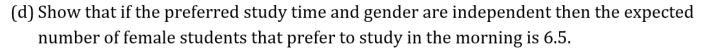
(b) State the null and alternative hypotheses.

[2 marks]

## Question 1c

The results of the survey are listed below.

| Female students | Male students |
|---|---|
| morning, evening, evening, afternoon, morning, morning, afternoon, evening, afternoon, morning, morning, evening, afternoon, morning, evening, evening, morning, afternoon, morning, morning | morning, afternoon, morning, evening, evening, evening, afternoon, morning, evening, evening, afternoon, evening, evening, afternoon, evening, morning, afternoon, evening, afternoon, evening |

(c) Create a contingency table of observed values.

[4 marks]

## Question 1d

(d) Show that if the preferred study time and gender are independent then the expected number of female students that prefer to study in the morning is 6.5.

[2 marks]

## Question 1e

The critical value for a chi-squared test performed at a significance level of 5% is 5.991.

(e) Using technology, calculate the $\chi^2$ statistic and comment on the conclusions of the test in the context of the question, giving a reason for your answer.

[4 marks]

## Question 2a

A group of university researchers are concerned that pollution from industrial activity is having an effect on the local river system.  In order to investigate whether runoff from local factories is affecting the water quality, samples were taken from next to the factories as well as from upstream.  The data from both locations is compared to test whether the average pH is higher next to the factories than it is upstream. The test is conducted at a significance level of 1%.  The pH of the samples is recorded in the table below.

| Factory Location (pH) | 7.8 | 8.1 | 8.0 | 7.5 | 7.9 | 8.1 | 7.8 | 7.8 | 7.8 | 7.7 |
|---|---|---|---|---|---|---|---|---|---|---|
| Upstream location (pH) | 7.2 | 7.7 | 7.6 | 7.8 | 7.6 | 7.5 | 7.9 | 7.8 | 7.3 | 7.6 |

(a) State the type of distribution of the data.

[1 mark]

## Question 2b

(b) State the null and alternative hypotheses.

[2 marks]

## Question 2c

(c) Calculate the $p$-value and state any conclusions of the test, giving a reason for your answer.

[5 marks]

## Question 2d

(d) Interpret the implications of the conclusion to the test for the researchers.

[1 mark]

## Question 2e

It is thought that agricultural runoff will also negatively affect the water quality. Samples have been taken near a number of farms and compared to the upstream samples at the same significance level. The $p$-value for this test is 0.0082.

(e) Comment on the result of this test and compare it to the initial test. Justify your answer.

[2 marks]

## Question 3a

A cat shelter looks after 328 cats that have been abandoned or neglected. As part of the initial health check of a new cat, its weight, $W$, in pounds is measured and recorded. The table below collects together the data for the cat weights.

| Weight | Frequency |
|---|---|
| $8 \leq W < 9$ | 17 |
| $9 \leq W < 10$ | 91 |
| $10 \leq W < 11$ | 141 |
| $11 \leq W < 12$ | 72 |
| $12 \leq W < 13$ | 7 |

The weight distribution of the general cat population can be described by $W \sim N(10.4, 0.9^2)$. The shelter wants to see if the weight distribution of the cats in their care fits that of the general cat population.

A chi-squared goodness of fit test is to be performed at the 10% significance level.

(a) Write down the null and alternative hypotheses.

[2 marks]

## Question 3b

(b) Draw a new table showing the expected frequencies if the shelter cat population were to match the general cat population weight distribution.

[4 marks]

## Question 3c

(c) Calculate the $p$-value.

[2 marks]

## Question 3d

(d) Comment on the results of the test and the conclusions that can be drawn, giving a reason for your answer.

[2 marks]

## Question 4a

A lottery machine contains a set of coloured balls: three red, five yellow, two green and two blue. When a button is pressed a ball is selected at random. After noting its colour, the same ball is returned to the machine.

(a) Calculate the probability of the machine selecting a yellow ball at random.

[1 mark]

## Question 4b

The machine is to be tested for bias and so the procedure of selecting a ball at random is repeated 180 times and the results are recorded in the table below.

| Colour | Red | Yellow | Blue | Green |
|---|---|---|---|---|
| **Frequency** | 41 | 72 | 29 | 38 |

(b) Draw a table showing the expected frequencies for the test.

