Topic: Statistical Testing (χ^2 Tests)
IB Math AI SL

Answer all questions. Show all working where appropriate. Total: 85 marks.

1. **[Paper 1 Style, Short Answer, Easy, 3 marks]**
Abhinav carries out a χ^2 test for independence at the 1% significance level to determine whether a person's gender is associated with their chosen professional field: Engineering, Medicine, or Law.

   (a) State the null hypothesis, H_0, for this test.

   (b) State the alternative hypothesis, H_1, for this test.

2. **[Paper 1 Style, Short Answer, Easy, 4 marks]**
Abhinav surveyed 220 people for his study in Question 1. The results are shown in the contingency table below.

	Engineering	**Medicine**	**Law**	**Total**
Male	55	30	25	110
Female	35	45	30	110
Total	90	75	55	220

   (a) Write down the number of degrees of freedom for this test.

   (b) Calculate the expected frequency of male engineers, assuming the null hypothesis is true. Show your working.

3. **[Paper 1 Style, Short Answer, Medium, 4 marks]**
Continuing with Abhinav's data from Question 2, a χ^2 test for independence is performed.

   (a) Use your graphic display calculator to find the p-value for this test.

   (b) State the conclusion for this test at the 1% significance level. Give a clear reason for your answer.

4. **[Paper 2 Style, Longer Question, Medium, 6 marks]**
Casanova restaurant offers a set menu where a customer chooses one of the following meals: Pasta, Fish, or Shrimp. The manager surveyed 150 customers and recorded the customer's age category and chosen meal.

	Pasta	**Fish**	**Shrimp**
Adults	24	25	32
Children	20	14	35

A χ^2 test for independence is performed at the 10% significance level.

(a) Write down the null hypothesis for this test.

(b) Show that the expected number of children who chose Shrimp is exactly 30.82.

(c) Use your graphic display calculator to find the χ^2 test statistic.

5. **[Paper 1 Style, Short Answer, Medium, 4 marks]**
Refer to the restaurant data in Question 4.

(a) Write down the p-value for the test.

(b) State the conclusion for this test. Give a reason for your answer by comparing your p-value to the 10% significance level.

6. **[Paper 1 Style, Short Answer, Easy, 4 marks]**
The call centre at the Austin Police Department receives hundreds of calls for service every week. To determine whether the phone calls are uniformly distributed among the days of the week, a random sample of 147 calls are studied.

(a) State the null hypothesis for a goodness-of-fit test on this data.

(b) Calculate the expected frequency of calls for Wednesday, assuming the null hypothesis is true.

7. **[Paper 2 Style, Longer Question, Medium, 6 marks]**
The observed frequencies for the 147 police calls from Question 6 are shown below.

Day	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Observed Freq	22	18	17	25	25	19	21

A χ^2 goodness-of-fit test is performed at a 5% significance level.

(a) Write down the number of degrees of freedom.

(b) Using your graphic display calculator, find the χ^2 test statistic and the associated p-value.

(c) Determine the conclusion for the test, providing a clear mathematical reason.

8. **[Paper 1 Style, Short Answer, Medium, 5 marks]**
A university canteen sells baskets of candies. Each basket contains a collectible sticker featuring a famous mathematician. The canteen claims the stickers are distributed according to the following percentages:

Mathematician	Euler	Pythagoras	Lagrange	Laplace	Fourier
Percentages (%)	50	20	15	10	5

A staff member buys 140 baskets of candy for his students.

(a) Calculate the expected frequency of finding a Pythagoras sticker.

(b) Calculate the expected frequency of finding a Fourier sticker.

(c) State the number of degrees of freedom for a χ^2 goodness-of-fit test.

9. **[Paper 2 Style, Longer Question, Hard, 7 marks]**
The staff member from Question 8 unboxes the 140 candies and records the following observed frequencies of the stickers:

Mathematician	Euler	Pythagoras	Lagrange	Laplace	Fourier
Observed Freq	72	25	20	15	8

He conducts a χ^2 goodness-of-fit test at a 5% significance level.

(a) Use your graphic display calculator to find the χ^2 test statistic.

(b) Find the p-value.

(c) Does the sample provide enough evidence to reject the canteen's claim? Justify your reasoning.

10. **[Paper 2 Style, Longer Question, Hard, 6 marks]**
An automobile manufacturer paints their vehicles according to specific colour proportions: Black (30%), White (30%), Red (20%), Blue (10%), and Green (10%). A local dealership randomly surveys 150 of their previous customers about their colour preference, obtaining the following results:

Colour	Black	White	Red	Blue	Green
Observed Responses	35	38	32	24	21

(a) State the null and alternative hypotheses to test whether customer preferences match the manufacturer's proportions.

(b) Calculate the expected number of customers who prefer Blue vehicles.

11. **[Paper 1 Style, Short Answer, Hard, 5 marks]**
Continuing with the automobile survey from Question 10, a χ^2 goodness-of-fit test is performed at a 5% significance level.

(a) Calculate the χ^2 test statistic and the p-value.

(b) Determine if the manufacturer should consider changing the colour distribution of vehicles sent to this dealership. Give a reason.

12. **[Paper 1 Style, Short Answer, Hard, 6 marks]**
A die is thrown 120 times with the following results:

Score	1	2	3	4	5	6
Observed Frequency	27	12	16	25	26	14

A χ^2 goodness-of-fit test is performed to investigate whether the die is fair.

(a) Write down the expected frequency for each score.

(b) Calculate the χ^2 test statistic.

(c) The test is conducted at the 5% level of significance. Explain what is meant by a "Type I Error" in the context of this question.

13. **[Paper 2 Style, Longer Question, Very Hard, 8 marks]**
A factory manufacturing computer chips has five conveyor belts. An engineer suspects that the proportion of defective chips is not evenly distributed across the five belts. On a given day, he records the following defectives:

Conveyor Belt	1	2	3	4	5
Defectives	57	32	48	55	78

He conducts a χ^2 goodness-of-fit test at a 1% level of significance.

(a) Calculate the total number of defectives recorded.

(b) Estimate the expected number of defectives from each conveyor belt if the defects are evenly distributed.

(c) Use your graphic display calculator to find the p-value.

(d) State the conclusion of the test, justifying your answer.

14. **[Paper 1 Style, Short Answer, Very Hard, 5 marks]**
A researcher tests the goodness-of-fit of a dataset against a Normal distribution model, sorting data into 7 classes (bins). The test yields an expected frequency of 3.5 for the final class interval.

(a) State why a χ^2 test should not be performed directly on these 7 classes.

(b) Describe the mathematical adjustment the researcher must make before performing the test.

(c) After making this adjustment, state the new number of degrees of freedom.

15. **[Paper 2 Style, Longer Question, Very Hard, 6 marks]**
The strength of earthquakes is measured on the Richter scale. A seismologist wishes to test if the location of an earthquake (Region A vs Region B) is independent of its severity (Low, Medium, High). The calculated p-value for the χ^2 test of independence is 0.0412.

(a) State the conclusion if the test is evaluated at the 5% significance level.

(b) State the conclusion if the test is evaluated at the 1% significance level.

(c) Explain the consequence of choosing a 5% significance level instead of 1% concerning the likelihood of committing a Type I error.