[2 marks]

## Question 4c

(c) Perform a goodness of fit test at a significance level of 5% and comment on your results, giving a reason for your answer.

[4 marks]

## Question 5a

Approximately 1.6 million pairs of twins are born each year world-wide.

(a) Calculate the probability that both babies in any random set of twins are male. State any assumptions you have made.

**[3 marks]**

## Question 5b

(b) Complete the probability table for the number of male babies born in a given set of twins.

| Number of male babies | 0 | 1 | 2 |
|---|---|---|---|
| Probability | | | |

**[2 marks]**

## Question 5c

In one particular hospital, there are 236 sets of twins born in one year.

(c) Complete the expected frequency table.

| Number of male babies | 0 | 1 | 2 |
|---|---|---|---|
| Expected frequency | | | |

[2 marks]

## Question 5d

The actual number of male babies born as part of a set of twins in this hospital during the year is shown in the table below.

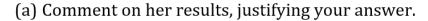| Number of male babies | 0 | 1 | 2 |
|---|---|---|---|
| Expected frequency | 46 | 149 | 41 |

(d) Perform a chi-squared goodness of fit test, given that the critical value for the test is is 5.991. Comment on the results of the test, giving a reason for your answer.

[4 marks]

## Question 6a

A group of 14 students are studying reaction times in their science lesson. They use a computer program to measure the time taken, in milliseconds, for a button to be pressed after a green light is shown. One of the students has compared the group's results to those of the expected normal distribution of reaction times and completed a chi-squared goodness of fit test at a significance level of 10%. The calculated $p$-value is 0.130.

(a) Comment on her results, justifying your answer.

[3 marks]

## Question 6b

The student noted that in the first test, the dominant hand was used to press the button. It was decided to repeat the experiment using the non-dominant hand. The calculated $p$-value for this test, which was also conducted with a significance level of 10%, is 0.0854.

(b) Compare the results of both tests. Justify your answer.

[3 marks]

## Question 6c

The data from both tests are then compared using a two-tailed $t$-test at a significance level of 10%. The $p$-value for this test is 0.0667.

(c) Explain the result of this test.

[3 marks]

## Question 7a

A study has noted that the top 20% of phone users spend more than 4.5 hours per day on their phones. An independent health charity wishes to investigate whether heavy phone usage is linked to the age of the user. A survey of 85 people is commissioned where the age group (child, teenager or adult) and the average number of hours, $h$, that are spent on the phone each day are noted.
The results are displayed in the table below.

| | Low usage $(0 \leq h < 2)$ | Medium usage $(2 \leq h < 4.5)$ | High usage $(4.5 \leq h)$ | Total |
|---|---|---|---|---|
| **Child** | 15 | 8 | 3 | 26 |
| **Teenager** | 6 | 18 | 8 | 32 |
| **Adult** | 9 | 12 | 6 | 27 |
| **Total** | 30 | 38 | 17 | 85 |

(a) Draw a contingency table of expected values.

**[3 marks]**

## Question 7b

(b) Perform a chi-squared test and comment on the results. You should state the null and alternative hypotheses and comment on any conclusions found.

Select the appropriate critical value from the table below.

| Degrees of freedom | Critical value (5%) |
|---|---|
| 1 | 3.841 |
| 2 | 5.991 |
| 3 | 7.815 |
| 4 | 9.488 |
| 5 | 11.070 |
| 6 | 12.592 |

**[7 marks]**

## Question 7c

(c) State two steps that could be taken to increase the validity of the conclusions drawn from the test.

[2 marks]

## Question 8a

Gorr and Greer are butchers and purchase beef from two different suppliers: Beefthor and Zeusbeef. Gorr and Greer use the beef to make 150 gram burgers. Gorr suspects that the mean amount of fat in the beef is different between the two suppliers. Gorr measures the amount of fat, in grams, in a sample of 150 gram burgers using beef from the two suppliers. The data is shown in the table below.

| Beefthor | 28.1 | 29.3 | 27.2 | 27.5 | 30.1 | 27.0 | 28.1 | 27.2 |
|---|---|---|---|---|---|---|---|---|
| Zeusbeef | 29.5 | 28.9 | 30.3 | 28.3 | 28.8 | 27.9 | 29.3 | |

Gorr uses a pooled two-sample $t$-test to test his suspicion using a 5% significance level.

(a) State two assumptions that Gorr has made about the distribution of the amount of fat in the burgers using beef from each of the two suppliers.

[2 marks]

## Question 8b

(b) Perform the hypothesis test to test Gorr's suspicion. State the hypotheses clearly and justify your conclusion.

**[4 marks]**

## Question 8c

(c) Greer suspects that the beef from Zeusbeef has more fat than the beef from Beefthor. Use the data from Gorr's sample to test Greer's suspicion using a 5% significance level. State the hypotheses clearly and justify your conclusion.

**[2 marks]**

## Question 8d

(d) Explain why Gorr and Greer's tests have different conclusions.

**[2 marks]**

## Question 8e

(e) Gorr and Greer use another sample to test their different suspicions using the same significance level. There is sufficient evidence from this sample to support Gorr's suspicion. Explain whether Greer's suspicion will also be supported using this sample.

**[2 marks]**

## Question 9a

Melanie is playing an online game which has two digital four-sided dice numbered 1 to 4. The pair of digital dice are rolled together three times and points are scored. High points are scored for rolling doubles this is when both dice land on the same number.

The following table shows the distribution of the number of doubles scored in each set of three throws when the Melanie plays the game 400 times.

| Number of doubles | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Frequency | 148 | 174 | 66 | 12 |

Melanie suspects that the data can be modelled using a binomial distribution with the probability of rolling a double being 0.25. A $\chi^2$ goodness of fit test is to be used with a 5% significance level to test Melanie's suspicion.

a)
Perform the test to show that the data can not be modelled accurately by the binomial distribution B(3, 0.25).

**[6 marks]**

## Question 9b

Melanie still suspects that the data can be modelled using a binomial distribution. A $\chi^2$ goodness of fit test is to be used with a 5% significance level to test Melanie's suspicion.

b)
Assuming that each dice roll is independent, use the table of results to estimate the probability of rolling doubles with the pair of dice.

**[2 marks]**

## Question 9c

c)
(i)
Write down the number of degrees of freedom for her test.
(ii)
Perform the $\chi^2$ goodness of fit test and state, with reason, a conclusion.

**[7 marks]**

## Question 10a

Gordon takes the bus to work at the same time each day and he records the time, in minutes, of each journey. He collates his data in the table below.

| Time ($t$ minutes) | Frequency |
|---|---|
| $10 \leq t < 15$ | 6 |
| $15 \leq t < 20$ | 11 |
| $20 \leq t < 25$ | 17 |
| $25 \leq t < 30$ | 21 |
| $30 \leq t < 40$ | 14 |
| $40 \leq t < 60$ | 1 |

Gordon wants to use a $\chi^2$ goodness of fit test with a 10% significance level to see whether the journey times can be modelled by a normal distribution.

a)
State the null and alternative hypotheses.

**[2 marks]**

## Question 10b

The mean time of the sample is 25.3 minutes and the standard deviation for the sample is 7.38 minutes.

b)
Write down the unbiased estimate for the population mean time and find the unbiased estimate for the population variance.

**[2 marks]**

## Question 10c

c)
Calculate the expected frequencies if the data were normally distributed using the unbiased estimates found in part (b).
Complete the table below giving the frequencies to 3 decimal places.

| Time ($t$ minutes) | Expected Frequency |
|---|---|
| $t < 15$ | |
| $15 \leq t < 20$ | 10.850 |
| $20 \leq t < 25$ | 17.219 |
| $25 \leq t < 30$ | 17.675 |
| $30 \leq t < 40$ | |
| $t \geq 40$ | 1.679 |

**[6 marks]**

## Question 10d

d)
Write down the number of degrees of freedom for the test.

[1 mark]

## Question 10e

e)
The critical value for the test is 4.605. State the conclusion of the test. Give a reason for your answer.

[4 marks]